Evaluating and using remote sensing data for studying economic activity and population change in China

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy

at

The University of Waikato

by

Xiaoxuan Zhang

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2024

Abstract

The use of remote sensing data, especially night-time lights (NTL) data, has flourished in recent decades as a proxy for measuring economic activity and the spatial distribution of population. The availability of these data can make up for the absence of and possible inaccuracy in traditional economic statistics, especially because NTL data are available at fine scale and with high frequency. The growing use of these data raises the question of how accurate is this night-time lights proxy, and for what purposes is it best suited when studying economic development and population changes at sub-national levels.

In the first part of this thesis, a 20-year time-series of GDP at China's third sub-national level (counties, county-level cities and districts) and census population counts at the same level from 2000, 2010 and 2020, are used as a benchmark to examine the performance of multiple sources of NTL data as proxies for local economic activity and inequality. I also test the accuracy of gridded population estimates, some of which rely on NTL data, for estimating inter-census changes in local population.

Based on the comprehensive evaluation of different NTL sources at different levels of spatial aggregation, the second part of the thesis uses NTL and other remote sensing data, in conjunction with the traditional GDP and population statistics, to study a set of economic issues that are pertinent to regional development in China. The impact of administrative upgrading of county-level units (a type of regional policy practiced in China) and of modern transport infrastructure is studied, using both GDP and NTL data to measure economic impacts. These studies use spatial econometric models that allow for spillovers. The changes in the city size distribution, in terms of land and people, is also studied because of the possibility that China's cities are physically expanding in places other than where the population is moving.

Overall, the thesis contributes a nuanced understanding of how NTL data and other remote sensing data can assist in studying complex social and economic problems, as well as raising some caveats about the limits to the uses of such data.

Acknowledgements

I would like to express my deepest gratitude and appreciation to all those who have contributed to the completion of this thesis. First and foremost, I would like to thank my Chief supervisor for his guidance, expertise, and unwavering support throughout the entire research process. I am also grateful to Supervisor Professor Frank Scrimgeour, and Geua-Boe Gibson for their kind assistance. During my research, they contributed to a rewarding graduate research experience by giving me intellectual freedom in my work, supporting my attendance at various conferences, engaging me in new ideas, and demanding a high quality of work in all my endeavours. I am very grateful for the opportunity the University of Waikato has given me to do my degree while receiving a doctoral scholarship. I would like to thank the Waikato Management School staff for their constant support and inspiration.

Publications generated in connection with this PhD research:

Chapter 2:

Zhang, X., and Gibson, J. (2022). Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. *Remote Sensing*, 14(5), 1282.

Chapter 3:

Zhang, X., Gibson, J., and Deng, X. (2023). Remotely too equal: Popular DMSP night time lights data understate spatial inequality. *Regional Science Policy and Practice*, 15(9), 2106-2126.

Chapter 5:

Zhang, X., Li, C., and Gibson, J. (2024). The role of spillovers when evaluating regional development interventions: evidence from administrative upgrading in China. *Letters in Spatial and Resource Sciences*, 17(1), 9.

Appendix 1:

Gibson, J., Jiang, Y., Zhang, X., and Boe-Gibson, G. (2024). Are disaster impact estimates distorted by errors in popular night-time lights data? *Economics of Disasters and Climate Change*, 1-26.

Table of Contents

Chapter 1: Introduction

Accurate and timely information on spatial and temporal dynamics of socio-economic parameters can reflect the development of regional economic growth (Pinkovskiy and Sala-i-Martin 2016a). Such information is essential for understanding socio-economic conditions of a country or region, and for making a scientific evaluation of the regional development (Feldmeyer et al. 2020; Liu et al. 2021). In particular, Gross Domestic Product (GDP) and resident population of subnational entities are important indicators to depict regional economic development, productivity and inequality (Henderson et al. 2012; Li and Gibson 2013; Ji et al. 2019). With detailed information on the spatial distribution of GDP and population researchers can better understand the dynamics of regional economic development and urbanization processes (Zhao, Currit and Samson 2011). Therefore, the question of how to obtain accurate and frequent data on resident population, GDP and other indicators is of great importance to socio-economic research. Traditional official statistics have been crucial, highly-regarded, data sources for compiling GDP and resident population (Kalkuhl and Wenz 2020). However, there are still many problems in GDP data collection via socioeconomic statistics, for example, infrequency, inconsistency of statistical procedures, and use of relatively coarse spatial units (Pérez-Sindín, Chen and Prishchepov, 2021). Due to the inherent shortcomings in economic surveys and administrative statistics, and weaknesses in calculation methods, there exist many limitations in the traditional way of collecting resident population and GDP data (Zhang et al. 2019).

Remote sensing provides promise in overcoming data errors and infrequency to study socioeconomic development (Donaldson and Storeygard, 2016). For example, satellite-detected night-time lights (NTL) data can support an alternative measure of the intensity of the socio-economic development of humans (Pinkovskiy and Sala-i-Martin 2016b; Son and Cho 2021; Pagaduan 2022; Kim et al. 2023). Compared to traditional economic activity statistics, the NTL data has the potential advantages of lower cost, comparability between countries irrespective of statistical capacity, and availability for spatial units below the level at which traditional statistical data are reported (Henderson et al. 2018). Previous studies show that NTL remote sensing data strongly correlate with social and economic variables, such as electricity consumption, income, population, GDP, land use, and urbanization (Michieka and Fletcher 2012; Miller et al. 2012; Elvidge et al. 2017; Carleton et al. 2022; Méndez and Van Patten 2022).

Numerous recent studies have used NTL data to study socioeconomic issues (Lee 2018; Moscona, Nunn and Robinson 2020; Puente-Ajovin, Sanso-Navarro and

Vera-Cabello 2024). Researchers generally find positive cross-sectional correlations of night lights intensities, measured by satellites from outer space, and levels of GDP (Bickenbach, Nunnenkamp and Söder 2016), especially for urban areas (Gibson et al., 2021) but weaker evidence for correlations with changes in economic activity (Asher et al. 2021; Gibson and Boe-Gibson 2021). In contrast, a widely cited study that mainly used national data suggested that the growth of night lights can reliably proxy for the growth of GDP in low-income countries where GDP data is frequently lacking or of poor quality (Henderson et al. (2012). Notably, that study used data from satellites developed in the 1960s for military purposes; in contrast, data from the newest systems have developed with research objectives in mind. In particular, more updated versions of NTL datasets have better spatial, temporal, and radiometric resolution; examples include the NASA Black Marble data (Román et al., 2018) and Luojia-1 (Zhong et al. 2019; Li et al. 2020). However, despite progress in using NTL data for socioeconomic studies, there has been an inconclusive debate about the validity of the key assumption that night-time lights can serve as a good proxy for economic development (which implies studying changes rather than just levels) of sub-national entities.

As the world's second-largest economy, China is undergoing a massive wave of infrastructure construction and continuous market economy development, and this is receiving growing research attention (Qiang and Jian 2020). To study this development a popular trend has been to use multisource NTL data for China (Clark, Pinkovskiy and Sala-i-Martin 2020; He 2023). In China, several studies have confirmed that the night-time light data can provide a valuable proxy for measuring GDP and resident population at the national or provincial scale. However, the predictive power of NTL data for smaller geographic units is still under debate even as applied studies use these data for research at very local levels, such as counties. Therefore, the growing use of the NTL data for research, and also of the related gridded population data, raises the question of how accurately these data can predict regional economic development.

In light of limited comparisons between multiple NTL data sources at local levels—especially in economics where dated Defense Meteorological Satellite Program (DMSP) data continue to be the most popular NTL data source (Gibson et al, 2024)—as well as in consideration of the regional heterogeneity of development in China, this thesis presents a series of regression-based studies to test predictive accuracy of NTL data and gridded population data. On the one hand, results in this thesis can provide a reference for studies elsewhere that use remote sensing data but don't have a benchmark

like China in terms of detailed county-level data for evaluating the remote sensing data. On the other hand, the results in this thesis that go beyond evaluating the remote sensing data can help to understand China's socio-economic development and provide scientific support for the government's planning decisions and related socio-economic research.

This thesis examines the performance of multiple sources of night-time lights data as proxies for local economic development. In the first part of the thesis predictive accuracy of multiple NTL data sources is studied, using county-level GDP data as the benchmark. China's official GDP statistical data has been doubted due to statistical difficulties, structural incomparability, and human manipulation but prior evaluations do not find these issues to be remarkable (Perkins, and Rawski 2008; Holz 2014). The relationships between multiple NTL datasets (and gridded population datasets) and socio-economic parameters (resident population and GDP) will be explored using regression methods for three administrative level units (provincial-, prefectural city- and county-level) in China. In general, my research will evaluate the performance of nighttime light data as a proxy for resident population and GDP in order to better reveal the development of counties' temporal and spatial characteristics. Moreover, based on the estimation at the county level, this thesis can explore some key economic issues to provide valuable recommendations for economic policy-makers.

The thesis is organized as follows. In the first part, which comprises Chapters 2, 3 and 4 and the two papers in the Appendix, I use census measures of local resident population and GDP from statistical yearbooks as a benchmark to study whether remote sensing data (NTL data and gridded population data) can predict the temporal changes in socio-economic indicators, contrasting the predictive performance in the time-series dimension with the cross-sectional accuracy. Chapter 2 uses multiple NTL datasets, with sensors of different vintages, to study GDP at the county/district level. Chapter 3 shows how blurring and top-coding in popular DMSP data distort apparent patterns of spatial inequality, and alter trends in this indicator. Chapter 4 examines use of gridded population estimates (some based on NTL data for spatialization) to proxy for counts of the resident population from the census, at three levels of spatial aggregation but with a particular focus at the county-level because the gridded data are increasingly used for very finely-scaled studies.

While the first part of the thesis has a comprehensive evaluation of different remote sensing data, and so the focus is on using China's administrative and census data as a benchmark, in the second part of the thesis these data are used to supplement the traditional data, to study a set of economic issues, covering regional policy, infrastructure and agglomeration, and land use and migration. Chapter 5 considers the impact of a popular regional development intervention in China—the administrative upgrading of counties to cities—using an estimation framework that allows for spatial spillovers, and using both GDP and NTL data to measure economic activity. A key driver of spatial spillovers is agglomeration effects, and transport infrastructure can facilitate development of these, so in Chapter 6 I use spatial econometric models for a panel of almost 2500 county-level units observed from 2012 to 2019, to study the impact of getting connected to the High-Speed Rail network, using both GDP and NTL data to measure economic activity. Finally, the data on resident population from three censuses are supplemented with remote sensing data on land cover to study the city size distribution in terms of people and land in Chapter 7 to examine the question of whether China's development of urban areas is occurring in the places where people are moving. The following paragraphs give a more detailed introduction to each chapter.

Chapter 2 examines the question of how good are nighttime lights data for studying differences in economic activity between areas and for studying changes in economic activity within areas, for 2657 county-level units in China, observed annually from 2012 to 2019. This goes beyond most existing studies that are at the national or regional level, and relies on an advantage of using data for China which is one of the few counties with GDP data down to the county-level. The estimation framework used considers both "within" and "between" elasticities of GDP with respect to four different NTL datasets, where the elasticities correspond to the time-series and cross-sectional uses of these data. A key result in this chapter is that the predictive power is roughly 50-times greater when NTL data are used to study cross-sectional differences in economic activity than when it is predicting time-series changes in activity. Amongst the four sources of NTL data considered, non-nested tests favour the new Black Marble

data, especially when using both snow-free and snow-covered nights so as to capture activity over more of the year for cold parts of China. The weak relationships between changes in nighttime lights data and changes in economic activity suggest that the NTL data are not a universally useful proxy for all types of socio-economic studies.

While NTL data are especially used in economics as an indicator for the level of, and growth in, economic activity another branch of the literature uses these data to study spatial inequality. This growing reliance on NTL data is because knowledge about inequality in the developing world is limited by poor quality and incomparability of traditional economic activity data, especially for finely-scaled spatial units. The DMSP data are popular in this inequality research but the flaws in these data due to images being blurred and top-coded have been ignored; these flaws induce errors with spatially mean-reverting structure making the DMSP data particularly unsuitable for studies of spatial inequality. In Chapter 3, I demonstrate this bias by using the county-level GDP data for China (and also for the USA to corroborate the findings) to estimate trends in spatial inequality over the last two decades. While newer NTL sources, whose images are not subject to blurring and top-coding because of their better sensor resolution, show trends in spatial inequality that are similar to the trends with the benchmark GDP data, the DMSP data greatly understate spatial inequality and also suggest that there was a downward trend in inequality in China which is not shown by the benchmark data. Thus, the continued reliance on the DMSP night-time lights data, which is motivated by the limited availability of county-level GDP data in most countries, may introduce some bias into estimates of the trends in, and levels of, spatial inequality.

The growing reliance of socio-economic research on remote sensing data is much broader than just studies that directly use NTL data. A key development over the last two decades has been the growing use of gridded population estimates, which are typically at the one kilometre level or even finer scale, where remotely sensed land cover and night-time lights data are often used to 'spread' (spatialize) population counts for more aggregated units into these small grid cells. Chapter 4 examines the question of how accurate are gridded population estimates, for studying differences in resident

population between areas and for studying the temporal changes in population. In China, the resident population statistics can only be collected by the census, which organizes every ten years. This chapter uses the population census data (2000, 2010, and 2020) as the benchmark, and uses four popular gridded population products to examine at three levels of spatial aggregation (county/district, prefectural city, and province) if the gridded data are good predictors of actual population change. The gridded population data did far worse at predicting time-series changes than at predicting the differences in resident population cross-sectionally, mirroring a pattern seen in Chapter 2 where the NTL data were used to predict county-level GDP. It was especially for the most local level units (counties, county-level cities and districts) where the gridded data products were inaccurate proxies for the actual changes in local population.

In Chapter 5, my focus switches from evaluating the data to, instead, using remote sensing data to supplement traditional administrative data for examining a key regional development policy, China's top-down administrative upgrading of counties to cities (either as county-level cities or as districts). This upgrading often has important effects because cities can convert more land from agriculture to urban use (a key source of local government financing) and have greater prominence than counties. Therefore, several studies examine whether economic activity grows faster in upgraded counties compared to their peers that were not upgraded. Yet most of these studies treat spatial units as independent of surrounding areas, which ignores potential spillovers. By using a 20-year panel of almost 2500 county-level units, with both GDP and NTL data used to indicate changes in economic activity, I find that positive direct effects on GDP and luminosity of a county being upgraded are amplified through positive indirect effects, especially in eastern China where economic activity and population are most densely concentrated. These spillovers also mean that the functional scope of a city may extend beyond the administrative boundaries; a potential advantage of using remote sensing data for studying these effects is that they are not limited to administrative areas.

Carrying on with this theme, there must be particular channels through which spillovers occur, such as infrastructure that enhances the flow of people and ideas. In

this regard, China has developed the largest and fastest growing High Speed Rail (HSR) network within the last 15 years. Chapter 6 explores effects on local economic activity from this substantial increase in HSR connections, building on an existing literature where scholars have used various methods to study HSR impacts in different regions and have found mixed evidence on impacts. It is natural to think about these impacts as having spatial spillovers, although these are often ignored in applied studies. The other issue that is often ignored is the endogenous placement of the network. I use a set of spatial econometric models for a panel for almost 2500 county-level units to study effects of connecting to HSR networks over the 2012-19 period, allowing for spillovers and endogenous placement. This study period also lets me use the newer generation night-time lights data to provide an alternative to GDP as an indicator of growth in local economic activity. The results do not show positive impacts on local economic activity of connections to the HSR network in the 2012-19 period.

In Chapter 7, I use population counts from the 2000, 2010 and 2020 censuses, in conjunction with remote sensing land cover information on built-up area, to compare China's changing size distribution patterns for cities, in terms of both people and land. One of the most robust empirical facts about the relative size of cities is that they follow either Pareto's or Zipf's Law which gives negative relationships between logarithms of city size and city rank within a country (with a slope of minus one for Zipf's Law). A common finding is that China's city sizes are becoming more evenly distributed over time (thus, the absolute value of the Pareto exponent is rising and moving further from what Zipf's Law implies). However, the population that these studies use are counts of the local hukou registered population, which miss the flocking of millions of migrants into a small number of China's cities. Therefore, in this chapter I use census data which is the only comprehensive source that allows a consistent definition of urban residents, while the city area is extracted from remote sensing global land cover data, and multiple sources of night-time lights data. Over time, China's cities are becoming less even in terms of population (contrary to what prior studies using the hukou population claimed), but more even in terms of city built-up area. These divergent patterns in land and people

suggest that growth in the resident population of the large cities is not being assisted by fast enough built-up area expansion, while the conversion of non-urban land to urban uses for the less populous cities is too fast for their slow growth in resident numbers. This study contributes evidence for the Chinese government to allow scientific decision-making basis for creating a reasonable city size distribution system.

Chapter 8 contains the conclusions of the thesis and highlights the overall contribution of the six main chapters. The theme that runs through all of these chapters is the need to evaluate the performance of NTL data and gridded population estimates. While these can be useful as a proxy for local socio-economic development they are not suitable indicators to use for all research designs (especially for designs that rely on temporal changes to identify their parameters of interest) so researchers need to be cautious when using these data.

References

Asher, S., Lunt, T., Matsuura, R., and Novosad, P. (2021). Development research at high geographic resolution: An analysis of night-lights, firms, and poverty in India using the shrug open data platform. *The World Bank Economic Review*, 35(4), 845-871.

Bickenbach, F., Bode, E., Nunnenkamp, P., and Söder, M. (2016). Night lights and regional GDP. *Review of World Economics*, 152, 425-447.

Carleton, T., Jina, A., Delgado, M., Greenstone, M., Houser, T., Hsiang, S., ... and Zhang, A. T. (2022). Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *The Quarterly Journal of Economics*, 137(4), 2037-2105.

Clark, H., Pinkovskiy, M., and Sala-i-Martin, X. (2020). China's GDP growth may be understated. *China Economic Review*, 62, 101243.

Donaldson, D., and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4), 171-198.

Elvidge, C. D., Hsu, F. C., Baugh, K. E., and Ghosh, T. (2017). Lighting tracks transition in Eastern Europe. *Land-Cover and Land-Use Changes in Eastern Europe after the Collapse of the Soviet Union in 1991*, 35-56.

Feldmeyer, D., Meisch, C., Sauter, H., and Birkmann, J. (2020). Using OpenStreetMap data and machine learning to generate socio-economic indicators. *ISPRS International Journal of Geo-Information*, 9(9), 498.

Gibson, J., and Boe-Gibson, G. (2021). Nighttime lights and county-level economic acitivity in the United States: 2001 to 2019. *Remote Sensing*. 13, 2741.

Gibson, J., Olivia, S., Boe-Gibson, G., and Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, 149, 102602.

Gibson, J., Jiang, Y., Zhang, X., and Boe-Gibson, G. (2024). Are disaster impact estimates distorted by errors in popular night-time lights data? *Economics of Disasters and Climate Change*, 1-26.

Han, G., Zhou, T., Sun, Y., and Zhu, S. (2022). The relationship between night-time light and socioeconomic factors in China and India. *Plos One*, 17(1), e0262503.

He, X. (2023). Dams, cropland productivity, and economic development in China. *China Economic Review*, 81, 102046.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, 102(2), 994-1028.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2018). The global distribution of economic activity: Nature, history, and the role of trade. *The Quarterly Journal of Economics*, 133(1), 357-406.

Holz, C. A. (2014). The quality of China's GDP statistics. *China Economic Review*, 30, 309-338.

Ji, X., Sun, J., Wang, Q., and Yuan, Q. (2019). Revealing energy over-consumption and pollutant over-emission behind GDP: a new multi-criteria sustainable measure. *Computational Economics*, 54, 1391-1421.

Kalkuhl, M., and Wenz, L. (2020). The impact of climate conditions on economic production. Evidence from a global panel of regions. *Journal of Environmental Economics and Management*, 103, 102360.

Kim, J., Kim, K., Park, S., and Sun, C. (2023). The economic costs of trade sanctions: Evidence from North Korea. *Journal of International Economics*, 145, 103813.

Lee Y. S. (2018). International isolation and regional inequality: Evidence from sanctions on North Korea. *Journal of Urban Economics*, 103, 34-51.

Li, C., and Gibson, J. (2013). Rising regional inequality in China: Fact or artifact? *World Development*, 47, 16-29.

Li, X., Zhou, Y., Zhao, M., and Zhao, X. (2020). A harmonized global nighttime light dataset 1992–2018. *Scientific Data*, 7(1), 168.

Liu, H., He, X., Bai, Y., Liu, X., Wu, Y., Zhao, Y., and Yang, H. (2021). Nightlight as a proxy of economic indicators: Fine-grained GDP inference around Chinese mainland via attention-augmented CNN from daytime satellite imagery. *Remote Sensing*, 13(11), 2067.

Méndez, E., and Van Patten, D. (2022). Multinationals, monopsony, and local development: Evidence from the United Fruit Company. *Econometrica*, 90(6), 2685-2721.

Michieka, N. M., and Fletcher, J. J. (2012). An investigation of the role of China's urban population on coal consumption. *Energy Policy*, 48, 668-676.

Miller, S. D., Mills, S. P., Elvidge, C. D., Lindsey, D. T., Lee, T. F., and Hawkins, J. D. (2012). Suomi satellite brings to light a unique frontier of nighttime

environmental sensing capabilities. *Proceedings of the National Academy of Sciences*, 109(39), 15706-15711.

Moscona, J., Nunn, N., and Robinson, J. A. (2020). Segmentary lineage organization and conflict in Sub-Saharan Africa. *Econometrica*, 88(5), 1999-2036.

Pagaduan, J. A. (2022). Do higher quality nighttime lights and net primary productivity predict subnational GDP in developing countries? Evidence from the Philippines. *Asian Economic Journal*, 36(3), 288-317.

Pérez-Sindín, X. S., Chen, T. H. K., and Prishchepov, A. V. (2021). Are night-time lights a good proxy of economic activity in rural areas in middle and low-income countries? Examining the empirical evidence from Colombia. *Remote Sensing Applications: Society and Environment*, 24, 100647.

Perkins, D. H., and Rawski, T. G. (2008). Forecasting China's economic growth to 2025. China's great economic transformation, 2008.

Pinkovskiy, M., and Sala-i-Martin, X. (2016a). Lights, camera… income! Illuminating the national accounts-household surveys debate. *The Quarterly Journal of Economics*, 131(2), 579-631.

Pinkovskiy, M., and Sala-i-Martin, X. (2016b). Newer need not be better: Evaluating the Penn World Tables and the World Development Indicators using nighttime lights (No. w22216). *National Bureau of Economic Research*.

Puente-Ajovín, M., Sanso-Navarro, M., and Vera-Cabello, M. (2024). Comparing city size distributions: Gridded population versus nighttime lights. *Journal of Regional Science*, 64, 1323–1358.

Qiang, Q., and Jian, C. (2020). Natural resource endowment, institutional quality and China's regional economic growth. *Resources Policy*, 66, 101644.

Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., ... and Masuoka, E. J. (2018). NASA's Black Marble nighttime lights product suite. *Remote Sensing of Environment*, 210, 113-143.

Son, S. H., and Cho, J. (2021). Are economic sanctions against North Korea effective? Assessing nighttime light in 25 major cities. *Pacific Affairs*, 94(3), 464-492.

Zhao, N., Currit, N., and Samson, E. (2011). Net primary production and gross domestic product in China derived from satellite imagery. *Ecological Economics*, 70(5), 921-928.

Zhong, X., Su, Z., Zhang, G., Chen, Z., Meng, Y., Li, D., and Liu, Y. (2019). Analysis and reduction of solar stray light in the nighttime imaging camera of Luojia-1 satellite. *Sensors*, 19(5), 1130

Zhang, G., Guo, X., Li, D., and Jiang, B. (2019). Evaluating the potential of LJ1-01 nighttime light data for modeling socio-economic parameters. *Sensors*, 19(6), 1465.

Chapter 2: Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019

Xiaoxuan Zhang and John Gibson

This paper has been published in the *Remote Sensing*, (2022), 14(5), 1282.

remote sensing

MDPI

Article

Using Multi-Source Nighttime Lights Data to Proxy for County-Level Economic Activity in China from 2012 to 2019

Xiaoxuan Zhang and John Gibson *

Department of Economics, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand; zhangxx24_simlab@163.com

* Correspondence: jkgibson@waikato.ac.nz; Tel.: +64-7-838-4289

Abstract: The use of nighttime lights (NTL) data to proxy for local economic activity is well established in remote sensing and other disciplines. Validation studies comparing NTL data with traditional economic indicators, such as Gross Domestic Product (GDP), underpin this usage in applied studies. Yet the most widely cited validation studies do not use the latest NTL data products, may not distinguish between time-series and cross-sectional uses of NTL data, and usually are for aggregated units, such as nation-states or the first sub-national level, yet applied studies increasingly focus on smaller and lower-level spatial units. To provide more updated and disaggregated validation results, this study examines relationships between GDP and NTL data for 2657 county-level units in China, observed each year from 2012 to 2019. The NTL data used were from three sources: the Defense Meteorological Satellite Program (DMSP), whose time series was recently extended to 2019; and two sets of Visible Infrared Imaging Radiometer Suite (VIIRS) data products. The first set of VIIRS products is the recently released version 2 (V.2 VNL) annual composites, and the second is the NASA Black Marble annual composites. Contrasts were made between cross-sectional predictions for GDP differences between areas and time-series predictions of economic activity changes over time, and also considered different levels of spatial aggregation.

Keywords: VIIRS; Black Marble; DMSP; GDP; nighttime lights; cross-sectional; time-series

Citation: Zhang, X.; Gibson, J. Using Multi-Source Nighttime Lights Data to Proxy for County-Level Economic Activity in China from 2012 to 2019. *Remote Sens.* **2022**, *14*, 1282. https://doi.org/10.3390/rs14051282

Academic Editor: Bailang Yu

Received: 22 January 2022
Accepted: 1 March 2022
Published: 5 March 2022

1. Introduction

The use of nighttime lights (NTL) data to proxy for economic activity is well-established in remote sensing and other disciplines [1–10]. This proxy enables research when traditional economic activity data, such as Gross Domestic Product (GDP), are either absent or are not trusted because of concerns about either measurement error or manipulation [11–13]. Potential advantages of NTL-based economic activity estimates are their timeliness, lower cost, comparability between countries irrespective of statistical capacity, and availability for spatial units below the level at which GDP data are reported.

Early studies focused on cross-sectional comparisons of nations and sub-national regions, but more recent studies use NTL data to track changes in economic activity. These fluctuations may be due to natural disasters, such as earthquakes, floods, hurricanes, and tsunami [14–18]; public health crises, such as COVID-19 [19–21]; or to various economic policies [22,23]. This use of NTL as a proxy for changes in local economic activity, plus ongoing cross-sectional use as a proxy for variation in economic performance [24], raises the question of how good are the NTL data for studying differences in economic activity between areas and for studying the temporal changes in economic activity within areas.

Extant evidence is that changes in nighttime lights data poorly predict temporal changes in economic variables despite NTL data being good cross-sectional predictors of differences in economic activity across space within these same studies [25,26]. These studies with negative findings on the performance of NTL data as a proxy for temporal changes in economic activity use Defense Meteorological Satellite Program (DMSP) data

and either contrast the shares of variation in economic indicators explained cross-sectionally and temporally, or else contrast cross-sectional and time-series elasticities of economic variables with respect to NTL data. Another approach with DMSP data estimates optimal weights for mixtures of NTL data and conventional economic statistics to best proxy for the true but unknown levels of, and changes in, economic activity [4,27]. Given that the DMSP data were not originally gathered for research purposes, in contrast to newer research-based products from the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi-NPP satellite [28], a relevant question is whether poor performance of NTL data for studying temporal changes in economic activity is a feature just of DMSP data or also applies to VIIRS data.

Some evidence suggests that poor predictive performance for time-series changes in local GDP also extends to the newer VIIRS data. In the United States, the share of variation in annual changes in GDP predicted by annual changes in VIIRS data is just 2% at the metropolitan level [29] or 12% at the county-level [30], yet VIIRS data predict almost 90% of the cross-sectional variation in GDP. However, the United States is a mature economy that has been highly urbanized for many decades, and so the scope for nighttime lights to change in response to economic fluctuations may be relatively low. In contrast, countries where NTL data are most useful for research tend to be undergoing rapid urbanization and electrification as part of their development process and so may have stronger relationships between changes in local economic activity and changes in NTL data. Therefore, this study examines relationships between GDP and NTL data for 2657 county-level units in China, observed each year from 2012 to 2019. The NTL data used are from three sources: DMSP stable lights annual composites [31], whose time-series was recently extended to 2019; and two sets of VIIRS data products—newly released version 2 (V.2 VNL) annual composites [32], and the NASA Black Marble annual composites [33]. Contrasts are made between cross-sectional predictions for GDP differences between counties and time-series predictions of economic activity changes over time. Recent monthly relationships between China's electricity consumption and the various NTL data products are also examined.

In order for the results to provide informative comparisons with prior literature, the estimation framework is deliberately based on that from a study on the United States [30], where panel data provide "within" and "between" elasticities of GDP with respect to NTL data that correspond to the time-series and cross-sectional uses of these data. This framework is also based on the approach used in the most widely cited study of the relationship between GDP and NTL data [5]. If a different estimation framework was used, it would be difficult to draw broader lessons because differences in results between studies may be due to the different methods. This matters particularly for validation studies, such as the current study, which provide elasticity estimates for applied studies to use when converting estimated effects on NTL data into equivalent effects on economic activity; if the validation studies use heterogeneous methods, it reduces the scope for this transferability.

2. Materials and Methods

2.1. Related Literature on NTL Validation Studies

Validation studies set out to estimate the nature of relationships between NTL data and benchmark traditional economic activity data. These studies can provide support for use of NTL data as a proxy in times and places where traditional economic activity data, such as GDP, are unavailable. Validation studies may be stand-alone [4,5,29,30,34,35], or else may be part of a broader study where the validation exercise aims to justify the use of NTL data in the main analysis [36,37]. The most widely cited study (with 2120 Google Scholar citations as of January 2022) used national-level DMSP data and GDP data for 188 countries from 1992 to 2008 to estimate an elasticity of 0.3 for temporal changes in GDP with respect to changes in the NTL data [5]. A similar highly cited study (with 730 Google Scholar citations) used DMSP data for 1500 regions (mostly at the first sub-national level) from 82 countries from 1992 to 2009 and estimated that the elasticity for temporal changes in GDP with respect to temporal changes in DMSP data was 0.4 [36].

Although widely cited, these studies may be less informative in the present era due to two factors: much of the literature has moved on from DMSP and now uses VIIRS data, and applied studies using NTL data to proxy for economic activity increasingly use spatial units that are lower in the sub-national administrative hierarchy, such as districts (or other second-level units), counties (or other similar third-level units), and even individual pixels. An important interpretation issue is posed by the fact that the widely cited validation studies are typically based on DMSP data for spatially aggregated units. Meanwhile, the applied studies that cite the validation studies and that may use elasticity estimates reported by the validation studies in order to convert effects on NTL data into economic activity terms are increasingly based on lower level spatial units and increasingly are using VIIRS data. The elasticity estimates from DMSP data may not be applicable to results estimated with VIIRS data, and, likewise, it may be incorrect to transfer elasticity estimates from spatially aggregated validation studies into spatially disaggregated applied studies. A problem with such a transfer of the elasticity estimates is that flaws in DMSP data, such as geo-location errors and blurring [38,39], seem to make predictive performance for GDP worse when using lower-level spatial units than when using aggregated spatial units [35].

The validation studies using VIIRS data to examine time-series and cross-sectional predictions of GDP have concentrated on the United States. Almost 90% of cross-sectional variation in metropolitan-level GDP was predicted by variation in the sum of VIIRS radiances, but just 2% of annual changes in GDP were predicted by the annual changes in the NTL data, albeit with only a three-year time series [29]. A longer time-series from 2014 to 2019, estimated at the county-level, found 12% of temporal variation in GDP predicted by temporal variation in the sum of VIIRS nighttime lights, using data products that were masked to remove outliers, while the NTL data were able to predict 86% of the cross-sectional variation in economic activity [30]. In terms of elasticities of county-level GDP with respect to NTL data, the cross-sectional estimates averaged 1.0 while time-series elasticities for annual changes were just one-tenth as large, averaging 0.1. Thus, not only are relationships between temporal changes in nighttime lights and changes in economic activity far weaker than are the cross-sectional relationships in the levels, the elasticities found with VIIRS data are far lower than the values of 0.3 or 0.4 reported in widely cited validation studies using DMSP data for the national or first sub-national level [5,36].

This difference between the elasticities estimated from temporal changes and those that are estimated from cross-sectional differences affects the interpretation of several recent applied studies. Applied studies estimate the effects on nighttime lights of a particular shock or intervention, such as flooding [15], tsunami [17], or social transfers [40]. In order to convert these effects into equivalent effects on economic activity, they use estimates from the literature on the elasticity of GDP with respect to NTL data (because these applied studies lack GDP data for their own setting). While a popular choice for the elasticity used is 0.3, based on annual changes in national GDP and national DMSP lights [5], in some cases, they use much higher elasticities and so get larger calculated effects on economic activity of the shock or intervention that they study. Some of these larger elasticity estimates are from cross-sectional studies when it is the far smaller time-series elasticities that are appropriate inputs into their calculation [26].

While many studies have linked NTL data to sub-national GDP data for China, these are typically either correlational [7] or involve regressions in a framework that does not allow the time-series elasticities for temporal changes to be estimated separately from the cross-sectional elasticities for differences across space with, instead, some mixture of the two types of parameters estimated [41]. Two studies using the framework that allows the time-series elasticities to be estimated separately, with DMSP data at the prefectural level (the second sub-national level), find elasticities of 0.25 for changes in GDP with respect to changes in DMSP stable lights over 1992–2005 [37], and an elasticity of 0.30 (for urban parts of prefectures and 0.19 for rural parts), over 2000–12 [42]. Thus, the prior results for China provide elasticities that are similar to the elasticities from the highly cited worldwide studies [5,36], and these serve as a point of comparison for the new results reported below.

2.2. Data

We use four types of data to test the relationships between China's nighttime lights and GDP. The first is annual information on the total GDP (in billions of Chinese Yuan, CNY) for each county-level unit from 2012 to 2019, which is obtained from three types of publications: the annual editions of the China Statistical Yearbook (county-level) (also named in Chinese as *Zhongguo Xianyu Tongji Nianjian*), annual editions of the China City Statistical Yearbook (also known in Chinese as *Zhongguo Chengshi Tongji Nianjian*), and annual editions of the Statistical Yearbook for each city (for example, the Beijing Statistical Yearbook) [43–45]. Each edition reports on GDP the previous year, so we use the 2013 to 2020 editions to obtain 2012 to 2019 GDP data. As of 2020, the county-level units in China were comprised of the following types: 965 municipal districts (that are highly urbanized), 387 county-level cities, 1323 counties, 117 autonomous counties, 49 banners, 3 autonomous banners, 1 special zone, and 1 forestry zone, and it is this diversity of types of units that necessitates using several different types of statistical yearbooks.

From a total of 2846 county-level units under the current administrative geography, we were able to create a balanced panel of GDP data for 2657 units (about 94% of the total) in each year from 2012 to 2019. The provinces with the largest number of spatial units for which we were not able to obtain the county-level GDP data were Tibet, Heilongjiang, and Xinjiang (Figure 1). Usually, the other occurrences of unavailable county-level GDP data were at most one or two per province, although for Qinghai, the two units with missing data are physically large and so occupy a prominent position on the map. Given the data that we have, our results can be thought of as representing almost 95% of China.

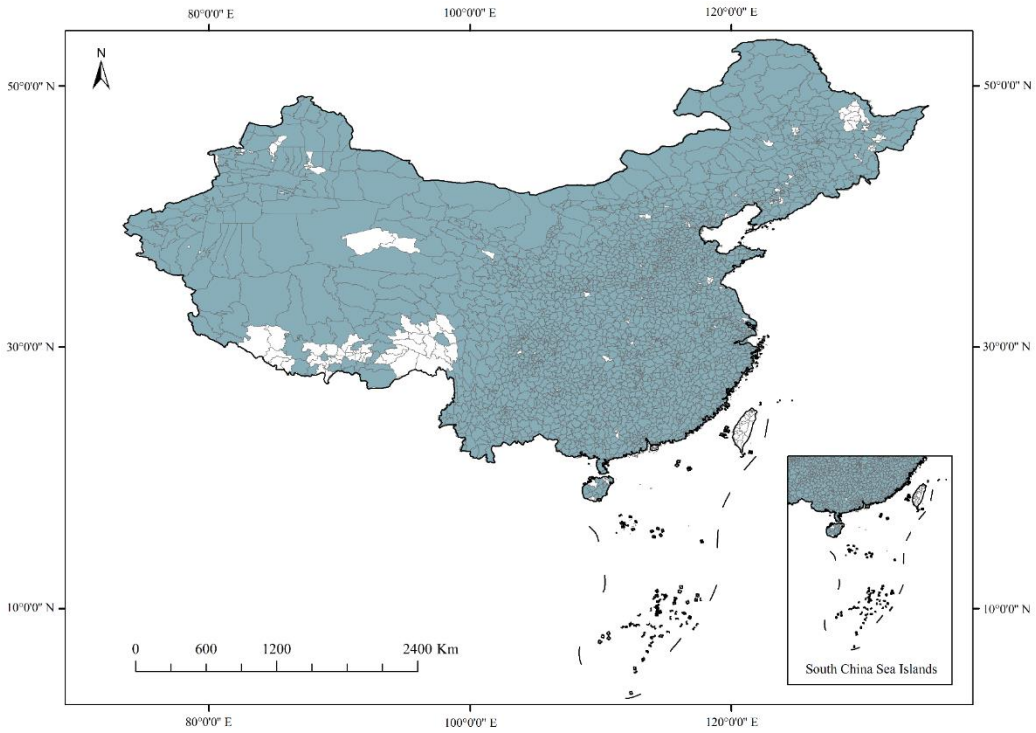


Figure 1. County-level spatial units in China (no GDP data if blank).

The second data source was annual composites from Defense Meteorological Satellite Program (DMSP) satellites F15 and F18. The stable lights product provides 6-bit digital numbers (DN) ranging from 0 to 63 for each 30 arc-second output pixel. Ephemeral lights, such as from fires and flaring, are removed, and processing excludes (at the pixel level) images for nights affected by clouds, moonlight, sunlight, and other glare. The usual stable lights product time-series ended in 2013 [31], but that has now been extended with images from F15 for 2014–19 [46]. The time of night that lighting is observed differs

between the extended time-series using F15 (ca. 3.30 a.m.) and the original series with F18 (ca. 8.30 p.m.), and there may also be sensor differences, as F15 has been in orbit for longer, so more depreciation may have affected the equipment. The estimation framework controls for year-by-year differences using time fixed effects (separate intercepts for each year). Given that there is no overlap in time between the F15 and F18 time series, with one ending in 2013 and the other starting in 2014, these fixed effects will adjust for satellite-specific measurement differences (and in this way provide a form of inter-calibration, especially because all variables in the regressions outlined below are in logarithms). Moreover, if the 2014 year fixed effect is interacted with either the DN value from DMSP or the radiance values from the other NTL data sources, there is no statistically significant interaction effect, which supports the sole use of the year fixed effect as a straightforward intercept shifter to synthesize the values from the two time-series.

The third data source was version 2 VIIRS nighttime lights (V.2 VNL) annual composites [32]. The V.2 VNL are produced from monthly cloud-free radiance averages, with initial filtering to remove extraneous features, such as fires and aurora, before the resulting rough annual composites are subjected to outlier removal procedures. To isolate lit grid cells from background, a data range threshold is based on a multiyear maximum median and a multiyear percent cloud-cover grid. In contrast, version 1 annual composites [47] used year-specific thresholds, which are less suitable for change detection (given that different average radiance values between years could be due to on-the-ground changes or to differences in the thresholds). The data are in units of nano Watts per square centimeter per steradian ($nW/cm^2/sr$) reported on a 15 arc-second output grid. We mainly use the average masked data product, which had the highest predictive power for GDP amongst all of the V.2 VNL data products in a study of the United States [30].

The fourth data source was NASA Black Marble VIIRS annual composites [33]. These use a bi-directional reflectance distributional function (BRDF) to remove the effects of extraneous artifacts, and processing also removes cloud-contaminated pixels. The data products are corrected for atmospheric, terrain, vegetation, snow, lunar, and stray light effects on VIIRS Day/Night Band (DNB) radiances and are calibrated across time and validated against ground measurements. The data are in units of nano Watts per square centimeter per steradian ($nW/cm^2/sr$) reported with 16-bit precision on a 15 arc-second output grid. We use the all-angle composites, which are built from the greatest number of nights per year, compared to either the near-nadir (view zenith angle of 0–20 degrees) composites or the off-nadir (view zenith angle of 40–60 degrees) composites.

For the three NTL data products described above—DMSP DN values, V.2 VNL masked average radiances, and radiance values from the various Black Marble products—the sum of lights in each cross-sectional unit (typically a county-level unit) in each year is used as the measure of luminosity. This corresponds to annual GDP; that is, the sum of all measured economic activity within the cross-sectional unit in the particular year.

Even though we use annual composites, of either DN values for DMSP or radiances for V.2 VNL and for Black Marble, the available files also provide a tally, for each output grid cell, of the number of nights per year that were used to form the annual composite. For example, this tally of the number of nights is available in the cf_cvg.tif file associated with the DMSP stable lights composite. These tallies of the number of nights provide an indication of the variation in the completeness of temporal coverage for each of the NTL data products in each grid cell in each year and also provide a diagnostic tool for helping with the choice of particular radiance composite data products to use.

A feature of the Black Marble annual composites is the separation into snow-free and snow-covered periods. For parts of northern Xinjiang, northern Heilongjiang, and from the Tibetan plateau through Qinghai and Gansu, one-fifth or more of the nights in the year available for forming annual composites are snow-covered (Figure 2). While only using the all-angle snow-free composite in southern and eastern China should not distort the analysis of annual data, given that almost no-nights are snow-covered in these areas, for other parts of China, the omission of snow-covered nights may matter. Overall for China,

ten percent of nights each year that meet quality control standards for inclusion in annual composites are snow-covered. Therefore, in addition to using existing snow-free annual composites, we created a weighted average of the snow-free and snow-covered composites. The weights are the number of snow-free days and of snow-covered days per year for each pixel. The weighted average should capture variation across the entire year in the same manner that annual GDP aggregates over all of the seasons.

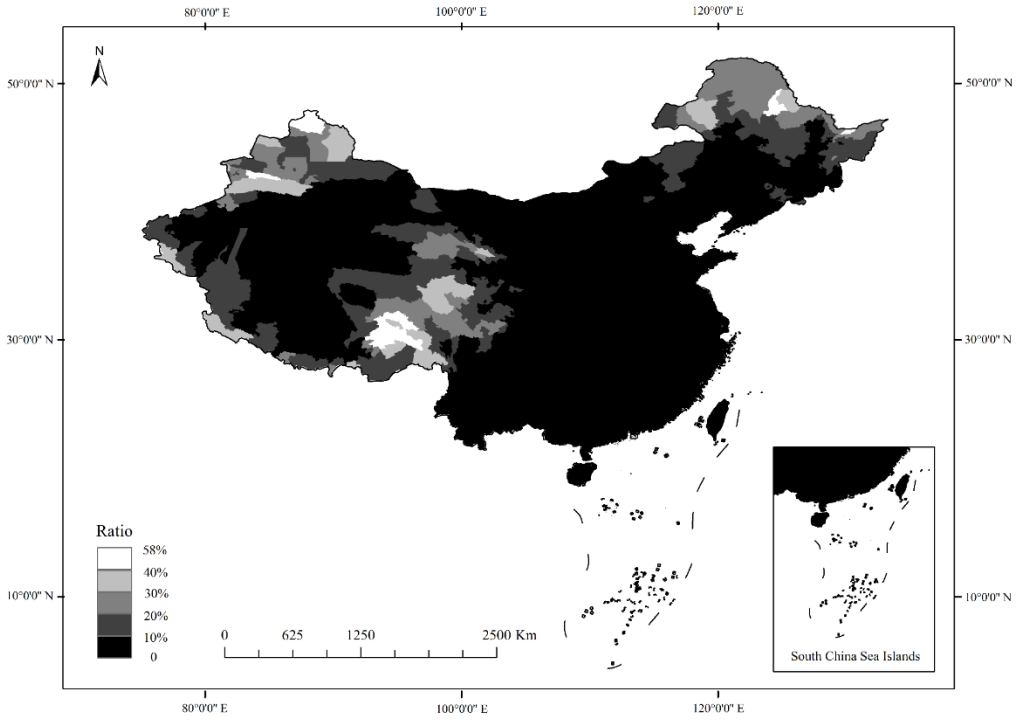


Figure 2. County-level average ratios of the number of snow-covered nights to the total number of nights with Black Marble data used for annual composites.

The average number of nights per year (averaged across all pixels in China) that are used to form the annual composites varies from around 50 nights per year for the DMSP stable lights composites to about 170 nights per year for the Black Marble composites that use weighted averages of the snow-free and snow-covered nights (Figure 3). On average, using the weighted average of snow-free and snow-covered composites gives about ten percent more nights per year than just relying on the snow-free composite, although as seen in Figure 2, the China-wide average disguises a lot of regional heterogeneity. Even though the V.2 VNL composites and the Black Marble composites rely on the same VIIRS source data, the different processing approaches result in 20–50% more nights being used in the formation of the Black Marble composites, particularly in the early years of the study period. All else the same, if more nights of the year are used to form an annual composite for an NTL data product, it should provide a more precise estimate of true luminosity. Moreover, to the extent that GDP covers economic activity over the entire course of the year, there may be a closer relationship between NTL data and GDP data for NTL products that are based on more nights of the year.

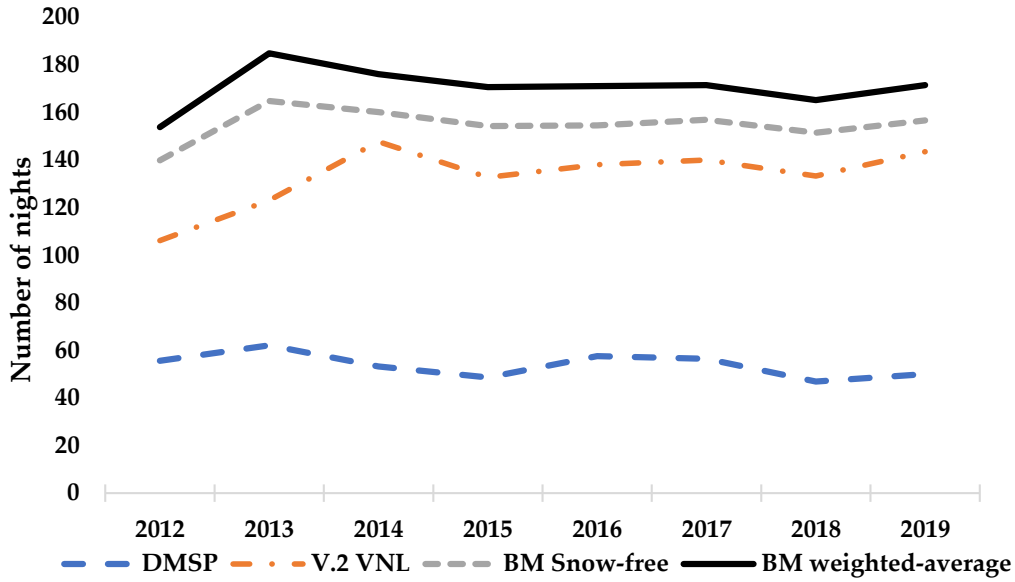


Figure 3. Average number of nights per year included in NTL composites for China.

The attributes of the three NTL data sources are summarized in Table 1. The top part of the table covers satellite and sensor attributes, which are the same for both V.2 VNL and Black Marble (so only one column is used), and the bottom part of the table covers the attributes of the various data products.

Table 1. Attributes of the various NTL data sources and data products.

	DMSP	V.2 VNL	Black Marble
Satellite/Sensor Attributes			
Operator	US DoD	NASA/NOAA	
Available years	1992–2019	2012–2020	
Spectral band	0.5–0.9 µm	0.5–0.9 µm	
Orbit type and altitude	Polar, 830 km	Polar, 830 km	
Spatial resolution at nadir	2.7 km	742 m	
Scan width	3000 km	3000 km	
Revisit time	12 h	12 h	
Pixel saturation	Saturated	Not saturated	
On-board calibration	No	Yes	
Data Products			
Creator of annual composites	EOG	EOG	NASA
Spatial resolution	30 arc second	15 arc second	15 arc second
Tiled	No	No	Yes, 648 tiles
Masking of ephemeral light sources	No	Yes	Yes
Stray-light correction	No	Yes, from 2014	Yes
User control over angle of detection	No	No	Yes
Treatment of snow	No	No	Yes

Note: DoD is Department of Defense, EOG is Earth Observation Group.

2.3. Estimation Framework

Our estimation framework relies on the following equation (where ln is logarithm):

$$\ln(GDP)_{it} = \alpha + \beta \ln(sum\ of\ lights)_{it} + \mu_i + \varphi_t + \varepsilon_{it} \quad (1)$$

where the i indexes cross-sectional units; the t indexes years; the μ_i are fixed effects for each cross-sectional unit; the φ_t are fixed effects for each year; and ε_{it} is the disturbance term. The parameter of greatest interest is β, the elasticity of GDP with respect to nighttime lights. The elasticity is a unit-free measure showing by what percentage the left-hand side

variable changes for each percentage change in the right-hand side variable. The fixed effects capture the influence of unobserved time-invariant features of each cross-sectional unit and the influence of spatially-invariant features of each time period.

This framework applied to panel data allows two different types of elasticities to be estimated. If Equation (1) is averaged over time, we get:

$$\overline{\ln(GDP)}_i = \beta^B \overline{\ln(sum\ of\ lights)}_i + \mu_i + \bar{\varepsilon}_i \quad (2)$$

where time-averaged values of (log) GDP and NTL are used in a cross-sectional regression estimated with Ordinary Least Squares, to yield β^B, which is known as the between estimator [48]. The elasticity in Equation (2) represents the effects of long-run differences between the cross-sectional units [49]. The second type of elasticity, β^W, is known as the within estimator, where the variation over time within each cross-sectional unit provides the basis for estimating:

$$\ln(GDP)_{it} - \overline{\ln(GDP)}_i = \beta^W \left[\ln(sum\ of\ lights)_{it} - \overline{\ln(sum\ of\ lights)}_i\right] + \varepsilon_{it} - \bar{\varepsilon}_i \quad (3)$$

Equation (3) is based on subtracting Equation (2) from Equation (1), and in doing so, it removes the effect of the unobservable fixed effects. Note that Equation (3) provides equivalent estimates to what would be obtained by estimating Equation (1) with separate intercepts for every cross-sectional unit. The between estimator β^B, enables examination of the cross-sectional GDP differences between areas, while the within estimator β^W, allows examination of the annual time-series fluctuations in GDP within areas.

If Equation (1) is estimated without fixed effects, or equivalently if time-demeaned data are not used, the resulting estimator for β will be a weighted average of the between estimator and the within estimator [50]. In other words, the resulting parameter will not identify either the long-run cross-sectional relationship, nor the effect of temporal changes. Thus, estimating Equation (1) without fixed effects will upwardly bias the estimator of β if it is being interpreted as a measure of the relationship between temporal changes in economic activity and temporal changes in NTL data. This upward bias is because the μ_i should positively correlate with both GDP and nighttime lights due to the fact that topography and other environmental factors that limit urbanization and electrification in particular places (and therefore limit the amount of nighttime light emitted) would also be expected to limit the value of GDP produced in such places.

This upward bias affects the validity of using NTL data to estimate impacts on the local economic activity of policy interventions or shocks (more generally, of 'treatments'). The product of two relationships: $(\partial GDP/\partial lights)\cdot(\partial lights/\partial treatment)$ is the basis of this approach to using NTL data for impact evaluation. In the settings of interest, typically, the $\partial GDP/\partial lights$ relationship is not estimated because there are no GDP data (as any available and trustworthy GDP data would already be used for the evaluation). Therefore, researchers may combine their estimate of the $(\partial lights/\partial treatment)$ effect with a $\partial GDP/\partial lights$ term that is taken from elsewhere, such as from a validation study, so as to allow interpretation of the effect they find in economic activity terms [26]. Therefore, if there is an upward bias in the $\partial GDP/\partial lights$ term, as would occur if the appropriate within estimator of the elasticity is not used, then the effect of the treatment on economic activity is likely to be overstated. Consequently, if the relationship between changes in GDP and changes in NTL data is very weak, in the sense of having a very small elasticity, it is hard to see how estimates of the $(\partial lights/\partial treatment)$ effect are informative about how the studied treatment impacts economic activity, which is what policymakers are likely to care about, rather than the impacts on nighttime lights, *per se*.

3. Results

3.1. Descriptive Statistics and Correlations

The means and standard deviations of GDP and the four NTL variables, along with the correlations between all the variables, are reported in Table 2. All variables are in

logarithms to facilitate the estimation of elasticities. The weighted average of the snow-free and snow-covered all-angles annual composites from Black Marble has the highest correlation with GDP, at 0.74, with the V.2 VNL masked average having the second highest correlation. The lowest correlation with GDP is for the DMSP stable lights composite.

Table 2. Descriptive statistics and correlation coefficients.

			Matrix of Correlation Coefficients			
	Mean	**Std Dev**	**DMSP**	**V.2 VNL**	**BMsf**	**BMwa**
DMSP	8.526	1.176				
V.2 VNL	7.999	1.247	0.850			
BMsf	10.478	1.254	0.791	0.872		
BMwa	10.500	1.226	0.857	0.964	0.903	
GDP	4.950	1.266	0.635	0.730	0.687	0.740

Notes: Based on n = 21,256 observations (2657 county-level units each observed 8 times). All variables are in logarithms. BMsf and BMwa are Black Marble snow-free and weighted average.

3.2. Regression Results at County-Level and Prefectural-Level

The results of using the four sets of annual NTL variables (DMSP stable lights, VIIRS V.2 VNL masked average radiance, VIIRS Black Marble all-angles snow-free radiance, and the weighted average of the snow-free and snow-covered Black Marble all-angles radiance) for a panel of 2657 county-level units observed each year from 2012 to 2019 are reported in Table 3. The top panel has the within-estimator's results, based on time-series fluctuations in annual NTL and in annual GDP, and the bottom panel has the between-estimator results that rely on cross-sectional differences in NTL and GDP.

Table 3. Relationships between multi-source NTL and county-level GDP: within and between-estimator results.

Independent Variables and Summary Statistics	**Annual NTL Data Product**			
	DMSP Stable Lights	**V.2 VNL Masked Average**	**Black Marble Snow-Free**	**Black Marble Weighted Average**
	Within-estimator, for annual GDP changes within each county			
ln (sum of lights)	0.109 ***	0.067 ***	0.022 ***	0.124 ***
	(0.010)	(0.015)	(0.004)	(0.016)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
R-squared (Within)	0.014	0.004	0.002	0.011
	Between-estimator, for average GDP differences between counties			
ln (sum of lights)	0.786 ***	0.764 ***	0.834 ***	0.780 ***
	(0.015)	(0.013)	(0.014)	(0.013)
R-squared (Between)	0.500	0.560	0.587	0.579

Notes: Results are based on a strongly balanced panel of 2657 county-level units, observed each year from 2012 to 2019, giving n = 21,256 observations. Standard errors are in parentheses (clustered by county for the within-estimator results), *** $p < 0.01$.

In line with prior evidence [25,26,29,30], the NTL data are far more powerful cross-sectional predictors of differences in GDP between areas than they are predictors of time-series changes in economic activity within areas. At best, just over one percent of the variation in annual fluctuations in GDP within China's county-level units is predicted by the annual changes in the NTL data. It is when using the DMSP stable lights and the Black Marble weighted average that there are the highest within-R^2 values (Table 3). In contrast, the cross-county differences in the (time-averaged) sum of lights predict about 58% of cross-county variation in time-averaged GDP, using either of the two Black Marble products, while the between R^2 is a little lower, at 56% when using masked average radiance from

the V.2 VNL data, and at 50% when using the DMSP stable lights data. In other words, the predictive performance of NTL data for cross-sectional differences in county-level GDP is around 50 times as high as the predictive power of annual changes in NTL data for predicting annual changes in county-level GDP.

In terms of the GDP-lights elasticity, for the within estimator, this ranges from 0.02 to 0.12 for annual changes in the two Black Marble NTL variables, with elasticities of 0.07 for the V.2 VNL masked average radiance and 0.11 for the DMSP stable lights composite. A reasonable summary of these values would be that annual time-series changes in county-level GDP in China have an elasticity of about 0.1 with respect to time-series changes in NTL data. This is similar to the county-level results for the United States using V.2 VNL masked composites [30], despite the far faster increases in luminosity in China due to rapid urbanization and electrification. Specifically, for V.2 VNL masked average radiance over the 2014–19 period, which matches the period and data product used for the United States study [30], there was zero trend growth in the sum of lights at the county-level in the United States while in China there was a 9.9% trend annual growth rate in the sum of lights at the county-level (with a standard error of $\pm$0.6%). Evidently, the small value of the estimated elasticity of annual GDP changes with respect to annual NTL changes from a county-level study in a mature, fully electrified economy (the United States) also seems to hold in a far more dynamic economy undergoing rapid urbanization and rising electricity consumption.

The within-estimator's results in Table 3 are based on Equation (3), where differences between (log) GDP in each year and time-averaged GDP are regressed on differences between annual NTL data and time-averaged NTL data. The identification comes from annual changes, and in this regard, it matches many applied studies that use a panel of NTL data whose time dimension is annual frequency in order to have the most observations, including some before (for testing the "parallel trends" assumption of the difference-in-differences estimator) and some after the time of the particular shock or intervention that is the focus of their study. However, panel data can also be used in another way to remove the spatial fixed effect, μ_i whose inclusion could (upwardly) bias elasticity estimates, which is to use "long differences". If (log) GDP at the start of the time-series is subtracted from the value at the end of the time series, and likewise for the NTL data, the regression of the long-differenced GDP on the long-differenced NTL data removes the effect of μ_i, and gives an estimate of medium-run relationships between changes in GDP and changes in NTL. In the seminal study using country-level DMSP data, long-differences were calculated using the average of the first two years and of the last two years of the time series, and yielded an elasticity of 0.32, which was about one-seventh higher than what was estimated using the annual changes based on an Equation (3) framework [5].

For China, using long differences (based on two-year averages at the start and end of the time series) rather than using annual changes has a bigger impact on the estimated elasticity. For the four NTL data products in Table 3, the elasticity using long-differences is two-thirds higher, on average, than when using annual changes. Especially if using DMSP data, the long-difference's elasticity is higher, at 0.22 $\pm$ 0.02 compared to 0.11 $\pm$ 0.01 when the identification of the elasticity comes from annual changes. For the Black Marble weighted average of snow-free and snow-covered periods, the long-differences elasticity is about one-half higher, at 0.19 $\pm$ 0.02 compared to 0.12 $\pm$ 0.02 for annual changes.

The between-estimator's estimates of the elasticity in the bottom panel of Table 3 range from 0.76 to 0.83. A reasonable summary of these values would be that cross-sectional time-averaged county-level GDP in China has an elasticity of about 0.8 with respect to the time-averaged NTL data. In other words, there is an eight-fold difference between the cross-sectional elasticities and the time-series elasticities using annual data. While this is a little smaller than in the United States, where there was a ten-fold difference, the results in Table 3 add to the growing set of studies [25,26,29,30] that contrast the close relationships between economic activity differences across areas and NTL differences across

areas with the far weaker relationships between temporal changes in annual GDP and temporal changes in annual NTL data.

The four models in Table 3 are non-nested, in the sense that no model can be obtained from any of the other models by imposing parametric restrictions. An appropriate likelihood ratio test to compare such models uses the Kullback–Leibler Information Criterion (KLIC). Intuitively, the KLIC is the log-likelihood function under the hypothesis of the true model minus the log-likelihood function for the (potentially misspecified) model under the assumption of the true model. A model is better than a competitor if it is closer to the truth under the KLIC [50,51]. The within-estimator model for annual GDP changes using the annual changes in weighted averages of the snow-free and snow-covered Black Marble all-angles composites is significantly closer to the truth than the models using either annual changes in the snow-free composite alone or changes in the V.2 VNL masked average. Evidently, for a country such as China where some areas have a non-trivial number of snow-covered nights, taking account of radiance measurement on these nights, as well as on the snow-free nights, is helpful for modeling annual GDP (which, by definition, measures economic activity in snow-free and snow-covered periods).

Prior NTL validation studies for China that use an estimation framework, such as Equation (1), to obtain the elasticity of the change in GDP with respect to the change in DMSP stable lights, are based on prefectural-level data (the second sub-national level), rather than on data from the county-level, such as those shown in Table 3. To provide links to these prior findings and also to document the effects of spatial aggregation, we examined how the results change when the county-level GDP and NTL data are aggregated to 347 prefectural-level units. The results are in Table 4 and have the same format as in Table 3.

Table 4. Relationships between multi-source NTL and prefectural-level GDP: within and between-estimator's results.

Independent Variables and Summary Statistics	Annual NTL Data Product			
	DMSP Stable Lights	**V.2 VNL Masked Average**	**Black Marble Snow-Free**	**Black Marble Weighted Average**
	Within-estimator, for annual GDP changes within each prefecture			
ln (sum of lights)	0.275 ***	0.038	0.025 *	0.135 **
	(0.061)	(0.046)	(0.013)	(0.068)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
R-squared (Within)	0.014	0.000	0.001	0.002
	Between-estimator, for average GDP differences between prefectures			
ln (sum of lights)	0.965 ***	0.924 ***	0.995 ***	0.947 ***
	(0.054)	(0.047)	(0.047)	(0.046)
R-squared (Between)	0.479	0.527	0.562	0.552

Notes: Based on a strongly balanced panel of 347 prefectural-level units, observed each year from 2012 to 2019, giving a sample of $n = 2776$ observations. Standard errors are in parentheses (clustered at prefectural level for the within-estimator results), * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The level of spatial aggregation matters greatly when the DMSP stable lights data are used, as seen by the within estimator providing an elasticity of 0.28 for changes in GDP with respect to changes in NTL (Table 4). Yet when the same underlying data were used at the county level, the elasticity was just 0.11. In contrast to this more than doubling of the elasticity when DMSP data are used, if Black Marble NTL data are used, the within-estimator's elasticities are about the same as their values in Table 3 (at about one-tenth larger, which is well within one standard error of the Table 3 values). The sensitivity of the estimated GDP-lights elasticities to the level of spatial aggregation when using DMSP data has been noted previously [35]. A plausible source of this sensitivity to spatial aggregation is that DMSP data have spatially mean-reverting errors [30,34], and so when spatial units are aggregated, it inherently causes reversion to the mean, and so DMSP data

perform relatively better with aggregated data than with spatially disaggregated data. This sensitivity matters because existing validation studies for China using the Equation (1) framework with DMSP data are at the prefectural level [37,42], and at this level, these data appear to provide higher elasticities than are found at the county level (as seen from comparing Tables 3 and 4). Moreover, the elasticities from the prior validation studies with DMSP data greatly exceed what is found here at either prefectural or county levels when using VIIRS data.

The sensitivity of DMSP-derived within-estimator's elasticities to spatial aggregation also shows up when estimation is based on long differences. The elasticity of the long-differenced GDP with respect to the long-differenced DMSP data is 0.41 at the prefectural level, which is 90% higher than what it was at the county level. In contrast, when the Black Marble weighted average of snow-free and snow-covered periods is used, the elasticity with long differences is 0.22, which is less than one-fifth higher than the long-difference estimate with this data product at the county level.

In contrast to the differences by the data source that emerge from spatial aggregation when the within estimator is used, the prefectural-level results from the between estimator, in the bottom panel of Table 4, have kept the same patterns seen in Table 3. While there is a uniformly higher elasticity, ranging from 0.92 to 0.99, compared to the 0.76 to 0.83 range in Table 3 when estimates were at the county level, all four of the NTL data sources provide between-estimator elasticities that have increased at about the same rate. The predictive power of the prefectural-level NTL data in the between-estimator results is slightly lower than it was in the county-level results, but with the same patterns between the four data sources.

3.3. Variation in GDP-Lights Relationships by Population Density

The county-level units used for the results in Table 3 cover a range of population densities. The mean population density in municipal districts is almost 3150 residents per square kilometer (as of the 2020 census), while in counties it is 380 and in the 'irregular' units, such as autonomous counties, banners, forestry zones, and so forth, the mean density is just 80 residents per square kilometer. A density-dependent variation in the strength of the relationship between GDP and NTL data is noted in previous studies [34]. This effect may be from variation in dominant types of economic activity, with low-density regions specializing in agriculture, which has less need for nighttime lights than light-requiring retail and distribution activities typically found in dense urban areas. For example, in the United States counties with an above-median share of agriculture in GDP, the within estimator of the GDP-lights elasticity is less than one-third as high as for counties where agriculture is less important [30]. In China's Chongqing municipality, which covers an area as large as Austria and has a mix of both counties and municipal districts, the relationship between GDP and VIIRS V.1 VNL data seems to be due to industry and services, with variation in primary sector economic activity unrelated to the variation in nighttime lights [35].

In order to examine these density effects, we interacted Equations (2) and (3) with a variable measuring population density (from China's 2020 population census). For ease of interpretation, the density variable is transformed to standard deviation units so that the coefficients on the interaction variables show how elasticities of GDP with respect to NTL data vary for a one standard deviation increase in population density.

For all three VIIRS data products used in Table 5, the within estimator of the elasticity of changes in GDP with respect to changes in NTL is statistically significantly higher, the higher the population density is. This density-dependency is particularly apparent with the V.2 VNL data, and the weighted-average Black Marble data, where a standard deviation higher population density proportionately increases the elasticity by about one-half, raising it from 0.08 to 0.12 for V.2 VNL and from 0.14 to 0.20 for the Black Marble weighted average. In contrast, the elasticity estimated with DMSP data is unaffected by population density, which is a pattern that has been observed previously [34]. Features of the DMSP sensors

and data management, such as top-coding and blurring [28,39], mean that the apparent brightness of densely populated big cities according to DMSP data is little different to the brightness of smaller towns, creating a mean-reverting error [34]. This insensitivity to density-related differences in luminosity is illustrated by the lack of a significant interaction term in Table 5.

Table 5. Effects of population density on the relationship between NTL variables and county-level GDP.

Independent Variables and Summary Statistics	Annual NTL Data Product			
	DMSP Stable Lights	**V.2 VNL Masked Average**	**Black Marble Snow-Free**	**Black Marble Weighted Average**
	Within-estimator, for annual GDP changes within each county			
ln (sum of lights)	0.111 ***	0.078 ***	0.021 ***	0.139 ***
	(0.011)	(0.015)	(0.004)	(0.016)
ln (sum of lights) × density	0.005	0.041 ***	0.004 **	0.063 ***
	(0.011)	(0.011)	(0.002)	(0.018)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
R-squared (Within)	0.014	0.005	0.002	0.012
	Between-estimator, for average GDP differences between counties			
ln (sum of lights)	0.767 ***	0.736 ***	0.805 ***	0.753 ***
	(0.014)	(0.013)	(0.014)	(0.013)
ln (sum of lights) × density	0.016	0.014	0.009	0.013
	(0.020)	(0.019)	(0.022)	(0.019)
R-squared (Between)	0.558	0.574	0.599	0.592

Notes: Density is in standard deviation units. Between-estimator specification also includes density as a level term, but coefficients were always insignificantly different from zero, so are not reported (for within-estimator results, the density variable is time-invariant, and so the level term drops out). ** $p < 0.05$, *** $p < 0.01$. For other notes, see Table 3.

The bottom panel of Table 5 shows that the between-estimator elasticities, for cross-sectional differences between county-level units, do not appear to be sensitive to variation in population density. Therefore, particularly for time-series studies of fluctuations in economic activity, users should be aware that relationships between NTL data and traditional indicators, such as GDP, that are estimated in one setting, with a particular population density, may not translate easily into other settings where the population density (and underlying economic activities that correlate with density) differ. The possibility of quite disparate GDP-NTL relationships for particular places and types of activity should be borne in mind when NTL data are used as a proxy measure of economic activity.

3.4. Relationships between Changes in Electricity Consumption and Changes in NTL Data

The results in Tables 3–5 show that temporal changes in NTL data are only weakly predictive of changes in GDP for China's county-level and prefectural-level units, even though predictive power for cross-sectional differences in GDP across spatial units is far higher. A large applied literature uses changes in NTL data as a proxy for changes in economic activity, so this negative finding may attract some scrutiny. One possibility is that the benchmark GDP data for China are not a reliable standard for comparison, although if that were so, it is puzzling that the cross-sectional predictive power is so high. Moreover, the result that time-series annual changes in county-level GDP in China have an elasticity of about 0.1 with respect to time-series annual changes in NTL data is similar to the county-level results for the United States, and the use of GDP data for the United States is not controversial.

In order to provide some corroboration for the main results, this section reports results that use an alternative benchmark—monthly electricity consumption—for assessing the predictive power of NTL data for studying short-run changes in economic activity. Many previous studies have used NTL data to proxy for spatiotemporal dynamics of electricity

consumption [52–54]. Therefore, the use of electricity data should be a widely accepted benchmark in the case of any doubts about using China's local GDP data as a benchmark. Moreover, the period we studied, from January 2020 onwards, was marked by considerable fluctuations in China's electricity consumption and by considerable lighting changes due to closure and containment responses to the COVID-19 pandemic [19]. Consequently, the use of monthly electricity data for China should be a good test of the predictive power of NTL data for modeling temporal changes in economic activity.

For the period from January 2020 to July 2021, national electricity consumption in China shows large monthly fluctuations (Figure 4). In February 2020, electricity demand was 28% lower than in the previous month, and in February 2021, it was 33% lower; in turn, March each year showed rebounds with monthly increases in electricity demand of 22–23%. The majority (two-thirds) of China's electricity demand comes from the secondary (industrial) sector [55].

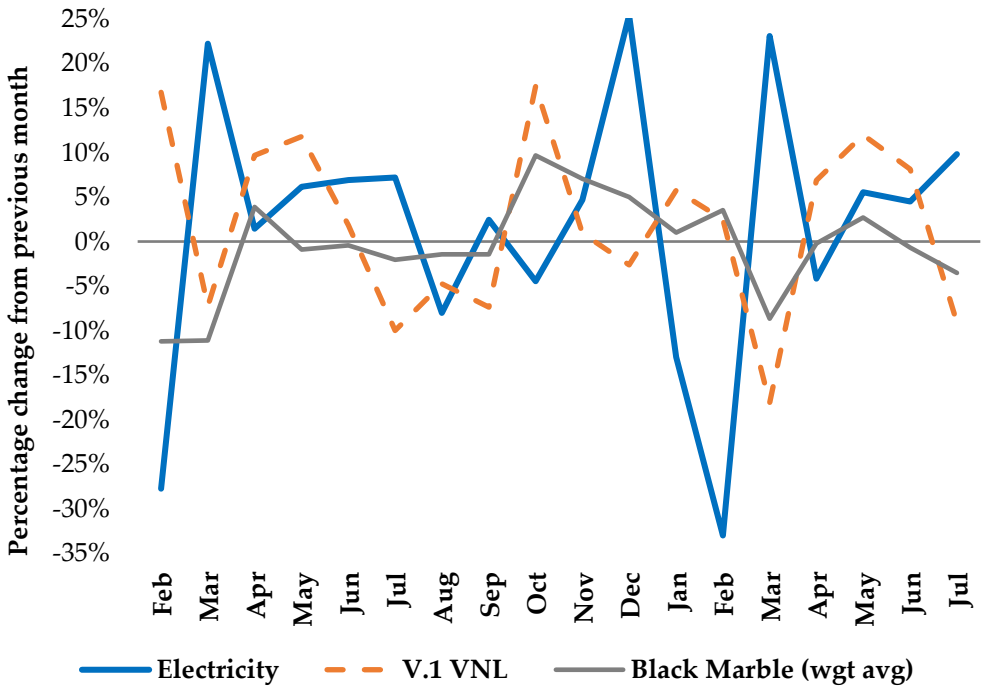


Figure 4. Monthly changes in China's electricity use and NTL composites: 2020–21.

A big contributor to these February declines and March rebounds in total electricity use is factory closures during the Chinese New Year and Spring Festival Golden Week national holidays, which started on 25 January in 2020 and on 12 February in 2021. Notably, despite the depressing effect on electricity demand of the holiday period in 2020 being lengthened by lockdowns instituted in some areas to control the spread of COVID-19 [19], the 'February-effect' of a decline in electricity use was even greater in 2021, which had no lockdown but which had the holiday period fully within the month rather than spread between late January and February. Thus, in the aggregate, the effect of seasonal holidays on China's electricity demand likely outweighs the effect of public health containment and closure measures. Even without the holiday-affected months of February and March, month-by-month changes in electricity consumption range from −13% to +25% and have an average absolute value of almost 8%.

The fluctuations in electricity consumption are not matched by monthly changes in the two NTL data products included in Figure 4: the monthly V.1 VNL cloud-free DNB composite and the weighted average of the Black Marble snow-free and snow-covered all-angles monthly composites. The V.1 VNL composite tends to show increases in NTL in months when total electricity use falls; the Pearson correlation coefficient between the

monthly changes in total electricity use and the monthly changes in NTL is −0.52. This negative relationship also holds for components of total electricity use, with correlations of −0.46 between changes in the V.1 VNL monthly composite and monthly changes in secondary sector electricity use and of −0.42 with monthly changes in residential electricity use. For the Black Marble monthly composite, the correlation coefficients are −0.11 with respect to changes in either total or secondary sector electricity use and −0.32 for changes in residential electricity use. If only the snow-free monthly composite is used, rather than the weighted average of snow-free and snow-covered composites, the correlations with changes in NTL data are hardly altered, at −0.08 for monthly changes in total electricity use and −0.34 for the monthly changes in residential electricity use.

4. Discussion

In this paper, we have used a comprehensive and updated set of annual nighttime lights composites from DMSP and two VIIRS products: V.2 VNL and Black Marble. In the core analysis, we examined relationships of NTL data with county-level economic activity in China over the 2012 to 2019 period, using an estimation framework that lets us separately examine cross-sectional relationships of luminosity with levels of economic activity and time-series changes in economic activity and in NTL data. We also link to prior results for China by examining relationships one step higher in the administrative hierarchy, at the prefectural level. Our aim in using multi-source NTL data, and multiple levels of the administrative hierarchy, stems from a concern that the most widely cited validation studies that assess NTL data as a proxy for economic activity are for DMSP data with aggregated spatial units, such as nations or the first sub-national level. However, applied studies increasingly use NTL data to proxy for economic activity at very local levels, such as the third sub-national level (e.g., counties) and below, and also increasingly use VIIRS data.

The previous evidence for China that uses the within estimator on DMSP data at the prefectural level, in order to estimate the elasticity of annual GDP changes with respect to annual changes in NTL data, provides elasticity estimates of about 0.25 [37,42]. These prior estimates used 1992–2005 and 2000–12 time series. Given that we estimate an elasticity of 0.28 using annual changes in DMSP data (with standard error of 0.06) at prefectural level (in Table 4), for the 2012 to 2019 period, it suggests that the lights-GDP elasticity has not shifted over time as this estimate is very close to the results for China in prior periods. Moreover, this estimate is very similar to the elasticity of 0.3 reported from DMSP data at the country level in the most widely cited validation study [5]. Given this proliferation of GDP-lights elasticity estimates of about 0.3, a common procedure in the applied literature is to estimate effects of the studied treatment (such as a disaster or a policy intervention) on the change in NTL data and to then use a GDP-lights elasticity of 0.3 (or thereabouts) from the literature to convert the estimated treatment-lights effect into an effect in terms of economic activity [15,17,40].

The apparent temporal stability of the GDP-lights elasticity is reassuring because it suggests that one can apply existing elasticity estimates, even if from a prior time period, to convert a contemporary estimate of the effect of treatment on nighttime lights into an estimate of the effect of treatment on GDP. However, a problem with this procedure that our findings highlight is that elasticities do not appear to be invariant to either the source of NTL data or (at least for DMSP data) to the level within the spatial hierarchy at which they are estimated. In terms of the spatial level, when we use annual changes in DMSP data and in GDP data at the county level, the elasticity is only 0.11 (Table 3), despite these data being the same as those used at the prefectural level where the elasticity appears to be 0.28. If long-differences are used, the DMSP-derived elasticities are also far higher when estimated at the prefectural level rather than the county level. Consequently, studies that combine an existing GDP-lights elasticity (estimated with more aggregated data) from the literature with their own estimate of the effect of a treatment on luminosity at a lower spatial level, such as at the county level, are likely to overstate the impacts of the treatment in terms of

economic activity. For example, studies using DMSP data at the city level and converting an estimated treatment effect in terms of luminosity into economic growth terms by using an assumed GDP-lights elasticity of 0.38 [22], or studies at the sub-district (third sub-national level) that assume that the GDP-lights elasticity ranges from 0.31 to 0.47 [17], are likely to overstate the treatment impacts on GDP changes.

In terms of the source of NTL data, at either the county level or the prefectural level, the elasticities of annual changes in GDP with respect to annual changes in VIIRS nighttime lights are far below 0.3 and would be better approximated by using a value of 0.1. Therefore, as applied studies increasingly switch from using DMSP data to using VIIRS data, the elasticities from existing validation studies based on DMSP data may no longer be applicable. Notably, the elasticity of 0.1, found by the within estimator with China's annual county and prefectural-level data, is very similar to what is found at the county level for the United States [30]. If this low value of the GDP-luminosity elasticity holds more generally (and noting that the within R^2 values are close to zero), then annual changes in NTL data may not be a good proxy for annual changes in local economic activity even though the levels of NTL data are good proxies for the levels of GDP. In other words, estimates of a treatment impact on changes in NTL data may be less informative of impacts on economic growth. Our analyses of China's monthly electricity use fluctuations in Figure 4 also show that changes in NTL data do not appear to be very good predictors of short-term changes in economic activity, at least in this particular context.

A general comment, which ties together the discussion of multi-source NTL data and multiple levels of spatial units in the administrative hierarchy, is that findings from validation studies in different settings, with different NTL data products, or at different levels of spatial aggregation, may not translate to other settings. Evidence for this point also comes through in Table 5, where elasticities of the change in GDP with respect to the change in NTL data vary with population density for all of the VIIRS data products that we use. In other words, an elasticity estimated from a densely populated urban area is unlikely to be suitable for a sparsely populated rural area [35], which is problematic because it is the sparsely populated rural areas where alternative approaches to the measurement of economic activity are most needed, given that these settings tend to be some of the most difficult places to gather data through traditional approaches, such as face-to-face surveys of rural enterprises and households.

5. Conclusions

In this paper, we have examined the question of how good are nighttime lights data for studying differences in economic activity between areas and for studying the temporal changes in economic activity within areas. In line with findings elsewhere, the nighttime lights data did far worse at predicting time-series changes in economic activity than at predicting differences in economic activity cross-sectionally. The predictive power is roughly 50-times greater for cross-sectional differences in economic activity than it is for modeling time-series changes in activity. The weak relationships between changes in nighttime lights data and changes in economic activity raise issues for interpreting applied studies that show effects of a treatment (e.g., a shock or intervention) on changes in nighttime lights. While such effects are interesting, they are unlikely to proportionately translate into economic activity terms, and it is economic activity rather than luminosity *per se* that is likely to be of interest to policymakers.

A reasonable summary of our estimates is that annual changes in county-level GDP in China have an elasticity of about 0.1 with respect to annual changes in NTL data. Interestingly, this elasticity is very similar to county-level results for the United States, even though the growth in nighttime lights in China has been far faster than in the United States over the period studied. It will be an important task for future research to find other settings with benchmark economic activity data for lower-level units in the administrative hierarchy, such as counties, in order to assess whether a GDP-lights elasticity of 0.1 for annual changes holds more widely. If elasticities with VIIRS data are found to be far lower

than the values of 0.3 or 0.4 reported in widely cited validation studies using DMSP data at national or first sub-national levels, it will suggest a need to revise the interpretation of various applied studies that use elasticity values from the literature to convert their findings into economic activity terms.

Beyond these broad conclusions, there are also some conclusions that may be more specific to China. In places where there are a non-trivial number of snow-covered nights per year, taking account of radiance measurements on these nights, as well as on the snow-free nights, is helpful for modeling annual GDP (which, by definition, measures economic activity in both snow-free and snow-covered periods). In particular, using the Black Marble data products, and including the snow-free and snow-covered composites, allowed us to have coverage of the greatest number of nights per year (about 10% more than just using a snow-free Black Marble composite and from 20–50% more nights per year than using V.2 VNL data). The resulting model with the weighted average of the Black Marble composites appeared to be closer to the truth than were other models of GDP changes, according to the non-nested testing. It will require further testing in other settings to see if these apparent advantages of the Black Marble data products hold more widely.

Author Contributions: Conceptualization and methodology, X.Z. and J.G.; Software, validation, data curation, and GIS analysis, X.Z.; Econometric analysis, preparation of original draft, and revision editing, J.G.; Visualization, X.Z.; Project administration and funding acquisition, J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Marsden Fund Grant UOW-1901.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The county-level GDP data are available from various issues of the China Statistical Yearbook and from the National Bureau of Statistics at http://www.stats.gov.cn/. DMSP stable lights annual composites are from: https://eogdata.mines.edu/products/dmsp/#download. The annual VNL V.2 data used in this study are available for download from the Earth Observation Group of the Colorado School of Mines at https://eogdata.mines.edu/products/vnl/#annual_v2. The annual Black Marble VIIRS data used in this study are available for download from: https://ladsweb.modaps.eosdis.nasa.gov/achive/allData/5000/VNP46A4/. All websites were last accessed on 28 January 2022.

Acknowledgments: We are grateful to Geua Boe-Gibson for assistance that helped to improve this paper. We also acknowledge the use of images and data processing by the Earth Observation Group, Payne Institute for Public Policy, Colorado School of Mines.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Elvidge, C.D.; Baugh, K.E.; Kihn, E.A.; Kroehl, H.W.; Davis, E.R.; Davis, C.W. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *Int. J. Remote Sens.* **1997**, *18*, 1373–1379. [CrossRef]
2. Doll, C.N.; Muller, J.P.; Morley, J.G. Mapping regional economic activity from night-time light satellite imagery. *Ecol. Econ.* **2006**, *57*, 75–92. [CrossRef]
3. Sutton, P.C.; Elvidge, C.D.; Ghosh, T. Estimation of Gross Domestic Product at sub-national scales using nighttime satellite imagery. *Int. J. Ecol. Econ. Stat.* **2007**, *8*, 5–21.
4. Chen, X.; Nordhaus, W.D. Using luminosity data as a proxy for economic statistics. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8589–8594. [CrossRef] [PubMed]
5. Henderson, J.V.; Storeygard, A.; Weil, D.N. Measuring economic growth from outer space. *Am. Econ. Rev.* **2012**, *102*, 994–1028. [CrossRef] [PubMed]
6. Shi, K.; Yu, B.; Huang, Y.; Hu, Y.; Yin, B.; Chen, Z.; Chen, L.; Wu, J. Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data. *Remote Sens.* **2014**, *6*, 1705–1724. [CrossRef]
7. Jing, X.; Shao, X.; Cao, C.; Fu, X.; Yan, L. Comparison between the Suomi-NPP Day-Night Band and DMSP-OLS for correlating socio-economic variables at the provincial level in China. *Remote Sens.* **2016**, *8*, 17. [CrossRef]

8. Zhao, M.; Cheng, W.; Zhou, C.; Li, M.; Wang, N.; Liu, Q. GDP spatialization and economic differences in South China based on NPP-VIIRS nighttime light imagery. *Remote Sens.* **2017**, *9*, 673. [CrossRef]
9. Wang, X.; Rafa, M.; Moyer, J.; Li, J.; Scheer, J.; Sutton, P. Estimation and mapping of sub-national GDP in Uganda using NPP-VIIRS imagery. *Remote Sens.* **2019**, *11*, 163. [CrossRef]
10. Liu, H.; He, X.; Bai, Y.; Liu, X.; Wu, Y.; Zhao, Y.; Yang, H. Nightlight as a proxy of economic indicators: Fine-grained GDP inference around mainland China via attention-augmented CNN from daytime satellite imagery. *Remote Sens.* **2021**, *13*, 2067. [CrossRef]
11. Clark, H.; Pinkovskiy, M.; Sala-I-Martin, X. China's GDP growth may be understated. *China Econ. Rev.* **2020**, *62*, 101243. [CrossRef]
12. Lee, Y.S. International isolation and regional inequality: Evidence from sanctions on North Korea. *J. Urban Econ.* **2018**, *103*, 34–51. [CrossRef]
13. Marx, A.; Rogers, M. Analysis of Panamanian DMSP/OLS nightlights corroborates suspicions of inaccurate fiscal data: A natural experiment examining the accuracy of GDP data. *Remote Sens. Appl. Soc. Environ.* **2017**, *8*, 99–104. [CrossRef]
14. Nguyen, C.N.; Noy, I. Measuring the impact of insurance on urban earthquake recovery using nightlights. *J. Econ. Geogr.* **2020**, *20*, 857–877. [CrossRef]
15. Kocornik-Mina, A.; McDermott, T.K.; Michaels, G.; Rauch, F. Flooded cities. *Am. Econ. J. Appl. Econ.* **2020**, *12*, 35–66. [CrossRef]
16. Zhao, N.; Liu, Y.; Hsu, F.; Samson, E.; Letu, H.; Liang, D.; Cao, G. Time series analysis of VIIRS-DNB nighttime lights imagery for change detection in urban areas: A case study of devastation in Puerto Rico from hurricanes Irma and Maria. *Appl. Geog.* **2020**, *120*, 102222. [CrossRef]
17. Heger, M.P.; Neumayer, E. The impact of the Indian Ocean tsunami on Aceh's long-term economic growth. *J. Dev. Econ.* **2019**, *141*, 102365. [CrossRef]
18. Zhao, X.; Yu, B.; Liu, Y.; Yao, S.; Lian, T.; Chen, L.; Yang, C.; Chen, Z.; Wu, J. NPP-VIIRS DNB daily data in natural disaster assessment: Evidence from selected case studies. *Remote Sens.* **2018**, *10*, 1526. [CrossRef]
19. Elvidge, C.; Ghosh, T.; Hsu, F.-C.; Zhizhin, M.; Bazilian, M. The dimming of lights in China during the COVID-19 pandemic. *Remote Sens.* **2020**, *12*, 2851. [CrossRef]
20. Xu, G.; Xiu, T.; Li, X.; Liang, X.; Jiao, L. Lockdown induced night-time light dynamics during the COVID-19 epidemic in global megacities. *Int. J. Appl. Earth* **2021**, *102*, 102421. [CrossRef]
21. Roberts, M. Tracking economic activity in response to the COVID-19 crisis using nighttime lights—The case of Morocco. *Dev. Eng.* **2021**, *6*, 100067. [CrossRef]
22. Eberhard-Ruiz, A.; Moradi, A. Regional market integration in East Africa: Local but no regional effects? *J. Dev. Econ.* **2019**, *140*, 255–268. [CrossRef]
23. Chodorow-Reich, G.; Gopinath, G.; Mishra, P.; Narayanan, A. Cash and the economy: Evidence from India's demonetization. *Q. J. Econ.* **2020**, *135*, 57–103. [CrossRef]
24. Henderson, J.V.; Squires, T.; Storeygard, A.; Weil, D. The global distribution of economic activity: Nature, history, and the role of trade. *Q. J. Econ.* **2018**, *133*, 357–406. [CrossRef] [PubMed]
25. Goldblatt, R.; Heilmann, K.; Vaizman, Y. Can Medium-Resolution Satellite Imagery Measure Economic Activity at Small Geographies? Evidence from Landsat in Vietnam. *World Bank Econ. Rev.* **2020**, *34*, 635–653. [CrossRef]
26. Asher, S.; Lunt, T.; Matsuura, R.; Novosad, P. Development research at high geographic resolution: An analysis of night-lights, firms and poverty in India suing the SHRUG open data platform. *World Bank Econ. Rev.* **2021**, *35*, 845–871. [CrossRef]
27. Nordhaus, W.; Chen, X. A sharper image? Estimates of the precision of nighttime lights as a proxy for economic statistics. *J. Econ. Geogr.* **2015**, *15*, 217–246. [CrossRef]
28. Elvidge, C.D.; Baugh, K.E.; Zhizhin, M.; Hsu, F.C. Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proc. Asia-Pac. Adv. Netw.* **2013**, *35*, 62. [CrossRef]
29. Chen, X.; Nordhaus, W.D. VIIRS nighttime lights in the estimation of cross-sectional and time-series GDP. *Remote Sens.* **2019**, *11*, 1057. [CrossRef]
30. Gibson, J.; Boe-Gibson, G. Nighttime lights and county-level economic activity in the United States: 2001 to 2019. *Remote Sens.* **2021**, *13*, 2741. [CrossRef]
31. Baugh, K.; Elvidge, C.D.; Ghosh, T.; Ziskin, D. Development of a 2009 stable lights product using DMSP-OLS data. *Proc. Asia-Pac. Adv. Netw.* **2010**, *30*, 114. [CrossRef]
32. Elvidge, C.D.; Zhizhin, M.; Ghosh, T.; Hsu, F.-C.; Taneja, J. Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sens.* **2021**, *13*, 922. [CrossRef]
33. Román, M.; Wang, Z.; Sun, Q.; Kalb, V.; Miller, S.; Molthan, A.; Schultz, L.; Bell, J.; Stokes, E.; Pandey, B.; et al. NASA's Black Marble nighttime lights product suite. *Remote Sens. Environ.* **2018**, *210*, 113–143. [CrossRef]
34. Gibson, J. Better night lights data, for longer. *Oxf. Bull. Econ. Stat.* **2021**, *83*, 770–791. [CrossRef]
35. Gibson, J.; Olivia, S.; Boe-Gibson, G.; Li, C. Which night lights data should we use in economics, and where? *J. Dev. Econ.* **2021**, *149*, 102602. [CrossRef]
36. Hodler, R.; Raschky, P.A. Regional favoritism. *Q. J. Econ.* **2014**, *129*, 995–1033. [CrossRef]
37. Storeygard, A. Farther on down the road: Transport costs, trade and urban growth in sub-Saharan Africa. *Rev. Econ. Stud.* **2016**, *83*, 1263–1295. [CrossRef]
38. Tuttle, B.; Anderson, S.; Sutton, P.; Elvidge, C.; Baugh, K. It used to be dark here. *Photogramm. Eng. Remote Sensing.* **2013**, *79*, 287–297. [CrossRef]

39. Abrahams, A.; Oram, C.; Lozano-Gracia, N. Deblurring DMSP nighttime lights: A new method using Gaussian filters and frequencies of illumination. *Remote Sens. Environ.* **2018**, *210*, 242–258. [CrossRef]
40. Villa, J. Social transfers and growth: Evidence from luminosity data. *Econ. Dev. Cult. Chang.* **2016**, *65*, 39–61. [CrossRef]
41. Ji, X.; Li, X.; He, Y.; Liu, X. A simple method to improve estimates of county-level economics in China using nighttime light data and GDP growth rate. *ISPRS Int. J. Geoinf.* **2019**, *8*, 419. [CrossRef]
42. Gibson, J.; Olivia, S.; Boe-Gibson, G. Night lights in economics: Sources and uses. *J. Econ. Surv.* **2020**, *34*, 955–980. [CrossRef]
43. National Bureau of Statistics (NBS). *China Statistical Yearbook (County-Level) [M]*; China Statistics Press: Beijing, China, 2013–2020.
44. National Bureau of Statistics (NBS). *China City Statistical Yearbook [M]*; China Statistics Press: Beijing, China, 2013–2020.
45. National Bureau of Statistics (NBS). *Beijing Statistical Yearbook [M]*; China Statistics Press: Beijing, China, 2013–2020.
46. Ghosh, T.; Baugh, K.; Elvidge, C.; Zhizhin, M.; Poyda, A.; Hsu, F.-C. Extending the DMSP Nighttime Lights Time Series beyond 2013. *Remote Sens.* **2021**, *13*, 5004. [CrossRef]
47. Elvidge, C.D.; Baugh, K.; Zhizhin, M.; Hsu, F.C.; Ghosh, T. VIIRS night-time lights. *Int. J. Remote Sens.* **2017**, *38*, 5860–5879. [CrossRef]
48. Wooldridge, J. *Introductory Econometrics: A Modern Approach*, 7th ed.; Cengage Publishing: Boston, MA, USA, 2020.
49. Kim, B.; Gibson, J.; Chung, C. Using panel data to estimate income under-reporting by the self-employed. *Manch. Sch.* **2017**, *85*, 41–64. [CrossRef]
50. Greene, W. *Econometric Analysis*, 4th ed.; Prentice Hall Publishing: Hoboken, NJ, USA, 2000.
51. Vuong, Q. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrical* **1989**, *57*, 307–333. [CrossRef]
52. Hu, X.; Huang, X. A novel locally adaptive method for modeling the spatiotemporal dynamics of global electric power consumption based on DMSP-OLS nighttime stable light data. *Appl. Energy* **2019**, *240*, 778–792. [CrossRef]
53. Zhao, F.; Ding, J.; Zhang, S.; Luan, G.; Song, L.; Peng, Z.; Du, Q.; Xie, Z. Estimating rural electric power consumption using NPP-VIIRS night-time light, toponym and POI data in ethnic minority areas of China. *Remote Sens.* **2020**, *12*, 2836. [CrossRef]
54. Sahoo, S.; Gupta, P.; Srivastav, S. Comparative analysis between VIIRS-DNB and DMSP-OLS night-time light data to estimate electric power consumption in Uttar Pradesh, India. *Int. J. Remote Sens.* **2020**, *41*, 2565–2580. [CrossRef]
55. National Energy Administration (NES). Available online: http://www.nea.gov.cn/ (accessed on 5 November 2021).

Chapter 3: Remotely too equal: Popular DMSP night-time lights data understate spatial inequality

Xiaoxuan Zhang, John Gibson and Xiangzheng Deng

This paper has been published in the *Regional Science Policy and Practice*, (2023), 15(9), 2106-2126.

Received: 31 March 2023 | Revised: 7 July 2023 | Accepted: 31 July 2023

DOI: 10.1111/rsp3.12716

ORIGINAL ARTICLE

Remotely too equal: Popular DMSP night-time lights data understate spatial inequality

Xiaoxuan Zhang[1] | John Gibson[1] | Xiangzheng Deng[2]

[1]Department of Economics, University of Waikato, Hamilton, New Zealand

[2]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

Correspondence
John Gibson, Department of Economics, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand.
Email: jkgibson@waikato.ac.nz

Funding information
Marsden Fund, Grant/Award Number: UOW 1901

Abstract

Regional science and economics studies increasingly use the Defense Meteorological Satellite Program (DMSP) night-time lights data to measure spatial inequality. These DMSP data are a poor proxy in this context because of their spatially mean-reverting errors, which yield significantly lower inequality estimates than what subnational GDP data show. Inequality estimates from DMSP are also lower than what newer, research-focused and more accurate satellites show. We demonstrate this bias using county-level data from China and the United States. The errors in the DMSP data distort estimates of both the level of and trend in spatial inequality.

KEYWORDS
DMSP, mean-reverting error, night lights, spatial inequality, VIIRS

JEL CLASSIFICATION
E01, R12

1 | INTRODUCTION

A growing literature in regional science and economics uses data on satellite-detected night-time lights to study spatial inequality. Over two dozen papers have come out, mostly in the last 4 years. The potential advantages of this remote sensing approach include greater comparability over space irrespective of countries' varying levels of statistical capacity, cheaper and more timely data, and the possibility of extending inequality estimates to spatial units

below the level at which GDP data or survey data are reported.[1] In contrast, the existing situation with traditional approaches is aptly described by Mirza et al., (2021:1) as one where 'our knowledge about inequality in the developing world is limited by paucity, poor quality, uncertainty, and incomparability of data'.

There is a potential flaw in this new research on inequality, because most studies rely on using Defense Meteorological Satellite Program (DMSP) data. This involves repurposing images originally gathered to detect clouds for short-term Air Force weather forecasts. While a seminal study in economics shows DMSP data can also be used for long-run observations of economic activity on earth (Henderson et al., 2012), that study only needed measurement errors in DMSP data to be independent of errors in reported GDP (because an optimal mix of the two types of noisy data is used to better proxy for true but unknown economic activity). In contrast, the recent inequality studies use the DMSP data directly, and because these data have spatially mean-reverting errors (Gibson, 2021), they are poorly suited for measuring spatial inequality, even if they are a reasonable proxy in the context where Henderson et al. (2012) used them (for studying growth in national GDP).

A consequence of these spatially mean-reverting errors is that DMSP data yield lower inequality estimates than what subnational GDP data show. The inequality estimates from DMSP are also lower than what newer, research-focused, and more accurate satellites show from their observations of the Earth at night. Moreover, our results here suggest that not just the apparent level of spatial inequality, but also the temporal trend, may be affected if DMSP data are used. Thus, even though using night-time lights data has the potential to overcome some constraints on our understanding of inequality, the data that have been most commonly used to date could introduce some new misunderstandings.

A few other recent studies critically reappraise use of DMSP data for studying small areas (Bluhm & McCord, 2022; Gibson et al., 2021; Goldblatt et al., 2020). Our study differs from these in at least three ways. First, we solely focus on inequality, looking at both levels and trend. These other papers largely do not consider inequality, and any discussion of DMSP errors and inequality estimates is just for level effects, while policymakers are often most interested in trends. Second, we also provide a visual example as well as statistical evidence, which may help make our findings accessible for practitioners. Third, we relate our findings to a detailed review of the literature, whereas the prior studies paid little attention to studies in the regional science literature.

In the next section we discuss the sources of spatially mean-reverting errors in DMSP data.[2] A visual example is given in Section III to show how DMSP blurring and top-coding distort apparent patterns of spatial variation in economic activity. Our literature review in Section IV shows that recent inequality studies ignore these spatially mean-reverting errors. Section V has county-level evidence from the United States and China on the understatement of spatial inequality when the DMSP data are used. Section VI has our conclusions.

2 | THE SPATIALLY MEAN-REVERTING ERRORS IN DMSP NIGHT LIGHTS DATA

The spatially mean-reverting errors in DMSP data have two main sources, and both reflect some inherent limitations of DMSP sensors and data management (Abrahams et al., 2018). First, DMSP images are blurred, so some light is attributed to places from where it was not emitted. Earlier remote sensing studies used terms such as 'blooming'

[1]In comparison, survey-based approaches may be incompatible over time and space, with measured inequality varying with survey welfare indicators (Deininger & Squire, 1996) and reference periods (Gibson et al., 2021). Surveys are also expensive: at up to $4,000 USD per household, they are fielded infrequently and can be slow to release data (Sharp et al., 2022). Moreover, samples are often just large enough to represent the first subnational level (e.g., provinces); survey-to-census imputation (e.g., Elbers et al., 2003) lets inequality estimates be formed at lower levels (e.g., counties) but limits temporal coverage because of census infrequency. The other traditional source of data for estimating spatial inequality is national and regional accounts data. Wide variation in statistical capacity and data availability impose limits on the usefulness of these data (Henderson et al., 2012).

[2]At least three studies empirically estimate strength of the mean reversion in DMSP data at different spatial scales, and these are discussed in Section II. The current paper builds on these prior estimates to highlight implications for measuring spatial inequality with DMSP data and also provides examples of how measurement errors affect both the apparent level and trend in spatial inequality in two large countries.

and 'overglow' to describe this pattern (e.g., Small et al., 2005) but that discussion did not pinpoint sources of the problem. Moreover, subsequent applied studies tended to focus on just one aspect of the earlier literature – that environmental factors such as snow and water may cause reflections and create the overglow – and inferred that the blurring problem should therefore matter in only a few places (e.g., Michalopoulos & Papaioannou, 2014). In fact, blurring is a universal feature of DMSP data that shows up everywhere.

The blurring occurs for at least three reasons. First, away from the nadir of the 3,000-km-wide scan, the DMSP sensor views the earth at an angle and this causes the field of view (FOV) to expand (by 400% at the scan edge).[3] All light from the expanded FOV gets (wrongly) attributed to a far smaller pixel in the center of the FOV. To limit the bias, the National Oceanic and Atmospheric Administration (NOAA) only uses pixels between the nadir and the half-scan (where the FOV is expanded by just 240%) when processing DMSP images, but this does not reverse the blurring. Second, the on-board computers cannot hold all the data, so to conserve memory the pixels are aggregated to 5×5 blocks. Third, random geolocation errors, with a mean of about 3 km, also spread recorded light away from its point of origin (Tuttle et al., 2013). A method to de-blur DMSP images was developed by Abrahams et al. (2018), but our review found no inequality studies using this method with DMSP data.

The net effect of the expanded FOV, the pixel aggregation, and the geolocation errors is that the ground footprint for the DMSP sensor is 25 km^2 at nadir (Elvidge et al., 2013). At the edge of the half-scan it may be up to 60 km^2 because the FOV expands when viewing the earth at an angle for off-nadir pixels. This spatial resolution is far coarser than the footprint of the sensors used by newer, research-orientated, night lights data sources such as VIIRS Nighttime Light (VNL) and NASA's Black Marble (BM).[4] A comparison of various night lights data sources is presented in Table 1. This comparison shows that the sensors used by VNL and BM have a constant 0.55 km^2 footprint, where this invariance is achieved by turning off some detector elements away from the nadir to counteract effects of viewing the earth at an angle. These newer data sources are at least 45 times more spatially precise than the DMSP images (Elvidge et al., 2013).

The 25 km^2 ground footprint of the DMSP sensor exceeds the 30 arc-sec (roughly 0.9 km $\times$ 0.9 km at the equator or 0.9 km $\times$ 0.6 km at 45° latitude) grid onto which DMSP data are allocated. The DMSP inequality literature does not distinguish between sensor resolution and output grid scale. For example, a recent study only says that DMSP has 'spatial resolution of 30 $\times$ 30 arc seconds (approximately 1 $\times$ 1 km)' (Sangkasem & Puttanapong, 2022: 830). In fact, spatial resolution of the sensor is far coarser than the output grid, so differences between each 30 arc-sec output pixel are blurred. The threat to spatial inequality estimates is easiest to explain with the pixel aggregation to conserve on-board memory. Consider an area with one brightly-lit pixel surrounded by 24 unlit pixels, which would suggest a high degree of local inequality. With pixel aggregation, luminosity attributed to the brightly-lit pixel will be dragged down to the mean for the 25 pixels, while for the 24 unlit pixels their values will be overstated; aggregation is inherently mean-reverting.

It is worth stating explicitly that blurring will affect a broad range of studies and not just those using pixel-level analysis.[5] Consider a spatial unit such as a state whose major city is near its boundary (a good example is Wisconsin in Section III). Blurred DMSP data attribute some light from this city to neighbors. The bias depends on spatial scale of analyses; it should matter more with smaller spatial units than larger ones. A good example is a recent study where North Korea's centrally controlled municipality (thus, like a province) of Kaesong had an industrial zone with 50,000 North Koreans employed by South Korean firms (Kim et al., 2022). The area of Kaesong is 180 km^2, divided into two districts, one with the industrial zone and one without. In both DMSP and VNL data, the district with the industrial zone was at the 99th percentile of the luminosity distribution for North Korea (out of 186 second subnational level

[3]To provide intuition for this effect, if a flashlight shines directly down onto the ground it illuminates a circle, but if it shines on a spot on the ground somewhere ahead it illuminates an ellipse that is far larger than the circle.

[4]Technical details on these two data sources, which both rely on the Suomi/NPP satellite but process the data in different ways, can be found in Elvidge et al. (2021) and Román et al. (2018).

[5]The inequality studies reviewed below build up from the pixel level to aggregations such as villages, grids, districts, provinces, and even countries. In contrast, some impact evaluation studies, such as Kocornik-Mina et al. (2020), directly use the pixel-level DMSP data.

TABLE 1 Attributes of DMSP and other night-time lights data sources.

	DMSP	VNL	Black Marble
Satellite/sensor attributes			
Operator	US DoD	NASA/NOAA	
Original purpose	Detect moon-lit clouds for weather forecasts	Earth observation for scientific research	
Available years	1992–2021	2012–2021	
Ground footprint at nadir	25 km^2	0.55 km^2	
Scan width	3,000 km (but only the center half is processed)	3,000 km	
Revisit time	12 h	12 h	
Pixel saturation	Saturated	Not saturated	
On-board calibration	No	Yes	
Data products			
Creator of annual composites	EOG	EOG	NASA
Spatial resolution of processed data	30 arc-sec grid	15 arc-sec grid	15 arc-sec grid
Quantization	6-bit ($n = 64$)	14 bit ($n = 16{,}384$)	16 bit ($n = 65{,}536$)
Masking of ephemeral light sources	No	Yes	Yes
Stray light correction	No	Yes, from 2014	Yes
User control over angle of detection	No	No	Yes

DMSP, Defense Meteorological Satellite Program; DoD, Department of Defense; EOG, Earth Observation Group; VNL, VIIRS Nighttime Lights.

units). However, VNL data showed the other district at just the 29th percentile, while DMSP data had it at the 79th percentile. The luminosity gap between the two districts is attenuated with DMSP data (seeming to be only 20 percentiles rather than 70 percentiles) because the blurred DMSP images attribute some lights from the industrial zone district to its neighbor. Thus, blurring matters even when working with spatial units far more aggregated than the pixel level.

The second source of mean-reverting error in DMSP data is top-coding, due to two factors. First, DMSP was developed to measure clouds rather than to measure lights on earth, and unrecorded sensor amplification occurs during the dark part of the lunar cycle when the cloud-tops are no longer visible in raw moonlight (Hsu et al., 2015). Given that the sensor has only a low dynamic range, when amplification is increased, the images for brightly lit parts of the Earth, such as central business districts, are saturated with light. Second, pixel aggregation to save memory sees the original 8-bit values divided by four and top-censored at 63, to give the widely used 6-bit DN (digital number). Top-coding shows up in anomalous patterns, such as key infrastructure such as Heathrow airport (the most brightly lit feature in England) seeming to be no brighter than the surrounding area (Gibson, 2021) and city centers seeming to be no brighter than lower density suburbs (Bluhm & Krause, 2022).

In contrast to DMSP, nightly images used by the VNL and BM data are not subject to saturation, so there is no top-coding problem with these data. The sensors for these newer data sources have far wider dynamic range, covering seven orders of magnitude of radiance (while DMSP only covers two orders of magnitude) so they can simultaneously observe dimly and bright lit areas. Moreover, when the signal is quantized, it is with 14-bit resolution (VNL) or 16-bit resolution (BM), providing up to 65,536 different values, as opposed to the coarse 6-bit resolution of DMSP data with only 64 possible values (Table 1). The VNL and BM data also have inbuilt masking of ephemeral light sources and correction for stray lights (such as from solar glare in long summer evenings) and users have some control over the angle of detection by restricting attention to either near-nadir or off-nadir pixels (BM only).

The improved quality of VNL and BM data versus the relatively crude DMSP data shows up when formal tests for mean-reverting error are conducted. The empirical basis for these tests (Black et al., 2000; Gibson & Kim, 2010) relates observed values of a variable of interest, y^* to the presumed true values, y by:

$$y^* = \theta + \lambda y + v \tag{1}$$

Textbook classical measurement error assumes $\theta = 0$, $\lambda = 1$, and $E(v) = cov(y,v) = 0$, so that just white noise is added to the true value. In contrast, if the measurement errors are mean-reverting, $0 < \lambda < 1$, and errors negatively correlate with true values $cov(y,v) < 0$.

At least three studies estimate Equation (1) with DMSP data as the noisy measure and the more accurate VNL or BM data as the truer values.[6] With pixel-level data for urban areas (mostly in developing countries), Alimi et al. (2022) estimate $\widehat{\lambda} = 0.3$. A higher estimate, of $\widehat{\lambda} = 0.7$, comes from regional data for Europe at the NUTS2 level (Gibson, 2021), which is typically the provincial level (or groups of counties in some countries). Kim et al. (2022) use the second subnational level for North Korea (which is counties in rural areas and districts in urban areas) and get $\widehat{\lambda} = 0.4$. Although three studies is not much basis for extrapolation, one explanation for variation in $\widehat{\lambda}$ is that mean reversion shows up more (as a lower $\widehat{\lambda}$) for smaller spatial units because the blurring has a proportionally larger effect on small spatial units.

3 | VISUAL EVIDENCE FROM WISCONSIN

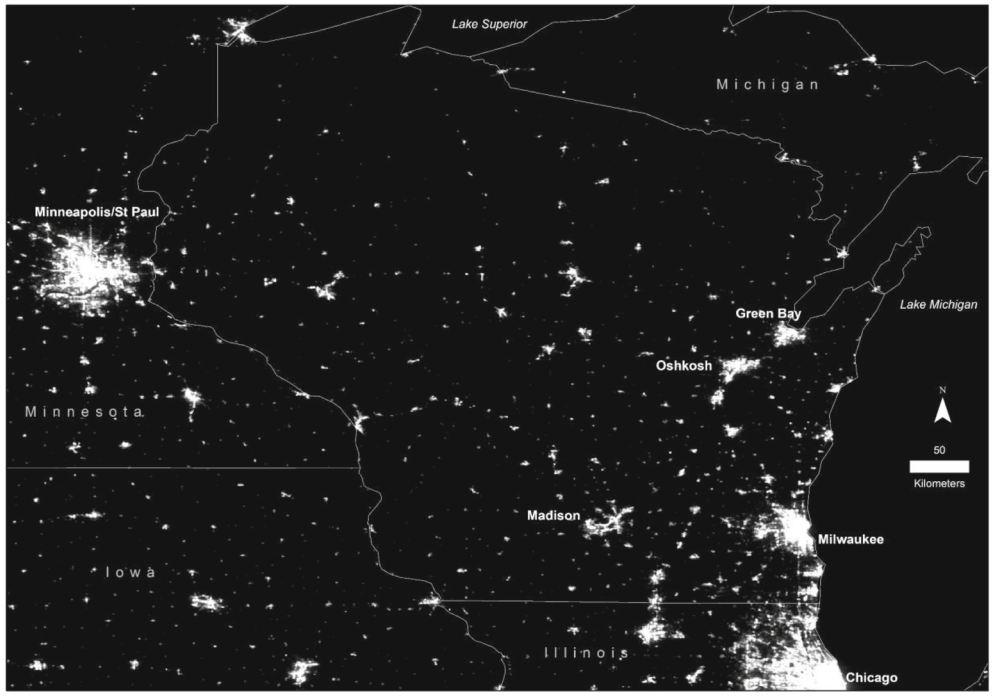
Mean-reverting errors in the DMSP data are illustrated in Figure 1 by comparing with VNL images for Wisconsin, United States. Wisconsin has a medium level of spatial inequality, ranking 21st of the 50 states for the Gini index of county GDP. It helps for this illustration that it is surrounded on two sides by water and by the sparsely populated upper peninsula of Michigan, minimizing visual distraction due to light from elsewhere (the other lights shown are mainly from Chicago and Minneapolis/St. Paul).

Much of the population and economic activity in Wisconsin is concentrated in the lower southeast corner of the state. Specifically, if one draws a straight line from Green Bay through Oshkosh and Madison down to the state border with Illinois, there are 22 counties (of 72 in the whole state) that are either intersected by that line or are southeast of it. While those 22 counties only cover 22% of the state area, they contributed 74% of state GDP (as of 2013), and were home to 69% of state population (as of the 2010 census). When VNL data shown in panel A of Figure 1 are used, the share of total state luminosity contributed by the counties in the southeast corner is similar to the GDP and population shares, with 66% of the state total of VNL radiance coming from these 22 counties.

Unlike the clear definition of urban areas shown by VNL data in panel A of Figure 1, the image using DMSP data in panel B is very blurred, with large areas that are unlit in the VNL image appearing illuminated according to DMSP. By attributing lights to unlit places, DMSP data revert toward the mean because unlit and poorly lit areas have their luminosity overstated. The other feature of DMSP that causes mean reversion is attenuated differences at the top of the lighting distribution, so that more brightly lit big cities seem no brighter than far smaller towns. For example, the brightness of Oshkosh (population 67,000) seems hardly less than that of Milwaukee, whose population is 10 times larger, when the DMSP data are used (the maximum available DMSP digital number of 63 is recorded for both cities). Yet VNL data that are not subject to top-coding show the most luminous parts of Milwaukee have radiance at least twice that of Oshkosh (at 240 nanowatts per square centimeter per steradian [nW/cm^2/sr] in Milwaukee versus 110 nW/cm^2/sr for Oshkosh).

[6]In support of the assumption that VNL or BM data provide truer values, Vuong (1989) non-nested tests based on the Kullback–Leibler information criterion show that models using VNL or BM data provide results that are closer to the truth than are models using DMSP data (Gibson, 2021; Zhang & Gibson, 2022).

(a) VIIRS Night Lights



(b) DMSP

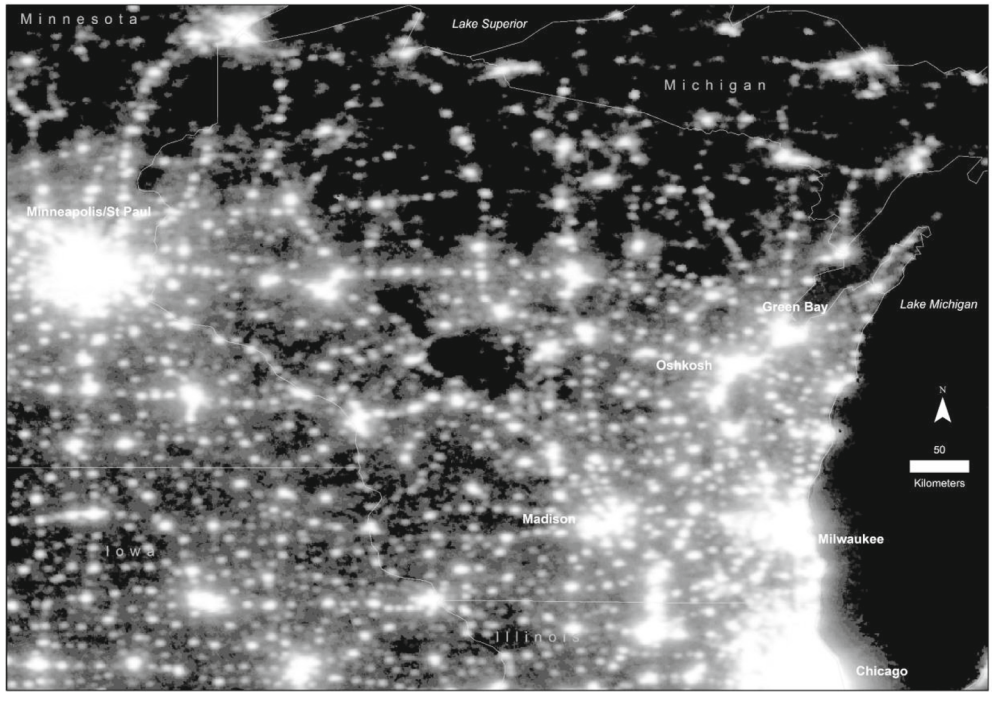


FIGURE 1 Wisconsin at night (annual composites for 2013).

TABLE 2 Summary of the studies reviewed.

	All studies	Recent studies (2018 onward)	Earlier studies (prior to 2018)
Number of studies	26	15	11
Number that use DMSP data	22	12	10
Number that mention mean-reverting errors in DMSP data	0	0	0
Number that de-blur DMSP data	0	0	0
Number that adjust for top-coding in DMSP data	1	1	0

Note: Details on the studies summarized in this table can be found in Appendix A.
DMSP, Defense Meteorological Satellite Program.

As a consequence of these mean-reverting errors in DMSP data, spatial inequality is greatly understated. Whereas GDP, population, and VNL data all show that at least two-thirds of Wisconsin totals come from the 22 counties in the southeast corner of the state, the DMSP data suggest these counties contribute a minority of total luminosity; 47% from the southeast corner and 53% from the rest of the state. By raising the share of lights seeming to come from the less economically active part of the state, and lowering the share seeming to come from the part with concentrated economic activity, DMSP data create a systematic understatement of spatial inequality. Consequently, the Gini index for Wisconsin based on county-level data is only 0.27 according to DMSP, while it is 0.51 if the VNL data are used (and slightly higher at 0.57 when using GDP data).

4 | INATTENTION TO SPATIALLY MEAN-REVERTING ERRORS IN THE LITERATURE

Despite the clear distortion seen in Figure 1, which also holds elsewhere (e.g., Gibson, 2021; Gibson et al., 2020), inequality studies using DMSP data have ignored implications of spatially mean-reverting errors. We reviewed 26 studies (15 published since 2018) that use night-time lights data to measure spatial inequality: 22 of these studies used DMSP data. There is no decline over time in the popularity of this data source, as 12 of the 15 most recent studies used DMSP data. Full details on the studies reviewed, including the data source(s), time scale and spatial units, and research objective, are provided in Appendix A, with summary tabulation of the review in Table 2.

We carefully reviewed each study to see whether authors mention spatially mean-reverting errors in the DMSP data. We also checked whether authors tried to de-blur DMSP data, such as with procedures outlined in Abrahams et al. (2018). Similarly, we checked if anything was done to adjust the data for top-coding problems.[7]

There were no mentions of spatially mean-reverting errors in the studies we reviewed. Likewise, none used procedures to de-blur the DMSP data (Table 2). In terms of top-coding, one study (Montalvo et al., 2021) adjusted top-coded observations relative to the distribution of observations in the locality that were not top-coded. Two other studies excluded certain places the authors assumed might be affected by top-coding; Chakravarty and Dehejia (2017) exclude five major metro areas as likely having top-coded data, and the same reason is used by Mendez and Santos-Marquez (2020) for excluding Singapore. However, it should be noted that top-coding issues extend far further down the distribution of DMSP digital numbers, due to the fact that various pixels may be top-coded on some nights of the year but not on others (given the variation in sensor amplification over the lunar cycle) and so annual

[7]Bluhm and Krause (2022) released Pareto-adjusted DMSP data for 1992–2013 for all pixels worldwide, which are meant to deal with top-coding (under the assumption that the true luminosity from the top-coded pixels follows a Pareto distribution). Those corrected data were available since the release of their first working paper in 2018. Likewise, a MATLAB script for the Abrahams et al. de-blurring approach has been available since 2015.

composites may have pixels that are partially top-coded even for annual average DN values as low as DN = 55 (Bluhm & Krause, 2022).

Another symptom of the inequality literature ignoring mean-reverting errors in DMSP data is that some studies used a time series that spliced together DMSP and VNL data (Chen & Zhang, 2023; Ivan et al., 2020). This splicing was to extend the length of the time series past the 2013 ending data of the most popular DMSP annual composites.[8] However, the mean-reverting error in DMSP data means that no simple adjustment factor can 'line up' the two data sources. A mean-reverting measurement error is not constant and instead will vary with the true, but unknown, luminosity (in other words, the error v in Equation (1) varies with the true value, contrary to the usual assumption that $\text{cov}(y,v) = 0$). The absence of discussion of this issue by inequality studies using spliced datasets is more evidence that implications of mean-reverting measurement errors in DMSP data seem to have been ignored.[9]

5 | COUNTY-LEVEL EVIDENCE FROM THE UNITED STATES AND CHINA

The example given in Section III suggests that inequality estimates from DMSP data may be subject to considerable downward bias. The discussion in Section II shows that the mean-reverting measurement errors that cause this downward bias in inequality estimates are due to inherent features of the DMSP sensors and data management; we can thus expect the bias to occur more widely beyond the particular example of Wisconsin used in Section III. In this section we demonstrate that the bias is more generalized, using data from all of the United States and also from China; the two largest countries reporting GDP at the county-level.[10] The advantage of working with data from countries that have county-level GDP data is that it gives a benchmark for the inequality estimates derived from night-time lights data; if these county-level GDP data were available more widely they would surely be used for spatial inequality studies.[11] The details on the GDP data sources we use for the two countries, and on the sources of the night-time lights data that we use, are provided in Appendix B.

Spatial inequality in the United States is shown in Figure 2 to have been quite stable over the last two decades according to either the Gini coefficient (panel A) or the Theil index (panel B). Specifically, the county-level GDP data provide Gini coefficients that hardly vary, ranging from a low of 0.71 in 2009 to a high of 0.72 in 2017 (with a mean of 0.713 across the 19 years with GDP data and a standard deviation across the annual estimates of 0.003). Along the same lines, annual values of the Theil index hardly varied, ranging from 0.98 to 1.01 (and the years with the minimum and maximum values are the same as for the Gini coefficient). The bootstrapped standard errors surrounding the inequality estimates for each year average 0.03 for the Gini coefficient and 0.12 for the Theil index.

In contrast to what is shown with the GDP data, the DMSP data on night-time lights suggest a significantly lower level of spatial inequality. The Gini coefficients vary from a high of 0.52 in 2006 to a low of 0.48 in 2010 (the mean is 0.50 and the standard deviation of the annual estimates is 0.011). There is no overlap between 95% confidence intervals of the DMSP-derived Gini coefficients and the GDP-based Gini coefficients. The gap between the

[8]An extension set of DMSP annual composites for 2013–2019 was provided by Ghosh et al. (2021), but these data do not seem to be used by any of the inequality studies. Unlike the original DMSP data from 1992 to 2013 that had early evening observations, the extended series uses pre-dawn readings and so the composition of lights is likely to be different, and this may affect apparent inequality, so these data are not used here. The extension series is available from https://eogdata.mines.edu/wwwdata/dmsp/extension_series/

[9]A reviewer suggested the revealed preference of inequality studies for using DMSP data is due to the longer time series and that authors probably knew of the measurement error issues because ChatGPT gives a good summary of the differences between DMSP and VNL according to what was embedded in online documents (as of 2021). If this were so, the studies would be expected to discuss the trade-offs faced when choosing a NTL data source – more accurate but shorter time series versus more error but longer time series. We saw no such discussions and so we assume that many authors were unaware of the mean-reverting errors in DMSP data.

[10]In Europe, GDP is only reported down to the NUTS2 level, which is groups of counties in some countries (e.g., the UK) and provinces in others. Japan only reports GDP down to the prefectural level. India and Indonesia only report to the district level.

[11]A reviewer suggested that population shrinkage in some urban areas could confound inequality measures from night-time lights data. If infrastructure persists in declining cities and areas of urban growth in emerging regions are newly illuminated, there may be some convergence in luminosity that is not seen with GDP or survey data. Examples of such urban shrinkage include northeastern China (Li & Mykhnenko, 2018) and movement from rustbelt to sunbelt cities in the United States. We do not pursue this issue here, but highlight it as a topic for the future.

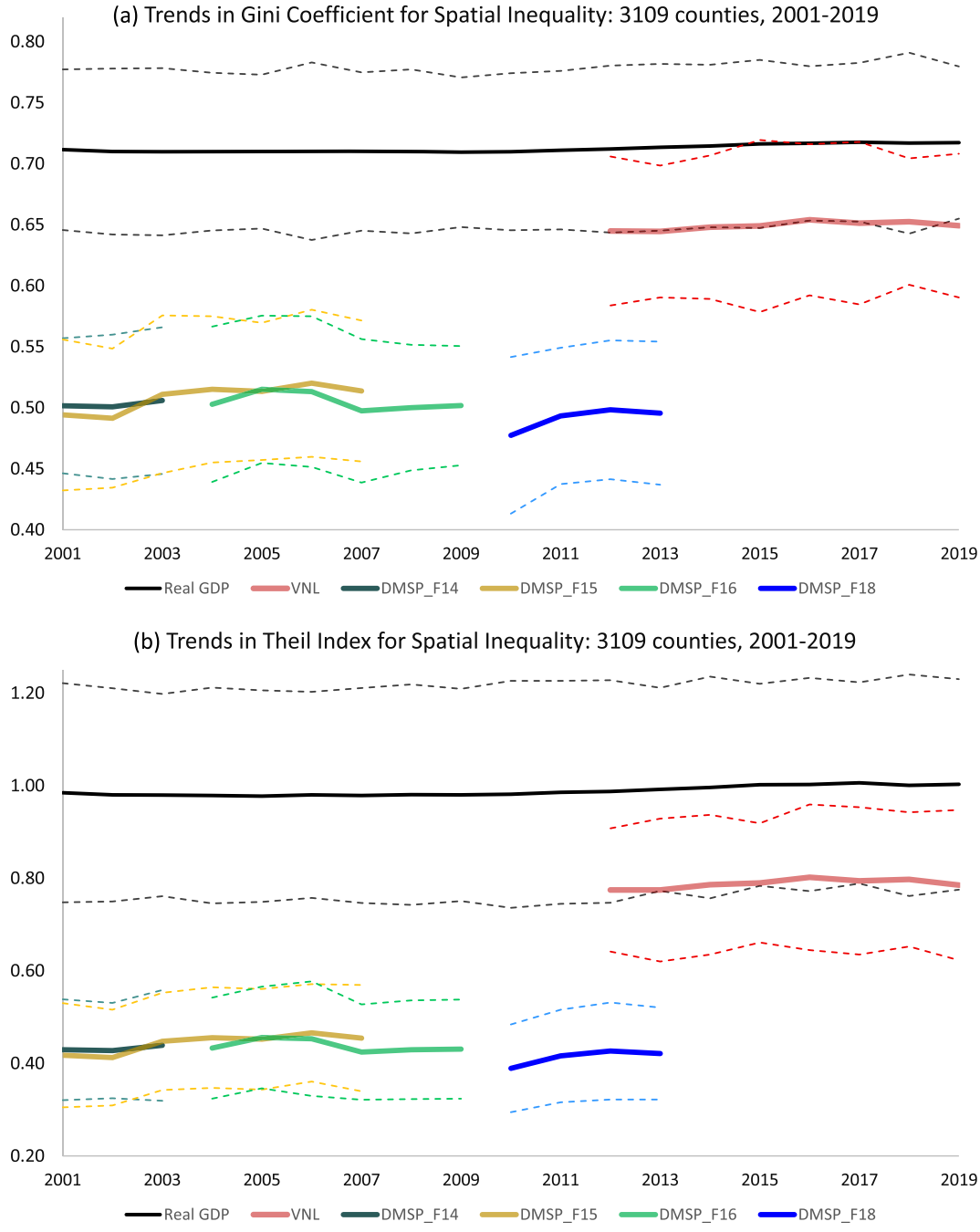


FIGURE 2 County-level spatial inequality in the United States. *Note*: 95% confidence intervals from bootstrapped standard errors are shown with dashed lines.

lights-based and GDP-based estimates is even more apparent for the Theil index; the mean of this index is 0.44 when DMSP is used, which is almost 60% below what the GDP data show. Moreover, the average gap between the GDP-based Theil index and the DMSP-based index is more than five times larger than the bootstrapped standard errors (for the Gini the average gap is six times larger than the standard errors). In other words, on the basis of the

patterns of economic activity that prevail across all counties in the United States, inequality measures derived from DMSP data are statistically significantly lower than the estimates derived from GDP data.

This gap in inequality estimates can be most correctly described as understatement when using DMSP data. The first reason for this claim is that GDP data would likely be the first choice of researchers wanting an economic activity measure, if such data were reliably available at a spatially disaggregated level elsewhere. In other words, the estimates derived from the GDP data can be sensibly thought of as a benchmark. Second, inequality estimates based on the VNL data show no significant differences from those based on GDP data.[12] The discussion in Section II on DMSP sensors and data management gave good reasons to expect that the DMSP data will be less accurate than VNL and BM data, and so close correspondence between the GDP-based and VNL-based estimates lends credence to the GDP-based results.

Two other features of Figure 2 warrant comment. First, the time series of inequality estimates based on DMSP data has discontinuities in the years when DMSP transitions from one satellite to another. For example, when the source satellite switched from F16 in 2009 to F18 in 2010, there was a 4-point drop in the Theil index, equivalent to a 10 % decline in the apparent level of inequality, even though GDP-derived inequality estimates showed no similar change in those years. These discontinuities likely stem from a lack of calibration of the DMSP satellites, which interferes with the consistency of the DMSP time series (Gibson et al., 2020).[13] This flaw also shows up as discrepancies when two satellites provide data in the same year (e.g., spatial inequality in 2007 was 7 % lower with data from satellite F16 than with what the data from satellite F15 indicated for the same year).

The other feature from the results of Figure 2 is that the understatement of inequality is far more apparent for the Theil index than for the Gini coefficient, probably because the Theil index is more sensitive to differences at the top of the distribution than the Gini coefficient is. There are at least two possible reasons why spatial variation at the top of the distribution is especially understated with the DMSP data: non-linearity in the GDP–luminosity relationship whereby DMSP night-time lights respond less to GDP in places where GDP is higher (Bluhm & McCord, 2022), and top-coding bias whereby DMSP data cannot distinguish brightly lit economic features, such as central business districts, from their less brightly lit surrounds (Bluhm & Krause, 2022; Gibson et al., 2021). A lack of uniform understatement of spatial inequality may distort the understanding of what drives some spatial inequality, given that it is not simply a parallel shift down in all inequality measures when the DMSP data are used.

The county-level inequality results for China are shown in Figure 3, and the patterns seen in the data for the United States are also exhibited for China. The VNL-based inequality estimates are not statistically significantly different from the GDP-based estimates; confidence intervals for the Gini coefficients with these two series almost completely overlap from 2012 to 2015, and overall the VNL-based Gini coefficients are just 4% (3 Gini points) below the GDP-based ones. In contrast, the DMSP-based inequality measures are 16% (37%) below the benchmark GDP-based estimates for the Gini coefficient (Theil index) over the 2000–2013 period that has both sets of data. The greater proportionate understatement of the Theil index (compared with the understatement of the Gini coefficient) if the DMSP data are used is in line with the pattern seen in Figure 2 for the United States.

In addition to the DMSP data indicating that the level of China's spatial inequality is lower than what either the GDP data or the more accurate VIIRS night-time lights data show, analyses of temporal trends in inequality could also be skewed by using the DMSP data. Over the 2000–2013 period, there was a slight fall in inequality according to GDP data – the Gini coefficient went from 0.80 to 0.79 and the Theil index from 1.45 to 1.36 – but far larger falls in inequality for that period are implied by the DMSP data. The DMSP-derived Theil index went from 1.02 in 2000 to 0.74 in 2013, which was a 28% fall (compared with a fall of just 6% in the GDP-based measure for the same period). Likewise, the DMSP-derived Gini coefficient went from 0.71 to 0.62, and the trend annual rate of decline in

[12]We do not show the inequality estimates from BM data in Figure 2 to avoid clutter, but these are very similar to the inequality estimates from the VNL data, and quite distinct to those from the DMSP data.

[13]Remote sensing researchers often use 'inter-calibrated' DMSP data in which a regression lines up results from the various satellites. This only alters the first moment of the distribution, so inequality estimates are unchanged.

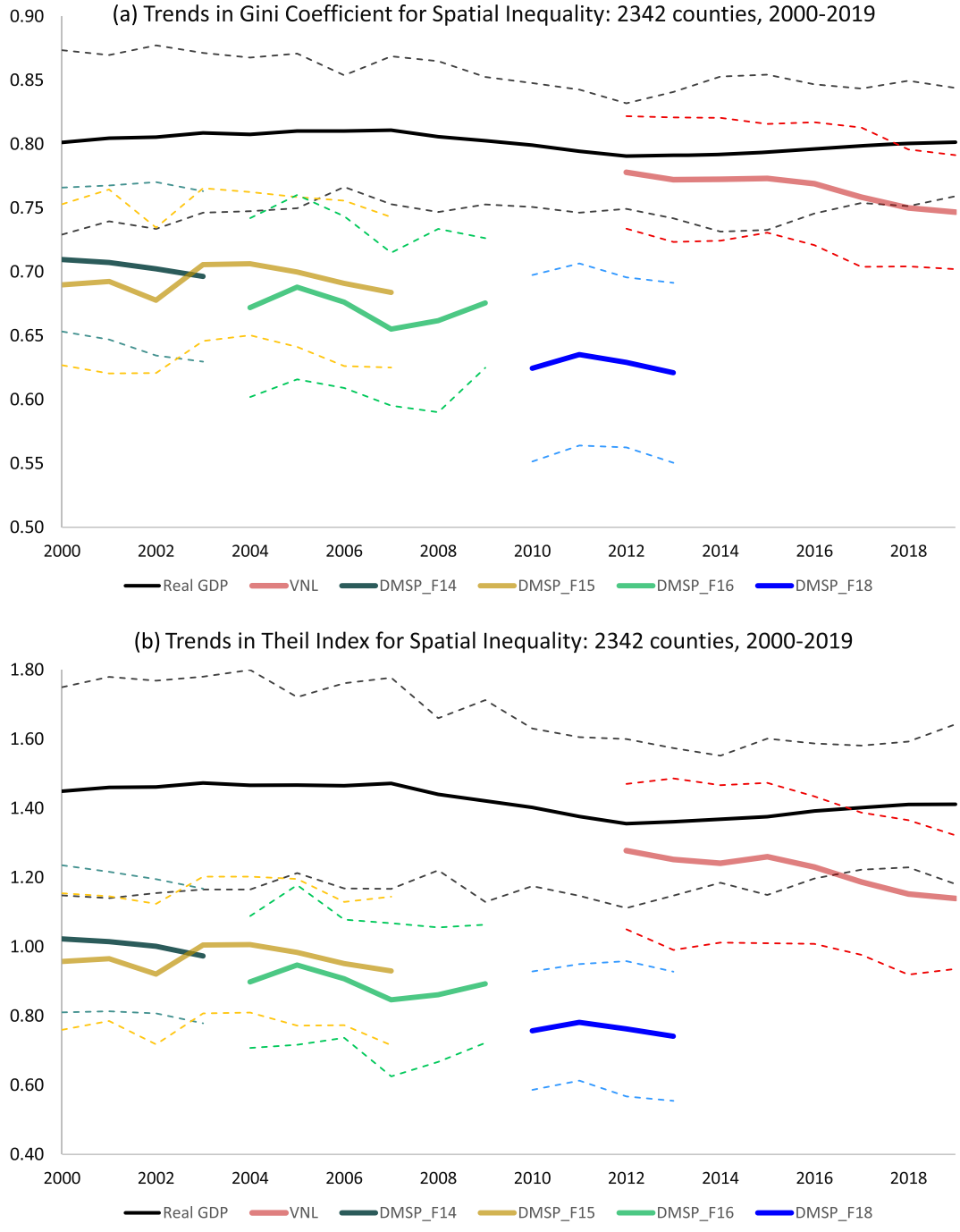


FIGURE 3 County-level spatial inequality in China. *Note*: 95% confidence intervals from bootstrapped standard errors are shown with dashed lines.

the Gini coefficient based on DMSP data for the 2000–2013 period was −1% (with standard error of −0.1%). This annual trend rate of decline with the DMSP data is statistically significantly ($p < 0.01$) greater (i.e., it implies an apparently faster decline) than is the annual trend in the inequality estimates that are derived from the GDP data for

China.[14] In other words, the impact that the errors in the DMSP data have on inequality estimates cannot only be treated as an intercept shift that affects all years (and all inequality measures) equally. Instead, these measurement errors in DMSP data also seem to affect evidence on the temporal trends in spatial inequality.

6 | CONCLUSIONS

The practice of regional science research is increasingly relying on satellite-detected data, especially of night-time lights. Most of these studies rely on the Defense Meteorological Satellite Program, whose original purpose was cloud observation for short-term weather forecasting. This is part of a growing 'big data' movement in social science research. Yet despite the popularity of these DMSP data, they have inherent measurement error features whose implications are largely ignored. In terms of measuring on-the-ground economic activity, DMSP data are flawed by blurring and top-coding due to intrinsic limitations of the sensors and data management. While prior authors have noted that even though DMSP data may be a noisy proxy, they can nevertheless still be useful (e.g., Henderson et al., 2012), the fact that the measurement errors are spatially mean-reverting, and the implications of this error pattern for particular uses of these data, are not emphasized. In particular, spatially mean-reverting errors make the DMSP data poorly suited for estimating spatial inequality, even if they may adequately proxy for the average levels of (or growth in) economic activity at the national or aggregated regional level in other research contexts.

In this paper we explain the sources of these mean-reverting errors and give examples of their impacts on spatial inequality estimates for the two largest countries with county-level GDP data available to use as a benchmark. For the United States we find that the Theil index of inequality is understated by almost 60% if the DMSP data are used, compared with what GDP data show. The understatement of the Gini coefficient is somewhat less but still statistically significant. For China, the Theil index appears to have fallen by 28% over the 2000–2013 period if the DMSP data are used, where this fall implies some leveling of China's spatial distribution of economic activity at a time when various government initiatives aimed to promote lagging regions. However, the GDP data do not show a similarly rapid reduction in spatial inequality; policymakers and regional science practice could thus be inadvertently misled by studies that use DMSP data.

In addition to these examples, which provide grounds to question use of DMSP data for estimating spatial inequality, we reviewed existing studies to see how measurement error issues are considered. Our review shows the literature appears to have ignored these issues. Along the same lines, de-blurring approaches for DMSP data available since 2015 (Abrahams et al., 2018) and data adjusted for top-coding (for all DMSP pixels from 1992 to 2013) available since 2018 (Bluhm & Krause, 2022) have not been used, despite many papers we review post-dating these studies.[15] That inequality studies continue to use DMSP data when more accurate VIIRS Nighttime Lights and Black Marble data are available also suggests that the measurement error problems with DMSP data have not been given sufficient attention (even noting possible trade-offs with length of the time series for newer, more accurate, night-time lights data). Thus, even though the availability of satellite-detected night-time lights data has the potential to overcome some data constraints that historically have limited the range of spatial inequality studies that can be carried out, uncritical use of DMSP data may introduce some new misunderstandings about levels of, and trends in, spatial inequality. It is especially true for places where conventional indicators of economic activity, such as GDP and

[14]The United States also showed this pattern, although it is less visually apparent in Figure 2. With DMSP data, the Gini coefficient was declining at a trend annual rate of −0.2% and the Theil index at −0.4% (with both trends statistically significant at $p < 0.05$) over the 2001–13 period. These declines were both statistically significantly faster (at the $p < 0.04$ level) than the trends in the GDP-based inequality measures. Running the same test for differences in time trends with the VNL-based inequality measures from 2012 to 2019 showed no difference in the VNL-based trend compared with what the GDP-based inequality measures showed over that period.

[15]A reviewer asked if blurring or top-coding matters more. The answer may depend on aggregation level and density. Gibson (2021) found Pareto-adjusted DMSP data understated spatial inequality in Europe almost as much as do the unadjusted DMSP data; Europe is more urbanized than China and more densely populated than the US so even less effect of this adjustment for top-coding should be expected in the current settings. While limited impact of top-coding corrections indirectly points to the importance of blurring, that may vary with how aggregated are the spatial units.

household survey data, are lacking that the measurement errors in the DMSP data may contribute to distorted regional policies, because data poverty limits opportunities to re-examine findings that relied on DMSP data.

ACKNOWLEDGMENTS

Financial support from the Marsden Fund project UOW1901, data and mapping assistance from Chao Li and Geua Boe-Gibson, helpful comments from the ERSA and ANZRSAI conferences and from seminars at Kyoto University and Waseda University, and the hospitality of the Institute of Economic Research at Hitotsubashi University are gratefully acknowledged. Open access publishing facilitated by The University of Waikato, as part of the Wiley - The University of Waikato agreement via the Council of Australian University Librarians.

ORCID

John Gibson https://orcid.org/0000-0003-3886-6873

REFERENCES

Abrahams, A., Oram, C., & Lozano-Gracia, N. (2018). Deblurring DMSP night-time lights: A new method using Gaussian filters and frequencies of illumination. *Remote Sensing of Environment*, *210*(1), 242–258. https://doi.org/10.1016/j.rse.2018.03.018

Alimi, O., Boe-Gibson, G., & Gibson, J. (2022). *Noisy night lights data: Effects on research findings for developing countries*Working Paper 12/22. Department of Economics, University of Waikato.

Black, D., Berger, M., & Scott, F. (2000). Bounding parameter estimates with nonclassical measurement error. *Journal of the American Statistical Association*, *95*(451), 739–748. https://doi.org/10.1080/01621459.2000.10474262

Bluhm, R., & Krause, M. (2022). Top lights: Bright cities and their contribution to economic development. *Journal of Development Economics*, *157*(1), 102880. https://doi.org/10.1016/j.jdeveco.2022.102880

Bluhm, R., & McCord, G. (2022). What can we learn from nighttime lights for small geographies? Measurement errors and heterogeneous elasticities. *Remote Sensing*, *14*(5), 1190. https://doi.org/10.3390/rs14051190

Chakravarty, P., & Dehejia, V. (2017). *India's income divergence: Governance or development model?* Briefing Paper Series. IDFC Institute.

Chen, G., & Zhang, J. (2023). Regional inequality in ASEAN countries: Evidence from an outer space perspective. *Emerging Markets Finance and Trade*, *59*(3), 722–736. https://doi.org/10.1080/1540496X.2022.2119810

Deininger, K., & Squire, L. (1996). A new data set measuring income inequality. *The World Bank Economic Review*, *10*(3), 565–591. https://doi.org/10.1093/wber/10.3.565

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, *71*(1), 355–364. https://doi.org/10.1111/1468-0262.00399

Elvidge, C., Baugh, K., Zhizhin, M., & Hsu, F.-C. (2013). Why VIIRS data are superior to DMSP for mapping night time lights. *Proceedings of the Asia-Pacific Advanced Network*, *35*(1), 62–69. https://doi.org/10.7125/APAN.35.7

Elvidge, C., Zhizhin, M., Ghosh, T., Hsu, F.-C., & Taneja, J. (2021). Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, *13*(5), 922. https://doi.org/10.3390/rs13050922

Ghosh, T., Baugh, K., Elvidge, C., Zhizhin, M., Poyda, A., & Hsu, F.-C. (2021). Extending the DMSP nighttime lights time series beyond 2013. *Remote Sensing*, *13*(24), 5004. https://doi.org/10.3390/rs13245004

Gibson, J. (2021). Better night lights data, for longer. *Oxford Bulletin of Economics and Statistics*, *83*(3), 770–791. https://doi.org/10.1111/obes.12417

Gibson, J., Huang, J., & Rozelle, S. (2001). Why is income inequality so low in China compared to other countries? The effect of household survey methods. *Economics Letters*, *71*(3), 329–333. https://doi.org/10.1016/S0165-1765(01)00357-3

Gibson, J., & Kim, B. (2010). Non-classical measurement error in long-term retrospective recall surveys. *Oxford Bulletin of Economics and Statistics*, *72*(5), 687–695. https://doi.org/10.1111/j.1468-0084.2010.00599.x

Gibson, J., Olivia, S., & Boe-Gibson, G. (2020). Night lights in economics: Sources and uses. *Journal of Economic Surveys*, *34*(5), 955–980. https://doi.org/10.1111/joes.12387

Gibson, J., Olivia, S., Boe-Gibson, G., & Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, *149*(1), 102602. https://doi.org/10.1016/j.jdeveco.2020.102602

Goldblatt, R., Heilmann, K., & Vaizman, Y. (2020). Can medium-resolution satellite imagery measure economic activity at small geographies? Evidence from Landsat in Vietnam. *The World Bank Economic Review*, *34*(3), 635–653. https://doi.org/10.1093/wber/lhz001

Henderson, V., Storeygard, A., & Weil, D. (2012). Measuring economic growth from outer space. *American Economic Review*, *102*(2), 994–1028. https://doi.org/10.1257/aer.102.2.994

Hsu, F.-C., Baugh, K., Ghosh, T., Zhizhin, M., & Elvidge, C. (2015). DMSP-OLS radiance calibrated night-time lights time series with inter-calibration. *Remote Sensing*, *7*(2), 1855–1876. https://doi.org/10.3390/rs70201855

Ivan, K., Holobâcă, I.-H., Benedek, J., & Török, I. (2020). Potential of night-time lights to measure regional inequality. *Remote Sensing*, *12*(1), 33.

Kim, B., Gibson, J., & Boe-Gibson, G. (2022). *How effective are sanctions on North Korea? Popular DMSP night-lights data may bias evaluations due to blurring and poor low-light detection*Working Paper Working Paper 06/22. Department of Economics, University of Waikato.

Kocornik-Mina, A., Michaels, G., McDermott, T., & Rauch, F. (2020). Flooded cities. *American Economic Journal: Applied Economics*, *12*(2), 35–66. https://doi.org/10.1257/app.20170066

Li, H., & Mykhnenko, V. (2018). Urban shrinkage with Chinese characteristics. *The Geographical Journal*, *184*(4), 398–412. https://doi.org/10.1111/geoj.12266

Mendez, C., & Santos-Marquez, F. (2020). Regional convergence and spatial dependence across subnational regions of ASEAN: Evidence from satellite nighttime light data. *Regional Science Policy & Practice*, *13*(6), 1750–1777.

Michalopoulos, S., & Papaioannou, E. (2014). National institutions and sub-national development in Africa. *Quarterly Journal of Economics*, *129*(1), 151–213.

Mirza, M. U., Chi, X., van Bavel, B., & Scheffer, M. (2021). Global inequality remotely sensed. *Proceedings of the National Academy of Sciences*, *118*(18), e1919913118. https://doi.org/10.1073/pnas.1919913118

Montalvo, J., Reynal-Querol, M., & Mora, J. C. M. (2021). *Measuring inequality from above*Working Papers 1252. Barcelona School of Economics.

Román, M., Wang, Z., Sun, Q., Kalb, V., Miller, S., Molthan, A., & Masuoka, E. (2018). NASA's black marble nighttime lights product suite. *Remote Sensing of Environment*, *210*(1), 113–143. https://doi.org/10.1016/j.rse.2018.03.017

Sangkasem, K., & Puttanapong, N. (2022). Analysis of spatial inequality using DMSP-OLS nighttime-light satellite imageries: A case study of Thailand. *Regional Science Policy & Practice*, *14*(4), 828–849. https://doi.org/10.1111/rsp3.12386

Sharp, M., Buffière, B., Himelein, K., Troubat, N., & Gibson, J. (2022). *Effects of data collection methods on estimated household consumption and survey costs. Policy research working paper no. 10029*. World Bank. https://doi.org/10.1596/1813-9450-10029

Small, C., Pozzi, F., & Elvidge, C. D. (2005). Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sensing of Environment*, *96*(3–4), 277–291. https://doi.org/10.1016/j.rse.2005.02.002

Tuttle, B., Anderson, S., Sutton, P., Elvidge, C., & Baugh, K. (2013). It used to be dark here. *Photogrammetric Engineering & Remote Sensing*, *79*(3), 287–297. https://doi.org/10.14358/PERS.79.3.287

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57*(2), 307–333. https://doi.org/10.2307/1912557

Zhang, X., & Gibson, J. (2022). Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. *Remote Sensing*, *14*(5), 1282. https://doi.org/10.3390/rs14051282

How to cite this article: Zhang, X., Gibson, J., & Deng, X. (2023). Remotely too equal: Popular DMSP night-time lights data understate spatial inequality. *Regional Science Policy & Practice*, *15*(9), 2106–2125. https://doi.org/10.1111/rsp3.12716

APPENDIX A: SELECTED STUDIES USING SATELLITE-DETECTED NIGHT-TIME LIGHTS TO MEASURE SPATIAL INEQUALITY (MOST RECENT FIRST)

Authors	Year	Data source	Spatial units	Time period	De-blur?	Adjust for top-coding?	Mention mean-reverting errors?	Research objective
Chen and Zhang	2023	DMSP (2000–2013); VIIRS (2014–2018)	337 regions in 10 ASEAN countries	2000–2018	No	No	No	Investigate trends in subnational inequality in ASEAN countries.
Sangkasem and Puttanapong	2022	DMSP	76 provinces in Thailand	1994–2013	No	No	No	Compare inequality based on night-time lights with survey-based index.
Galimberti et al.	2021	DMSP	234 countries	1992–2013	No	No	No	Study economic inequality drivers and dynamics
Mirza et al.	2021	DMSP	Global and 50 states in the United States	Five-yearly from 1990	No	No	No	Test if remotely sensed luminosity reflects economic inequality.
Montalvo et al.	2021	DMSP	186 countries	2000–2013	No	Yes	No	Develop new measure of night-time lights-based inequality.
Weidmann and Theunissen	2021	VIIRS	37 countries in Africa	2012–2019	No	No	No	Compare inequality based on night-time lights with survey estimates.
Ivan et al.	2020	DMSP (1992–2013); VIIRS (2014–2018)	42 administrative counties in Romania	1992–2018	No	No	No	To measure regional inequality in Romania.
Mendez-Guerra and Santos-Marquez	2020	DMSP	274 subnational regions in ASEAN	1998–2012	No	No	No	Evaluate subnational convergence and regional inequality.
Singhal et al.	2020	DMSP	Districts in 23 states in India	1999–2008	No	No	No	Test if regional inequality based on night-time lights has Kuznets curve.
Elvekjær	2019	VIIRS	402 regions in 36 OECD countries	2015	No	No	No	Compare models that predict GDP and inequality at five spatial scales.
Mukhopadhyay and Urzainqui	2018	DMSP	Villages, urban blocks, and districts in India	2004 and 2011	No	No	No	Decompose inequality into within- and between-area components.

Authors	Year	Data source	Spatial units	Time period	De-blur?	Adjust for top-coding?	Mention mean-reverting errors?	Research objective
Mveyange	2018	DMSP	623 regions in 48 African countries	1992–2012	No	No	No	Small-area imputation: Lights-based inequality used to predict income inequality where unreported.
Puttanapong and Zin	2018	DMSP	14 provinces in Myanmar	1992–2013	No	No	No	Estimate trends in regional inequality in Myanmar.
Sangkasem	2018	DMSP	76 provinces in Thailand	1992–2013	No	No	No	Reveal the clusters of poverty and inequality in Thailand.
Wu et al.	2018	VIIRS	341 prefectures and 31 provinces in China	2014–2017	No	No	No	Estimate trends in Theil index and in inequality components
Chakravarty and Dehejia	2017	DMSP	676 districts in India	1992–2013	No	No	No	Test for intra-state divergence in economic activity
Hu et al.	2017	DMSP	13 prefectural cities in Beijing-Tianjin-Hebei	2000–2012	No	No	No	Estimate trends in spatial inequality and identify spatial clusters.
Lessmann and Seidel	2017	DMSP	3,163 regions in 180 countries	1992–2012	No	No	No	Study convergence and determinants of regional inequality.
Li et al.	2017	DMSP	41 regions in five cent Asian countries	1993–2012	No	No	No	Identify poor areas and measure regional development inequality.
Weidmann and Schutte	2017	DMSP	39 less developed countries	2003–2012	No	No	No	Test if luminosity can predict local wealth, as detected by surveys.
Alesina et al.	2016	DMSP	Ethnic group areas in 173 countries	1992–2009	No	No	No	Examine linkages between ethnicity, inequality, and economic development.
Mveyange	2015	DMSP	423 regions in 32 African countries	1995, 2000, 2005	No	No	No	Test if luminosity can be used to proxy for regional inequality.

(Continues)

Authors	Year	Data source	Spatial units	Time period	De-blur?	Adjust for top-coding?	Mention mean-reverting errors?	Research objective
Xu et al.	2015	DMSP	8 regions, 31 provinces, 354 prefectures in China	2005 and 2010	No	No	No	Measure regional differences in inequality of public services.
Zhou et al.	2015	VIIRS	30 provinces and municipalities inmainland China	2012	No	No	No	Study socioeconomic inequality patterns in China using luminosity data.
Gennaioli et al.	2014	DMSP	1,528 regions from 83 countries	1992–2010	No	No	No	Investigate how within-country regional inequality drives regional convergence across countries.
Elvidge et al.	2012	DMSP	Global, national (189), subnational (4550), and gridded	2006	No	No	No	Explore potential of night-time lights data for estimating disparities in income distribution.

DMSP, Defense Meteorological Satellite Program.

REFERENCES FOR APPENDIX A (MOST RECENT FIRST)

Chen, G., & Zhang, J. (2023). **Regional inequality in ASEAN countries: Evidence from an outer space perspective**. Emerging Markets Finance and Trade, 59(3), 722–736.

Sangkasem, K., & Puttanapong, N. (2022). **Analysis of spatial inequality using DMSP-OLS nighttime-light satellite imageries: A case study of Thailand**. Regional Science Policy & Practice, 14(4), 828–849.

Galimberti, J., Pichler, S., & Pleninger, R. (2021). *Measuring inequality using geospatial data. KOF working papers 493*. KOF Swiss Economic Institute.

Mirza, M. U., Chi, X., van Bavel, B., & Scheffer, M. (2021). **Global inequality remotely sensed**. Proceedings of the National Academy of Sciences, 118(18), e1919913118.

Montalvo, J., Reynal-Querol, M., & Mora, J. C. M. (2021). *Measuring inequality from above*. Working Papers 1252. Barcelona School of Economics.

Weidmann, N., & Theunissen, G. (2021). **Estimating local inequality from nighttime lights**. Remote Sensing, 13(22), 4624.

Ivan, K., Holobâcă, I.-H., Benedek, J., & Török, I. (2020). **Potential of night-time lights to measure regional inequality**. Remote Sensing, 12(1), 33.

Mendez, C., & Santos-Marquez, F. (2020). **Regional convergence and spatial dependence across subnational regions of ASEAN: Evidence from satellite nighttime light data**. Regional Science Policy & Practice, 13(6), 1750–1777.

Singhal, A., Sahu, S., Chattopadhyay, S., Mukherjee, A., & Bhanja, S. (2020). **Using night time lights to find regional inequality in India and its relationship with economic development**. PLoS ONE, 15(11), 1–18.

Elvekjaer, N. M. (2019). *Disaggregating economic inequality from space*. Thesis. Department of Physical Geography and Ecosystem Science, Lund University.

Mukhopadhyay, A., & David, G. U. (2018). *The dynamics of spatial and local inequalities in India*. WIDER Working Paper. United Nations University (UNU), World Institute for Development Economics Research (WIDER).

Mveyange, A. (2018). *Measuring and explaining patterns of spatial income inequality from outer space: Evidence from Africa*. World Bank Policy Research Working Paper No. 8484. Development Economics Research Group, The World Bank.

Puttanapong, N., & Zin, S. Z. (2018). **Spatial inequality in Myanmar during 1992–2016: An application of spatial statistics and satellite data**. The Social Science Review: Special Issue on Mekong Economy, 156, 161–182.

Sangkasem, K. (2018). Poverty and inequality assessment using DMSP/OLS nighttime light satellite imageries at provincial level in Thailand. Thesis: Thammasat University.

Wu, RongWei, Yang, Degang, Jiefang Dong, Lu Zhang, and Fuqiang Xia (2018) **Regional inequality in China based on NPP-VIIRS night-time light imagery**. Remote Sensing, 10(2): 240.

Chakravarty, P., & Dehejia, V. (2017). *India's income divergence: Governance or development model?* Briefing Paper Series. IDFC Institute.

Hu, Y., Peng, J., Liu, Y., Yueyue, D., Li, H., & Jiansheng, W. (2017). **Mapping development pattern in Beijing-Tianjin-Hebei urban agglomeration using DMSP/OLS nighttime light data**. Remote Sensing, 9(7), 760.

Lessmann, C., & André, S. (2017). **Regional inequality, convergence, and its determinants—A view from outer space**. European Economic Review, 92, 110–132.

Li, S., Zhang, T., Yang, Z., Li, X., & Huimin, X. (2017). **Night time light satellite data for evaluating the socioeconomics in Central Asia**. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42(1237–1), 243.

Weidmann, N. B., & Schutte, S. (2017). **Using night light emissions for the prediction of local wealth**. Journal of Peace Research, 54(2), 125–140.

Alesina, A., Michalopoulos, S., & Papaioannou, E. (2016). **Ethnic inequality**. Journal of Political Economy, 124(2), 428–488.

Mveyange, A. (2015). *Night lights and regional income inequality in Africa. WIDER working paper*. The United Nations University World Institute for Development Economics Research (UNU-WIDER).

Xu, H., Yang, H., Li, X., Jin, H., & Li, D. (2015). **Multi-scale measurement of regional inequality in mainland China during 2005–2010 using DMSP/OLS night light imagery and population density grid data**. Sustainability, 7(10), 13469–13499.

Zhou, Y., Ma, T., Zhou, C., & Tao, X. (2015). **Nighttime light derived assessment of regional inequality of socioeconomic development in China**. Remote Sensing, 7(2), 1242–1262.

Gennaioli, N., La Porta, R., De Silanes, F. L., & Shleifer, A. (2014). **Growth in regions**. Journal of Economic Growth, 19(3), 259–309.

Elvidge, C., Baugh, K., Anderson, S., Sutton, P., & Ghosh, T. (2012). **The night light development index (NLDI): A spatially explicit measure of human development from satellite data**. Social Geography, 7(1), 23–35.

APPENDIX B: DETAILS ON THE DATA SOURCES FOR THE COUNTY-LEVEL ANALYSIS IN SECTION 5

The county-level GDP data for the United States are from the Bureau of Economic Analysis (BEA) at: https://www.bea.gov/data/gdp/gdp-county-metro-and-other-areas. We used chained series in 2012 US dollars. Annual estimates are provided separately for each county for the 2001–2019 period, with three broad exceptions. In Alaska the BEA combine some census areas in their reporting, and in Hawaii they combine Maui and Kalawao counties. Virginia has the most adjustments because the BEA creates 23 combination areas where one or two independent cities whose population in the 1980 census was less than 100,000 are combined with an adjacent county. Overall there are $n = 3{,}109$ counties and combination areas (we refer to all of these as county-level units or just counties) with data available in each year.

In contrast to the United States, subnational GDP data for China are not provided by a single unified source. Instead we used three types of publications from the National Bureau of Statistics (NBS) to build our GDP database: annual editions of the China Statistical Yearbook (county-level) (in Chinese, *Zhongguo Xianyu Tongji Nianjian* [*Xianshi Juan*]), annual editions of the China City Statistical Yearbook (in Chinese, *Zhongguo Chengshi Tongji Nianjian*), and annual editions of the Statistical Yearbook for each city or province (for example, the Beijing Statistical Yearbook) (NBS, various dates). Each edition reports on GDP the previous year, so we use the 2001–2020 editions to obtain annual GDP data from 2000 to 2019. When any county (the typical spatial unit in rural areas) or district (the typical unit in urban areas) was subsequently merged, we enforce the same aggregation on earlier years to have a consistent 2000–2019 time series for each spatial unit. Overall, we have annual GDP data for each of $n = 2{,}342$ units at the third level of the subnational administrative hierarchy, where these units maintain a consistent spatial definition from 2000 to 2019. We refer to these as either county-level units or just counties.

The DMSP data are annual composites from satellites F14, F15, F16, and F18 that collectively cover each year from 2000 to 2013 (with some years having two satellites providing data). The stable lights product (where 'stable' simply means ephemeral lights such as from fires are removed, it does not necessarily imply temporal consistency) provides 6-bit digital numbers (DN) ranging from 0 to 63. The technical details on the steps used to create these data are in Baugh et al. (2010) and the data are available at: https://eogdata.mines.edu/products/dmsp/.

The VNL data are version 2.1 annual composites for 2012–2019 from Elvidge et al. (2021) on the basis of monthly cloud-free radiance averages coming from the Suomi/NPP satellite. These data undergo an initial filtering to remove extraneous features such as fires and aurora before the resulting rough annual composites have further outlier removal procedures applied. The lit grid cells are isolated from background noise using multi-year thresholds that make these data better for change detection than were the earlier vintages of VNL data that used single-year thresholds. The data are in units of nanowatts per square centimeter per steradian ($nW/cm^2/sr$) that are presented on a 15 arc-second output grid available from: https://eogdata.mines.edu/nighttime_light/annual/v21/.

The NASA Black Marble annual composites are derived from the same satellite as VNL V2.1 data but are processed differently (Román et al., 2018). The data products are corrected for atmospheric, terrain, vegetation, snow, lunar, and stray light effects on the radiance values, which are calibrated across time and are also validated against ground measurements. The data are in units of nanowatts per square centimeter per steradian ($nW/cm^2/sr$), with 16-bit precision on a 15 arc-sec output grid. We use the all-angle composites for snow-free nights, available from: https://ladsweb.modaps.eosdis.nasa.gov/achive/allData/5000/VNP46A4/.

REFERENCES

Baugh, K., Elvidge, C., Ghosh, T., & Ziskin, D. (2010). **Development of a 2009 stable lights product using DMSP-OLS data**. Proceedings Asia-Pacific Advanced Networks, 30, 114. https://doi.org/10.7125/APAN.30.17

Elvidge, C., Zhizhin, M., Ghosh, T., Hsu, F.-C., & Taneja, J. (2021). **Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019**. Remote Sensing, 13, 922. https://doi.org/10.3390/rs13050922

National Bureau of Statistics [NBS]. (2001-2020). *China statistical yearbook (county-level)[M]*. China Statistics Press.

National Bureau of Statistics [NBS]. (2001-2020). *China city statistical yearbook[M]*. China Statistics Press.

National Bureau of Statistics [NBS]. (2001-2020). *Beijing statistical yearbook[M]*. China Statistics Press.

Román, M., Wang, Z., Sun, Q., Kalb, V., Miller, S., Molthan, A., Schultz, L., Bell, J., Stokes, E., Pandey, B., & Seto, K. (2018). **NASA's black marble nighttime lights product suite**. Remote Sensing of the Environment, 210, 113–143. https://doi.org/10.1016/j.rse.2018.03.017

DOI: 10.1111/rsp3.12716

Resumen. Los estudios económicos y de las ciencias regionales utilizan cada vez más los datos de las luces nocturnas del Programa de Satélites Meteorológicos de Defensa (DMSP, por sus siglas en inglés) para medir la desigualdad espacial. Estos datos del DMSP son un pobre indicador indirecto en este contexto debido a sus errores de reversión espacial a la media, que arrojan estimaciones de desigualdad significativamente más bajas que las que muestran los datos del PIB subnacional. Las estimaciones de desigualdad del DMSP son también inferiores a las que muestran los satélites más nuevos, orientados hacia la investigación y más precisos. Se demuestra este sesgo utilizando datos a nivel de condado de China y de Estados Unidos. Los errores en los datos del DMSP distorsionan las estimaciones tanto del nivel de desigualdad espacial como de la tendencia.

抄録: 地域科学と経済学の研究では、空間的不平等を測定するために、防衛気象衛星計画 (Defense Meteorological Satellite Program:DMSP)の夜間光データを使用することが多くなっている。これらのDMSPデータは、その空間的平均回帰誤差のため、この文脈では貧弱なプロキシであり、これは地方GDPデータが示すものよりも著しく低い不平等推定値をもたらす。DMSPからの不平等推定値も、より新しい、研究に焦点を当てた、より正確な衛星が示すものよりも低い。中国と米国の郡レベルのデータを用いてこのバイアスを実証した。DMSPデータの誤差は、空間的不平等のレベルと傾向の両方の推定値を歪めるものである。

Chapter 4: How well do gridded population estimates proxy for actual population changes? Evidence from four gridded data products and three population censuses for China

Xiaoxuan Zhang and John Gibson

Abstract

High-resolution gridded population estimates are increasingly used to support public health, disaster, and socio-economic research. These gridded data allow phenomena to be studied at a finer spatial scale than the usual survey or administrative data (spatialization) and with higher frequency than typical decadal census data allow (temporal interpolation). However, little is known about how accurately these gridded data follow actual changes in population. Therefore, we use China's census data for 2000, 2010, and 2020 to test predictive accuracy of four popular gridded population data products, conducting our tests at three spatial levels (county/district, prefectural city, and province). The gridded population data are accurate cross-sectional predictors at all three spatial levels, with less than five percent of variation unexplained. They far less accurately predict temporal changes in population, especially for disaggregated spatial units (counties and districts) where just one fifth of the variation in population changes is predicted by the gridded data. Predictive performance of gridded data for population changes has fallen substantially in the last decade. We illustrate how these inaccurate predictions could distort analyses that examine trends in spatial inequality. Overall, our results suggest that caution is required in using these gridded data products as proxies for the actual changes in local population.

JEL Codes

R12

Keywords

gridded population data
cross-sectional
time-series
Census
China

Acknowledgements

Financial support from the Marsden Fund project UOW1901.

Introduction

High-resolution gridded population estimates are increasingly used to support a wide range of research. These grids are more flexible than relying on administrative boundaries, as they are often one km^2 or smaller, and are available at higher frequency than census data (Xu et al. 2021). Since the first globally available Gridded Population of the World (GPW) introduced in 1995, many global or regional gridded population datasets have been produced, mirroring the growing role of remotely sensed data in socio-economic research.[1] These datasets provide gridded population information for different temporal and spatial scales, and have been widely used for disaster assessments and risk management (Opdyke, and Fatima 2024), for analyzing land use change (Balk et al. 2019), in public health research and management (Colón-González et al. 2023; Dodd et al. 2024), for analyzing ecological and environmental change (Jing et al. 2024, Dasgupta et al. 2024), and for other socio-economic research (Ma et al. 2024; Melchiorri et al. 2024; Zeng et al. 2024). A motivation for using these gridded data is that analyses should be inclusive, omitting no-one from consideration (Freire et al. 2020).

The accuracy of gridded population estimates relies on source data, spatialisation methods, and environmental features (Chen et al. 2020; Thomson et al. 2022). For example, some gridded products, such as WorldPop and GRUMP, generate population projections over time and space based on remotely sensed night lights data to model the spatial distribution of population (Lloyd et al. 2019; Ye et al. 2019; Liu et al. 2024). Yet the extant evidence is that luminosity data poorly extract demographic and socioeconomic characteristics in complex environments (Zhao et al. 2019; Gibson et al. 2021). For example, Pérez-Sindín et al. (2021) find that luminosity predicts worst in low density places with high forest cover, while Chen and Nordhaus (2015) find that DMSP nighttime lights data cannot detect dim lights in parts of Africa with low population density and low productivity. Specifically for China, Gibson et al. (2024) shows that satellite-detected luminosity data are about six-times more responsive to changes in build-up area in cities than for the same extent of built-up area in villages; hence,

[1] A non-exhaustive list includes the Global Rural-Urban Mapping (GRUMP) project, LandScan Population data, Global Urban Footprint (GUF), High-Resolution Settlement Layer (HRSL), WorldPop datasets, the Global Human Settlement Population Layer (GHS-POP), Global Resource Information Database (UNEP/GRID), OpenPopGrid, and the China Gridded Population (CnPop) datasets. Our quantitative review is in Section II.

gridded data for non-urban areas are likely to be far less accurate than for urban areas.

A further concern is that luminosity data poor predict local changes in economic activity (which correlate with population change), even where the cross-sectional prediction is good (Goldblatt et al, 2020; Sun et al. 2020; Asher et al. 2021). For example, Zhang and Gibson (2022) found the newest and most accurate satellite-detected luminosity data could account for about 60 percent of the cross-county variation in GDP in China, but only about one percent of the annual changes. Likewise, Dong et al. (2021) showed that the changes in luminosity data are inconsistent with the long-term changes in local population in China. These findings raise the question of whether some of these problems in the uses of luminosity data carry over into the gridded population estimates.

It is therefore important to test how well gridded population data predict temporal changes, which requires a setting with on-the-ground changes in population location. China provides a good setting for such a test because of shifting migration patterns from 2000-2010 to 2010-2020. In particular, population growth recently shifted from the largest urban areas towards smaller cities. As an example of this trend, one of three previous hotspots for inward migration, the Beijing-Tianjin region, is no longer a hotspot and Tianjin actually had fewer resident migrants in the 2020 census than in the 2010 census (Li et al. 2024). In contrast, for 2000 to 2010 the Yangtze River Delta (YRD), Pearl River Delta (PRD), and Beijing-Tianjin were the main migration destinations (Cao et al. 2018). These shifts in choices of destination cities and migrant motivations (Wang and Zhang, 2022; White et al. 2024) have tended to favour smaller cities for quality of life reasons in the last decade (Li and Xiao, 2022; Yang et al. 2022), reflecting a new multipolar era of migration patterns (Cao et al. 2023).

In this paper we examine how accurately gridded population estimates can capture the sharp change in population locations in China over the last two decades. Existing studies have neither included a systematic review of the different gridded population products nor a comparison of their performance in predicting population change. Because the gridded data are used for both cross-sectional and time-series studies, we also compare predictive accuracy in these two dimensions. We use China's population census data (2000, 2010, and 2020) as a benchmark to examine whether four popular gridded population datasets are good predictors

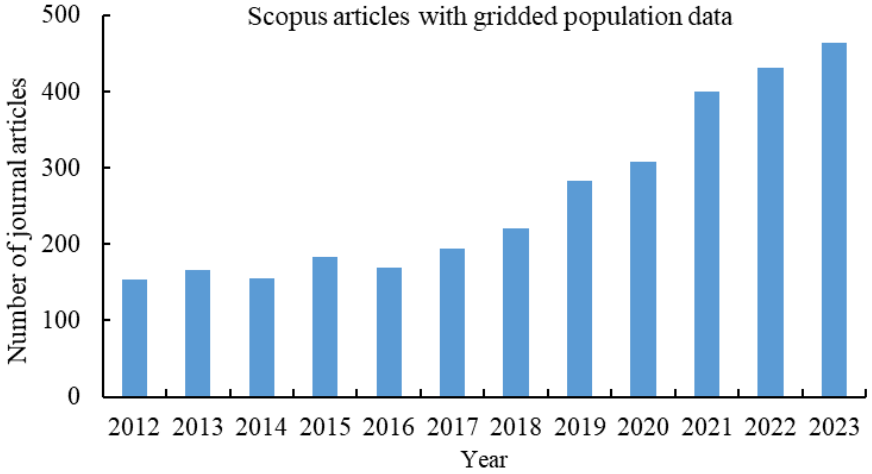
of population change. We conduct our tests at three spatial levels (county/district, prefectural city, and province) so that future users of these gridded data may be better informed about the spatial scale where these datasets can be best used as proxies for actual population change.

This paper is organized as follows: Section 2 reviews the application of gridded population datasets and discusses the main contribution of the study. Section 3 give the background of the changes in China's population migration patterns. Section 4 introduces the data sources, and the evaluation framework used in the analysis. Section 5 presents the empirical results, comparing the prediction accuracy of multiple gridded population datasets at three administrative levels. Finally, Section 6 includes the main conclusions.

Literature Review

To provide quantitative evidence on the growing use of gridded population data in research in various disciplines we conducted a search in the Scopus database for journal articles that had any mention of GPW, GHS-POP, LandScan or WorldPop. About 150 articles per year were published using these gridded data products a decade ago but the output is now far higher, at about 500 articles per year (Fig 1). The growing use of these gridded data is particularly apparent in the field of environmental sciences, and earth and planetary sciences.

Figure 1. Annual output of articles using gridded population data



Notes: Search of Scopus database on 31 July, 2024. These are annual counts, rather than cumulative.

Amongst journal articles using gridded data products, several exploit the underlying cross-sectional variation in population density to study issues like accumulated ground heat (Benz et al. 2022), flood vulnerability (Yang et al. 2024), noise threshold values (Yasin and Alistair, 2019), and ecosystem services (Peng et al. 2024). These articles are especially prominent in the environmental science field, which is the discipline (under Scopus subject headings) that makes the most use of gridded population data. For example, using GPW data as a measure of population density, Unfried, Kis-Katos and Poser (2022), and Rathore et al. (2024) assessed the influence of irrigation expansion and water scarcity in Africa and the United States.

A reliance on the spatial variation in gridded population datasets is also seen in some of the economics studies using these data. For example, Dijkstra et al. (2022) use the WorldPop data to provide a harmonized measure of global urbanization. Using the LandScan data and GPW data, Bosker et al. (2021) and Gollin et al. (2023) study agglomeration effects in Indonesia and Africa. González et al. (2023) used GPW data in their study of regional favouritism in Argentine districts and its impact on economic development. These population datasets are also used for modelling studies that aim to extend infrastructure to the unreached; for example, Oughton et al. (2023) used WorldPop data to evaluate the necessary investment requirements to achieve affordable universal broadband.

Studies relying on cross-sectional variation make up the majority of articles using gridded population estimates but there are also studies of *changes*, which inherently rely on the time-series variation. For example, Muhwezi et al. (2021) use the WorldPop estimates of annual population changes to explore drivers of rising electricity consumption in Kenya from 2010 to 2015. Maldonado (2023) studied rural poverty trends from 2000 to 2020 for states and municipalities in Venezuela, using annual gridded population and luminosity datasets. Likewise, gridded data are used by Patias et al. (2022) to study socioeconomic inequalities at various geographical levels in Britain for a 40-year period from 1971 to 2011. The changing flood risks due to hurricanes in Puerto Rico are studied using annual population density grids from WorldPop estimates for 2005 to 2016 (Archer et al. 2024). Changes in CO_2 emissions and air pollution are related to changing population density in GHS-POP and GPW gridded

estimates (Castells-Quintana et al. 2021; Wilmot et al. 2024). Likewise, annual gridded population density estimates from WorldPop are used by Joseph (2022) to estimate impacts of the 2010 earthquake in Haiti on economic growth and recovery while Bourget et al. (2024) used GPW data to provide climate change risk assessments of pluvial flooding for insurance portfolios for Canada and the United States.

In summary, rapid advancement of remote sensing has enabled a profusion of gridded population estimates, which are increasingly used in different research fields. Our quantitative review suggests that WorldPop and GPW are the most widely used gridded data products, due to their wide coverage, ease of use (and the annual frequency for WorldPop facilitates study of short-term changes). GHS-POP estimates are also widely used, partly due to their fine spatial resolution (100m), and the annual frequency of LandScan estimates also make them a popular data source. However, little attention has been paid to the accuracy of these gridded products, particularly for studying time series changes, which motivates our interest to test how well the gridded population datasets can predict the actual population changes over time, at different spatial scales.

The test setting: China's changing spatial population distribution

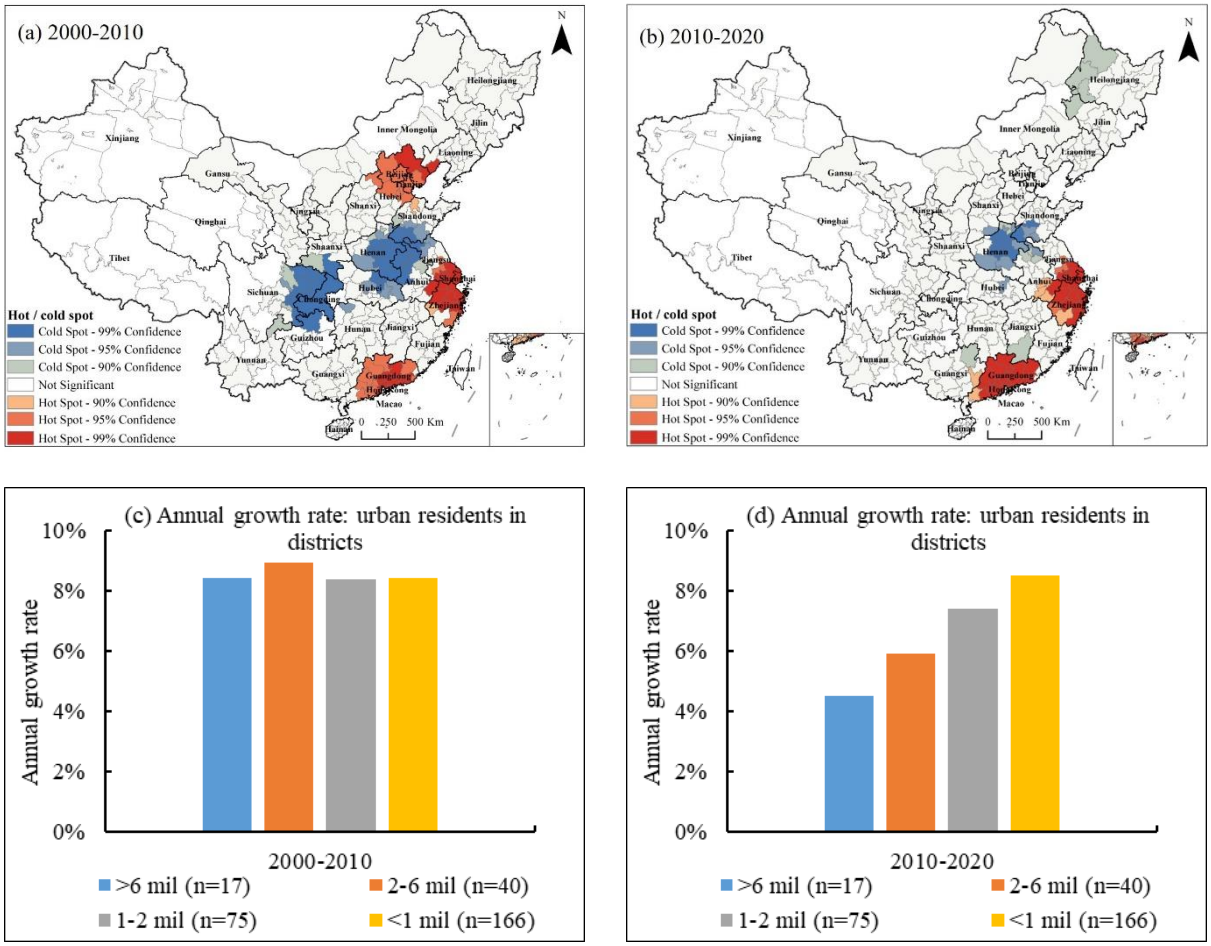
China is a useful setting for testing the accuracy of the gridded data products because over the last several decades it is home to the largest migration flow in history (Gu et al. 2020; Niu, 2022). In 1955, China established an internal registration ("Hukou") system, to control rural migration to urban areas (Chan 2009); this produced a population that was both too rural and too inland to take advantage of new economic opportunities in coastal regions after China's reform and opening up in 1978 (Shi et al. 2024). Once mobility controls were less stringently enforced, a massive population flowed out of the relatively underdeveloped central and western provinces (Zhou et al. 2024). Taking advantage of convenient transport facilities and information channels, the scale of population migration has increased in the past decade (Yang et al. 2022). As of China's Seventh National Census in 2020, 376 million people were living in a different prefecture than their place of registration (inter-provincial 125 million, intra-provincial 251 million) (Cheng and Duan, 2021). Compared to the sixth census in 2010, this is a significant increase of 155 million (almost 70% growth) more migrants.

Adding to the advantage of using China to test the predictive accuracy of gridded population products is a sharp change in destinations of this migrant flow, from 2000-2010 to 2010-2020. In the earlier decade, the Yangtze River Delta (YRD), Pearl River Delta (PRD), and Beijing-Tianjin-Hebei regions were all hot spots for inward migration, while cold spots (sources for outward migration) were less developed central regions such as Chongqing and parts of Henan, Anhui, and Sichuan provinces (Figure 2a). This pattern changed in the most recent decade, with the Beijing-Tianjin–Hebei region ceasing to be a migration hotspot (in fact the 2020 census recorded fewer migrants in Tianjin than there were in 2010). The cold spots also receded to mainly be centered on Henan (Figure 2b).

These changes in the origins and destinations of the migrant flows from 2000-2010 to 2010-2020 are also seen in changes in the patterns of population growth rates by city size category (Figures 2c and 2d).[2] Between 2000 and 2010 large cities and small cities all grew at about the same rate of eight percent per annum; notably this growth is almost entirely due to migration rather than natural increase and so it implies that cities across a range of sizes were equally attractive to migrants (thresholds for the size categories are chosen to roughly double the number of cities in the next group when moving down one category). However, the most recent decade saw the largest cities (with > 6 million residents) growing at only one-half the rate of the smallest cities (with < 1 million residents). Instead of migrants going into the big urban areas like Beijing-Tianjin they were increasingly moving to smaller cities. These sharp changes in the sources and destinations for migration flows, and the consequent changes in the patterns of population growth rates by city size should provide a challenging test for the ability of the gridded population products to predict these on-the-ground changes.

[2] The spatial units in panels a and b of Figure 2 are prefectural-level cities, but these are not 'cities proper' because they often include low density, largely rural, counties. In contrast, the units in panels c and d of Figure 2 are 'districts' (or *shiqu*); these are one of three types of 3rd level sub-national units and have a mean population density that is almost ten-times higher than for counties, which are the other common 3rd level unit.

Figure 2. Hot/cold spots of China's net migration (panels a and b) and population growth rates by city size category (panels c and d), by decade



Data and Methods

Data Source

We use census data (and GDP for the inequality analysis) for all of the parts of China that are organized as prefectural-level cities (rather than other 2nd level units, such as banners). This covers more than 94% of China's resident population according to the seventh population census in 2020. In this sample there are a total of 31 provinces, 297 prefectural-level cities, and 2851 county-level units (districts, counties, and county-level cities). The analyses are carried out at all three spatial levels, to test whether the predictive performance of the gridded data varies with the level of spatial aggregation.

Most gridded datasets use dasymetric models to estimate the spatial distribution of population, combining census data with other spatial data (e.g., land cover) to disaggregate

population counts across grid cells. The “top-down” approach spreads population counts into small grid cells, as used, for example, by GHS-POP. Simple top-down approaches assume a uniform distribution of population within spatial units (e.g., GPWv4), while more complex approaches incorporate ancillary data (e.g., land cover, night-time lights) to generate weights for allocating population. In contrast, “bottom-up” approaches typically rely on micro-census samples and build geo-statistical relationships between micro-census population density and the built environment to predict population counts across grid cells of unsampled areas (e.g., LandScan). Considering the availability and diversity of the gridded population datasets, we chose the GPW, GHS-POP, WorldPop, and LandScan data to evaluate the performance of gridded population data sets in capturing the changes in population patterns. The census data for 2000, 2010, and 2020 are used as the benchmark. Table 1 presents some characteristics of the four gridded population datasets that we work with.

The GPW dataset, developed by the NASA Socioeconomic Data and Applications Center (SEDAC), is now in its fourth version (GPW4). This data product has gridded data on total population counts and densities and other key demographic variables, globally at a nominal spatial resolution of 1 km. In contrast, WorldPop was developed collaboratively by multiple organizations and institutions to meet needs of a wide range of users. In this study we used the unconstrained individual country UN adjusted population count from the WorldPop dataset with a spatial resolution of 100-m for 2000, 2010, and 2020.

The Global Human Settlement Population Grid (GHS-POP) is the latest released global gridded population dataset based on remotely sensed data developed by the EU Joint Research Centre, and this depicts the distribution and density of the total population as the number of people per grid cell (Schiavina et al. 2023). Finally, we also use the LandScan dataset for 2000, 2010, and 2020, obtained from the Oak Ridge National Laboratory (ORNL), and providing a high-resolution global population distribution dataset that has been used for a wide range of applications (Sims et al. 2023). Leveraging state-of-the-art spatial modelling techniques and advanced geospatial data sources, LandScan provides detailed information on population counts and density at a 30 arc-second resolution (ca. 1 km^2), enabling precise and up-to-date insights into human settlement patterns across the globe.

Table 1. Summary of gridded population datasets used in this study

Datasets	WorldPop	GPWv4.11	GHS-POP	LandScan
Temporal range	2000-2020	2000, 2005, 2010, 2015, 2020	1975-2020	1998 and 2000-2019
Resolution	100 m	1 km	100 m	1 km
Coverage	Global	Global	Global	global
Frequency	Annual	Every 5 year	Every 5 year	Annual
Predictor variables	Roads, Land Cover, Built structures, NTL, Infrastructure, Environmental data, Water Bodies	Water Bodies	Built structures	Roads, Land Cover, Built structures, Infrastructure, Environmental data, Water Bodies
Sources	WorldPop, University of Southampton	SEDAC	GHSL project	ORNL
Access Link	https://www.worldpop.org/	https://sedac.ciesin.columbia.edu/	https://www.resdc.cn/	https://landscan.ornl.gov

Note: GPW is Gridded population of the world, SEDAC is the Socioeconomic Data and Applications Center, RESDP is Resource and Environmental Science Data Platform, ORNL is the Oak Ridge National Laboratory, GHS is Global Human Settlement, GHSL is the Global Human Settlement Layer.

Estimation Framework

To test the predictive accuracy of gridded population estimates the following regression model is used, where it is estimated separately at county-level, prefectural city level, and at province level, and separately for each of the four gridded data products:

$$\ln(census\_pop)_{it} = \alpha + \beta \ln(grid\_pop)_{it} + \mu_i + \varphi_t + \varepsilon_{it} \quad (1)$$

where the *i* indexes cross-sectional units; the *t* indexes years; the μ_i are fixed effects for each cross-sectional unit; the φ_t are fixed effects for each year; and ε_{it} is the disturbance term. The parameters of greatest interest are β, the elasticity of census population with respect to the gridded population estimate, and also the R^2 values showing the predictive performance.

The fixed effects control for the influence of unobserved time-invariant features of each cross-sectional unit, and the influence of spatially-invariant features of each time period. The equation is estimated with a 3-period panel, using data for 2000, 2010, and 2020.

The estimation framework allows two different types of elasticities and R^2 which relate to the cross-sectional and time-series variation in panel data. If equation (1) is averaged over time, we get:

$$\overline{\ln(census\_pop)}_i = \beta^B \overline{ln(gridded\_pop)}_i + \mu_i + \bar{\varepsilon}_i \quad (2)$$

where time-averaged values of (log) census population and gridded population are used in a cross-sectional Ordinary Least Squares regression, to yield β^B, the *between estimator*. The relationship in equation (2) refers to the long-run and shows how well time-averaged gridded data might predict the location of the population.

The second type of elasticity, β^W, is the *within estimator*, where the variation over time within each cross-sectional unit provides the basis for estimating:

$$\ln(census_{pop})_{it} - \overline{\ln(census_{pop})}_i =$$

$$\beta^W\left[\ln(gridded\_pop)_{it} - \overline{ln(gridded\_pop)}_i\right] + \varepsilon_{it} - \bar{\varepsilon}_i \quad (3)$$

Equation (3) is based on subtracting equation (2) from equation (1); in doing so it removes effects of spatially invariant fixed effects (e.g. topography). Note that equation (3) provides equivalent estimates to what would be obtained by estimating equation (1) with separate intercepts for every cross-sectional unit. The within estimator allows examination of the time-series fluctuations in population within areas and specifically provides a basis for assessing how well gridded estimates can predict *changes* in the location of population. Further details on the within and between estimators can be seen in Zhang and Gibson (2022).

The within estimator is based on the changes in population from 2000 to 2010, and from 2010 to 2020, for the same places. However, it does not allow a direct comparison of the predictive performance in these two decadal sub-periods, so we also use the following regression, where *T*=1 for the second decade and zero otherwise:

$$\Delta(census\_pop)_{it} = b_0 + d_0 * T + b_1 T * \Delta(gridded\_pop)_{it} \quad (4)$$

This equation allows a Chow test to be estimated, for whether there is a change between the 2000-2010 period and the 2010-2020 period, in the degree of predictability of the changes in population between censuses, using the changes in the gridded population estimates as the predictor variable.

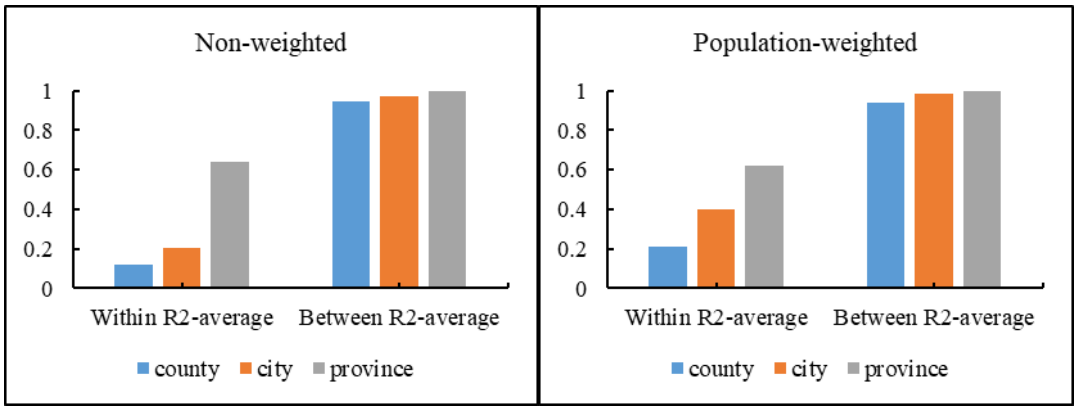
Results

Within and Between Estimator Results at Three Aggregation Levels

Tables 2-4 have within and between estimator results, using (log) census population in 2000, 2010 and 2020 as the dependent variable, and the four sets of gridded population estimates (WorldPop, GPW, GHS, and Landscan) as the predictor variables. The results for each administrative level (county-level units, prefectural cities, and provinces) are each in their own table, with within estimator results presented first and then the between estimator results. We also present population-weighted results in case impressions about predictive accuracy are affected by patterns for spatial units with not many people, noting the large imbalance in the population density between eastern and western regions of China.

We begin with a figure that summarizes patterns of predictive performance as the level of spatial aggregation changes. In order to reduce clutter, the R^2 values from each model for each gridded data product are averaged, because spatial aggregation patterns are common across the four data products. The level of spatial aggregation does not make much difference to the predictive performance for the between estimator, with R^2 values close to 1.0 at the county, prefectural city and province level (the values rise slightly, from 0.95 to 0.99, using the unweighted results). However, for the within estimator, predictive performance of the gridded data, in accounting for inter-census changes in population, gets much worse as spatial scales become more local. Specifically, the average within R^2 at the county level is just 0.12 (using the unweighted regressions, or 0.21 using population-weighted regressions). Average within R^2 values are somewhat higher, at about 0.2 (unweighted) or 0.4 (weighted) at the prefectural-city level, and are higher still, at about 0.6, at the province level.

Figure 3. Averages of the within-R-squared, and between-R-squared (averaging over four models, one per gridded population product) by three spatial aggregation levels.



The poor predictive performance for population changes from gridded estimates at the county local level is concerning because gridded estimates are especially used as a proxy for changes in local population. There are other sources of population estimates at provincial level in the years between a census; for example, surveys have representative samples at this aggregation level even if they are not representative at county-level for all counties. So, the gridded data are rather less needed for provincial level analyses and much more needed for county-level or even more local level analyses. Therefore, neither the good predictive performance for gridded data in cross-sectional analyses nor a reasonable predictive performance (with R^2 values of about 0.6) for time-series changes at the provincial level are relevant to the poor performance in predicting changes in population at the most local levels.

Turning to the detailed regression results at the county-level (Table 2), within-R^2 values are about one-eighth of the between-R^2 although the use of population weights raises the predictive performance of the within estimator. To be specific, the within-R^2 values are about 0.1 (unweighted) and 0.2 (weighted), while between-R^2 values are all higher than 0.9. Within estimator elasticities from three of the gridded datasets are relatively similar (slightly higher for GHS-POP) at around 0.7 but are far lower for LandScan at around 0.4. In contrast, the between elasticity values are more alike for all four gridded population datasets (ranging from 0.96 to 0.98), which is much higher than the within estimator elasticities.

The prefectural city-level populations are more predictable, both cross-sectionally and for time-series changes. The basic pattern of the between-R^2 being higher (by up to five times) than the within-R^2 still holds. The elasticities are also higher, especially for the within estimator, where they are above 0.9 for WorldPop and GHS-POP (but are only from 0.6-0.7 for LandScan). Also, unlike the pattern in the county-level results, using population weights does not raise within estimator elasticities although it does approximately double the within-R^2 values, to around 0.4. In terms of cross-sectionally predicting the locations of census populations, the between-R^2 values all exceed 0.96 and the elasticities are close to 1.0.

The gridded data have their highest predictive accuracy at the provincial level, with the within-R^2 values ranging from 0.55 to 0.7. At this level, population weighting makes little difference (the variation is much less than at the 3rd subnational level). The GHS-POP data have the closest relationships with changes in census population, with elasticities of 0.70 (unweighted) or 0.67 (weighted) while LandScan data has the weakest relationships and the lowest elasticities. However, even at this aggregation level, the time-series relationships for population changes are far weaker than the cross-sectional ones (Table 4).

Table 2 (a). Relationships between gridded population datasets and census population: within and between estimator results (2814 county-level units, non-population weighted).

Dependent variable: ln(pop_census)	**Gridded population dataset**			
	WorldPop	**GPW**	**GHS-POP**	**LandScan**
	Within-estimator, for annual pop changes within each county			
ln(pop_grid)	0.683***	0.691***	0.797***	0.388***
	(0.043)	(0.043)	(0.040)	(0.024)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
R-squared (Within)	0.113	0.114	0.140	0.118
	Between-estimator, for average pop differences between counties			
ln(pop_grid)	0.972***	0.972***	0.965***	0.976***
	(0.004)	(0.005)	(0.004)	(0.005)
R-squared (Between)	0.959	0.940	0.949	0.937

Table 2 (b). Relationships between gridded population datasets and census population: within and between estimator results (2814 county-level units, population weighted).

Dependent variable: ln(pop_census)	**Gridded population dataset**			
	WorldPop	**GPW**	**GHS-POP**	**LandScan**
	Within-estimator, for annual pop changes within each county			
ln(pop_grid)	0.774***	0.780***	0.805***	0.418***
	(0.034)	(0.034)	(0.036)	(0.026)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
R-squared (Within)	0.212	0.214	0.208	0.210
	Between-estimator, for average pop differences between counties			
ln(pop_grid)	0.960***	0.957***	0.957***	1.005***
	(0.004)	(0.004)	(0.004)	(0.006)
R-squared (Between)	0.947	0.943	0.942	0.913

Table 3 (a). Relationships between gridded population datasets and census population: within and between estimator results (297 prefectural city-level units, non-population weighted).

Dependent variable: ln(pop_census)	**Gridded population dataset**			
	WorldPop	**GPW**	**GHS-POP**	**LandScan**
	Within-estimator, for annual pop changes within each city			
ln(pop_grid)	0.915***	0.876***	0.957***	0.677***
	(0.086)	(0.080)	(0.099)	(0.078)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
R-squared (Within)	0.226	0.205	0.173	0.210
	Between-estimator, for average GDP differences between cites			
ln(pop_grid)	0.973***	0.978***	0.994***	1.012***
	(0.007)	(0.010)	(0.009)	(0.012)
R-squared (Between)	0.983	0.973	0.977	0.963

Table 3 (b). Relationships between gridded population datasets and census population: within and between estimator results (297 prefectural city-level units, population weighted).

Dependent variable: ln(pop_census)	**Gridded population dataset**			
	WorldPop	**GPW**	**GHS-POP**	**LandScan**
	Within-estimator, for annual pop changes within each city			
ln(pop_grid)	0.864***	0.866***	0.967***	0.629***
	(0.072)	(0.072)	(0.081)	(0.093)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
R-squared (Within)	0.443	0.442	0.366	0.338
	Between-estimator, for average pop differences between cites			
ln(pop_grid)	1.004***	1.004***	1.012***	1.033***
	(0.006)	(0.006)	(0.007)	(0.012)
R-squared (Between)	0.990	0.988	0.987	0.963

Table 4 (a). Relationships between gridded population datasets and census population: within and between estimator results (31 provincial -level units, non-population weighted).

Dependent variable: ln(pop_census)	**Gridded population dataset**			
	WorldPop	**GPW**	**GHS-POP**	**LandScan**
	Within-estimator, for annual pop changes within each province			
ln(pop_grid)	0.673***	0.674***	0.837***	0.777***
	(0.061)	(0.060)	(0.088)	(0.066)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
R-squared (Within)	0.649	0.661	0.701	0.546
	Between-estimator, for average pop differences between provinces			
ln(pop_grid)	1.007***	1.039***	1.028***	0.991***
	(0.007)	(0.012)	(0.009)	(0.016)
R-squared (Between)	0.999	0.996	0.997	0.993

Table 4 (b). Relationships between gridded population datasets and census population: within and between estimator results (31 provincial -level units, population weighted).

Dependent variable: ln(pop_census)	**Gridded population dataset**			
	WorldPop	**GPW**	**GHS-POP**	**LandScan**
	Within-estimator, for annual pop changes within each province			
ln(pop_grid)	0.718***	0.718***	0.870***	0.744***
	(0.095)	(0.095)	(0.123)	(0.053)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
R-squared (Within)	0.654	0.656	0.666	0.588
	Between-estimator, for average pop differences between provinces			
ln(pop_grid)	1.018***	1.025***	1.020***	0.992***
	(0.009)	(0.010)	(0.008)	(0.022)
R-squared (Between)	0.998	0.997	0.998	0.986

Regression Results for Population Changes

The results in Tables 2-4 consistently show the gridded data products are far better predictors for cross-sectional analysis than for time-series studies where variation comes from changes in population. In this section we drill down into the time-series variation by using regressions to test if there has been a decline over time in the predictive performance for the population changes. Specifically, we test whether the predictions of changes in population became less accurate from 2000-10 to 2010-20. We carry out this test at the county/district level, at the prefectural city level, and at the provincial level.

Consistent with Table 2, that shows predicting inter-census changes in county-level population is the task that the gridded population estimates do least well, the R^2 values in Table 5(a) for the county-level population changes show that only a small percentage of the variation is predicted by the changes in the gridded data. Specifically, for the first decade the R^2 averages 0.53 for three gridded products (WorldPop, GPW and GHS-POP) and is even lower, at just 0.31, for LandScan. There was a sharp fall in predictive accuracy in the second decade, for changes from 2010 to 2020, with R^2 values that average just 0.10. This substantial weakening in the relationship is also shown in the coefficients, which fell from around 1.2 to just 0.5 for WorldPop, GPW and GHS-POP, and from about 0.6 to about 0.4 for LandScan. Ideally for the specification we use, coefficients would be close to 1.0, so that each extra person in a county by next census date that is suggested by the gridded population product is matched with an actual extra person in that same place according to the two censuses.

For the results at the prefectural city level, in Table 5(b), the coefficients in the first decade were close to the ideal value of 1.0 for WorldPop, GPW and GHS-POP, but they fell to being only two thirds as large in the second decade. For LandScan the fall in coefficients was from 0.84 to 0.75. Likewise, the R^2 values for the first three gridded products fell from 0.65 to the 0.2-0.3 range (while for LandScan they stayed just over 0.3 for both decades). So, even at the more aggregated level there was a decline in predictive performance. These falls also show up in the Chow tests, which suggest that regressions for the 2010-to-2020 decade are significantly different to the regressions for the earlier decade.

At the provincial level, the regressions using WorldPop, GPW or GHS-POP had coefficients and R^2 values close to 1.0 for the population changes from 2000 to 2010. Yet for the second decade the R^2 values had fallen by about one-half. Moreover, the coefficients suggested that the gridded data were overstating population change (each one extra person in a gridded dataset results in just 0.8 extra persons in the census data for 2020).

Table 5 (a). Relationships between changes of gridded population datasets and census population: 2000-2010 versus 2010-2020 (2814 county-level units).

Dependent variable: d(pop_census)	**Gridded population dataset**			
	WorldPop	**GPW**	**GHS-POP**	**LandScan**
	pop changes from 2000 to 2010			
d(pop_grid)	1.232***	1.260***	1.244***	0.563***
	(0.022)	(0.022)	(0.022)	(0.016)
R-squared	0.527	0.532	0.531	0.307
	pop changes from 2010 to 2020			
d(pop_grid)	0.517***	0.501***	0.412***	0.394***
	(0.026)	(0.025)	(0.027)	(0.020)
R-squared	0.120	0.127	0.076	0.120
Chow test	28.02***	32.52***	70.25***	1.15

Table 5 (b). Relationships between changes of gridded population datasets and census population: 2000-2010 versus 2010-2020 (297 city-level units).

Dependent variable: d(pop_census)	**Gridded population dataset**			
	WorldPop	**GPW**	**GHS-POP**	**LandScan**
	pop changes from 2000 to 2010			
d(pop_grid)	0.940***	0.962***	0.952***	0.837***
	(0.041)	(0.041)	(0.041)	(0.070)
R-squared	0.641	0.647	0.646	0.326
	pop changes from 2010 to 2020			
d(pop_grid)	0.619***	0.605***	0.648***	0.745***
	(0.053)	(0.049)	(0.078)	(0.063)
R-squared	0.315	0.337	0.185	0.323
Chow test	3.39*	4.43*	2.12	0.08

Table 5 (c). Relationships between changes of gridded population datasets and census population: 2000-2010 versus 2010-2020 (31 provincial -level units).

Dependent variable: d(pop_census)	**Gridded population dataset**			
	WorldPop	**GPW**	**GHS-POP**	**LandScan**
	pop changes from 2000 to 2010			
d(pop_grid)	1.038***	1.058***	1.047***	0.866***
	(0.023)	(0.018)	(0.019)	(0.112)
R-squared	0.985	0.991	0.990	0.660
	pop changes from 2010 to 2020			
d(pop_grid)	0.771***	0.784***	0.986***	0.845***
	(0.144)	(0.123)	(0.200)	(0.157)
R-squared	0.479	0.570	0.438	0.483
Chow test	1.79	3.49*	0.03	0.01

Note: The Chow test results for the difference in the models between the two periods is shown in the last row, with statistical significance at the 1%, 5% and 10% level denoted by ***, **, *.

An Example: How Gridded Population Data Distort Apparent Trends in Spatial Inequality

The results thus far suggest that the gridded population datasets may be useful for spatialization but are less useful for studies that involve population changes, especially at the most local level. In this section we provide an example of how an analysis may go astray when changes in the gridded population estimates are used as a proxy for changes in actual population. The example concerns spatial inequality, which is an ongoing subject of study in China (Zhang et al., 2023) where population estimates are needed as the weight for constructing inequality statistics from the annual data on local GDP (essentially comparing shares of income versus population at the various quantiles). Chen (2025) is an example of just such a study, estimating trends in spatial inequality in China over the last two decades by combining county-level annual GDP data with LandScan annual gridded population estimates.

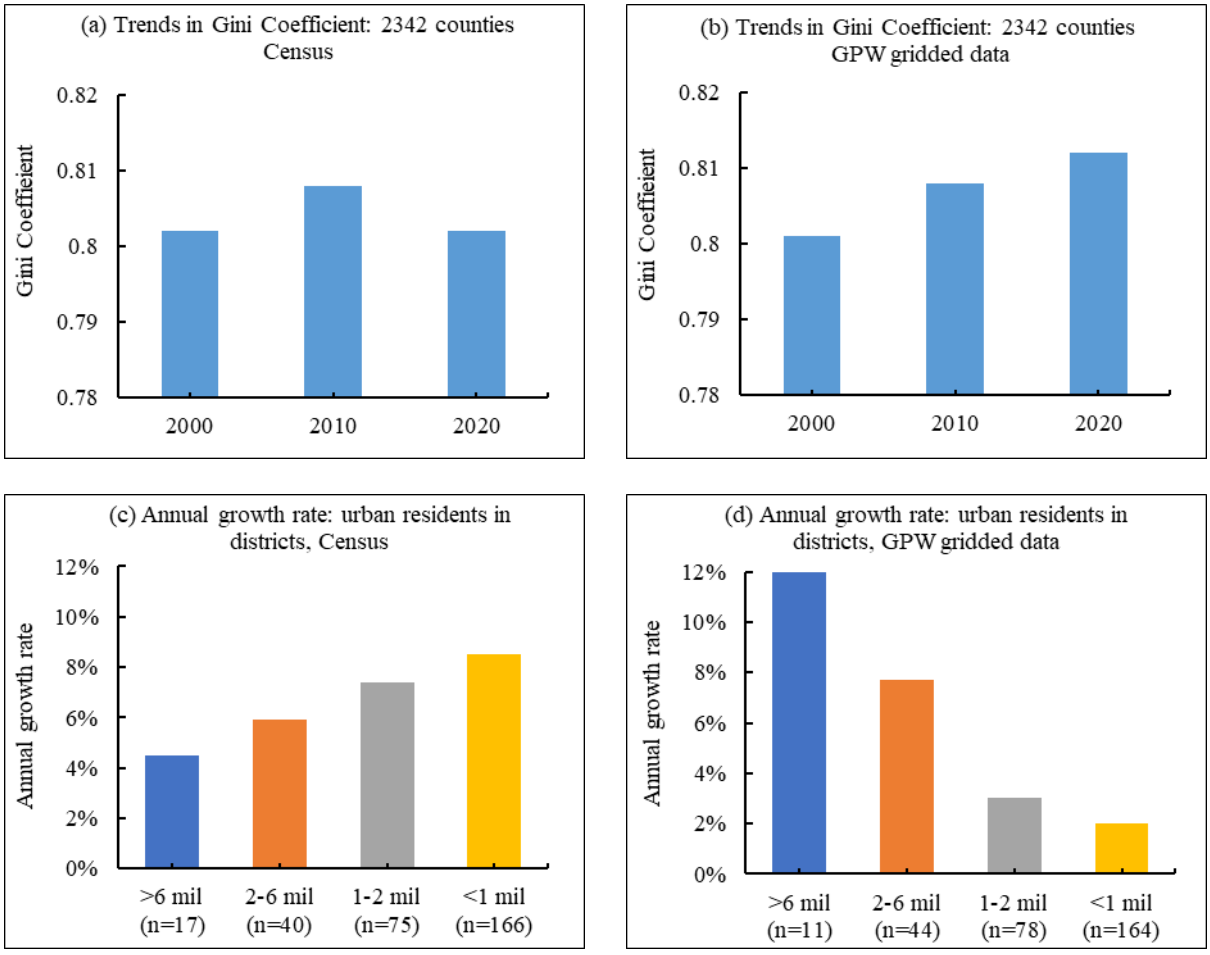
In order to present the gridded population estimates in the best possible light we use the GPW estimates that had the best predictive performance at the county level in Table 5(a). The county-level GDP data and census counts for 2000, 2010, and 2020 are used to estimate a Gini index of inequality.[3] We then repeat the calculations using GPW gridded data aggregated to the county level. When the census data are used it is apparent that there was a rise in inequality from 2000 to 2010, that then reversed between 2010 and 2020 with the Gini index returning to the initial value (Figure 4(a)). In contrast, when the GPW estimates are used there appears to be an ongoing rise in spatial inequality, with the Gini index always higher than in the previous census year (Figure 4(b)). In other words, the gridded data product gives the impression that spatial inequality in China has continued to rise over the most recent decade which distorts the patterns shown by the actual population counts from the census.

One explanation for this distortion of the spatial inequality trends is that the GPW estimates miss the sharp change in patterns of population growth rates according to city size. The bar chart in panel (c) of Figure 3 reproduces the annual population growth rates for urban districts (the high-density parts of prefectural level cities) calculated from the 2010 and 2020

[3] One difference from the analyses in Tables 2 and 5 is that the county-level GDP data for individual districts that are adjacent to each other within the core urban area of a prefectural city have their GDP reported as a single entity rather than on a district-by-district basis in the statistical yearbooks we use; these aggregations are why the sample size falls to n=2342 county/district-level units from the previous n=2814 sample size.

census counts. Recall that prior to 2010 the large and small urban districts were all growing at a rate of about eight per cent per year, but from 2010 to 2020 the most populated urban districts had growth rates just one-half as fast as for the least populated ones, as the migrant destinations shifted away from big cities to smaller urban areas. This shift has impacts on spatial inequality because the previous migration hotspots like the Beijing-Tianjin region had higher GDP per capita than other migration destinations and so spatial inequality rose as these regions gained population. In contrast, in the most recent decade spatial inequality fell as population growth shifted towards smaller cities, with lower per capita GDP, because of the changes in migration patterns.

Figure 4. The comparison between census population and gridded population datasets.



The last figure, in panel (d) shows that the gridded population estimates miss this sharp change in the patterns of population growth rates. Rather than the largest cities (urban districts with more than six million residents) having annual population growth rates that were

only half as fast as for the smallest (< 1 million residents) districts – what the census data show – the gridded estimates suggest that the most highly populated urban districts were growing far faster than the least populated ones. Consequently, the rise in spatial inequality that was shown with the census data from 2000 to 2010 is incorrectly shown to have continued on into the 2010-20 decade because the gridded population estimates miss the sharp change in migration patterns.

Conclusions

With the development of remote sensing technology, several recent efforts have produced global- and continental-extent gridded population estimates. These data are available at fine spatial scales, with grid sizes of 1 km or smaller, and are often updated with annual frequency. In contrast, actual population counts from censuses are often available just once per decade and may have publicly available data only disaggregated down to the county/district level so study at very local levels may be limited. Given these constraints, the gridded population estimates appear to be an attractive data source, which may account for growing use of these data in various branches of research. While our review suggests that more of the studies using these data rely upon their spatialization aspects (and so cross-sectional accuracy matters), there are also a growing number of studies that rely on estimated population changes, and in this case the time-series accuracy of these gridded estimates matters.

In this paper, we examined how well the gridded population estimates can predict the differences in census population between places and between time periods. Our testing used a set of regression analyses, with four popular sources of gridded population estimates used to predict census counts for China in 2000, 2010, and 2020, at three subnational administrative levels: county/district, prefectural level city, and province. Our key finding is that gridded data are better at predicting cross-sectional population differences than at predicting the changes in population over time, especially at the most local level. For example, at the county level, most of the temporal variation in the changes in population is not predicted by changes in the gridded estimates; only about ten percent is predicted using the unweighted regressions, or 20 percent using population-weighted regressions. In contrast, at this same spatial level, about 95 percent of cross-sectional variation in census population is predicted by the gridded estimates. It also

seems that the predictive performance for temporal changes has fallen over time, being worse for changes in the most recent decade compared to the prior decade.

The predictive performance for temporal changes in population is better at higher levels of the subnational administrative hierarchy, such as prefectural level cities and provinces, but at these higher levels – especially for provinces – there are other sources of annual population estimates in the years between a census, such as from sample surveys. Hence, weak predictive performance of gridded population estimates is especially a problem for the type of spatially disaggregated units where the gridded data would seem most useful given that other data may not be available at the same 1 km (or smaller) scale. Moreover, even for the higher-level spatial units, there was a clear decline in predictive performance in the most recent decade.

Although our study is only focused on China, there are broader implications for the uses of these gridded population products in research. If a research design is based on temporal variation in the subject under study, relying on gridded population estimates may be a source of vulnerability, especially if the locations of population change have shifted sharply, as they did for China between the most recent decade and the prior decade. For example, where there is a large-scale migration underway in the presence of substantial regional disparities in living standards, population movements will also be associated with changes in spatial inequality. To the extent that the gridded estimates miss some of the on-the-ground changes they may give a misleading impression about the trends in spatial inequality and could lead to inappropriate policy interventions that are based on a misreading of the actual situation. Overall, our findings suggest that researchers should be cautious in relying on gridded population estimates as a proxy for the actual temporal changes in population that are occurring at fine spatial scales.

References

Archer, L., Neal, J., Bates, P., Vosper, E., Carroll, D., Sosa, J., and Mitchell, D. (2024). Current and future rainfall-driven flood risk from hurricanes in Puerto Rico under 1.5 and 2℃ climate change. *Natural Hazards and Earth System Sciences*, 24(2), 375-396.

Asher, S., Lunt, T., Matsuura, R., and Novosad, P. (2021). Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform. *The World Bank Economic Review*, 35(4), 845-871.

Balk, D. L., Nghiem, S. V., Jones, B. R., Liu, Z., and Dunn, G. (2019). Up and out: A multifaceted approach to characterizing urbanization in Greater Saigon, 2000–2009. *Landscape and Urban Planning*, 187, 199-209.

Benz, S. A., Menberg, K., Bayer, P., and Kurylyk, B. L. (2022). Shallow subsurface heat recycling is a sustainable global space heating alternative. *Nature Communications*, 13(1), 1-11.

Bosker, M., Park, J., and Roberts, M. (2021). Definition matters. metropolitan areas and agglomeration economies in a large-developing country. *Journal of Urban Economics*, 125, 103275.

Bourget, M., Boudreault, M., Carozza, D. A., Boudreault, J., and Raymond, S. (2024). A data science approach to climate change risk assessment applied to pluvial flood occurrences for the United States and Canada. ASTIN Bulletin: *The Journal of the IAA*, 1-23.

Cao, Y., Hua, Z., Chen, T., Li, X., Li, H., and Tao, D. (2023). Understanding population movement and the evolution of urban spatial patterns: An empirical study on social network fusion data. *Land Use Policy*, 125, 106454.

Cao Z, Zheng X, Liu Y, Li Y, Chen Y. Exploring the changing patterns of China's migration and its determinants using census data of 2000 and 2010. *Habitat Int*. (2018) 82:72–82.

Castells-Quintana, D., Dienesch, E., and Krause, M. (2021). Air pollution in an urban world: A global view on density, cities and emissions. *Ecological Economics*, 189, 107153.

Chan, K. W. (2009). The Chinese hukou system at 50. *Eurasian Geography and Economics*, 50(2), 197-221.

Chen, Y. (2025). Remotely sensed spatial inequality: Luminosity evidence from Chinese cities and counties. *Asian Development Review*, 42(2): 1-29.

Chen, X., and Nordhaus, W. (2015). A test of the new VIIRS lights data set: Population and economic output in Africa. *Remote Sensing*, 7(4), 4937-4947.

Chen, R., Yan, H., Liu, F., Du, W., and Yang, Y. (2020). Multiple global population datasets: Differences and spatial distribution characteristics. *ISPRS International Journal of Geo-Information*, 9(11), 637.

Cheng, M., and Duan, C. (2021). The changing trends of internal migration and urbanization in China: new evidence from the seventh National Population Census. *China Population and Development Studies*, 5(3), 275-295.

Colón-González, F. J., Gibb, R., Khan, K., Watts, A., Lowe, R., and Brady, O. J. (2023). Projecting the future incidence and burden of dengue in Southeast Asia. *Nature Communications*, 14(1), 5439.

Dasgupta, S., Blankespoor, B., and Wheeler, D. (2024). Estimating extinction threats with species occurrence data from the Global biodiversity information facility (No. 10822). The World Bank.

Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., and Schiavina, M. (2021). Applying the degree of urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics*, 125, 103312.

Dodd, R., Awuor, A. O., Garcia Bardales, P. F., Khanam, F., Mategula, D., Onwuchekwa, U., ... and McGrath, C. J. (2024, March). Population enumeration and household utilization survey methods in the Enterics for Global Health (EFGH): Shigella surveillance study. In Open Forum Infectious Diseases (Vol. 11, No. Supplement_1, pp. S17-S24). US: Oxford University Press.

Dong, B., Ye, Y., You, S., Zheng, Q., Huang, L., Zhu, C., ... and Wang, K. (2021). Identifying and classifying shrinking cities using long-term continuous night-time light time series. *Remote Sensing*, 13(16), 3142.

Freire, S., Schiavina, M., Florczyk, A. J., MacManus, K., Pesaresi, M., Corbane, C., Borkovska, O., Mills, J., Pistolesi, L., Squires, J., and Sliuzas, R. (2020). Enhanced data

and methods for improving open and free global population grids: putting ‘leaving no one behind’ into practice. *International Journal of Digital Earth*, 13(1), 1–17.

Gibson, J., Olivia, S., Boe-Gibson, G., and Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, 149, 102602.

Gibson, J., Zhang, X., Park, A., Yi, J., and Xi, L. (2024). Remotely measuring rural economic activity and poverty: Do we just need better sensors? (No. 2023-08). Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University.

Goldblatt, R., Heilmann, K., and Vaizman, Y. (2020). Can medium-resolution satellite imagery measure economic activity at small geographies? Evidence from Landsat in Vietnam. *The World Bank Economic Review*, 34(3), 635-653.

Gollin, D., Kirchberger, M., and Lagakos, D. (2021). Do urban wage premia reflect lower amenities? Evidence from Africa. *Journal of Urban Economics*, 121, 103301.

González, F. A. I., Cantero, L. S., and Szyszko, P. A. (2023). Inequality and economic activity under regional favoritism: evidence from Argentina. *Review of Regional Research*, 43(2), 343-361.

Gu, H., Liu, Z., and Shen, T. (2020). Spatial pattern and determinants of migrant workers' interprovincial hukou transfer intention in China: Evidence from a national migrant population dynamic monitoring survey in 2016. *Population, Space and Place*, 26(2), e2250.

Jing, R., Heft-Neal, S., Chavas, D. R., Griswold, M., Wang, Z., Clark-Ginsberg, A., ... and Wagner, Z. (2024). Global population profile of tropical cyclone exposure from 2002 to 2019. *Nature*, 626(7999), 549-554.

Joseph, I. L. (2022). The effect of natural disaster on economic growth: Evidence from a major earthquake in Haiti. *World Development*, 159, 106053.

Li, Z., Li, C., Gibson, J., and Deng, X. (2024). Two decades of inter-city migration in China: The role of economic, natural and social amenities. *Working Paper* No. 24/05, Department of Economics, University of Waikato.

Li, H., and Xiao, Z. (2022). Comparisons and predictions of intercity population migration propensity in major urban clusters in China: based on use of the Baidu index. *China Population and Development Studies*, 6(1), 55-77.

Liu, L., Cao, X., Li, S., and Jie, N. (2024). A 31-year (1990–2020) global gridded population dataset generated by cluster analysis and statistical learning. *Scientific Data*, 11(1), 124.

Lloyd, C. T., Chamberlain, H., Kerr, D., Yetman, G., Pistolesi, L., Stevens, F. R., ... and Tatem, A. J. (2019). Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data*, 3(2), 108-139.

Ma, S., Li, S., Luo, Q., Yu, Z., and Wang, Y. (2024). Revisiting the relationships between energy consumption, economic development and urban size: A global perspective using remote sensing data. *Heliyon*, 10(5).

Maldonado, L. (2023). Living in darkness: rural poverty in Venezuela. *Journal of Applied Economics*, 26(1), 2168464.

Melchiorri, M., Freire, S., Schiavina, M., Florczyk, A., Corbane, C., Maffenini, L., ... and Kemper, T. (2024). The multi-temporal and multi-dimensional global urban centre database to delineate and analyse world cities. *Scientific Data*, 11(1), 82.

Muhwezi, B., Williams, N. J., and Taneja, J. (2021). Ingredients for growth: Examining electricity consumption and complementary infrastructure for Small and Medium Enterprises in Kenya. *Development Engineering*, 6, 100072.

Opdyke, A., and Fatima, K. (2024). Comparing the suitability of global gridded population datasets for local landslide risk assessments. *Natural Hazards*, 120(3), 2415-2432.

Oughton, E. J., Amaglobeli, D., and Moszoro, M. (2023). What would it cost to connect the unconnected? Estimating global universal broadband infrastructure investment. *Telecommunications Policy*, 47(10), 102670.

Patias, N., Rowe, F., and Arribas-Bel, D. (2022). Trajectories of neighbourhood inequality in Britain: Unpacking inter‐regional socioeconomic imbalances, 1971-2011. *The Geographical Journal*, 188(2), 150-165.

Peng, Y., Welden, N., and Renaud, F. G. (2024). Incorporating ecosystem services into comparative vulnerability and risk assessments in the Pearl River and Yangtze River Deltas, China. *Ocean and Coastal Management*, 249, 106980.

Pérez-Sindín, X. S., Chen, T. H. K., and Prishchepov, A. V. (2021). Are night-time lights a good proxy of economic activity in rural areas in middle and low-income countries?

Examining the empirical evidence from Colombia. *Remote Sensing Applications: Society and Environment*, 24, 100647.

Rathore, L. S., Kumar, M., Hanasaki, N., Mekonnen, M. M., and Raghav, P. (2024). Water scarcity challenges across urban regions with expanding irrigation. *Environmental Research Letters*, 19(1), 014065.

Schiavina, M., Freire, S., Carioli, A., and MacManus, K. (2023). GHS-POP R2023A–GHS population grid multitemporal (1975-2030). In European Commission. Joint Research Centre (JRC).

Shi, Q., Liu, T., and Feng, C. (2024). The changing geography of interprovincial migration in China: history and new trends. Eurasian Geography and Economics, 1-23.

Sims, K., Reith, A., Bright, E., Kaufman, J., Pyle, J., Epting, J., Gonzales, J., Adams, D., Powell, E., Urban, M., and Rose, A. (2023). LandScan Global 2022 [Data set]. Oak Ridge National Laboratory.

Sun, J., Di, L., Sun, Z., Wang, J., and Wu, Y. (2020). Estimation of GDP using deep learning with NPP-VIIRS imagery and land cover data at the county level in CONUS. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1400-1415.

Tatem, A. J. (2017). WorldPop, open data for spatial demography. Scientific Data, 4(1), 1-4.

Thomson, D. R., Leasure, D. R., Bird, T., Tzavidis, N., and Tatem, A. J. (2022). How accurate are WorldPop-Global-Unconstrained gridded population data at the cell-level?: A simulation analysis in urban Namibia. *Plos one*, 17(7), e0271504.

Unfried, K., Kis-Katos, K., and Poser, T. (2022). Water scarcity and social conflict. *Journal of Environmental Economics and Management*, 113, 102633.

Wang, J., and Zhang, Y. (2022). Destination-to-origin differences and settlement intentions of Chinese internal migrants: A birth cohort analysis. *Population, Space and Place*, 28(5), e2544.

White, M., Sun, L., and Jiang, L. (2024). Changes in migration determinants along the urban hierarchy in China. *Asian Population Studies*, 20(1), 81-103.

Wilmot, T. Y., Lin, J. C., Wu, D., Oda, T., and Kort, E. A. (2024). Toward a satellite-based monitoring system for urban CO2 emissions in support of global collective climate mitigation actions. *Environmental Research Letters*, 19(8), 084029.

Xiao, C., Feng, Z., You, Z., and Zheng, F. (2022). Population boom in the borderlands globally. *Journal of Cleaner Production*, 371, 133685.

Xu, Y., Ho, H. C., Knudby, A., and He, M. (2021). Comparative assessment of gridded population data sets for complex topography: A study of Southwest China. *Population and Environment*, 42, 360-378.

Yang, L., Ji, X., Li, M., Yang, P., Jiang, W., Chen, L., ... and Li, Y. (2024). A comprehensive framework for assessing the spatial drivers of flood disasters using an optimal Parameter-based geographical Detector–machine learning coupled model. *Geoscience Frontiers*, 101889.

Yang, S., Shu, T., and Yu, T. (2022). Migration for better jobs or better living: Shifts in China. *International Journal of Environmental Research and Public Health*, 19(21), 14576.

Yasin, S., and Alistair, B. (2019). The impacts of pollution for new high-speed railways: the case of noise in Turkey. *Acoustics Australia*, 47(2), 141-151.

Ye, T., Zhao, N., Yang, X., Ouyang, Z., Liu, X., Chen, Q., ... and Jia, P. (2019). Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. *Science of the Total Environment*, 658, 936-946.

Zeng, P., Liu, Y., Tian, T., Che, Y., and Helbich, M. (2024). Geographic inequalities in park visits to mitigate thermal discomfort: A novel approach based on thermal differences and cellular population data. *Urban Forestry and Urban Greening*, 128419.

Zhang, X., and Gibson, J. (2022). Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. *Remote Sensing*, 14(5), 1282.

Zhang, X., Gibson, J., and Deng, X. (2023). Remotely too equal: Popular DMSP night-time lights data understate spatial inequality. *Regional Science Policy and Practice*, 15(9), 2106-2126.

Zhao, M., Zhou, Y., Li, X., Cao, W., He, C., Yu, B., ... and Zhou, C. (2019). Applications of satellite remote sensing of nighttime light observations: Advances, challenges, and perspectives. *Remote Sensing*, 11(17), 1971.

Zhou, Y., Chen, H., and Fang, T. (2024). Spatial analysis of intercity migration patterns of China's rural population: Based on the network perspective. *Agriculture*, 14(5), 655.

Chapter 5: The role of spillovers when evaluating regional development interventions: evidence from administrative upgrading in China

Xiaoxuan Zhang, Chao Li and John Gibson

This paper has been published in the *Letters in Spatial and Resource Sciences*, (2024), 17(1) 9.

Letters in Spatial and Resource Sciences (2024) 17:9
https://doi.org/10.1007/s12076-024-00371-1

ORIGINAL PAPER

The role of spillovers when evaluating regional development interventions: evidence from administrative upgrading in China

Xiaoxuan Zhang[1] · **Chao Li**[2] · **John Gibson**[1]

Received: 8 November 2023 / Accepted: 8 January 2024

Abstract

Direct effects of regional development interventions on targeted areas may be amplified by positive spillovers from elsewhere or offset by negative spillovers. Yet spillovers are often ignored in the applied literature, where impact analyses based on difference-in-differences typically treat spatial units as independent of their neighbours. We study spatial spillovers from a popular regional development intervention in China—converting counties to cities. China's top-down approach lets only central government bestow city status on an area, with over ten percent of counties upgraded to cities in the last two decades. A growing literature estimates impacts of these conversions, with spatial units typically treated as independent of their neighbours. In contrast, our spatial econometric models use a 20-year panel for almost 2500 county-level units to allow indirect spillover effects on indicators of local economic activity. The positive direct effects on GDP and luminosity of a county being upgraded are amplified through positive indirect effects, especially in the eastern regions of China where economic activity and population are more densely concentrated. The models without spatial lags that ignore spillovers give estimated effects of converting counties to cities that are only two-fifths to two-thirds as large as the estimated effects coming from the spatial models.

Keywords County upgrading · Luminosity · Regional development · Spatial spillovers · China

JEL Classification R12

✉ Xiaoxuan Zhang
jkgibson@waikato.ac.nz

[1] Department of Economics, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand

[2] University of Auckland, Auckland, New Zealand

Published online: 13 February 2024

1 Introduction

The direct effects of regional development interventions may be amplified by positive spillovers or offset by negative spillovers. Yet these spillovers are often ignored in applied studies by popular impact evaluation frameworks that treat the spatial units as independent of surrounding areas. An emerging set of studies finds that spillovers should be considered. For example, De Castris et al (2023) examine whether European cohesion policy targeted at less developed areas increases subsequent local economic growth (at the NUTS-3 level); allowing for spillovers gives total (positive) impacts about one-fifth larger than direct effects. Gambina and Mazzola (2023) find a similar result for Italian provinces, and national cohesion projects have even larger indirect effects. Spillovers also occur at finer spatial scale. For example, Qiu and Tong (2021) study effects of Light Rail Transit (LRT) on housing prices in Edmonton, Canada. A spatial difference-in-differences (SDID) model yields a negative direct effect of C$20,000 lower prices for homes within 2 km of a LRT station, which is exacerbated by a further C$5000 fall from indirect effects that result from property price spillovers. A model that ignores spillovers yields estimated treatment effects that are only two-thirds as large.

A particularly salient place to study these issues is China, which uses many regional development interventions, and whose pace of growth means that samples of just a couple of decades can observe changes that may have taken far longer to unfold in other places. Under China's hierarchical governance, these interventions often involve altering administrative ranks. While cities elsewhere may arise organically in a bottom-up way from agglomeration forces, under China's top-down approach only the central government may bestow city status on a particular area. Therefore, an ongoing aspect of urbanization in China is administrative conversions, with over 250 counties upgraded to city status (as either districts or county-level cities) in the last two decades, which is more than ten percent of all counties (Fig. 1).[1] Potentially important impacts follow from this upgrading because cities have larger quotas for converting land from agricultural use to urban use, can levy higher taxes on urban construction and keep a larger share of revenue from land sales, have higher local public sector employment and generally more prominence for investment.

In light of the importance of this particular type of regional development intervention, a growing empirical literature uses econometric models to evaluate impacts of administrative upgrading. These studies use panel data to test if various indicators of economic activity, such as GDP or night-time lights, grow faster in the upgraded counties compared to their peers that were not upgraded. This literature has yielded mixed findings; an early study found counties upgraded from 1994 to 1997 performed no better than those that stayed as counties, for an evaluation period ending in 2004 (Fan et al 2012). A more recent study of the same episode found positive

[1] Counties, county-level cities, and districts are all *de jure* third-level units in China's administrative ranks (with provinces and prefectures the first two levels) but counties de facto rank lower (see Table 1 for reasons).

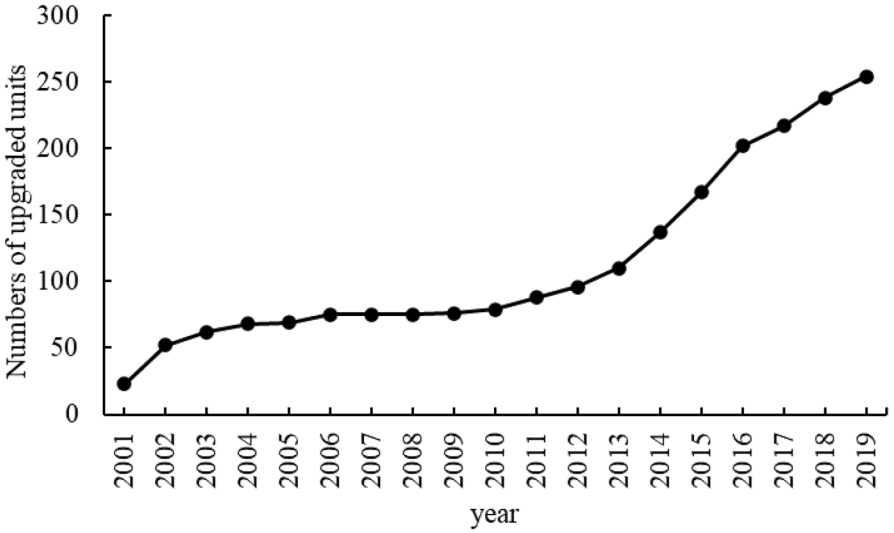


Fig. 1 Cumulative number of counties upgraded: 2001–2019

impacts becoming apparent after 2004, especially in more densely populated eastern parts of China where initial agglomeration forces were stronger (Tang 2021).[2]

There are several empirical challenges confronting these studies and the one that we examine concerns potential spillovers. These could occur if mobile resources, such as labour and capital, move from neighbouring counties into newly upgraded cities, causing a negative effect (Chen and Partridge 2013). This matters not just distributionally but also empirically; if comparison groups are subject to a negative spillover it can exaggerate apparent treatment effects if the spillover effect is ignored (Tang 2021). Typically, the studies in this literature use difference-in-differences frameworks that generally rely on a stable unit treatment value assumption (SUTVA) of independence between the spatial units (Delgado and Florax 2015).

The spillovers could also be positive if upgraded counties become growth poles after relaxation of their land-related constraints (Lichtenberg and Ding 2009), allowing economic activity in hinterlands to also flourish. County-to-district upgrades often lower transport costs from newer, better, infrastructure that has network efficiencies, improving market access and firm productivity (Tang and Hewings 2017). Indeed, classic development theories emphasize a role for positive spillovers when explaining the variation in economic performance; these may be between industrial sectors (Rosenstein-Rodan 1943), through demand-side linkages (Park and Johnston 1995) but also through geographic externalities (Ravallion 2005).

[2] Likewise, Tang and Hewings (2017) find larger effects after four years, for a different type of upgrading, from county to district (*che xian she qu*) that often involves consolidation and then expansion of enlarged urban areas.

Table 1 Potential benefits for a county in getting upgraded to city or district status

Category	Benefits	Source
County-to-county-level city		
Administrative power	County-level cities have greater economic and social autonomy, for investment projects, for flexibility when establishing branches of customs offices and large state-owned banks, and for governing urban-related affairs	Landry (2008), Su (2022), Tang (2021)
Fiscal benefits	County-level cities have powers to collect extra taxes and fees related to urban construction and management (7% versus 5%) and receive more fiscal subsides	Lichtenberg and Ding (2009), Fan et al. (2012)
Land-related	County-level cities have more autonomy over land use as well as more construction land quota allocated by higher level governments. They can also retain a larger share of revenues from land sales	Chung and Lam (2004), Fan et al. (2012)
Policy priority	County-level cities are more likely to have policy priority than are counties, as they will be listed separately in provincial plans	Shen (2007)
Rank and salary	In some cases, the bureaucratic rank and salary of officials is raised after upgrading	Li (2011), Mukim and Zhu (2018)
Reputation	County-level cities enjoy higher prestige compared with counties, making them more attractive to outside investors	Chung and Lam (2004)
County-to-district		
Administrative power	Facilitates centralization of administrative powers, for unifying urban planning and industrial layout which is conducive to market integration and optimal allocation of resources	Cao et al. (2022), Tang and Hewings (2017), Tian et al. (2020), Lu and Wang (2023)
Fiscal benefits	More fiscal transfers from prefecture-level governments to the newly established districts	Lu and Wang (2023)
Land-related	Governments development strategies have full use of land resources previously controlled by county government, for citywide planning and rural–urban integration, aiding decentralization of urban industries, and reducing hurdles for the expansion of urban built-up areas	Ma (2005)
Policy priority	Mitigates issues of regionalism in economic policy, balances the spatial distribution of resources between municipal districts and counties, and delays city shrinkage	Chen et al. (2020), Deng et al. (2022)
Reputation	Upgraded status from county to district makes the county more economically connected to the city, in sharing reputational benefits	Lu and Wang (2023)

Given the volatility of Chinese policies, the benefits listed here are continuously changing over time, and not all of them are necessarily practiced contemporaneously

Despite the potential importance of spillovers for understanding effects of the county upgrading process in China the literature to date has mainly ignored them or else used only partial or informal approaches. For example, Tang (2021) drops counties that are closer than 50 kms to upgraded counties (amounting to 16% of county-year observations), in case the nearby counties had been subject to some spillover from the treated units but he otherwise uses an econometric framework that assumes that spatial units are independent. Along the same lines, Zeng and You (2022) drop 46%, 57% and 66% of their county-year observations based on removing those within bands of 50 km, 60 km, and 70 km from upgraded counties in case of spatial spillovers in their outcome measure. The particular distance thresholds do not have an underlying theoretical basis, such as a spatial econometric model that could allow for both local and global spillovers. Relatedly, Bo (2020) includes the treatment status of contiguous neighbours in his prefectural-level study, and also includes the treatment status of same-province prefectures as a covariate. However, as discussed below, only local spillovers will have their effects observed with this specification, which is effectively a spatial lag of the covariate; if there also are global spillovers these will operate through the spatial lag of the outcome variable, and so these will be missed with this more restricted framework.

Given this limited and only partial evidence on the role of spillovers, the current study applies spatial econometric models to a 20-year panel of almost 2500 county-level units to examine effects of converting China's counties to cities (either as county-level cities or as districts). We use GDP and night-time lights as our indicators of local economic activity. We start with a very general spatial autoregressive model with spatial autoregressive errors (SARAR) that encompasses some popular models, such as the spatial Durbin model, spatial lag model, and spatial error model. In this particular setting and for the period we study, the spatial lags appear to be relevant for the outcomes, the covariates and the errors. The positive direct effects on economic activity of a county being upgraded are amplified through positive indirect effects. The indirect spillovers are especially apparent in the eastern regions of China where economic activity and population are more densely concentrated. The models without spatial lags give estimated effects of county upgrading that are only two-fifths to two-thirds as large as the estimated total effects coming from the spatial models. Given that models without spatial lags have been almost exclusively what is used to date, these findings suggest that the literature studying county upgrading in China might benefit from more widespread use of spatial models that explicitly allow spillover effects to be studied. More generally, our example adds to the growing list showing the importance of allowing for spatial spillovers.

In the next section we provide details on China's administrative hierarchy, describe some pathways through which counties can be upgraded, and show that prior studies in this literature have not used spatial econometric models. Our data and methods are discussed in Section III. Section IV has the results, using both county-level GDP and night-time lights as indicators of economic activity. Section V has the discussion and conclusions.

2 Background and context

2.1 China's administrative hierarchy

In China's administrative hierarchy four levels of local government exist: province, prefecture, county, and town. The central government can bestow city status at any of the first three levels (Fan et al 2012). For example, four cities directly under the central government—Beijing, Tianjin, Shanghai and Chongqing—have equivalent status to provinces. Notably, even though they are designated as cities, this is an administrative rather than a functional characterization: Chongqing, for example, is as large as the country of Austria and has third-level units (districts and counties) whose population density ranges from 26,000 people per square kilometer (Yuzhong district, which is the main business area) to as low as 60 people per square kilometer (Chengkou county). Likewise, the second level units designated as prefectural-level cities (which have a rising share over time, as shown below) consist not only of core urban areas, whose subsidiary spatial units are districts, but also may include counties that are largely rural, and county-level cities that have some modest urban development.

The variation in population density of the three main types of third-level spatial units illustrates their differing degree of urbanity, even though they all have the same *de jure* rank. In the 2020 census, the median population density of districts was 930 residents per km^2, the median for county-level cities was 345 persons per km^2 and the median density for counties was 135 persons per km^2. In other words, county-level cities are almost three times more densely populated than counties, and districts are almost three times more densely populated than county-level cities.[3]

In addition to density, the important differences between the three types of third-level spatial units relate to administrative power, fiscal benefits, land-related issues, reputation and the priority they may receive in the plans of decision makers at higher-levels (Table 1). For example, more land is available for urban construction and higher taxes can be imposed on this activity once a county becomes a county-level city (Lichtenberg and Ding 2009; Fan et al 2012) while conversion to district status also reduces land-related hurdles to the expansion of urban built-up area (Ma 2005). China has a complex system of inter-governmental fiscal transfers (Tan and Tan 2024), and transfers from prefectural-level governments to districts generally exceed those going to counties (Lu and Wang 2023). There are also more subtle benefits of being a district or county-level city that include factors related to prestige and policy priority (Chung and Lam 2004; Shen 2007; Deng et al 2022).

[3] We use census data, which consistently define the usual resident population. Other population data for China's subnational units are still often based on the *hukou* registered population, which can be a misleading indicator because millions of people now reside away from where they are registered (Gibson and Li 2017).

2.2 Pathways to upgrading

There are several way for a county to be upgraded to become either a county-level city or a district. Some counties adjacent to existing big cities are converted to districts to aid expansion of the existing urban area and relieve pressure on higher density districts nearer to the central city. For example, between 2001 and 2015 five counties on the outskirts of Beijing were converted to districts and one of these (Daxing) subsequently was chosen for Beijing's second international airport. In the neighbouring Tongzhou District, which was a county until 1997, the new Beijing Municipal Administrative Center is being developed by transferring many of the city government offices out of the central city, along with movements of the headquarters of some centrally-administered state-owned enterprises.

In less urbanized regions of China, a driving force for the conversion of counties into districts has been the growing territorial coverage of prefectural-level cities (given that there must be at least one district per prefectural-level city). At the time of the 2000 census, spatial units at the second subnational level that were organized in the form of prefectural-level cities covered territory that contained just under 87% of China's resident population (Fig. 2, top panel). By the time of the census in 2020, 94% of China's population resided in the areas organized as prefectural-level cities, with expansion especially in southwest and northern China (Fig. 2, bottom panel). The difference between the two maps is 36 prefectural-level cities in 2020 that did not have that status in the year 2000; the need for each prefectural level city to have at least one district is a driving force for some counties to be converted (as shown below).[4] While there are still some physically large areas not organized as prefectural-level cities, as seen in the map in the bottom panel of Fig. 2, these large areas (especially in Xinjiang) have only a low share of population and economic activity.

A good example of this second pathway comes from Haidong prefectural-level city in Qinghai province, to the east of the provincial capital, Xining. When this prefectural-level city was created in 2013, one of Haidong's six counties, Ledu, got upgraded to become a district (as prefectural-level cities are required to have at least one district). Two years later, adjacent Ping'an county, which borders Xining to the west, also got upgraded. This sequence is shown in Fig. 3, with the administrative status of each of the constituent spatial units in Haidong shown in 2012, 2014, and 2016. The maps also show a non-administrative indicator of urban areas, based on DMSP night-time lights (with red for the brightest areas, orange and green for less bright areas and white for unlit (completely rural) areas).[5] The main urban axis through Haidong is on a slightly tilted east-to-west orientation, along the Huangshui river (a tributary of the Yellow river); a route also used by China's G6 Beijing–Lhasa Expressway. The upgrading for Ledu and Ping'an reflects this increasing urban activity linked to Xining, 40 km to the west. It is also notable that the

[4] Strategic objectives, such as maintaining border populations, may influence prefectural-level city upgrading.

[5] DMSP is the Defense Meteorological Satellite Program. This is the only source of satellite-detected luminosity data available for time-series that begin prior to 2012.

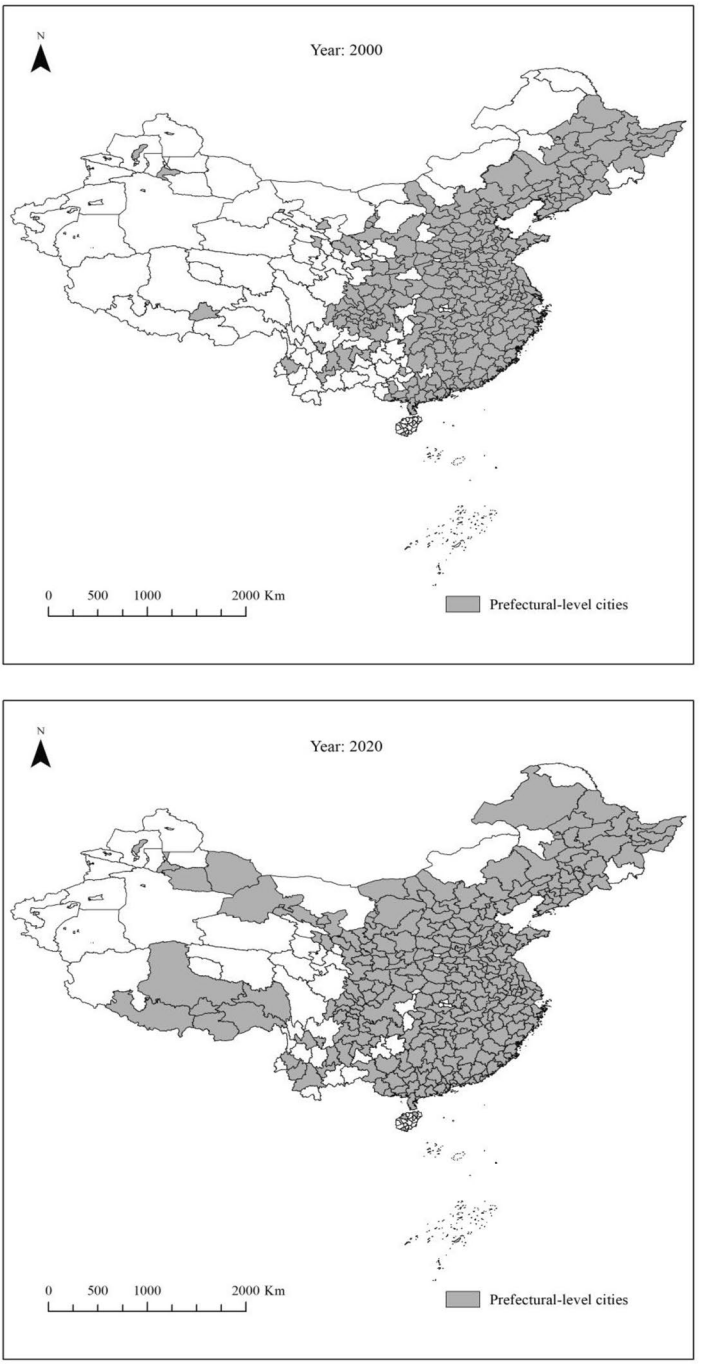


Fig. 2 The growing spread of prefectural-level cities from 2000 to 2020 in China

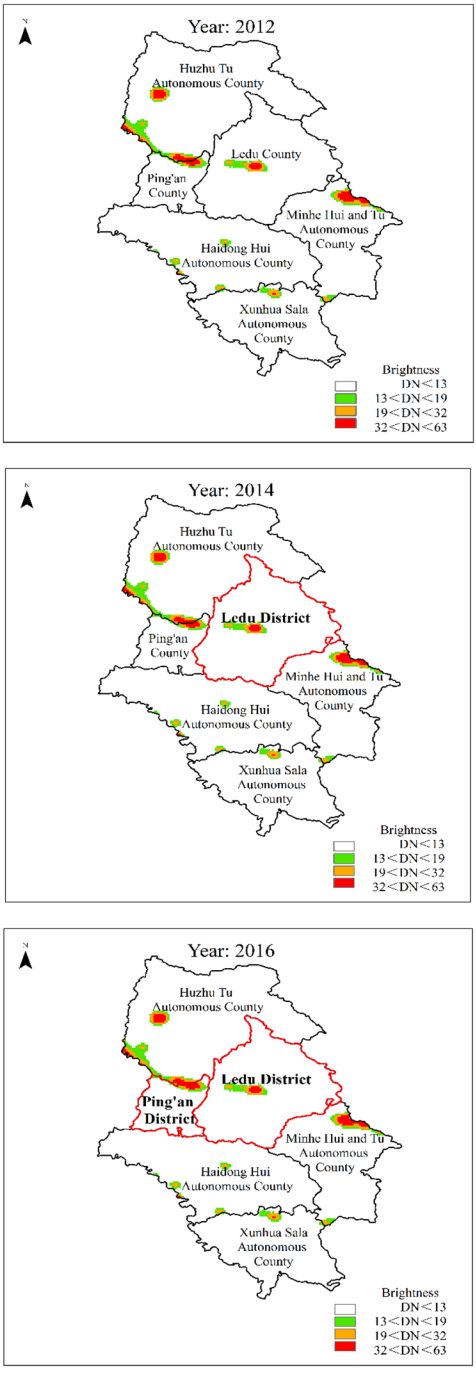


Fig. 3 County-to-district upgrading in Haidong City, Qinghai Province, China

upgrading of one county was quickly followed by the upgrading of an adjacent one; indicating some non-independence between spatial units.

The other common type of county upgrading is to county-level city; about 50 counties underwent this type of upgrade in the last two decades, contributing to about one-fifth of the total shown in Fig. 1. The county to county-level city conversion had been practiced more widely in the 1980s, when China's urbanization strategy prioritized the development of small cities. According to the standards in place at the time, almost one-half of all counties met the criteria for conversion in terms of their share of non-agricultural *hukou* population, economic activity (GDP and industrial output) and local budgetary revenues (Chung and Lam 2004). Yet standards may not have been uniformly applied given that some counties without much agglomeration potential still got upgraded (Fan et al 2012). This type of upgrading then fell out of favour for a decade from the mid-1990s, as shown clearly in year-by-year counts of upgrades recorded by China's Ministry of Civil Affairs (see Fig. 2, in Tang 2021).

2.3 Prior research evaluating effects of county upgrading

The economic effects of de facto upgrading of counties to county-level city or district status has attracted interest from researchers. The studies in this literature apply panel data econometric models to various samples (usually of third-level spatial units), to test if there is faster post-upgrading growth in indicators of economic activity; typically these studies use local GDP as the economic activity indicator but a couple use luminosity. Our summary of a selection of these studies is in Table 2, which also includes a few where some of the results are for spatial units at the prefectural level (e.g. Deng et al 2022) because of the similarity of data and methods and also because the urbanization processes studied are similar. Typically a difference-in-differences framework is used as the non-experimental evaluation approach, given that the counties subject to upgrading are unlikely to be a random selection.

A key feature of the difference-in-differences approach is that it typically treats spatial units as independent of neighbours. In other words, a stable unit treatment value assumption (SUTVA) is relied upon. While alternative difference-in-difference estimators that relax this assumption of independence have been developed using spatial econometric models (e.g., Delgado and Florax 2015) these have not been used in the studies of upgrading in China. In fact, none of the studies we reviewed seem to use spatial econometric models (as seen from column 3 of Table 2) although these is some discussion of possible spillovers in a few of the papers. For example, Tang (2021) and Zeng and You (2002) drop some untreated counties if they are near to the upgraded counties, as sensitivity analyses in case outcomes for the nearby counties had been affected by spillovers coming from the treated units. Perhaps the closest to using a spatial model is Bo (2020) who includes treatment status of contiguous neighbours; a spatial lag of a covariate can generate

Table 2 Selected studies analysing upgrading of county-level units in China

Authors	Year	Spatial models used?	Estimation time period	Spatial units used	Treatment variable period	Data source
Deng et al.	2022	No	1997–2016	1137 counties	2000–2006	GDP per capita
He	2022	No	1999–2018	153 units	2003–2018	GDP
Jiang et al	2022	No	2003–2018	283 units	2003–2018	GDP
Zeng and You	2022	No	1990–2018	2013 county-level units	2005–2018	GDP
Zhao andXi	2022	No	1996–2016	240 cities	1998–2016	GDP
Zhan et al.	2022	No	2011–2018	403 districts	2011–2018	GDP
Tang	2021	No	1976–2018	1546 counties	1992–2012	DMSP[a]
Wang	2021	No	1990–2010	1991 counties	1990–2010	GDP
Wang and Yeh	2020	No	1984–2010	82 counties	1994–2010	GDP
Bo	2020	No	1983–2003	2861 county-level units	1983–2003	GDP
Chen et al.	2020	No	1992–2012	9 districts in Hangzhou	1992–2012	GDP
Liu et al.	2019	No	1992–2008	52 districts	1992–2008	DMSP[a]
Mukim and Zhu	2018	No	1993–2004	793 counties	1998–2004	GDP
Tang and Hewings	2017	No	1995–2010	1757 counties	1995–1999 and 2005–2010	GDP and DMSP[a]
Fan et al.	2012	No	1993–2004	95 counties	1993–2004	GDP
Li	2011	No	1993–1997	1629 counties	1993–1997	GVIAO[b]
Fan et al.	2009	No	1990–2000	1537 counties	1994–1997	GDP

[a]DMSP is the Defence Meteorological Satellite Program, which provides luminosity data

[b]GVIAO is the Gross Value of Industrial and Agricultural Output

local but not global spillovers (LeSage and Pace 2009) so this specification restricts the types of spillovers.[6]

3 Data and methods

Our review of studies evaluating effects of giving selected subnational units in China city status shows that spatial econometric models have not been used, even for studies with a (limited) discussion of spillover effects. In this section, the data and estimation framework used here to shed some light on the nature of these spillovers are described. In contrast to the prior studies, we are less interested in quasi-causal analyses of particular treatments (such as converting counties to county-level cities). Instead, our interest is in using a comprehensive modelling framework to help to establish where and how spillovers may occur.[7] In doing so, we adopt the language of spatial econometrics (especially LeSage and Pace (2009)) in terms of direct, indirect, and total "impacts" even though some branches of economics increasingly reserve the term "impacts" for contexts with explanatory variables subject to either explicit manipulation (as in randomized control trials) or that vary due to some naturally occurring (partial) randomization that allows use of instrumental variables (Gibson et al 2023). Even if the patterns that we find are considered as correlations these patterns of spatial spillovers should still be present (and so should be accounted for) in other studies that use more formal quasi-causal models. In other words, the results reported below are partly an exercise in allowing the data to speak, to see whether modelling decisions that a priori rule out spatial spillovers have a sound basis.

3.1 Data sources

In order to empirically analyse spillovers we use two measures of economic activity—local GDP and local luminosity. China reports GDP for different types of spatial units in a variety of publications from the National Bureau of Statistics (NBS). We rely on three main sets of these: (i) annual editions of the China Statistical Yearbook (county-level) (in Chinese it is *Zhongguo Xianyu Tongji Nianjian (Xianshi Juan)*), (ii) annual editions of the China City Statistical Yearbook (known as *Zhongguo Chengshi Tongji Nianjian*), and (iii) annual editions of the Statistical Yearbook for each city or province (for example, the Beijing Statistical Yearbook) (NBS, various dates). These different types of publications are needed because, for example, annual GDP of districts comes out in a different report than the annual GDP of counties. Each edition reports on GDP the previous year, so we use the 2001 to 2020 editions to obtain annual GDP data from 2000 to 2019. Overall, we have a

[6] This variable was not statistically significant so Bo (2020) concluded that that there were no spillovers.

[7] This is one reason for combining together conversions from county to county-level city, and from county to district. As a practical matter it would be difficult to use spatial models to study either pathway in isolation because the spatial weights matrix would then be incomplete.

balanced panel of annual GDP data for each of $n = 2342$ units at the 3rd level of the sub-national administrative hierarchy, where these units maintain a consistent spatial definition from 2000 to 2019.

The luminosity data are DMSP annual composites from satellites F14, F15, F16 and F18 that collectively cover each year from 2000 to 2019.[8] The stable lights product used here (where 'stable' simply means ephemeral lights from sources such as from fires are removed) provides 6-bit digital numbers (DN) that range from 0 to 63 (with higher values for greater luminosity). The technical details on the steps used to create these data are in Baugh et al (2010) and Ghosh et al. (2021).[9] We use the sum of lights for each third-level spatial unit in each year, following previous research for China that finds that this luminosity indicator is the best proxy for local economic activity (Zhang and Gibson 2022).

To find when counties were converted to either county-level city or district status we compared administrative settings of all spatial units listed in the 2000 population census with the listings for the same places in the 2010 and 2020 population censuses. These comparisons identified counties whose status had changed, and we then went to the specific page for each of those places in the Baidu Encyclopaedia (baike.baidu.com). Within the place-specific page the administrative history section gave the precise data of any change in administrative status.

The summary statistics for the two outcome measures and for the indicator of whether a county was upgraded are reported in Appendix Table 1 We show results for observations that are year-by-spatial unit, and also time-averaged results for the ever-upgraded subset of the spatial units. The baseline differences shown by this comparison are controlled for by the spatial fixed effects in the estimation framework discussed below.

3.2 Estimation framework

Spatial econometric methods let us examine the nature of possible spillovers, and are used for this purpose in many contexts (Krisztin et al 2020; Asyahid and Pekerti 2022). A key aspect of these models aiding the study of spillovers is that possible interactions between spatial units are summarized with a $N \times N$ spatial weights matrix, W. In this study we use a row normalized contiguity weights matrix that has values of one for neighbours and zero otherwise, with a diagonal of zeros because a spatial unit cannot neighbour itself. At the level of spatial disaggregation that we use, the average spatial unit in China has six neighbours.

In what follows, the proxy for economic output in spatial area i in year t is denoted as O_{it}, where the two proxy variables we use are log GDP in our main specification, and the log of the sum of night-time lights in our sensitivity analysis. The indicator

[8] Each satellite orbits for many years and so two or more DMSP satellites may be in orbit simultaneously. The time that they observe earth gets earlier as they age and so there is often only one that is well timed to capture luminosity. We use images from F14 from 2000 to 03, F16 for 2004–09, F18 for 2010–12, and F15 for 2013–19.

[9] The DMSP data are available from: https://eogdata.mines.edu/products/dmsp/.

for whether a spatial unit has been upgraded is D_{it} the μ_i are time-invariant fixed effects for each spatial unit, the ϑ_t are year fixed effects, and e_{it} is a random error. By using the spatial weights matrix we can allow for spatial lags, which are averages of these variables over the neighbouring units.

Our starting point is a very general model, which is a spatial autoregressive model with spatial autoregressive errors (SARAR).[10] This model allows for spatially lagged dependent variables, spatially lagged independent variables and spatially lagged errors:

$$O_{it} = \lambda WO_{it} + \beta_1 D_{it} + \beta_2 WD_{it} + \mu_i + \vartheta_t + \rho Wu_{it} + e_{it} \tag{1}$$

The SARAR model allows for changes in an outcome variable in a given area to have effects on contemporaneous outcomes in other areas (via the autoregressive spatial lag of the dependent variable, if $\lambda \neq 0$). It also allows changes in independent variables (such as being converted from county to city status) to affect not only own-area outcomes but also outcomes in neighbouring areas $\left(\text{if } \beta_2 \neq 0\right)$. The ρWu_{it} term allows for spatial autocorrelation, where errors for a given area correlate (ρ) with a weighted average of the errors from surrounding areas. Equation (1) nests a spatial Durbin model if $\rho = 0$, a spatial auto-regressive model (*aka* spatial lag model) where only the dependent variable is spatially lagged if $\beta_2 = \rho = 0$, a spatial error model where only the errors are spatially lagged (if $\lambda = \beta_2 = 0$), and the most restrictive of all, which is an aspatial model with no spatial lags $\left(\text{ if } \lambda = \beta_2 = \rho = 0\right)$. The aspatial model has been the approach underpinning the previous studies of county upgrading (Table 2). The encompassing nature of Eq. (1) allows for a general-to-specific model selection strategy which appears to be more robust than the reverse simple-to-general selection strategy, especially if they are any anomalies in the Data Generating Process (Mur and Angulo 2009).

An important feature of spatial econometric models is that lags of either the outcome variable or of independent variables (but not of errors) mean that total effects of changes in an independent variable—e.g. whether a county is upgraded—may be quite different to what the regression coefficient on the dummy variable for being upgraded shows. Thus, while $\hat{\beta}_1$ is the object of interest in the typical model without spatial lags, in the spatial models when either the spatial lags of outcomes or the spatial lags of independent variables are non-zero then $\hat{\beta}_1$ does not capture the total effect of a change in the administrative status of a county. A useful decomposition of the more complex spatial relationships that occur relies on rewriting Eq. (1) in matrix notation (for simplicity, subscripts are dropped and fixed effects and error terms combined in v because the errors do not affect this decomposition) as:

$$O = (I - \lambda W)^{-1}\left(D\beta_1 + WD\beta_2\right) + (I - \lambda W)^{-1} v \tag{2}$$

Following Elhorst (2012), the $N \times N$ matrix of partial derivatives can be written (noting that diagonal elements of W are zero) as:

[10] This is estimated by quasi-maximum likelihood (Lee and Yu 2010), using the Drukker et al (2013) estimator.

$$\frac{\partial O}{\partial D_k} = (I - \lambda W)^{-1}\left(\beta_{1k} I_N + \beta_{2k} W\right) \tag{3}$$

where D_k is the upgrade status in spatial unit k. The total marginal effect on output that is associated with a county being upgraded has two components, a direct one and an indirect one, that may both vary over space. The estimator that we use follows LeSage and Pace (2009) in reporting a single direct effect, that averages the diagonal elements of the matrix in (3) and a single indirect effect that averages the row sums of the non-diagonal elements of that matrix. Indirect effects arise not just from adjacent area units, if $\beta_{2k} \neq 0$, but also from (potentially) all areas through the spatial autoregressive effect if $\lambda \neq 0$. Thus, there can be both local and global spillovers and when these are accounted for, averages from the matrix of derivatives $\partial O/\partial D_k$ may be quite different to the estimated direct impact effect, $\hat{\beta}_1$.

4 Results

The results of estimating Eq. (1) and then imposing various restrictions on the parameters and estimating the nested models are given in Table 3 (for GDP) and Table 4 (for luminosity). The nesting restrictions are rejected in all cases, so that the SARAR models appear to be the most data-acceptable models for both GDP and luminosity. The discussion therefore concentrates mostly on the results in column (1) for the SARAR model, and then by way of contrast on column (5) for the standard two-way fixed effects panel data model that does not allow for any spatial lags. The column (5) models are similar to the models used in the literature to date (such as some of those studies summarized in Table 2).

The lack of data acceptability for any of the nesting restrictions indicates that for the spatial units and period we study, the interactions are occurring through the spatial lags of the outcomes, the lags of the treatment, and the lags of the errors. There are two consequences of this pattern. First, partial approaches that only allow for local spillovers, such as including the treatment status of nearby spatial units (e.g., as in Bo 2020), may not reveal the full pattern of spillovers. Second, the regression coefficients by themselves do not tell the full story and the matrices of marginal effects based on Eq. (3) need to be taken into consideration.

The results of the marginal effects calculations are reported in the "average impacts" rows of Table 3 and 4, using the decomposition due to LeSage and Pace (2009). These results indicate that the positive direct relationship between a county being upgraded and economic activity is amplified by positive indirect effects. For example, from column (1) of Table 3 it is apparent that the upgrading is associated with GDP being 13% higher, comprised of seven percent as the direct effect and six percent as the indirect effect (thus, the spillovers and feedback effects provide a component that is almost as large as the direct effect).

When luminosity is used as the proxy for local economic activity, as a sensitivity analysis in case of mistrust in China's GDP figures, upgrading is associated with 18% higher activity (10 percent direct and eight percent indirect). One caveat to this result is that DMSP data on night-time lights are subject to blurring, and so tend

Table 3 Relationships between county upgrading and the change in economic output (log GDP) in China: 2000 to 2019

	(1)	(2)	(3)	(4)	(5)
	Spatial lag of errors, covariate and outcome	Spatial lag of the covariate and outcome	Spatial lag of the outcome	Spatial lag of the errors	Standard panel model analysis
County is upgraded	0.076***	0.053***	0.052***	0.053***	0.056***
	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)
Average impacts:					
Direct	0.070***	0.054***	0.057***	0.053***	0.056***
	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)
Indirect	0.056***	0.024	0.076***	n.a	n.a
	(0.013)	(0.034)	(0.010)		
Total	0.126***	0.079**	0.133***	n.a	n.a
	(0.017)	(0.039)	(0.018)		
County fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Spatial lag: upgraded	Yes	Yes	No	No	No
Spatial lag: output	Yes	Yes	Yes	No	No
Spatial lag: errors	Yes	No	No	Yes	No
All covariates = 0	χ^2 = 34,414***	χ^2 = 870,721***	χ^2 = 870,644***	χ^2 = 116,667***	χ^2 = 569,235***
Nesting restrictions	n.a	χ^2 = 35,586***	χ^2 = 35,638***	χ^2 = 1665***	χ^2 = 59,884***

The sample period is 2000–2019, for 2342 county-level units, giving an estimation sample of n = 46,840. Coefficients for the fixed effects and the spatial lags are not reported. The decomposition of average impacts into direct, indirect and total components is based on LeSage and Pace (2009). The nesting restrictions are imposed on the SARAR model in column (1) to derive the models in columns (2–5) Standard errors are in (), with statistical significance at the 1%, 5% and 10% level denoted by ***, **, *

Table 4 Relationships between county upgrading and the change in luminosity (log DMSP) in China: 2000 to 2019

	(1)	(2)	(3)	(4)	(5)
	Spatial lag of errors, covariate and outcome	Spatial lag of the covariate and outcome	Spatial lag of the outcome	Spatial lag of the errors	Standard panel model analysis
County is upgraded	0.111***	0.093***	0.095***	0.079***	0.124***
	(0.013)	(0.012)	(0.012)	(0.011)	(0.014)
Average impacts:					
Direct	0.100***	0.111***	0.104***	0.079***	0.124***
	(0.012)	(0.013)	(0.013)	(0.011)	(0.014)
Indirect	0.082***	0.279***	0.140***	n.a	n.a
	(0.021)	(0.057)	(0.017)		
Total	0.183***	0.391***	0.244***	n.a	n.a
	(0.027)	(0.065)	(0.030)		
County fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Spatial lag: upgraded	Yes	Yes	No	No	No
Spatial lag: output	Yes	Yes	Yes	No	No
Spatial lag: errors	Yes	No	No	Yes	No
All covariates = 0	χ^2 = 4573***	χ^2 = 73,633***	χ^2 = 73,632***	χ^2 = 8291***	χ^2 = 38,300***
Nesting restrictions	n.a	χ^2 = 46,752***	χ^2 = 46,757***	χ^2 = 2441***	χ^2 = 70,072***

The sample period is 2000–2019, for 2342 county-level units, giving an estimation sample of n = 46,840. Coefficients for the fixed effects and the spatial lags are not reported. The decomposition of average impacts into direct, indirect and total components is based on LeSage and Pace (2009). The nesting restrictions are imposed on the SARAR model in column (1) to derive the models in columns (2–5) Standard errors are in (), with statistical significance at the 1%, 5% and 10% level denoted by ***, **, *

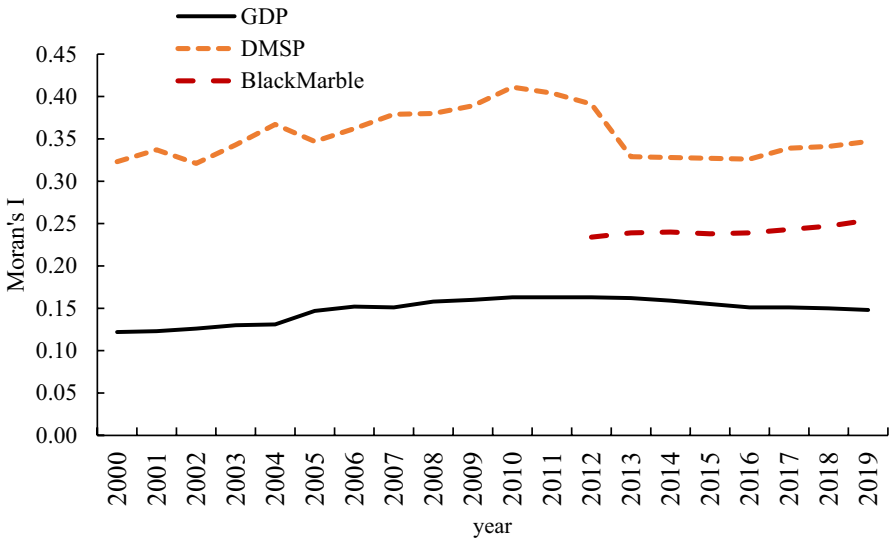


Fig. 4 Spatial autocorrelation in GDP and night-time lights data

to overstate similarity for nearby areas (Zhang et al 2023).[11] To provide some evidence on this issue, Fig. 4 has time-series of Moran's *I* statistics from all $n = 2342$ county-level spatial units each year. The *I* statistic is a popular measure of spatial autocorrelation, ranging from -1 to + 1 (higher values show greater similarity). The Moran's *I* for the GDP of the third-level spatial units is around 0.15 but it is far higher, at around 0.4, for DMSP night-time lights.[12] The higher *I* statistic most likely is an overstatement; more spatially precise luminosity data that are free of blurring issues—from NASA's Black Marble series—have *I* statistics about halfway between values for DMSP and for GDP. Hence, there is likely to be an exaggerated similarity of nearby areas with the DMSP data, which could contribute to the appearance of larger spillovers. Thus, we consider the results in Table 4 to be an upper bound for the contribution of indirect spillover effects.

If results in column (1) are compared with those in column (5) for Table 3 and 4 it provides some insight into possible distortions in the literature on effects of county upgrading in China. If spatial lags are not included in the models (a restriction that is inconsistent with the data, according to the nesting tests), GDP appears to be just under six percent higher in the upgraded counties, post-upgrading (based on column (5) of Table 3). Yet the SARAR model results, which are the ones supported by the data, show total upgrading effects of 13% higher GDP. In other words, the standard model that a priori rules out spillovers only shows about two-fifths of

[11] We use these data even acknowledging this flaw because alternative luminosity indicators are only available from 2012 onwards and so we would lose a majority of our observations if we used these alternatives.

[12] The derived standard errors for the *I* statistics in Fig. 4 are about 0.01, so the difference between values for DMSP and for GDP are statistically significant.

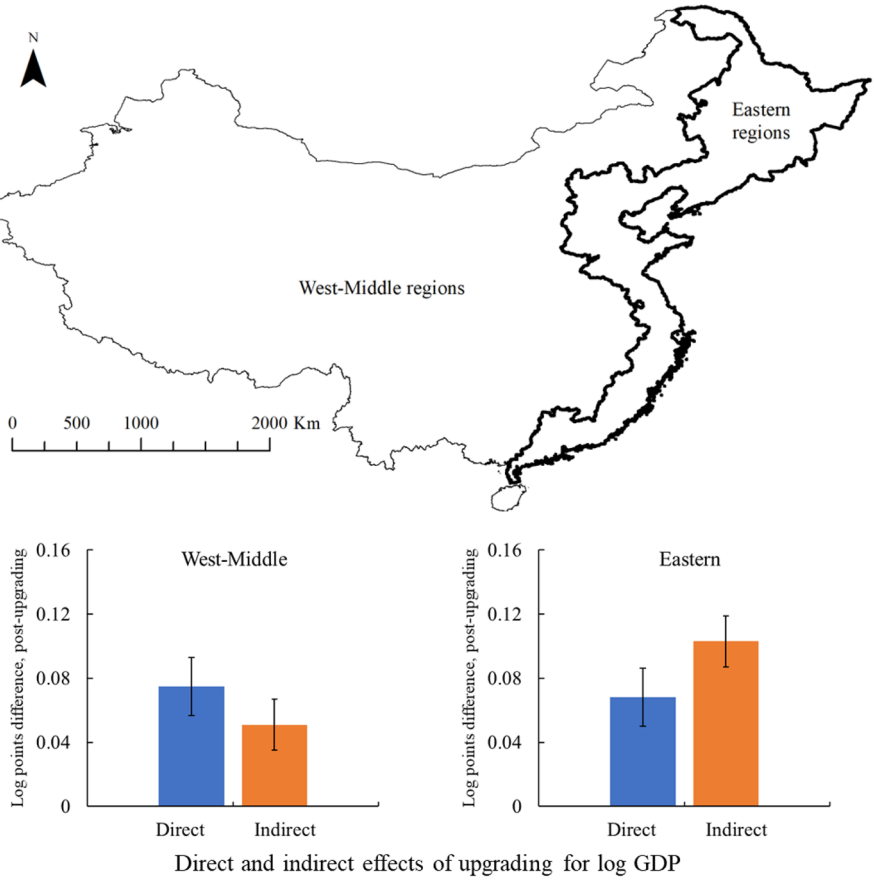


Fig. 5 Spillovers appear to especially matter in China's Eastern regions

the effect. If we, instead, use luminosity as the economic activity indicator the same pattern is visible with the aspatial model in column (5) of Table 4 giving an effect that is only about two-thirds of what the data-consistent model in column (1) shows. In other words, it is possible that prior studies have understated the economic effects of converting counties to cities because they have not allowed for positive spillovers.

4.1 Heterogeneity analysis

There is huge variation across space in the density of economic activity in China and so some geographic heterogeneity in the strength of the spillovers is likely. We explore this in Fig. 5, which presents results for the Eastern region of China versus the rest. The usual disaggregation divides China into Eastern, Middle, and Western regions as the top-level geographic breakdown but these regions are quite unequal in economic size. The Eastern region produces 56% of GDP, is home to 46% of residents (as of the 2020 census) but has only 17% of the land area. While the direct

effects on GDP of converting counties to cities are of similar magnitude in the East and the rest (at just under eight percent higher GDP), the indirect spillover effects are twice as large in the East as they are in the rest of China (ten percent versus five percent). The results in Fig. 5 are consistent with what Tang (2021) finds, of stronger total effects of upgrading in the East where the initial agglomeration forces are stronger.

There are at least two implications of stronger indirect spillovers in the high density Eastern region. In terms of measurement, China lacks a Metropolitan Statistical Area (MSA) designation, for large population nuclei plus adjacent communities that are highly integrated, typically seen in commute-to-work patterns (Forstall and Chan 2015). Elsewhere, such as the United States, the MSA concept helps to study the spatial patterns of economic development without limits imposed by data based on administrative boundaries. The links between nearby areas in Eastern China seen in the strong spillovers suggest a MSA approach could be useful, even if other parts do not require this statistical innovation (because of the weaker spillovers). Relatedly, in terms of policy, China in the last decade saw dispersed urbanization, as smaller cities experienced faster growth in resident population (comprised of both *de jure* holders of local *hukou* plus de facto residents with *hukou* from elsewhere), than what big cities, that are especially in the Eastern region, experienced. This was a sharp change from the prior period (between the 2000 and 2010 censuses), where big and small cities both had urban population growth rates of above 8% per annum. A dispersed urbanization may potentially forego some agglomeration gains that could produce up to 20% higher per capita GDP under concentrated patterns of urban growth (MGI, 2009). The stronger spillovers in the Eastern region favour a concentrated form of urbanization there, if the gains from agglomeration are to be exploited.

5 Discussion and conclusions

In this paper we have used econometric models that explicitly allow the pattern of spatial spillovers to be revealed through the estimated coefficients on the various spatial lags. In contrast, prior analyses of the regional development intervention we study—administrative conversion of some of China's counties into cities—use estimation frameworks where the observations are treated as independent of their neighbours and so do not allow the pattern of spillovers to be freely revealed. We find that all three types of spatial lags—of the outcomes, the treatment, and the errors—are statistically relevant, with restrictions to arbitrarily set any of these lags to zero rejected by the data. Consequently, positive direct relationships between a county being upgraded and post-upgrading local economic growth are amplified by positive indirect relationships. These positive indirect relationships operate both locally and globally. If these spillovers are ignored, the upgrading effects seem only two-fifths as large, according to the economic activity indicator that we consider more reliable, which is the reported GDP of China's third level sub-national units (counties, county-level cities, and districts).

Our objective in reporting these results is to inform the growing literature that sets out to evaluate impacts of China's county upgrading process. While there has been some limited discussion of spatial spillovers in this literature, the empirical approaches that have been used are informal, such as dropping observations within various distance bands of treated units in case the observations were affected by spillovers. One study did include the upgrade status of adjacent spatial units (Bo 2020) but that approach implicitly restricts spillovers to operate locally rather than globally. Instead, the focus of much of the existing literature has been to provide reassurance to readers that trends in economic activity for counties whose status had not changed were a reasonable counterfactual for what might have happened in the upgraded counties if they hadn't been upgraded. In some ways our approach is the reverse; we use a modelling framework that lets the data speak freely about the nature of the spatial spillovers, even if the patterns revealed may be considered as just conditional correlations. We have no reason to believe that the spillover patterns would be different if a quasi-causal framework had been used that involved testing for parallel trends, trimming on propensity scores and so on. In other words, we expect that the spillover patterns when the data are allowed to speak freely should also show up in quasi-causal analyses as long as those analyses do not a priori limit the nature of the spillovers that can be revealed.

There are several ways to extend the current analysis. A more nuanced way to allow for agglomeration forces that generate spatial spillovers might use asymmetric weights so that a more populous urban area has a greater effect on an adjacent county than the reverse. In the Section II example of Ledu District, in Haidong prefectural-level city of Qinghai, it is likely no coincidence that the next county upgraded was Ping'an, to the west, rather than the Minhe Hui and Tu Autonomous County to the east. The provincial capital of Xining, with a population of two million, is just 40 km to the west; in the other direction the next big city—Lanzhou in Gansu province—is almost 200 km away. So, a gravitational pull for urban development in Haidong is probably more towards the west, where the nearest large market is, and a weights matrix that allows for asymmetries would be one way to recognise such patterns. Presumably, such a weights matrix could be combined with a difference-in-differences setup to extend the estimator proposed by Delgado and Florax (2015) that allows for spatial interactions.

Another extension would consider treatment intensity. In our results, all converted counties get the same value of the treatment indicator (= 1). A more flexible approach would allow for a sequence of treatments, such as from county to county-level city and then to district. It remains to be seen whether these extensions would alter our core finding of the need to allow for spillovers when evaluating China's regional development interventions. To the extent that the finding of positive spillovers persists, it does suggest that upgrading has successfully created agglomeration effects, especially in the Eastern region. In contrast, some of the early evaluations, such as Fan et al (2012) reached more pessimistic conclusions about upgrading and so there may be grounds to revisit these earlier findings.

Moving beyond China, our study also has implications for the international literature. First, we provide another example to add to the emerging evidence from Europe and North America that spillovers should be considered in studies that

evaluate regional development interventions. If these spillovers are a priori ruled out by using empirical approaches that rely on stable unit treatment value assumptions the results may provide a misleading guide to the actual impact of the intervention. Second, our results raise a methodological question, for the developers of new spatial estimation approaches to think about. There are at least two sets of modelling decisions faced in applied research studying regional development interventions; the approach to causal effects estimation (including just estimating conditional correlations as we do), and the framework for allowing any spatial spillovers to reveal themselves. Some recent causal effects studies allowing spatial spillovers start with particular spatial models, such as a spatial Durbin model or a spatial lag model, and generalize the DID model with the particular structure of spatial lags embedded in their choice of spatial model. Instead, here we started with a more general spatial model and tested to see if nested models, like the spatial Durbin model, were data-acceptable; restrictions to nest the special case models were rejected in all cases. So the question arises as to which decision to focus on first—the causal effects estimation framework or the spillover effects framework, and whether the two decisions have an interaction, as may occur if first settling upon a particular spatial model so as to implement a particular causal effects approach obscures some spatial effects that otherwise could have been revealed if a different causal effects framework had been used.

Appendix

See Table 5.

Table 5 Descriptive statistics for upgrading, log GDP and log luminosity

	All spatial units		Upgraded (during 2000–19)		Not upgraded	
	Mean	Std Deviation	Mean	Std Deviation	Mean	Std Deviation
Upgraded	0.046	0.210	1	n.a	0	n.a
log (GDP)	4.093	1.524	5.338	1.059	4.033	1.517
log (sum DMSP DN values)	8.393	1.405	9.378	0.832	8.346	1.410

There are 46,840 observations (2342 third-level spatial units, each observed for 20 years). Of these, there were 2163 post-upgrade observations (with spatial units that were already upgraded prior to 2000 coded as having unchanged status). Full period means for the ever-upgraded (within the 2000–19 period) spatial units are 4.65 for log (GDP) and 9.09 for log (sum DMSP DN values), versus 4.03 and 8.31 for never-upgraded units, showing that the post-upgrading differences exceed the time-average differences

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. Financial support from the Marsden Fund project UOW1901 and helpful comments from participants at the 14th Geoffrey J.D. Hewings Workshop at WIFO, from a seminar audience at Beijing Technology and Business University and from two anonymous reviewers are gratefully acknowledged.

Conflict of interest The authors declare that they have no conflict of interest.

References

Asyahid, E.A., Pekerti, I.S.: Economic impact of natural disasters, spillovers, and role of human development: case of Indonesia. Lett. Spat. Resour. Sci.resour. Sci. **15**(3), 493–506 (2022)

Baugh, K., Elvidge, C., Ghosh, T., Ziskin, D.: Development of a 2009 stable lights product using DMSP-OLS data. Proceed. Asia-Pacific Adv. Netw. **30**, 114 (2010)

Bo, S.: Centralization and regional development: evidence from a political hierarchy reform to create cities in China. J. Urban Econ. **115**, 103182 (2020)

Cao, Z., Wang, L., Zhang, Y.: Environmental effects of city-county mergers in China: strengthening governance or aggravating pollution? Sustainability. **14**(9), 5522 (2022)

Chen, A., Partridge, M.D.: When are cities engines of growth in China? Spread and backwash effects across the urban hierarchy. Reg. Stud. **47**(8), 1313–1331 (2013)

Chen, Y., Wang, K., Wang, F.: Economic growth mechanism of county-to-district conversion and its dialectical relationship with city shrinkage: case study of county-to-district conversion in Hangzhou, China. J. Urban Plan. Dev. **146**(4), 05020029 (2020)

Chung, J.H., Lam, T.C.: China's city system in flux: explaining post-Mao administrative changes. China Q. **180**, 945–964 (2004)

De Castris, M., Di Gennaro, D., Pellegrini, G.: Do spatial spillovers of regional policies aid the reduction of regional inequalities in Europe? In Capello, R. and Conte, A., (editors). *Cities and Regions in Transition*. FrancoAngeli, (2023), pp. 265–288.

Delgado, M.S., Florax, R.J.: Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction. Econ. Lett. **137**, 123–126 (2015)

Deng, N., Feng, B., Partridge, M.D.: A blessing or curse: the spillover effects of city–county consolidation on local economies. Reg. Stud. **56**(9), 1571–1588 (2022)

Drukker, D.M., Prucha, I.R., Raciborski, R.: Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. Stata J. **13**(2), 221–241 (2013)

Elhorst, J.P.: Dynamic spatial panels: models, methods, and inferences. J. Geogr. Syst. **14**(1), 5–28 (2012)

Fan, S., Li, L., Zhang, X.: Rethinking China's under-urbanization: an evaluation of its county-to-city upgrading policy (No. 875). International Food Policy Research Institute (IFPRI), (2009)

Fan, S., Li, L., Zhang, X.: Challenges of creating cities in China: Lessons from a short-lived county-to-city upgrading policy. J. Comp. Econ. **40**(3), 476–491 (2012)

Forstall, R.L., Chan, K.W.: Urban places: statistical definitions. In: Wright, J. (ed.) International Encyclopedia of the Social and Behavioral Sciences, pp. 854–861. Elsevier (2015)

Gambina, D., Mazzola, F.: The impact of spatial spillovers on Cohesion Funds' effectiveness: A spatial panel analysis for the Italian provinces. In Capello, R. and Conte, A., (editors). *Cities and Regions in Transition*. FrancoAngeli, (2023), pp. 245–263.

Ghosh, T., Baugh, K.E., Elvidge, C.D., Zhizhin, M., Poyda, A., Hsu, F.C.: Extending the DMSP nighttime lights time series beyond 2013. Remote Sens. **13**(24), 5004 (2021)

Gibson, J., Li, C.: The erroneous use of China's population and per capita data: a structured review and critical test. J. Econ. Surv. **31**(4), 905–922 (2017)
Gibson, J., Jiang, Y., Susantono, B.: Revisiting the role of secondary towns: how different types of urban growth relate to poverty in Indonesia. World Dev. **169**, 106281 (2023)
He, J., Jaros, K.: The multilevel politics of County-to-District mergers in China. J. Contemp. China **32**(144), 932–950 (2023)
Jiang, M., Chen, W., Yu, X., Zhong, G., Dai, M., Shen, X.: How can urban administrative boundary expansion affect air pollution? Mechanism analysis and empirical test. J. Environ. Manage. **322**, 116075 (2022)
Krisztin, T., Piribauer, P., Wögerer, M.: The spatial econometrics of the coronavirus pandemic. Lett. Spat. Resour. Sci. Resour. Sci. **13**(1), 209–218 (2020)
Landry, P.F.: Decentralized authoritarianism in China: the Communist Party's control of local elites in the post-Mao era, vol. 1. Cambridge University Press, New York (2008)
Lee, L., Yu, J.: Estimation of spatial autoregressive panel data models with fixed effects. J. Econom. **154**(1), 165–185 (2010)
LeSage, J., Pace, R.K.: Introduction to Spatial Econometrics. Chapman and Hall and CRC Press, Boca Raton (2009)
Li, L.: The incentive role of creating cities in China. China Econ. Rev. Econ. Rev. **22**(1), 172–181 (2011)
Lichtenberg, E., Ding, C.: Local officials as land developers: urban spatial expansion in China. J. Urban Econ. **66**(1), 57–64 (2009)
Liu, X., Zeng, J., Zhou, Q.: The chosen fortunate in the urbanization process in China? evidence from a geographic regression discontinuity study. Rev. Dev. Econ. **23**(4), 1768–1787 (2019)
Lu, S., Wang, H.: Regional integration by administrative division adjustment: City-County consolidation and layout of land development in China. J. Urban Plan. Dev. **149**(3), 04023017 (2023)
Ma, L.J.: Urban administrative restructuring, changing scale relations and local economic development in China. Polit. Geogr.geogr. **24**(4), 477–497 (2005)
McKinsey Global Institute [MGI]: *Preparing for China's urban billion*. (2009) https://www.mckinsey.com/featured-insights/urbanization/preparing-for-chinas-urban-billion
Mukim, M., Zhu, T.J.: Empowering cities: good for growth? Evidence from the people's republic of China. Asian Dev. Rev. **35**(1), 175–195 (2018)
Mur, J., Angulo, A.: Model selection strategies in a spatial setting: Some additional results. Reg. Sci. Urban Econ. **39**(2), 200–213 (2009)
National Bureau of Statistics [NBS]. *Beijing Statistical Yearbook*. China Statistics Press, Beijing (2001–2020)
National Bureau of Statistics [NBS]. *China City Statistical Yearbook*. China Statistics Press, Beijing (2001–2020)
National Bureau of Statistics [NBS]. *China Statistical Yearbook (county-level)*. China Statistics Press, Beijing (2001–2020)
Park, A., Johnston, B.: Rural development and dynamic externalities in Taiwan's structural transformation. Econ. Dev. Cult. Change **44**(1), 181–208 (1995)
Qiu, F., Tong, Q.: A spatial difference-in-differences approach to evaluate the impact of light rail transit on property values. Econ. Model. **99**(1), 105496 (2021)
Ravallion, M.: Externalities in rural development: Evidence for China. In: Kanbur, R., Venables, A. (eds.) Spatial Inequality and Development, pp. 137–162. Oxford University Press, Oxford (2005)
Rosenstein-Rodan, P.N.: Problems of industrialisation of eastern and south-eastern Europe. Econ. J. **53**(210–211), 202–211 (1943)
Shen, J.: Scale, state and the city: urban transformation in post-reform China. Habitat Int. **31**(3–4), 303–316 (2007)
Su, X.: Unpacking administrative rank: intercity competition and the remaking of local state space in China. Polit. Geogr.geogr. **92**, 102518 (2022)
Tan, X., Tan, Y.: Equalization through China's intergovernmental fiscal system: The effectiveness of central and provincial transfers. *Asian Dev. Rev*. **41**(1), (2024)
Tang, W.: Decentralization and development of small cites: evidence from county-to-city upgrading in China. China Econ. Quarterly Int. **1**(3), 191–207 (2021)
Tang, W., Hewings, G.J.: Do city–county mergers in China promote local economic development? Econ. Transit. **25**(3), 439–469 (2017)
Tian, C., Ji, W., Chen, S., Wu, J.: The time and spatial effects of a City-County merger on housing prices—Evidence from Fuyang. Sustainability. **12**(4), 1639 (2020)

Wang, J., Wang, J.: Effects of urban administrative system on urban system development in China. Role State China's Urban Syst. Dev. Gov. Capacit. Inst. Policy. pp. 161-185 (2021)
Wang, J., Yeh, A.G.: Administrative restructuring and urban development in China: effects of urban administrative level upgrading. Urban Stud. **57**(6), 1201–1223 (2020)
Zeng, Z., You, C.: The price of becoming a City: decentralization and air pollution—the evidence from the policy of County-to-City upgrade in China. Int. J. Environ. Res. Public Health **19**(23), 15621 (2022)
Zhang, X., Gibson, J.: Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. Remote Sens. **14**(5), 1282 (2022)
Zhang, S., Zhang, M.A., Sun, W.: Administrative division adjustment and housing price co-movement: evidence from City-County mergers in China. Chin. World. Econ. **30**(4), 149–173 (2022)
Zhang, X., Gibson, J., Deng, X.: Remotely too equal: popular DMSP night-time lights data understate spatial inequality. Reg. Sci. Policy Pract.pract. **15**(9), 2106 (2023)
Zhao, B., Xi, X.: Economic effects of conversion from county (or county-level city) to municipal district in China. PLoS ONE **17**(9), e0272267 (2022)

Chapter 6: Local economic effects of connecting to China's high-speed rail network: Evidence from spatial econometric models

Xiaoxuan Zhang and John Gibson

Abstract

China's high-speed rail (HSR) has quickly expanded to over 40,000 km of lines operating and another 10,000 km under construction. This is over 10-times longer than the networks in long-established HSR countries like France, Germany or Japan. While fewer than 100 county-level units had stations on the HSR network in the first years of operation, the eight years from 2012-19 saw almost 400 more county-level units connect to the HSR network. Effects on local economic activity from this substantial increase in connections to the HSR network remain contested. Some prior studies find either insignificant effects on local economic growth or even negative effects in peripheral regions. In light of this debate we use spatial econometric models for a panel for almost 2500 county-level units to study effects of connecting to the HSR network. We especially concentrate on the 2012-19 period that has high quality night-time lights data to provide an alternative to GDP as an indicator of growth in local economic activity. Our spatial econometric models allow for spatial lags of the outcomes, of the covariates, and of the errors. We also address potential endogeneity of the HSR networks and connections, using an instrumental variables strategy. Across a range of specifications, we generally find that growth in local economic activity is lower following connection to the HSR network, with this effect especially apparent when using high quality night-time lights data for the 2012-19 period. Hence, expansion of the HSR network may not boost China's economic growth.

JEL Codes

R12

Keywords

High-speed rail
infrastructure
luminosity
spatial spillovers
China

Acknowledgements

Financial support from the Marsden Fund project UOW1901.

I. Introduction

Transportation infrastructure is generally regarded as a key driver of regional economic growth and development (Ma, Chen, and Yang 2020; Egger, Loumeau, and Loumeau 2023). High-speed rail (HSR) is a significant transport innovation (Wang, Jiang, and Miao 2023; Li, Yu, and Ma 2024), with advantages of speed, efficiency, energy saving, safety, punctuality, and stability (Li et al. 2021; Wang et al. 2023; Liu, Diao, and Lu 2024). Since the first HSR route opened in Japan in 1964, more than 20 countries have commenced HSR operations (Wang et al. 2024). Although China was a latecomer to HSR construction, it has developed extremely fast (Mei, and Zhang, 2021). Starting with the opening of the Beijing-Tianjin intercity HSR in 2008, China developed the world's longest HSR network with the highest transportation density (Banerjee et al. 2020; Wang, Jiang, and Miao 2023; Tang et al. 2024). According to the State Council Information Office, China's HSR network exceeded 43,700 km by the end of November 2023, which ranks first in the world (Ren et al. 2024). In addition, China's HSR network is still expanding. The China National Railway Group Limited stated that China's railway network will eventually span approximately 200,000 kilometers, including about 70,000 kilometers of HSR, by 2035 (Li and Li, 2024).

A growing literature studies local economic development impacts of HSR networks. For example, Ahlfeldt and Feddersen (2018) find that HSR stimulated the economic vitality of cities in Germany. Carbo et al. (2019) studied economic effects of HSR in Spain from 1995 to 2014 using difference-in-differences (DID) and found positive effects on urban economies. Based on a Computable General Equilibrium (CGE) model, Kim and Yi (2019) found that HSR benefitted Gross Regional Product (GRP) growth in South Korea. Likewise, positive effects of HSR on economic activity are found in studies from Italy (Cascetta et al., 2020), France (Facchinetti-Mannone, 2019) and Japan (Miwa, Bhatt, and Kato, 2022).

For China, scholars have used various methods to study HSR impacts in different regions. The empirical evidence on impacts tends to be mixed. Most studies show significant positive economic effects on cities along HSR routes. For example, using DID, Diao (2018) and Wang et al. (2022) found HSR had a positive role in promoting provincial economies by increasing GDP and accessibility. With a CGE model, Yang et al. (2023) found that HSR could stimulate GDP growth and reduce regional disparities in eastern China. Relatedly, Jin et al. (2024) focus on accessibility and connectivity features of HSR, and the role of the HSR network in alleviating urban–rural income disparity. Li and Li (2024) find that the opening of HSR lines positively stimulates economic growth, primarily through market size.

However, some researchers have found that HSR may have negative economic effects. For example, Qin (2017) found that GDP of counties along China's HSR routes are reduced by 3%-5% on average, with this decline operating through decreased fixed asset investments. Using a DID model, Gao et al. (2020) estimate a long-lasting negative effect of HSR on GDP per capita within the Yangtze River Delta region, of approximately 10%, mainly due to the population reallocation. Liu et al. (2022) found that large scale inter-regional HSR reduced labor productivity by almost 13% and the productivity gap between central and peripheral cities in the network widened after connecting to the HSR network.

Despite this growing body of empirical research, a key aspect of the potential impact on local economic activity from connecting to the HSR network remains largely unexamined. Most previous studies only explore the direct impact of HSR opening on regional economic development. This may be insufficient given that transport infrastructure, and especially HSR, shortens travel time and improves inter-city accessibility, and thus contributes to the flow of capital, ideas, population, and other factors between cities (Shi, and Wang, 2024). Consequently, the effects induced by transport infrastructure are likely to extend to the neighbouring areas. These spillovers are often ignored in applied studies, which use impact evaluation frameworks that treat the spatial units as independent of their surrounding areas. Hence, the results from prior studies may not represent the complete impacts if these spatial spillover effects are not considered.

To overcome this gap in the literature, the current study uses a comprehensive set of spatial econometric models to examine the impacts that connecting to the HSR network has on local economic activity. We use a decade-long panel of almost 2500 county-level units in China to examine HSR impacts on GDP and on luminosity as indicators of local economic activity. The research framework not only considers the spillover effects of connecting to the HSR network but also allows for the endogenous placement of HSR routes. In comparison to much of the existing research, our focus is on economic activity at a more local level, but with our sample covering almost all of China. Second, previous studies often use the DID model to test the impacts of HSR operation but do not consider spatial factors such as spillovers. In contrast, spatial spillover effects of HSR can be captured with spatial econometric models. In our framework, we start with a very general spatial autoregressive model with spatial autoregressive errors (SARAR) that encompasses popular models like the spatial Durbin model, the spatial lag model, and the spatial error model. Lastly, by using an instrumental variables (IV) strategy, we can assess whether the endogenous placement issue interferes with estimation

of the impact of HSR on the regional economy. Our analysis not only advances the understanding of how HSR affects economic growth at the county level in China but also provides insight into the planning of future HSR networks.

The subsequent parts of the paper are structured as follows. Section 2 reviews the literature related to HSR and discusses the main contribution of the study. Section 3 describes the data used in this study and the econometric modelling framework, including the spatial econometric models, and the instrumental variables strategy used to allow for the endogenous placement of HSR. Section 4 presents the empirical results, using both county-level GDP and night-time lights as indicators of economic activity. Finally, Section 5 concludes the study, and discusses the limitations.

II. Literature review

In this section, we review literature on the impact of high-speed rail (HSR) on local economic development in China. We focus especially on the empirical methods and the indicators of local economic activity. The reviewed studies cover a wide range of topical issues, including whether to take into account the spatial econometric models and the HSR endogenous placement issues.

Rail transport is a crucial mode of inter-region transportation in China, a large developing country with a widespread geographical distribution of natural resources and population (Ke et al. 2017). HSR has become the fastest-growing transportation mode in China due to the advantages of efficiency, punctuality, and stability (Xu, and Huang, 2019; Tian et al., 2021). As a result of the rapid development, research on the impacts of HSR networks has become a hot topic in China (Luan, Guo, and Liang, 2024). However, the impact of HSR on the regional economic development of peripheral cities along its line is still under debate. Researchers use a variety of spatial scales for these studies, and different conclusions concerning the promotion, inhibition, or insignificance of HSR on local economic growth have been reported.

Studies of the impact of HSR on regional economic growth (both within and beyond China) are mostly based on methods such as the synthetic control method (Albalate, Campos, and Jiménez, 2023), computable general equilibrium (CGE) model (Kim and Yi, 2019), and difference-in-differences (DID) model (Chen, Lv, and Zhang, 2024). For example, Yu et al. (2021) applied the synthetic control method and found that HSR raises a county's total GDP.

Using a dynamic CGE model, Chen (2019) found that HSR infrastructure development in China has generated a positive regional economic impact. Compared to the above methods, the DID model has been used more broadly to investigate HSR's impacts on socioeconomic conditions. For example, Chen et al. (2023), and Liang et al. (2020) used county-level panel data to study the economic impacts of HSR, which showed that the opening of HSR led to economic growth along the route. Wang et al. (2023) used GDP data to investigate the heterogeneous economic effects of HSR on prefecture-level cities. In addition to GDP, nighttime lights data are recognized as an effective proxy index related to human activities and urban economic development. Guo et al. (2023) adopted nighttime light data as a proxy for regional economic activity and found that HSR had a positive impact.

Considering that transport infrastructure in a region not only affects local economic growth but also economic activities in adjacent areas, some studies used spatial econometric models to examine the aggregate growth effect and potential spatial spillover effect of HSR. Liu, Tang, and Wang (2024) showed that the construction of HSR led to a notable increase in income through a positive spatial spillover effect. Based on the Spatial Durbin model (SDM), Sun et al. (2023) and Huang (2021) analyzed the direct and indirect effects of HSR on economic development. However, not all uses of these models find positive economic spillovers for regional economic development. Jin et al. (2020) found that the spillover effect is insignificant based on the SDM model.

The selection of high-speed rail stations is usually considered to be non-random and biased towards cities with better economic development. Given these endogeneity issues, many studies used an Instrumental variables (IV) strategy. To construct an IV for actual HSR connections, previous studies either referred to historical information or used the straight-line strategy. Historical transport routes were mainly located based on the original local geological conditions due to the then-technical restrictions, and the original geological conditions were highly related to the construction of transportation infrastructure, such as for post-cities. Hence, post-cities from the Ming Dynasty (1368—1644) were chosen as the instrumental variables for transportation infrastructure in the regression models of Guan, Chen, and Li (2023), and Wang, Jiang, and Miao (2023). Similar practices are followed by Chen, Cheng, and Zhang (2023), and Guo et al. (2021), who used China's historical transportation network in 1961 as the IV for the current HSR connections.

In contrast to the historical instruments, the straight-line strategy is based on connecting large central cities, using the shortest distance (which should also result in fastest

journeys notwithstanding potentially higher construction costs). The instrumental variable obtained by the minimum spanning tree should be highly relevant to the actual HSR line and thus is used to construct IVs for the actual transport connection. For example, Qiu, Liu, and Liao (2023), and An et al. (2022) constructed an IV by drawing straight lines among major cities, to predict the location of the current HSR connection. Shen, Li, and Li (2023), and Chong, Chen, and Qin (2019) used both the spatial econometric model and an IV strategy to explore the economic impacts of HSR, finding that the impacts were not significant.

To help summarize this large and growing literature on the impacts of HSR on regional economic activity in China, 30 recent studies are described classifying them along two dimensions: whether they allow for or ignore spatial spillovers and whether they treat the placement of the HSR networks as endogenous or exogenous (Table 1). Despite broad discussions about the effects of HSR connection on regional economic development, it can be seen that the spillover effects of HSR have received inadequate attention, with far more studies in Table 1 ignoring spillovers (n=23) than allowing for spillovers (n=7). The combination of allowing for spillovers and allowing for endogenous placement is very rare (n=2). Our paper adds to this small literature because we consider both of these issues in our econometric models.

In summary, existing studies on the economic impact of high-speed rail have not yet reached consistent conclusions, perhaps due to the differences in their spatial scope and research methods. In terms of spatial scale, there is little systematic research focused on the county-level economic impacts. Most of the previous HSR impact studies are based on panel data from provinces or prefectural cities, while panel data from county-level areas are less commonly used. This paper aims to address these gaps.

Table.1 Selected studies analyzing HSR impacts

Authors	Year	Time period	Spatial units	Spatial approach	IV approaches	Indicator(s)	Result	Objective
Panel A: Studies that assume exogenous placement and no spatial spillovers								
Chen et al.	2023	2008-2017	328 counties	/	/	GDP	+ve	To assess impacts of HSR opening on the inter-county economic gap
Wang, Yu, and Zhang	2023	1999-2013	333 prefectures	/	/	GDP	+ve	To find the heterogeneous economic effects of distance to HSR
Wang et al.	2023	2007-2018	62 prefecture-level cities	/	/	GDP	+ve	To estimate direct impacts of opening HSR on regional economic development
Yu et al.	2021	2001-2017	54 counties in Hubei Province	/	/	GDP	+ve	To analyse impacts of HSR on economic growth
Guo et al.	2020	2002-2018	24 cities	/	/	Luminosity	+ve	To investigated the impact of the HSR on urban economic development using NTL
Li, Wu, and Zhao	2020	2001-2017	40 prefecture-level cities	/	/	GDP	+ve	To investigate the economic effect of HSR
Liang et al.	2020	2012-2017	632 counties	/	/	GDP	+ve	To explore impacts of HSR on economic growth along the route
Chen	2019	2002-2013		/	/	GDP	+ve	To access the regional economic impacts of HSR
Panel B: Studies that assume exogenous placement but allow for potential spatial spillovers								
Liu, Tang, and Wang	2024	2000-2018	286 prefecture-level cities	SDM	/	Income	+ve	To investigate the impact of HSR on urban residents' income
Sun et al.	2023	2003-2018	285 prefecture-level cities	SDM	/	GDP	+ve	To examine joint impacts of HSR on the urban economy
Huang	2021	2008-2018	281 cities	SDM	/	GDP	+ve	To discuss the spatial–temporal heterogeneity of the relationship between HSR and the urban economy

Yu	2021	2005-2013	47 prefecture-level cities	SDID	/	GDP	+ve	To investigate the influence of Beijing-Shanghai HSR on regional economy
Jin et al.	2020	2002-2016	285 cities	SDM	/	GDP	+ve	To examine the role of HSR in influencing economic growth and disparity
Panel C: Studies that allow for endogenous placement but do not allow for potential spatial spillovers								
Chen, Cheng, and Zhang	2023	2003-2014	351 cities	/	1961 train station	GDP	+ve	To study the effects of HSR construction on firms' cross-regional development
Guan et al.	2023	2011-2019	41 cities in YRD region	/	Ming Dynasty post station	GDP	+ve	To test the impact of HSR on high-quality economic development
Qiu, Liu, and Liao	2023	2003-2019	284 prefecture-level cities	/	Straight lines	GDP	+ve	To investigate the impact of railway express on urban carbon emissions reduction
Wang, Jiang, and Miao	2023	2003-2019	277 prefecture-level cities	/	Courier roads in the Ming Dynasty	GDP	+ve	To examine HSR's impact on urban resilience
Wang, Zhou, and Guo	2023	2010-2018	34 prefecture level cities	/	Railway stations in 1993	Urban tourism incomes	+ve	To explore the spatial impact of HSR on tourism economy
Yang and Ma	2023	2001-2020	285 cities	/	Railway stations in 1912s	GDP	+ve	To study the impact of the border effect on HSR links
An et al.	2022	2005-2018	40 prefecture cities in YRD region	/	Straight lines	GDP	+ve	To investigate the effects of HSR on local economic development
Chen and Wang	2022	2003-2019	291 prefecture-level cities	/	Straight lines	GDP	+ve	To study the impact of HSR on the consumer service industry
Liu et al.	2022	2001-2015	236 prefecture-level cities	/	Straight lines	GDP	-ve	To explore effects of HSR on labor productivity
Wang, Wu, and Liu	2022	2001-2016	219 prefecture-level cities	/	Straight lines	GDP	+ve	To explore effects of HSR connection on urban innovation

Dong et al.	2021	2000-2020	180 towns	/	Straight lines	GDP	+ve	To explores the determinants of HSR new towns' economic growth
Guo et al.	2021	2004-2010	20 cities	/	1961 China railway map	GDP	+ve	To investigate the impact of HSR on regional water environment
Kuang, Liu, and Zhu	2021	2004-2017	141 cities	/	1962 China railway map	GDP	+ve	To study the relationship between HSR connection and firm performance
Gao et al.	2020	2006-2015	211 county-level units in YRD region	/	Straight lines	GDP	-ve	To investigate effects of connecting to an HSR line on the local economy
Yao et al.	2019	2010-2014	285 cities	/	China's historical railway network	GDP	+ve	To study the HSR impacts on economic growth
Panel D: Studies that allow for endogenous placement and also allow for potential spatial spillovers								
Shen, Li, and Li	2023	2011-2020	229 cities	GS2SLS	The post cities in the Ming Dynasty	GDP	ns	To study the influence of transportation infrastructure on digital economy
Chong, Chen and Qin	2019	2008-2015	268 cities	SAR, SEM, SDM	Historical routes	GDP	+ve	To evaluate the economic benefits of HSR network from the perspective of connectivity improvement

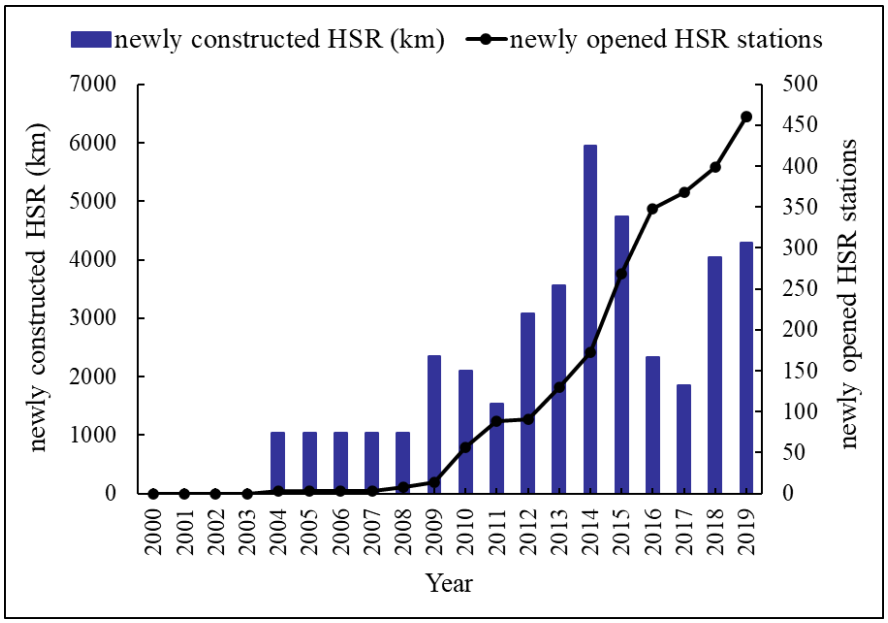
Notes: Abbreviations for spatial approach: GS2SLS (generalized spatial two stage least squares), SAR (spatial autoregressive model), SDM (spatial Durbin model), SEM (spatial error model), SDID (spatial difference-in-differences); +ve refers to positive effects, -ve refers to negative effects, ns refers to insignificant effects.

III. Data and Methods

Data Sources

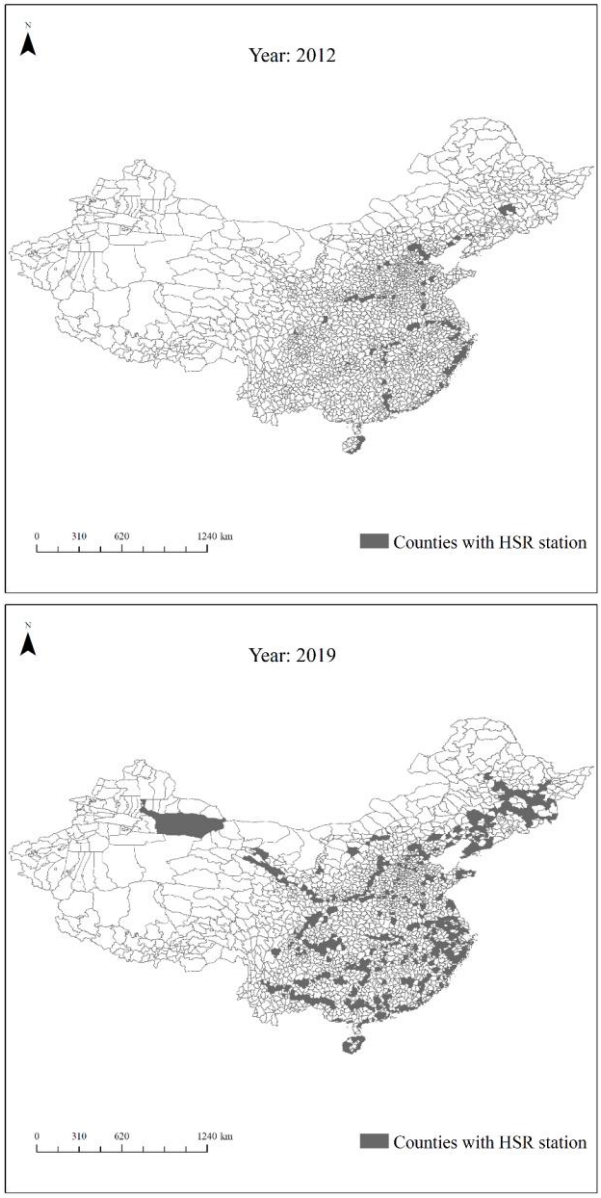
High-speed rail refers to high-speed trains with a design speed of over 250 km/h or passenger dedicated intercity lines with an average speed of 200 km/h or more. The HSR data were manually collected and include information such as the opening dates of each of China's HSR corridors in operation by the end of 2019, and the new length for each year. The sources for high-speed rail data include the National Railway Corporation (www.china-railway.com.cn) and the 12306 China Railway (http://www.12306.cn). For stations opening later than July 1 in a year, the opening date is set as the next year (the dummy variable HSRopen = 1). The development of HSR in China began with the opening of the first HSR line at the end of 2008. Since then, the HSR network has experienced rapid expansion, particularly starting from 2012 (as shown in Fig. 1). Thus we adopt the research period from 2012 to 2019 (which also coincides with the availability of high quality luminosity data). During the research period, a total of 460 counties (out of 2342 counties in all of China) had HSR stations opened. The spatial distribution of HSR stations in China in 2012 and in 2019 is shown in Fig. 2. The expansion westwards, and to the northeast, and the increased density in central and southeast China is evident.

Figure 1. The growth of China's high-speed rail (HSR) network



Note: The data are sourced from the National Railway Administration of the people of China (http://www.nra.gov.cn) and National Bureau of Statistics (www.stats.gov.cn)

Figure 2. HSR stations in China as of 2012, and 2019



Note: The darkest areas, which take the value of 1 in the econometric models, are the counties and districts connected by HSR lines; the areas taking the value of 0 are counties and districts not connected by HSR lines.

To empirically analyze impacts we use two measures of local economic activity—GDP and luminosity. The GDP data were obtained from three main products of the National Bureau of Statistics (NBS): (i) annual editions of the China Statistical Yearbook (county-level) (in Chinese it is *Zhongguo Xianyu Tongji Nianjian* (*Xianshi Juan*)), (ii) annual editions of the China City Statistical Yearbook (known as *Zhongguo Chengshi Tongji Nianjian*), and (iii) annual editions of the Statistical Yearbook for each city or province (for example, the Beijing Statistical Yearbook) (NBS, various dates). Each edition reports on GDP the previous year, so we use the 2013 to 2020 editions to obtain annual GDP data from 2012 to 2019. Overall, we have a balanced panel of annual GDP data for each of n = 2342 units at the 3rd level of the sub-

national administrative hierarchy, where these units maintain a consistent spatial definition from 2012 to 2019.

The luminosity data we select to represent the level of regional economic growth are the annual composites from Defense Meteorological Satellite Program (DMSP) satellites and from the Visible Infrared Imaging Radiometer Suite (VIIRS) from the Suomi-NPP satellites. The DMSP annual composites from satellites F15 and F18 collectively cover each year from 2012 to 2019. The stable lights product provides 6-bit digital numbers (DN) from 0–63 for each 30 arc-second output pixel (approximate 1.04 km×1.04 km at China's latitude). Ephemeral lights, such as from fires and flaring, are removed and processing excludes (at pixel level) images for nights affected by clouds, moonlight, sunlight, and other glare. The VIIRS data we used is the version 2.1 VIIRS Day/Night Band nighttime lights (VNL), which are available beginning in 2012. The VNL is produced from monthly cloud-free radiance averages, with initial filtering to remove extraneous features such as fires and aurora before the resulting rough annual composites are subjected to outlier removal procedures (Elvidge et al., 2021). The data are in units of nano Watts per square centimeter per steradian ($nW/cm^2/sr$) reported on a 15 arc-second output grid (approximate 0.52 km×0.52 km). We mainly use the average masked data product with background noise, biomass burning, and aurora zeroed out, which had the highest predictive power for GDP amongst all of the VNL data products in a prior study at the same county-level resolution that the current study uses (Zhang, and Gibson, 2022). Our constructed panel dataset contains annual observations for 2342 county-level units (which includes districts, county-level cities and counties) in China from 2012 to 2019. Table 2 shows the descriptive statistics.

Table 2. Descriptive statistics of variables

Variable	Sample size	Mean	Standard deviation	Minimum	Maximum
HSRopen	18736	0.119	0.324	0	1
lGDP	18736	4.883	1.284	0.262	10.508
lDMSP	18736	8.400	1.353	0	12.829
lVNL	18736	8.017	1.329	1.816	13.214

Notes: Statistics are for n=2342 spatial units, annually observed in a balanced panel from 2012 to 2019.

Estimation Framework

Spatial econometric models let us examine the nature of possible spillovers, and are used for this purpose in many contexts (LeSage and Pace, 2009). A key aspect of these models aiding the study of spillovers is that possible interactions between spatial units are summarized with a $N \times N$ spatial weights matrix, *W*. In this study we use a row normalized contiguity weights matrix that has values of one for neighbours and zero otherwise, with a diagonal of zeros because a spatial unit cannot neighbour itself. At the level of spatial disaggregation that we use, the average spatial unit in China has six neighbours.

In what follows, the proxy for economic output in spatial area *i* in year *t* is denoted as O_{it}, where the two proxy variables we use are log GDP in our main specification, and the log of the sum of night-time lights in our sensitivity analysis. The indicator for whether a spatial unit has a HSR station is D_{it} the μ_i are time-invariant fixed effects for each spatial unit, the ϑ_t are year fixed effects, and e_{it} is a random error. By using the spatial weights matrix we can allow for spatial lags, which are averages of these variables over the neighbouring units.

Our starting point is a very general model, which is a spatial autoregressive model with spatial autoregressive errors (SARAR). This model allows for spatially lagged dependent variables, spatially lagged independent variables and spatially lagged errors:

$$O_{it} = \lambda W O_{it} + \beta_1 D_{it} + \beta_2 W D_{it} + \mu_i + \vartheta_t + \rho W u_{it} + e_{it} \quad (1)$$

The SARAR model allows for changes in an outcome variable in a given area to have effects on contemporaneous outcomes in other areas (via the autoregressive spatial lag of the dependent variable, if $\lambda \neq 0$). It also allows changes in independent variables (such as getting connected to a HSR network) to affect not only own-area outcomes but also outcomes in neighbouring areas (if $\beta_2 \neq 0$). The $\rho W u_{it}$ term allows for spatial autocorrelation, where errors for a given area correlate (ρ) with a weighted average of the errors from surrounding areas. Equation (1) nests a spatial Durbin model if $\rho = 0$, a spatial auto-regressive model (*aka* spatial lag model) where only the dependent variable is spatially lagged if $\beta_2 = \rho = 0$, a spatial error model where only the errors are spatially lagged (if $\lambda = \beta_2 = 0$), and the most restrictive of all, which is an aspatial model with no spatial lags (if $\lambda = \beta_2 = \rho = 0$). The aspatial model has been the approach underpinning many previous studies of HSR impacts on the regional economy (as shown in Table 1). The encompassing nature of equation (1) allows for a general-to-specific model selection strategy which appears to be more robust than the

reverse simple-to-general selection strategy, especially if they are any anomalies in the Data Generating Process (Mur and Angulo, 2009).

An important feature of spatial econometric models is that lags of either the outcome variable or of independent variables (but not of errors) mean that total effects of changes in an independent variable—e.g. whether a county gets connected to a HSR network—may be quite different to what the regression coefficient on the dummy variable for being connected shows. Thus, while $\hat{\beta}_1$ is the object of interest in the typical model without spatial lags, in the spatial models when either the spatial lags of outcomes or the spatial lags of independent variables are non-zero then $\hat{\beta}_1$ does not capture the total effect of a change in the administrative status of a county. A useful decomposition of the more complex spatial relationships that occur relies on rewriting equation (1) in matrix notation (for simplicity, subscripts are dropped and fixed effects and error terms combined in *v* because the errors do not affect this decomposition) as:

$$O = (I - \lambda W)^{-1}(D\beta_1 + WD\beta_2) + (I - \lambda W)^{-1}v \qquad (2)$$

Following Elhorst (2012), the $N \times N$ matrix of partial derivatives can be written (noting that diagonal elements of *W* are zero) as:

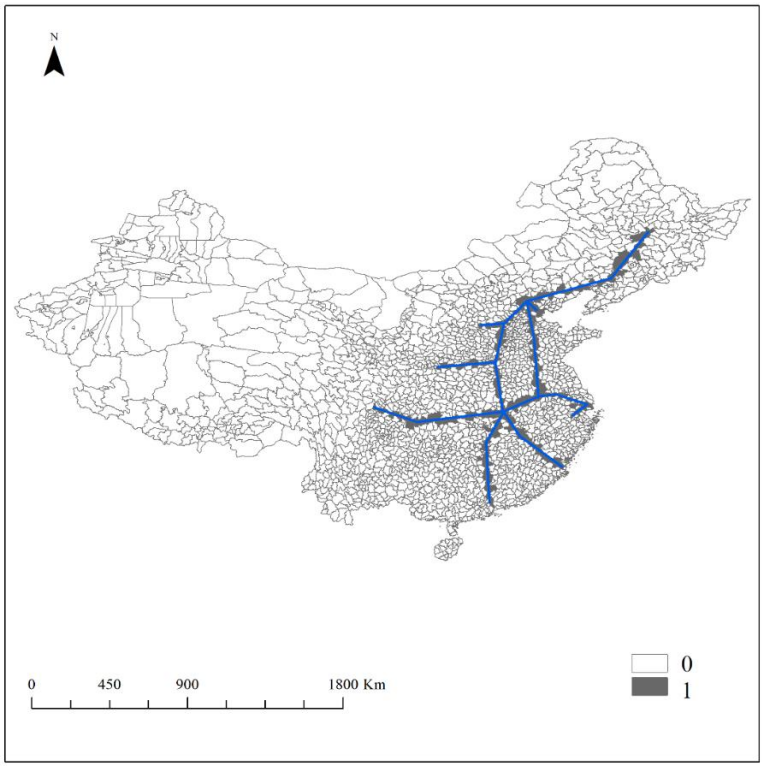
$$\frac{\partial O}{\partial D_k} = (I - \lambda W)^{-1}(\beta_{1k} I_N + \beta_{2k} W) \qquad (3)$$

where D_k is the HSR connection status in spatial unit *k*. The total marginal effect on output that is associated with a county getting connected has two components, a direct one and an indirect one, that may both vary over space. The estimator that we use follows LeSage and Pace (2009) in reporting a single direct effect, that averages the diagonal elements of the matrix in (3) and a single indirect effect that averages the row sums of the non-diagonal elements of that matrix. Indirect effects arise not just from adjacent area units, if $\beta_{2k} \neq 0$, but also from (potentially) all areas through the spatial autoregressive effect if $\lambda \neq 0$. Thus, there can be both local and global spillovers and when these are accounted for, averages from the matrix of derivatives $\partial O / \partial D_k$ may be quite different to the estimated direct impact effect, $\hat{\beta}_1$.

To further address the endogeneity of HSR route placement, we combine the IV method with equation (1). We used the straight-line strategy to construct the potential HSR connection variables, as the IV for the actual HSR connections. Compared with highways and traditional railways, early stage HSR especially aims to reduce travel time between targeted central cities. To best realize this aim, a designer might hypothetically just draw a straight line between two targeted cities. Thus, those counties and districts falling on the straight line are

potential counties connected to the HSR network and this design choice can be used to construct the IV for the actual HSR connection variable. The straight-line strategy tends to produce valid IV because counties lying on the straight line of two HSR-targeted cities are more likely to have actual HSR connections, but whether or not a county is on the line is exogenous. Specifically, for each HSR line in operation, we draw a straight line between two targeted provincial capitals. Those counties on the line are potential counties for HSR connection, assigned to be in the same year the actual HSR line started to operate (see Figure 3 for an example).

Figure 3. The straight-line strategy for the instrumental variable



Note: This map illustrates the IV strategy based on potential HSR lines (shown in blue) only for 2012. The darkest areas, which take the value of 1 for the instrumental variable in the first-stage model, are the potential counties and districts connected by HSR lines based on straight lines between capital cities.

IV. Results

Empirical results

The results of estimating Eq. (1) and then imposing various restrictions on the parameters and estimating the nested models are given in Table 3 (for GDP), Table 4 (for luminosity: DMSP), and Table 5 (for luminosity: VNL). The nesting restrictions are rejected in all cases, so that the SARAR models appear to be the most data-acceptable models for both GDP and luminosity. The discussion therefore concentrates mostly on the results in column (1) for the SARAR model. Nevertheless, we also devote some discussion to the results in column (5) of each table for the standard two-way fixed effects panel data model that does not allow for any spatial lags. Those sorts of aspatial models are often used in the literature (such as some of those studies summarized in Panel A and Panel C of Table 1).

Table 3: Relationships Between HSR station opening and the Change in Economic Activity (log GDP) in China: 2012 to 2019

	(1) Spatial lag of errors, covariate and outcome	(2) Spatial lag of the covariate and outcome	(3) Spatial lag of the outcome	(4) Spatial lag of the errors	(5) Standard panel model analysis
HSR station opened	-0.005	-0.004	-0.007	-0.004	-0.025***
	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)
Average impacts:					
Direct	-0.004	-0.008	-0.008	-0.004	-0.025***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
Indirect	-0.005	-0.052	-0.016	n.a.	n.a.
	(0.010)	(0.030)	(0.012)		
Total	-0.009	-0.060	-0.024	n.a.	n.a.
	(0.011)	(0.033)	(0.018)		
County fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Spatial lag: HSRopen	Yes	Yes	No	No	No
Spatial lag: output	Yes	Yes	Yes	No	No
Spatial lag: errors	Yes	No	No	Yes	No
All covariates = 0	χ^2=2693***	χ^2=45069***	χ^2=45071***	χ^2=2780***	χ^2=16335***
Nesting restrictions	n.a.	χ^2=54935***	χ^2=54953***	χ^2=2898***	χ^2=70852***

Note: The sample period is 2012-2019, for 2342 county-level units, giving an estimation sample of n=18736. Coefficients for the fixed effects and the spatial lags are not reported. The decomposition of average impacts into direct, indirect and total components is based on LeSage and Pace (2009). The nesting restrictions are imposed on the SARAR model in column (1) to derive the models in columns (2) to (5) Standard errors are in ( ), with statistical significance at the 1%, 5% and 10% level denoted by ***, **, *.

Table 4: Relationships Between HSR station opening and the Change in Luminosity (log DMSP) in China: 2012 to 2019

	(1) Spatial lag of errors, covariate and outcome	(2) Spatial lag of the covariate and outcome	(3) Spatial lag of the outcome	(4) Spatial lag of the errors	(5) Standard panel model analysis
HSR station opened	-0.006	-0.003	-0.007	-0.001	-0.027**
	(0.012)	(0.012)	(0.011)	(0.012)	(0.013)
Average impacts:					
Direct	-0.001	-0.004	-0.007	-0.001	-0.027**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)
Indirect	-0.035*	-0.105**	-0.009	n.a.	n.a.
	(0.021)	(0.044)	(0.014)		
Total	-0.036	-0.108**	-0.016	n.a.	n.a.
	(0.024)	(0.048)	(0.026)		
County fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Spatial lag: HSRopen	Yes	Yes	No	No	No
Spatial lag: output	Yes	Yes	Yes	No	No
Spatial lag: errors	Yes	No	No	Yes	No
All covariates = 0	χ^2=1667***	χ^2=21837***	χ^2=21829**	χ^2=2794***	χ^2=11908***
Nesting restrictions	n.a.	χ^2=16119***	χ^2=16134***	χ^2=1016***	χ^2=23655***

Note: The sample period is 2012-2019, for 2342 county-level units, giving an estimation sample of n=18736. Coefficients for the fixed effects and the spatial lags are not reported. The decomposition of average impacts into direct, indirect and total components is based on LeSage and Pace (2009). The nesting restrictions are imposed on the SARAR model in column (1) to derive the models in columns (2) to (5) Standard errors are in ( ), with statistical significance at the 1%, 5% and 10% level denoted by ***, **, *.

Table 5: Relationships Between HSR station opening and the Change in Luminosity (log VNL) in China: 2012 to 2019

	(1) Spatial lag of errors, covariate and outcome	(2) Spatial lag of the covariate and outcome	(3) Spatial lag of the outcome	(4) Spatial lag of the errors	(5) Standard panel model analysis
HSR station opened	-0.049***	-0.043***	-0.044***	-0.044***	-0.056***
	(0.010)	(0.009)	(0.009)	(0.009)	(0.010)
Average impacts:					
Direct	-0.046***	-0.047***	-0.048***	-0.044***	-0.056***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
Indirect	-0.019	-0.063*	-0.053***	n.a.	n.a.
	(0.016)	(0.034)	(0.011)		
Total	-0.065***	-0.110**	-0.101***	n.a.	n.a.
	(0.019)	(0.037)	(0.020)		
County fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Spatial lag: HSRopen	Yes	Yes	No	No	No
Spatial lag: output	Yes	Yes	Yes	No	No
Spatial lag: errors	Yes	No	No	Yes	No
All covariates = 0	χ^2=1822***	χ^2=21950***	χ^2=21950***	χ^2=2180***	χ^2=12775***
Nesting restrictions	n.a.	χ^2=15956***	χ^2=15957***	χ^2=1090***	χ^2=22394***

Note: The sample period is 2012-2019, for 2342 county-level units, giving an estimation sample of n=18736. Coefficients for the fixed effects and the spatial lags are not reported. The decomposition of average impacts into direct, indirect and total components is based on LeSage and Pace (2009). The nesting restrictions are imposed on the SARAR model in column (1) to derive the models in columns (2) to (5) Standard errors are in ( ), with statistical significance at the 1%, 5% and 10% level denoted by ***, **, *.

The lack of data acceptability for any of the nesting restrictions indicates that for the spatial units and period we study, the interactions are occurring through the spatial lags of the outcomes, the lags of the treatment, and the lags of the errors. There are two consequences of this pattern. First, partial approaches that only allow for local spillovers, such as including the treatment status of nearby spatial units, may not reveal the full pattern of spillovers. Second, the regression coefficients by themselves do not tell the full story, and the matrices of marginal effects based on equation (3) need to be taken into consideration.

The results of the marginal effects calculations are reported in the "average impacts" rows of Tables 3, 4, and 5, using the decomposition due to LeSage and Pace (2009). These results do not indicate that HSR expansion had a positive effect on the regional economy. From column (1) of Table 3 we cannot rule out that the HSR connection has zero direct, indirect and total effect on local GDP. For the aspatial model that does not consider the spatial spillovers, column (5) of Table 3, suggests that the HSR connection is associated with post-connection local GDP being -2.5% lower compared to what it would have been based on the GDP growth of the counties without connections. The inhibiting effect on local GDP growth of a county being connected with HSR is seen very clearly with the aspatial model.

The lack of any apparent effect of HSR connections on local GDP growth should not be due to insufficient variation in the data or a research design that is somehow 'under-powered'. The same annual GDP data, for the same spatial units, and with the same set of estimators as in Table 3, has recently been used by Zhang et al (2024) to study direct and indirect impacts on economic activity from China's counties being administratively upgraded. In the table in that study that corresponds to Table 3 here, nine of the 11 average impact estimates were sufficiently precisely estimated to be statistically significant at $p<0.01$ and another was significant at the $p<0.05$ level. So it seems to be something more about the lack of impact of HSR connections, rather than issues with the data and estimation framework, that accounts for the imprecise estimates.

When luminosity is used as the proxy for local activity, as a sensitivity analysis in case of mistrust in China's GDP figures, connection to the HSR network is associated with lower rates of growth in local economic activity. Using DMSP data, for the standard two-way fixed effects panel model without considering any spatial spillovers, the results in column (5) of Table 4 suggest an impact of almost minus three percent ($p<0.05$). This effect is even larger with the spatial Durbin model (in column (2)); luminosity is 11 percent lower

after connection to the HSR network compared to the luminosity growth exhibited by the counties that remain unconnected. This negative effect operates almost entirely through the indirect impacts, suggesting that there are negative spillovers from the expansion of the HSR network. The predominance of the indirect negative impacts also shows up with the SARAR results in column (1) of Table 4, but each type of impact is surrounded by wide standard errors so that indirect and total effects would only be statistically significant at $p<0.09$ and $p<0.13$.

The clearest evidence on the negative impacts of HSR connection is with the VNL data, which are known to be far more precise and accurate than DMSP data.[1] The impact is almost minus six percent with the standard two-way fixed effects panel model without considering any spatial spillovers (Table 5, column (5)). In fact, across all five estimators in Table 5, the estimates of average direct impacts tightly group, varying only between -4.4% and -5.6%. Variation in estimates of total impacts comes from the indirect effects that range from a statistically insignificant -2% for the SARAR model to weakly significant -6.3% for the spatial Durbin model and a very precisely estimated -5.3% with the spatial lag model. Overall, no result in Table 5 gives support to a hypothesis of greater local economic activity after connection to the HSR network, but there is consistently strong evidence of negative effects from the expansion of HSR connections between 2012 and 2019.

Results from instrumental variables

The results reported thus far are potentially biased, if the issue of endogenous placement of the HSR network matters. We therefore used an instrumental variables approach based on straight-lines which connect the major provincial cities, similar to An et al. (2022) and Liu et al. (2022). In the first stage of the two-stage least square (2SLS) results we examined the strength of this instrumental variable, with a value of the F-statistic of 1038 for omitting it from the first stage equation. Therefore, we should not have any weak instruments problem. We then repeated the models from columns (1) and (5) of Tables 3 and 5, for GDP and VNL luminosity, but using the predicted HSR connections as instruments for actual connections.

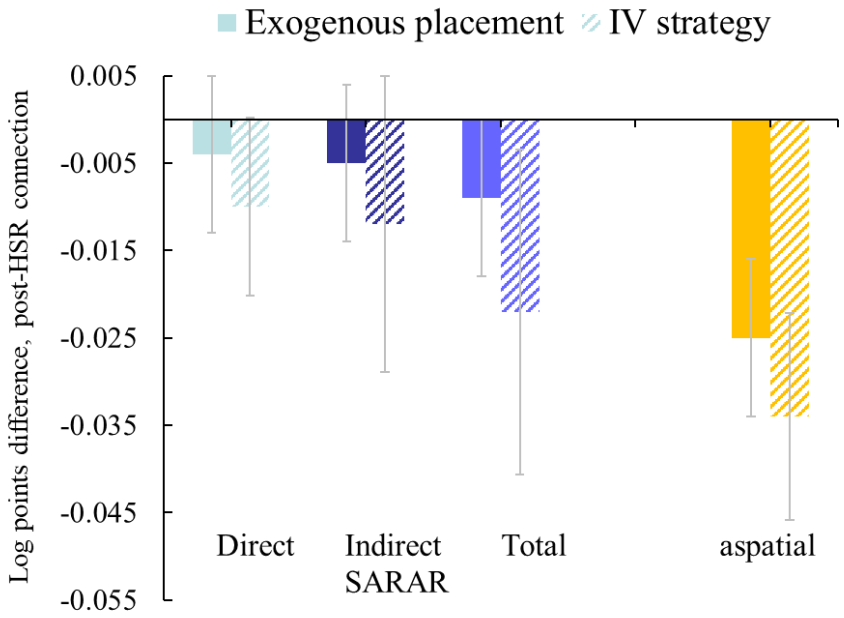
[1] Prior study for China shows that county-level inequality measured with VNL data is far closer to benchmark estimates than what DMSP data show (Zhang et al, 2023) and a similar result is shown at the district-level in India (Mathen et al, 2024). In terms of evaluating treatment effects of spatially targeted interventions, Kim et al (2024) show that VNL data provide far more precise treatment effect estimates than do DMSP data.

The results are shown in Figure 4, with the GDP results in the top panel and the VNL results in the bottom panel.

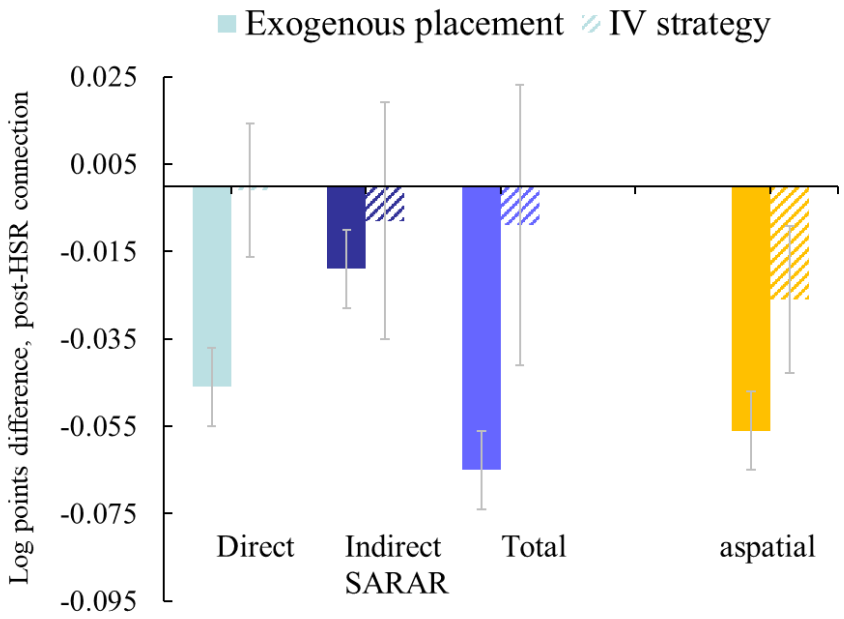
The finding of negative average total impacts on local GDP from connecting to the HSR network is amplified when the instrumental variables method is used to deal with the endogenous placement issue. The estimate of the average total impacts is more than doubled (thus, more negative) and even allowing for the larger standard errors with instrumental variables, this is a statistically significant ($p<0.05$) finding. Likewise, with aspatial two-way fixed effects, which already yielded a precisely estimated impact of -2.5% ($p<0.01$) when assuming exogenous placement, the effect when using instrumental variables becomes even more negative, at -3.4% (Fig 4a). Conversely, when luminosity (log VNL) is used to indicate local economic activity the use of instrumental variables causes the estimated impacts to become smaller (thus, less negative). This attenuation of the estimated negative impacts occurs with both the SARAR model and also the aspatial model. With the IV results for luminosity, one could not rule out a hypothesis of there being no effect on local economic activity from connecting to the HSR network (Fig. 4b).

Figure 4. Comparing HSR impacts with exogenous versus endogenous placement

(a) log GDP as outcome variable



(b) log VNL luminosity as outcome variable



Notes: Error bars show 95% CI. The IV strategy uses hypothetical straight-line connections between city pairs. Exogenous placement results are from Tables 3 and 5, column (1) and (5)

V. Discussion and Conclusions

A dramatic expansion of high-speed rail in China presents an important empirical setting to study the impact of transport infrastructure enhancement on economic growth. Given this importance, a growing literature with dozens of studies has examined high-speed rail impacts in China. What we add to this literature is a consideration of both spatial spillovers and the endogenous placement of the high-speed rail network. Most prior studies ignore any spatial spillovers, as shown in our review of 30 papers. Indeed, we are aware of only two studies that consider both the endogeneity and the spillover issue. The other feature of our study is that we cover all parts of China, using panel data for 2342 county-level units. Many of the prior studies have focused only on a particular region, or have used more spatially aggregated data for prefectural-level cities. The growing national scope of China's high-speed rail network, and the ambition to almost double the network length over the next decade, makes a national scope for the analysis more salient.

In the period under study, the number of county-level units in China that are connected to the high-speed rail network more than tripled, with almost one-fifth of all county-level units now connected to the network. Despite this substantial increase in connectivity, our results do not show positive impacts on local economic activity. Instead, we predominantly find negative impacts, or cannot rule out nil impacts, whether allowing for spatial spillovers or not, whether allowing for endogeneity or not, and using both GDP and luminosity data as our measures of local economic activity. In this regard, our findings echo Qin (2017) and Gao et al. (2020), who note that connecting to the high-speed rail network can impede economic growth through the spatial reallocation of the economically active population from peripheral areas to more central areas, and through the spatial restructuring of industries.

Of course the priority given to investments in high-speed rail may not just be because of expected positive impacts on local GDP. The Chinese government proactively promotes indigenous technology innovation, with the high-speed rail industry a prominent example of innovation-driven development (Sun, 2015). By developing high-speed rail technology, the Chinese state successfully leapfrogged into the high-speed land transportation age, fostering innovation, and reducing dependency on foreign suppliers (Yan, 2023). These technologies serve as a fruitful starting point for China in promoting infrastructure construction in a globalized economy, such as through the Belt and Road Initiative. Moreover, even if it does not (yet) show up in local GDP, the growth of high-speed rail may improve corporate financing

efficiency and optimise the efficiency of resource allocation in capital markets (Jin, Zhang, and Xin, 2020). For example being connected to the network significantly increases a city's capital mobility (Duan et al. 2021), and the resource redistribution between regions (Meng, Lin, and Zhu 2018).

Notwithstanding these considerations, the apparent negative impacts on county-level economic growth, and potentially negative spillover effects on neighbouring spatial units does pose a problem as China adapts to lower expectations of future economic growth. Before planning and constructing future additions to the high-speed rail network, and future stations on the network, policy makers may need to do more thorough evaluation of economic impacts on the targeted county-level units and also on their neighbours.

Future research could delve into the heterogeneous effects of high-speed rail on economic growth and explore broader impacts on the environment and on society. For example, some studies have proposed that impacts of high-speed rail spatially vary along the route, tending to benefit developed cities with several rail stations and higher accessibility rather than least-developed and developing cities with remote locations and backward economies (Liang et al., 2020; Zou and Chen, 2024). Furthermore, any spillover effects may be examined using a broader set of weighting matrices to what we have used, such as allowing for asymmetric effects that reflect greater agglomeration forces in some directions rather than others, or in some regions rather than others (as discussed, for example, in Zhang et al, 2024). These potential refinements provide a direction and focus for our future research.

References

Ahlfeldt, G. M., and Feddersen, A. (2018). From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography*, 18(2), 355-390.

Albalate, D., Campos, J., and Jiménez, J. L. (2023). Local tourism effects of HSR in small cities: three synthetic control case studies. *Current Issues in Tourism*, 26(14), 2301-2316.

An, Y., Wei, Y. D., Yuan, F., and Chen, W. (2022). Impacts of high-speed rails on urban networks and regional development: A study of the Yangtze River Delta, China. *International Journal of Sustainable Transportation*, 16(6), 483-495.

Banerjee, A., Duflo, E., and Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, 145, 102442.

Carbo, J. M., Graham, D. J., Anupriya, Casas, D., and Melo, P. C. (2019). Evaluating the causal economic impacts of transport investments: evidence from the Madrid–Barcelona high speed rail corridor. *Journal of Applied Statistics*, 46(9), 1714-1723.

Cascetta, E., Cartenì, A., Henke, I., and Pagliara, F. (2020). Economic growth, transport accessibility and regional equity impacts of high-speed railways in Italy: Ten years ex post evaluation and future perspectives. *Transportation Research Part A: Policy and Practice*, 139, 412-428.

Chen, Z. (2019). Measuring the regional economic impacts of high-speed rail using a dynamic SCGE model: the case of China. *European Planning Studies*, 27(3), 483-512.

Chen, H., Cheng, K., and Zhang, M. (2023). Does geographic proximity affect firms' cross-regional development? Evidence from high-speed rail construction in China. *Economic Modelling*, 106402.

Chen, M., Liu, X., Xiong, X., and Wu, J. (2023). Has the opening of high-speed rail narrowed the inter-county economic gap? The perspective of China's state-designated poor counties. *International Review of Economics and Finance*.

Chen, Y., Lv, Y., and Zhang, Y. (2024). The impact of high-speed railway construction on regional economic development in Hubei Province. *Management System Engineering*, 3(1), 3.

Chen, Q., and Wang, M. (2022). Opening of high-speed rail and the consumer service industry: Evidence from China. *Economic Analysis and Policy*, 76, 31-45.

Chong, Z., Chen, Z., and Qin, C. (2019). Estimating the economic benefits of high-speed rail in China. *Journal of Transport and Land Use*, 12(1), 287-302.

Diao, M. (2018). Does growth follow the rail? The potential impact of high-speed rail on the economic geography of China. *Transportation Research Part A: Policy and Practice*, 113, 279-290.

Dong, L., Du, R., Kahn, M., Ratti, C., and Zheng, S. (2021). "Ghost cities" versus boom towns: Do China's high-speed rail new towns thrive? *Regional Science and Urban Economics*, 89, 103682.

Duan, L., Niu, D., Sun, W., and Zheng, S. (2021). Transportation infrastructure and capital mobility: Evidence from China's high-speed railways. *The Annals of Regional Science*, 67, 617-648.

Egger, P. H., Loumeau, G., and Loumeau, N. (2023). China's dazzling transport infrastructure growth: Measurement and effects. *Journal of International Economics*, 142, 103734.

Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F.-C., and Taneja, J. (2021). Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, 13(5), 922.

Facchinetti-Mannone, V. (2019) A methodological approach to analyze the territorial appropriation of high-speed rail from interactions between actions and representations of local actors, *European Planning Studies*, 27(3), pp. 461–482.

Gao, Y., Song, S., Sun, J., and Zang, L. (2020). Does high-speed rail connection really promote local economy? Evidence from China's Yangtze River Delta. *Review of Development Economics*, 24(1), 316-338.

Guan, C., Chen, L., and Li, D. (2023). Does the opening of high-speed railway improve high-quality economic development in the Yangtze River Delta, China? *Land*, 12(8), 1629.

Guo, H., Chen, C., Dong, X., and Jiang, C. (2021). The evolution of transport networks and the regional water environment: the case of Chinese high-speed rail. *Regional Studies*, 55(6), 1084-1110.

Guo, Y., Yu, W., Chen, Z., and Zou, R. (2020). Impact of high-speed rail on urban economic development: An observation from the Beijing-Guangzhou line based on night-time light images. *Socio-Economic Planning Sciences*, 72, 100905.

Huang, Y. (2021). Spatial and temporal heterogeneity of the impact of high-speed railway on urban economy: Empirical study of Chinese cities. *Journal of Transport Geography*, 91, 102972.

Jin, M., Gu, R., Li, K. X., Shi, W., and Xiao, Y. (2024). Heterogeneous impacts of the high-speed railway network on urban–rural income disparity: Spatiotemporal evidence from Yangtze River Delta of China. *Transportation Research Part A: Policy and Practice*, 183, 104050.

Jin, Z., Zhang, L., and Xin, Q. (2020). Transportation infrastructure and resource allocation of capital market: Evidence from high-speed rail opening and company going public. *China Journal of Accounting Studies*, 8(2), 272-297.

Ke, X., Chen, H., Hong, Y., and Hsiao, C. (2017). Do China's high-speed-rail projects promote local economy?—New evidence from a panel data approach. *China Economic Review*, 44, 203-226.

Kim, E., and Yi, Y. (2019). Impact analysis of high-speed rail investment on regional economic inequality: a hybrid approach using a transportation network-CGE model. *Journal of Transport Economics and Policy*, 53(3), 314-333.

Kim, B., Gibson, J., and Boe-Gibson, G. (2024). Measurement errors in popular night lights data may bias estimated impacts of economic sanctions: Evidence from closing the Kaesong Industrial Zone. *Economic Inquiry*, 62(1), 375-389.

Kuang, C., Liu, Z., and Zhu, W. (2021). Need for speed: High-speed rail and firm performance. *Journal of Corporate Finance*, 66, 101830.

LeSage, J., and Pace, R. K. (2009). Introduction to spatial econometrics. Chapman and Hall/CRC.

Li, H., Dong, X., Jiang, Q., and Dong, K. (2021). Policy analysis for high-speed rail in China: Evolution, evaluation, and expectation. *Transport Policy*, 106, 37-53.

Li, M., and Li, H. (2024). Impacts of high-speed rail on regional economic growth: an empirical analysis of the moderated mediation effects. *Transportation Research Record*, 2678(3), 476-492.

Li, X., Wu, Z., and Zhao, X. (2020). Economic effect and its disparity of high speed rail in China: A study of mechanism based on synthesis control method. *Transport Policy*, 99, 262-274.

Li, J., Yu, Q., and Ma, D. (2024). Does China's high-speed rail network promote inter-city technology transfer?——A multilevel network analysis based on the electronic information industry. *Transport Policy*, 145, 11-24.

Liang, Y., Zhou, K., Li, X., Zhou, Z., Sun, W., and Zeng, J. (2020). Effectiveness of high-speed railway on regional economic growth for less developed areas. *Journal of Transport Geography*, 82, 102621.

Liu, Z., Diao, Z., and Lu, Y. (2024). Can the opening of high-speed rail boost the reduction of air pollution and carbon emissions? Quasi-experimental evidence from China. *Socio-Economic Planning Sciences*, 92, 101799.

Liu, X., Li, H., Sun, Y., and Wang, C. A. (2022). High-speed railway and urban productivity disparities. *Growth and Change*, 53(2), 680-701.

Liu, Y., Tang, D., and Wang, F. (2024). Research on the spatial spillover effect of high-speed railway on the income of urban residents in China. *Humanities and Social Sciences Communications*, 11(1), 1-13.

Luan, Z., Guo, S., and Liang, W. (2024). Can the construction of a high-speed rail alleviate haze pollution: an empirical analysis based on social networks and dynamic spatial econometric models. *Environmental Science and Pollution Research*, 1-17.

Ma, G. R., Chen, X. M., and Yang, E. Y. (2020). How transportation infrastructure promotes capital flow: A study based on the opening of high-speed rail and the off-site investment of listed companies. *China Industrial Economics*, 2020(06), 5-23.

Mathen, C., Chattopadhyay, S., Sahu, S., and Mukherjee, A. (2023). Which nighttime lights data better represent India's economic activities and regional inequality? *Asian Development Review*, 41(2), (online ready)

Mei, L., and Zhang, N. (2021). Catch up of complex products and systems: lessons from China's high-speed rail sectoral system. *Industrial and Corporate Change*, 30(4), 1108-1130.

Meng, X., Lin, S., and Zhu, X. (2018). The resource redistribution effect of high-speed rail stations on the economic growth of neighbouring regions: Evidence from China. *Transport Policy*, 68, 178-191.

Miwa, N., Bhatt, A., and Kato, H. (2022). High-speed rail development and regional inequalities: Evidence from Japan. *Transportation Research Record*, 2676(7), 363-378.

Mur, J., and Angulo, A. (2009). Model selection strategies in a spatial setting: Some additional results. *Regional Science and Urban Economics*, 39(2), 200-213.

Qin, Y. (2017). 'No county left behind?' The distributional impact of high-speed rail upgrades in China. *Journal of Economic Geography*, 17(3), 489-520.

Qiu, Z., Liu, D., and Liao, Q. (2023). Can the opening of China railway express reduce urban carbon emissions? A difference-in-differences analysis in China. *Frontiers in Environmental Science*, 10, 2686.

Ren, J. J., Liu, W. G., Lai, J. L., Ye, W. L., Deng, S. J., Liu, X. Y., and Tan, B. (2024). Performance deterioration and structural state diagnosis of slab tracks for high-speed railways: A review. *Engineering Failure Analysis*, 107955.

Shen, S., Li, H., and Li, M. (2023). Transportation infrastructure and digital economy—Evidence from Chinese Cities. *Sustainability*, 15(22), 16024.

Shi, K., and Wang, J. (2024). The influence and spatial effects of high-speed railway construction on urban industrial upgrading: Based on an industrial transfer perspective. *SocioEconomic Planning Sciences*, 93, 101886.

Sun, Z. (2015). Technology innovation and entrepreneurial state: the development of China's high-speed rail industry. *Technology Analysis and Strategic Management*, 27(6), 646-659.

Sun, X., Yan, S., Liu, T., and Wang, J. (2023). The impact of high-speed rail on urban economy: Synergy with urban agglomeration policy. *Transport Policy*, 130, 141-154.

Tang, H., Zhang, J., Fan, F., and Wang, Z. (2024). High-speed rail, urban form, and regional innovation: A time-varying difference-in-differences approach. *Technology Analysis and Strategic Management*, 36(2), 195-209.

Tian, M., Li, T., Ye, X., Zhao, H., and Meng, X. (2021). The impact of high-speed rail on service industry agglomeration in peripheral cities. *Transportation Research Part D*, 93, 102745.

Wang, C., Chen, J., Li, B., Chen, N., and Wang, W. (2023). Impact of high-speed railway construction on spatial patterns of regional economic development along the route: A case study of the Shanghai–Kunming high-speed railway. *Socio-Economic Planning Sciences*, 87, 101583.

Wang, K. L., Jiang, W., and Miao, Z. (2023). Impact of high-speed railway on urban resilience in China: Does urban innovation matter? *Socio-Economic Planning Sciences*, 87, 101607.

Wang, F., Liu, Z., Xue, P., and Dang, A. (2022). High-speed railway development and its impact on urban economy and population: A case study of nine provinces along the yellow river, China. *Sustainable Cities and Society*, 87, 104172.

Wang, J., Wang, L., Zhang, Y., Zhang, J., Li, J., and Li, S. (2024). A georeferenced dataset for mapping and assessing subgrade defects in China's high-speed railways. *Scientific Data*, 11(1), 293.

Wang, C. A., Wu, J., and Liu, X. (2022). High-speed rail and urban innovation: based on the perspective of labor mobility. *Journal of the Asia Pacific Economy*, 1-26.

Wang, B., Yu, F., and Zhang, M. (2023). The heterogeneous economic impacts of distance to high-speed railway stations in China. *International Journal of Business and Management*, 16(2), 1-75.

Wang, S., Zhou, Y., Guo, J., and Mao, J. (2023). Did high speed rail accelerate the development of tourism economy?–Empirical analysis from Northeast China. *Transport Policy*, 143, 25-35.

Xu, W., and Huang, Y. (2019). The correlation between HSR construction and economic development-empirical study of Chinese cities. *Transportation Research Part A: Policy and Practice*, 126, 24–36.

Yan, K. (2023). Market-creating states: rethinking China's high-speed rail development. *Review of International Political Economy*, 30(4), 1220-1237.

Yang, Y., and Ma, G. (2023). How can HSR promote inter-city collaborative innovation across regional borders? *Cities*, 138, 104367.

Yang, Q., Wang, Y., Liu, Y., Liu, J., Hu, X., Ma, J., and Tao, S. (2023). The impact of China's high-speed rail investment on regional economy and air pollution emissions. *Journal of Environmental Sciences*, 131, 26-36.

Yao, S., Zhang, F., Wang, F., and Ou, J. (2019). Regional economic growth and the role of high-speed rail in China. *Applied Economics*, 51(32), 3465-3479.

Yu, L. (2021). Study on treatment effects and spatial spillover effects of Beijing–Shanghai HSR on the cities along the line. *The Annals of Regional Science*, 67(3), 671-695.

Yu, J., Zhou, Y., Huang, Q., Li, X., Hou, Y., and Wang, X. (2021). Analysis of the impact of high-speed railway on county economic development based on the synthetic control method: The Hubei province in China. *Mathematical Problems in Engineering*, 2021, 1-16.

Zhang, X., and Gibson, J. (2022). Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. *Remote Sensing*, 14(5), 1282.

Zhang, X., Gibson, J., and Deng, X. (2023). Remotely too equal: Popular DMSP night-time lights data understate spatial inequality. *Regional Science Policy and Practice*, 15(9), 2106-2125.

Zhang, X., Li, C., and Gibson, J. (2024). The role of spillovers when evaluating regional development interventions: evidence from administrative upgrading in China. *Letters in Spatial and Resource Sciences*, 17(1), 9.

Zou, W., and Chen, L. (2024). The impact of high-speed railway on firms' productivity. *International Review of Economics and Finance.*

Chapter 7: China's City Size Distribution: Diverging in terms of people and converging in terms of area

Xiaoxuan Zhang, Chao Li and John Gibson

Abstract

The size distribution of cities in China matters for both economic and environmental reasons. The legacy of central planning, such as mobility restrictions from hukou registration, hampered evolution of the city size distribution, creating the need for rapid adjustment in city sizes in the more market-oriented era. In this study we use the three most recent censuses of population (for 2020, 2010 and 2000) which allow a consistent definition of urban residents. Over time, cities are becoming less even in terms of population, but more even in city built-up area (based on remote sensing data).

JEL Codes

R12, O15

Keywords

City Size
resident population
land use
Pareto's Law
China

Acknowledgements

Financial support from the Marsden Fund project UOW1901.

Introduction

The size distribution of China's cities is of ongoing interest to researchers (Au and Henderson, 2006). The legacy of central planning, and particularly the mobility restrictions from the *hukou* household registration system, hampered evolution of the city size distribution. Consequently, the size and location of cities had to rapidly adjust in the more market-oriented era so China embarked upon an unprecedented building boom.[1] The ensuing demand for resources affected many other countries and is likely to continue to do so given forecasts of China's continued urban growth (Fernández, 2007). Another byproduct of this building boom was that converting agricultural land into urban uses put stress on China's food security (He et al., 2017).

There are several ways to assess city size distributions and what their implications are for economic development, social change, and resource use (Mulligan and Carruthers 2021). A large branch of the literature applies power laws to city sizes: the rank associated with some size *S* is proportional to *S* to some negative power, so the rank-size relationship is linear in the logarithms. The rank-size relationships are usually in terms of population (Su 2020; Li et al., 2022) but they may also hold for city area (Rozenfeld et al., 2011). A particular focus of many studies is empirically testing whether Zipf's law holds, which specifies that city size follows a Pareto distribution with the parameter equal to one. More generally, changes in rank-size coefficients over time can show impacts from major policy changes, such as economic reform and fertility regulation (Anderson and Ge, 2005).

Some previous studies suggest that the size distribution of China's cities has become more equal, so the rank-size coefficients were rising over time. However, these studies used population data reported in annual *Statistical Yearbooks*, based on *hukou* registrations for each city rather than counts of actual residents (Gan et al., 2006; Xu and Zhu, 2009). Only the census fully counts the urban resident population (using a 6-month residency rule) and because of the scale of non-*hukou* migration (people moving

[1] For example, China used more cement in three years (from 2011 to 2013) than the United States did in the entire 20th Century (https://www.washingtonpost.com/news/wonk/wp/2015/03/24/how-china-used-more-cement-in-3-years-than-the-u-s-did-in-the-entire-20th-century/).

to a place other than where they are registered) the distributions of residents and of *hukou* registered population are quite different.[2] In a study using China's census data from 2000 and 2010, Gibson and Li (2020) showed that rank-size coefficients were falling (and moving closer to Zipf's law) if city size was measured in terms of urban residents; thus, China's cities were becoming less equal in terms of their populations contrary to the prior claims of the size distribution becoming more equal.

In this paper we show evidence from China's last three censuses, in 2000, 2010 and 2020, of China's city-size distribution becoming more unequal in terms of people (with the rank-size coefficients falling over time). Thus, the process identified by Gibson and Li (2020) has continued, in contrast to earlier claims (using the *hukou* registered population) that the city size distribution was becoming more equal. Much of China's urban growth is from a funnelling of people leaving many places—over two-thirds of prefectures have fewer residents than their registered population (Li et al, 2024)—and congregating in a small number of destination cities. Therefore, the more unequal distribution is apparent when analyses use the number of urban residents as the measure of city size rather than using the *hukou* registrations of people, many of whom now live elsewhere.

While the city-size distribution is becoming less equal in terms of people it is becoming more equal in terms of urbanized area. To show this we use satellite-detected remote sensing data on night-time lights and on urban land-use types, with both types of data showing a trend of rank-size coefficients rising over time. This divergence, where the population distribution of cities is becoming more unequal and the urbanized area becoming more equal, suggests that population growth of big cities is not being aided by fast enough urban area expansion there, while urban area elsewhere expands faster than required by the slower growth (or even decline) in the number of residents. We relate these divergences to changes in housing prices and find the largest price increases in cities where growth in urbanized area is far slower than growth in the

[2] At the time of China's Seventh National Census in 2020, 376 million people lived in a different prefecture than their place of registration (125 million in a different province, 251 million elsewhere in the same province).

resident population; in contrast, cities with only a small rise in nominal housing prices had urbanized area expand far faster than the growth in the resident population. Thus, the failure of the city-size distribution for urbanized area to evolve in a way that matched trends in the size distribution in terms of people may have contributed to a misalignment of house prices in China and resulting macroeconomic fragility (Rogoff and Yang, 2024).

Data and Methods

Our review of data quality issues indicated that the most reliable information on city population is from censuses; these count residents as those people living in an area for at least six months. This is a more realistic measure of city size than the count of people whose *hukou* registration is from there but who may live elsewhere. A further advantage of census data is that they report on each individual district within a prefectural-level city; notably, the contiguous districts of a prefectural-level city are the best proxy for an urban core in China (Roberts et al., 2012).[3]

We give an example of how urban cores are identified in Figure 1, based on the province of Fujian. The nine prefectural-level cities in Fujian range in size from 26,000 km^2 for Nanping (about the size of Albania or the US state of Massachusetts) to just 1700 km^2 for Xiamen. Most prefectural cities include large parts that are not urbanized (these areas are usually counties); for example, average population density in Nanping is just 100 persons per km^2. Consequently, the prefectural-level cities are not a suitable spatial unit for studying the city-size distribution because their populations include large unurbanized areas. While some prefectural-level cities, such as Xiamen, consist only of districts, and so have a far higher population density (e.g. over 3000 people per km^2 for Xiamen), the district-only prefectures are home to only a fraction of China's urban population. Instead, the urban core for each prefectural-level city is made up of the contiguous districts (e.g., population in the six districts of Xiamen is merged into a single value). In some prefectural-level cities (e.g. Nanping) the urban

[3] Prefectural cities are the second level of China's subnational hierarchy (below provinces). At the third level are districts, county-level cities, and counties. On average, districts are about three times more densely populated than county-level cities, and about nine times more densely populated than counties (Zhang et al, 2024).

core is just a single district. Overall, our sample has 265 urban cores, which hold over 98% of China's urban population.

Figure 2 shows how relationships between log size and log rank for these 265 cities have changed over time. At the top of the distribution, cities with the same log-rank had large rises in their resident population. For example, there were 10.6 million urban residents in the 2010 census in Guangzhou (the 5th ranked city) but 16.1 million by 2020 (a 51% rise). The slope of the Figure 2 relationship is the Pareto coefficient, where we can express it as:

$$\ln(Rank) = \alpha - \beta \ln(Size) + \varepsilon \quad (1)$$

where ε is a random error and the special case of $\hat{\beta}$ =1 is Zipf's law. If log-size increases but log-rank stays the same (as with Guangzhou, for example), the value of $\hat{\beta}$ must fall, so a falling Pareto coefficient from the rank-size regression denotes a distribution becoming more unequal, while a rising coefficient suggests moves towards a more equal distribution. We use Chow tests to establish if the changes in $\hat{\beta}$ between census years are statistically significant.

The relationship between log-rank and log-size can also be estimated in terms of urban area. A growing literature estimates urban area from satellite-detected night-time lights for just this purpose (Düben and Krause 2021). One advantage of remote sensing data is that they break the link with administrative boundaries; for example, some districts in Fujian in Figure 1 will include lower density less urbanized areas, and with night-time lights data it is possible to set brightness thresholds that exclude those areas from being considered 'urban'. To create natural boundaries we use an algorithm to "grow" each city from a fixed point (e.g. the central railway station) that is expected to be surrounded by brightly lit pixels; as the algorithm moves in each direction it looks for pixels below the brightness threshold and then sets natural boundaries to the built up area (Olivia et al., 2018). For this boundary-setting we use the Luojia-01 satellite

that measures night-time lights at a spatial resolution of 130 meters (while other sources of the night-time lights data have a resolution of about 500 meters, at best).[4]

Within these luminosity-based boundaries we then calculate illuminated area, using a harmonized series of night-time lights data that links VIIRS images (available since 2012) with DMSP images (available back to 1992). This harmonized series has a spatial resolution of about 1 km^2, where this coarseness reflects the reality that remote sensing data for earlier periods are typically of lower resolution because of the ongoing improvement in sensors. Given the rate at which China's illuminated area is growing, even with these relatively low resolution data we should be able to observe changes if they exist because of the decadal gap between the estimates. We use 2019 rather than 2020 night-time lights, because otherwise the apparent changes in urbanized area may be affected by the dimming of lights following China's lockdowns for dealing with COVID-19 from February 2020 (Deng et al., 2022). We create two estimates of the illuminated area in each year: one counts any pixel with non-zero luminosity and the other only counts pixels above a lower limit of 20% of the maximum luminosity in order to measure the area with more concentrated lights (thresholds are often used for lights-based measures).

Our final data source is land cover data from the European Space Agency Climate Change Initiative (ESA CCI), which is a continuous global land cover product with 300 m grid resolution which is temporally consistent with baselines prior to the year 2000. Previous studies evaluating the accuracy of ESA CCI data show that they provide an improved measure of land cover distribution compared to other fine resolution data such as Landsat (Sun et al. 2022). There are 37 land use types classified in the ESA CCI data and for this study we measure the area within each city that is of the land use types that are closely related to urban activities.

[4] The Luojia-01 satellite was only launched in 2018, so the boundaries for urbanized area are set towards the end of our study period. This should not cause a bias because China's illuminated area expands at a rate consistent with a doubling every decade (Olivia et al., 2018) so boundaries in 2019 should be larger than in earlier years.

Results

The main results are shown in Table 1. Over time, China's cities have become less equal in terms of their resident population. The Pareto coefficients from the rank-size relationship are becoming statistically significantly smaller, and getting closer to the values implied by Zipf's Law. The Pareto coefficient was 1.46 in 2000 and fell to 1.44 in 2010 and to 1.36 in 2020. These declines in coefficient values are statistically significant over both the full sample period and for the most recent decade. Indeed, the rate of decline in the Pareto coefficient has increased in the most recent inter-census period.

The results in panel B of Table 1 for the Pareto coefficients for urbanized area show the opposite pattern; the rising value of the coefficients indicates a more equal distribution of city sizes, according to their brightly-illuminated area. Once again, the changes over time are shown by the Chow test results to be statistically significant. In other words, both city population and city area are approaching the value implied by Zipf's law, but from opposite directions.

The same pattern, of urbanized areas becoming more equally distributed, also shows up when two alternative ways of measuring urban area are used (Appendix Table 1). If all of the illuminated pixels within a city are counted, as the way to measure urbanized area, the Pareto coefficient rose from 0.76 in 2000 to 0.79 in 2010 and 0.83 in 2020 (with these changes over time all statistically significant). If day-time images on land cover data are used to measure the urbanized area the Pareto coefficients rose from 0.82 to 0.84 between 2000 and 2020 (albeit with the changes being less statistically significant than for night-time lights-based estimates).

Discussion and Conclusions

The changing parameters of the rank-size relationships are a convenient way to summarize China's evolving city-size distribution (Li and Lu 2021). To further compare the different pattern in resident population and urban area, the change in area and population of three specific cities shown in Fig 2. In general, it can be seen that the large cities where the expansion in resident population has been far faster than the

expansion in land area, and also the reverse, of small cities in terms of population that had only slow growth in residents but large growth in city area. Specifically, as the median city among the samples, the land and population growth rates have been roughly the same over the past two decades in Hangzhou city. The growth pattern in Xiamen City shows a big mismatch, that is, the growth of urban area can not support the excessive influx of people (the annual growth rate during the period 2000-2020 is 23% vs 46%). In contrast, some cities like Suining city expand their urban area rapidly but did not attract as much as population, which results in a waste of land resources.

Conclusions

An important phenomenon was found in this study: China's city size distribution is diverging in terms of the urban resident population but converging in terms of city area. These divergent patterns in land and people suggest that growth in the resident population of the large cities is not being assisted by fast enough built-up area expansion, while the conversion of non-urban land to urban uses for the less populous cities is too fast for their slow growth in resident numbers. Consequently more rural land is converted into built up area than may be needed, with some of this urban conversion likely to be in the wrong place, creating ongoing environmental issues. Hence cities are getting more equal in terms of urbanized area but less equal in terms of resident numbers. This study contributes evidence for the Chinese government to allow scientific decision-making basis for creating a reasonable city size distribution system.

Discussion

The results show that the city size distribution of the above-territory city system in China during 2000–2020 is consistent with the law of dislocation size, but deviates from the Zipf ideal. This indicates that there are some irrationalities in China's city size system and needs to adjust the future urban development strategy. In future studies, we will further examine the intra-regional variation in city size. In addition, the long intervals between census data can neglect changes in city size distribution in the

intervening years. We will build a better combination of data and methods for a more rigorous study.

References

Anderson, G., and Ge, Y. (2005). The size distribution of Chinese cities. *Regional Science and Urban Economics*, *35*(6), 756-776.

Au, C. C., and Henderson, J. V. (2006). Are Chinese cities too small? *The Review of Economic Studies*, *73*(3), 549-576.

Deng, M., Lai, G., Li, Q., Li, W., Pan, Y., and Li, K. (2022). Impact analysis of COVID-19 pandemic control measures on nighttime light and air quality in cities. *Remote Sensing Applications: Society and Environment*, 27, 100806.

Düben, C., and Krause, M. (2021). Population, light, and the size distribution of cities. *Journal of Regional Science*, 61(1), 189-211.

Fernández, J. E. (2007). Resource consumption of new urban construction in China. *Journal of Industrial Ecology*, *11*(2), 99-115.

Gan, L., Li, D., and Song, S. (2006). Is the Zipf law spurious in explaining city-size distributions? *Economics Letters*, *92*(2), 256-262.

Gibson, J., and Li, C. (2020). Pareto's law and city size in China: Diverging patterns in land and people. In J. Poot & M. Roskruge (ed) *Population Change and Impacts in Asia and the Pacific*, pp. 29-47, Springer: Singapore.

He, C., Liu, Z., Xu, M., Ma, Q., and Dou, Y. (2017). Urban expansion brought stress to food security in China: Evidence from decreased cropland net primary productivity. *Science of the Total Environment*, *576*, 660-670.

Li, H., Chen, J., Chen, Z., and Xu, J. (2022). Urban population distribution in China: Evidence from internet population. *China Economic Review*, 74, 101808.

Li, P., and Lu, M. (2021). Urban systems: Understanding and predicting the spatial distribution of China's population. *China and World Economy*, 29(4), 35-62.

Li, Z., Li, C., Gibson, J., and Deng, X. (2024). Two decades of inter-city migration in China: The role of economic, natural and social amenities. *Working Paper* No. 24/05, Department of Economics, University of Waikato.

Mulligan, G. F., and Carruthers, J. I. (2021). City-size distribution: the evolution of theory, evidence, and policy in regional science. *Handbook of Regional Science*, 41-59.

Olivia, S., Boe-Gibson, G., Stitchbury, G., Brabyn, L., and Gibson, J. (2018). Urban land expansion in Indonesia 1992–2012: evidence from satellite-detected luminosity. *Australian Journal of Agricultural and Resource Economics*, *62*(3), 438-456.

Roberts, M., Deichmann, U., Fingleton, B., and Shi, T. (2012). Evaluating China's road to prosperity: A new economic geography approach. *Regional Science and Urban Economics*, *42*(4), 580-594.

Rogoff, K., and Yang, Y. (2024). Rethinking China's growth. *Economic Policy*, *39*(119), 517-548.

Rozenfeld, H., Rybski, D., Gabaix, X., and Makse, H. (2011). The area and population of cities: New insights from a different perspective on cities. *American Economic Review*, *101*(5), 2205-2225.

Su, H. L. (2020). On the city size distribution: A finite mixture interpretation. *Journal of Urban Economics*, 116, 103216.

Sun, W., Ding, X., Su, J., Mu, X., Zhang, Y., Gao, P., and Zhao, G. (2022). Land use and cover changes on the Loess Plateau: A comparison of six global or national land use and cover datasets. *Land Use Policy*, 119, 106165.

Wang, X., Shi, R., and Zhou, Y. (2020). Dynamics of urban sprawl and sustainable development in China. *Socio-Economic Planning Sciences*, 70, 100736.

Xu, Z., and Zhu, N. (2009). City size distribution in China: are large cities dominant? *Urban Studies*, *46*(10), 2159-2185.

Zhang, X., Li, C., and Gibson, J. (2024). The role of spillovers when evaluating regional development interventions: evidence from administrative upgrading in China. *Letters in Spatial and Resource Sciences, 17*(1), 9.

Figure 1. Example of forming urban cores: Fujian Province

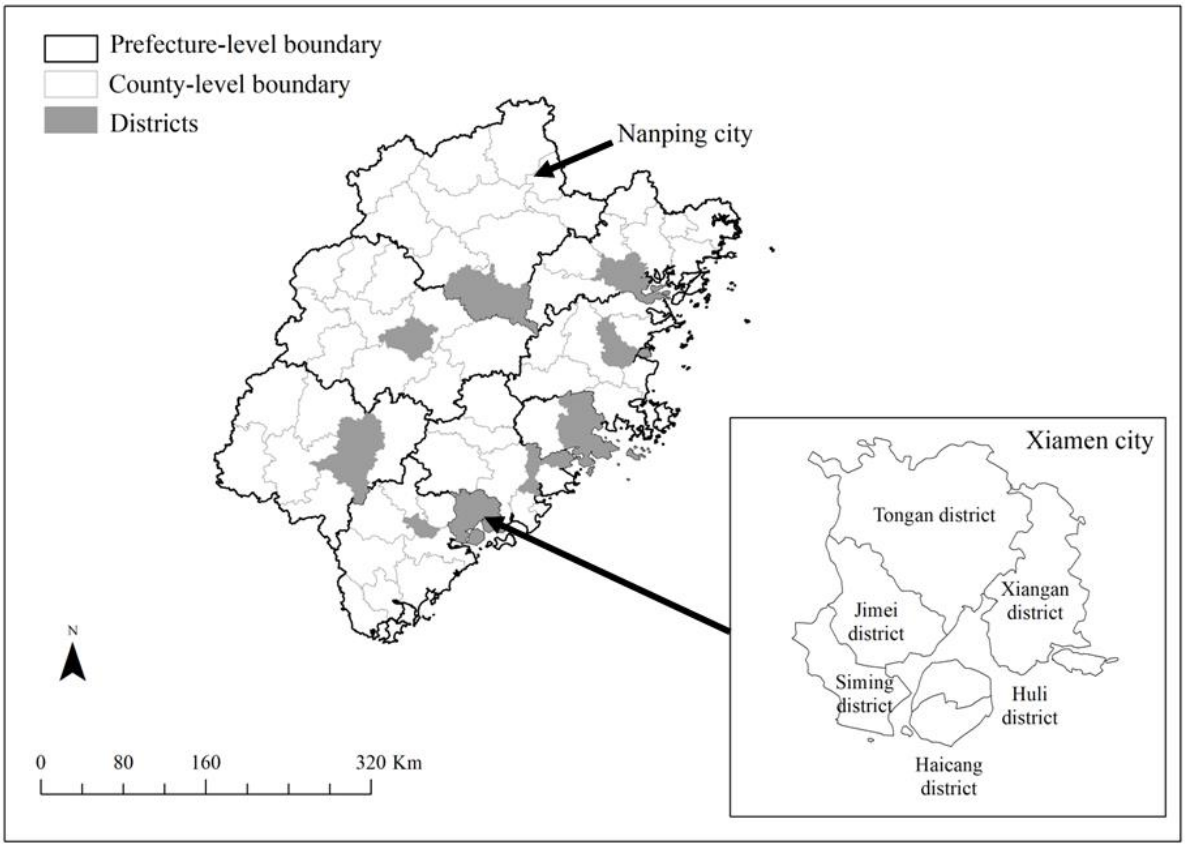


Figure 2. Rank size plot for resident population showing the slope becoming flatter over time, and giving the example of Guangzhou

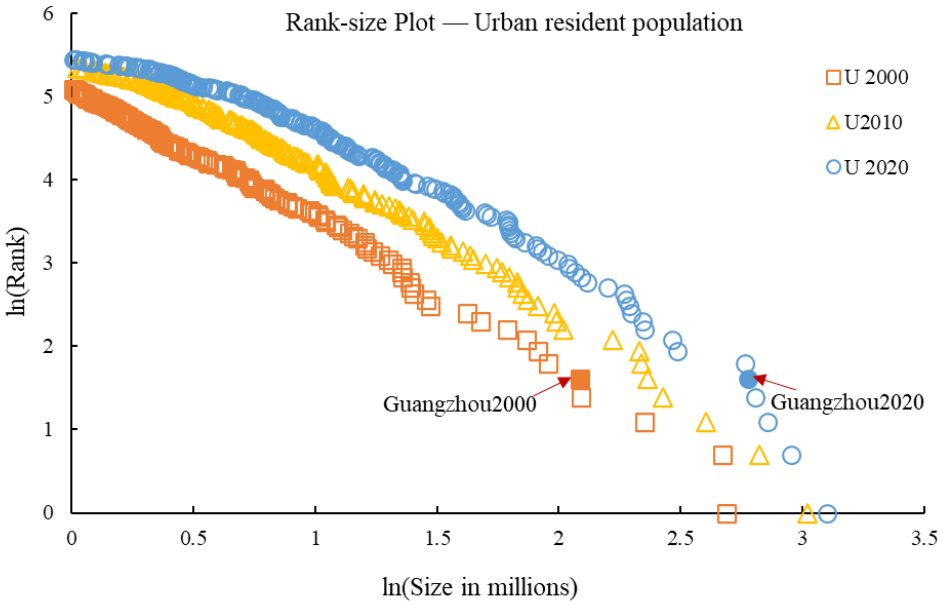
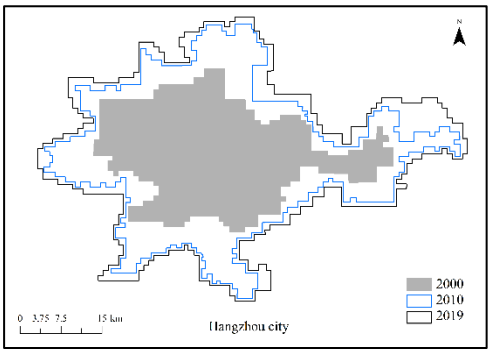
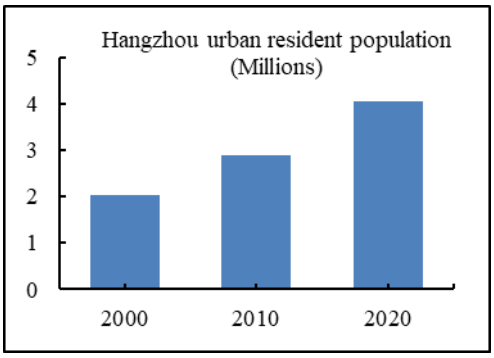
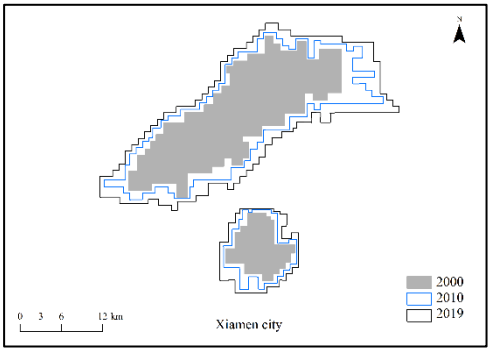
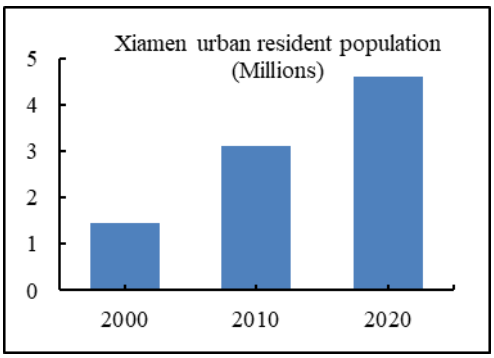
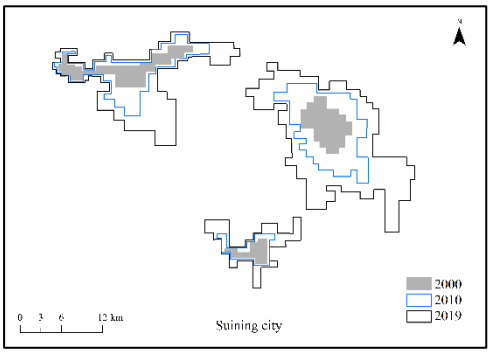
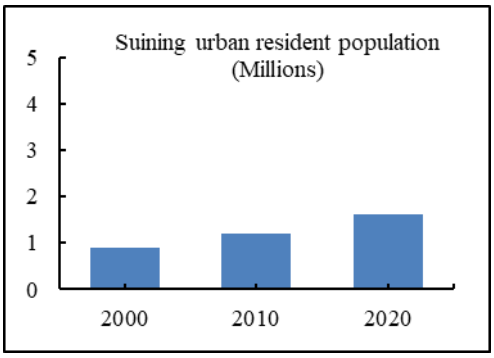


Figure 3. Various combinations of growth in city population and area

	Changes in city area	Changes in resident population
Median city		
	g=33.5%	*g*=35.1%
Fast population growth but slow area growth		
	g=23.8%	*g*=47.0%
Slow population growth but faster area growth		
	g=86.2%	*g*=21.8%

Note: g refers to the decadal growth rate.

Table 1. Changes over time in Pareto coefficients for city population and urbanized area

Pareto coefficient	*Panel A Population*		
	2000	2010	2020
2000	1.463*** (0.018)		
2010		1.444*** (0.020)	
2020			1.345*** (0.022)
Adjusted R^2	0.964	0.954	0.937
Chow test:			
2020-2000			9.49***
2020-2010			21.74***
	Panel B City area (using a 20% luminosity threshold)		
	2000	2010	2019
2000	0.957*** (0.011)		
2010		0.973*** (0.011)	
2019			0.981*** (0.011)
Adjusted R^2	0.968	0.968	0.968
Chow test:			
2020-2000			30.54***
2020-2010			4.58*

Note: The Pareto coefficients and their standard errors in ( ) are reported in the cell values. The results for city area use the area that is illuminated above a threshold (set at 20% of the maximum luminosity) to exclude the area with stray light or with sufficiently dim light as to be unlikely to come from urban area. * significant at 10%; ** at 5%; *** at 1%. The Chow test is for the difference in the coefficients between the two periods shown in the last two rows.

Appendix Table 1. Comparison of Pareto coefficient based on NTL data and land use data

Pareto coefficient	*Panel A City area (using a 20% luminosity threshold)*		
	2000	2010	2019
2000	0.957*** (0.011)		
2010		0.973*** (0.011)	
2019			0.981*** (0.022)
Adjusted R^2	0.968	0.968	0.968
Chow test:			
2020-2000			30.54***
2020-2010			4.58*
	Panel B City area (all NTL)		
	2000	2010	2019
2000	0.756*** (0.014)		
2010		0.785*** (0.015)	
2019			0.831*** (0.015)
Adjusted R^2	0.918	0.921	0.926
Chow test:			
2020-2000			19.03***
2020-2010			17.83***
	Panel C LUCC		
	2000	2010	2020
2000	0.816*** (0.014)		
2010		0.838*** (0.015)	
2020			0.840*** (0.015)
Adjusted R^2	0.928	0.928	0.926
Chow test:			
2020-2000			1.09

2020-2010	0.02

Note: The city area in this table is extracted based on the nighttime light data and land use data respectively. The nighttime lit area used here refers to the lit area of the NTL images (set a threshold of the digital values) to exclude the impact of stray light, lightning, lunar illumination, and cloud cover. Standard errors in parentheses: * significant at 10%; ** at 5%; *** at 1%. The Chow test is for the difference in the coefficients between the two periods shown in the last two rows.

Chapter 8: Conclusion

8.1 Overview

China's economic growth has drawn much attention over recent decades (Fan and Hao 2020; Zhou et al. 2021). At the same time, regional science research placed growing relying on satellite-detected data, especially of night-time lights (Ji et al. 2019; Mathen et al. 2024). In particular, in economics and other fields nighttime lights (NTL) data are used as a proxy for economic activity when time series data on gross domestic product (GDP, GDP per capita or gross regional product) are either unavailable or of poor quality (Ustaoglu, Bovkır, and Aydınoglu 2021; Zhao et al. 2023). The NTL data are increasingly applied to study the effects of shocks (e.g. floods) or policy interventions (Kocornik-Mina et al. 2020; Bluhm, and McCord 2022). Many of these studies cover China in their samples. However, uncertainties in nighttime light data can bias the results when assessing regional economic development, especially when working with data for small spatial units, leading to unreliable economic conclusions for policymakers, and thus hindering its use (Xiao, Wang, and Zhou 2021). In the context of accelerating regional economic integration in China, studying the performance of NTL data as an indicator of socio-economic activity is important.

Accurate spatialized data on GDP and resident population aid investigation of regional economic dynamics. Yet the most widely cited validation studies do not use the newest NTL data products, may not distinguish time-series from cross-sectional uses of NTL data, and usually are for aggregated units, such as nation-states or the first sub-national level (Cui et al. 2020; Chen et al. 2022; Su 2022), while applied studies increasingly focus on smaller and lower-level spatial units. Through a case study of China, this thesis focused on the estimation and spatial distribution of GDP and resident population using multiple sources of NTL data, gridded population data, statistical data on economic output and census data on population. One specific purpose of this thesis was to examine the performance of NTL data in proxying for socio-economic indicators at the county-level in China. Though the NTL data can effectively reflect spatial differences in economic activities it cannot appropriately reflect the temporal change of GDP at the county level due to complexity of the economy and natural conditions.

This thesis, therefore, seeks to fill a gap in the literature by comprehensively testing the prediction accuracy of NTL data and gridded population data as indicators of county-level economic development, based on comparisons with long-term GDP and population data. Empirical results for this aspect of the thesis are presented as standalone papers in Chapter 2 through 4. The thesis also analyzes a series of key regional economic issues for China using multiple NTL data and other remote sensing data to supplement the official GDP and population data, with the results presented as standalone papers in Chapters 5 to 7.

8.1.1 Key Results

This thesis focuses on the use of NTL data as a proxy for socioeconomic indicators in economic research. A particular aim of this thesis was to consider prediction accuracy of NTL data for county-level economic development and expand its application to some regional economy issues, for example, regional inequality, regional policy, land use, infrastructure, and so on. In going beyond most existing studies on NTL data, this study contributes to the literature in the following three ways.

First, this study considers the effects of the saturation issue of NTL data, which cannot be overlooked when applied on a smaller scale, such as at the county level. Second, this study focuses on the analysis at finer spatial level data in China (for example county and district level units). Third, this study comparatively uses multiple sources of nighttime light data to improve upon the year-by-year estimates of economic activity and the decadal estimates of the resident population from each census. This research contributes to the existing literature by using an econometric approach. This thesis provides a new perspective in the research on China's local economies, enabling the visualization of economic development trends on a spatially continuous surface through spatialized simulations of China's county population and GDP per capita.

Indeed, as shown in chapters 2-4, this study uses a set of reliable estimates of resident population and GDP to test how well the NTL data and gridded population data can predict the temporal change in socio-economic indicators, especially at the most

spatially disaggregated county-level. Then Chapters 5-7 apply the NTL data to analyze economic issues, including regional policy, infrastructure and agglomeration, and land use and migration. Overall, these various empirical studies help to objectively and accurately show development status and process of relevant socio-economic activities and provide scientific support for the government's planning and decision-making and relevant socio-economic research.

Chapter 2 conducts a comprehensive regression-based analysis of four widely used NTL data to provide more updated and disaggregated validation results. This chapter examines the relationships between annual GDP and the NTL data for 2657 county-level units in China from 2012 to 2019. It uses the same estimation framework as a previous study on the United States (Gibson, and Boe-Gibson 2021) to draw broader lessons through the comparison. In line with findings elsewhere, it concludes that the nighttime light data has a far worse performance in predicting the time-series changes in economic activity than the cross-sectional differences. The annual changes in county-level GDP in China have an elasticity of about 0.1 with respect to annual changes in NTL data. The results also show that using the weighted average of readings from snow-free and snow-covered nights from the NASA Black Marble NTL composites appeared to be closer to the truth than were other models of GDP changes that used other NTL data sources. The weak relationships shown in temporal changes of economic activity and nighttime light data raise issues for interpreting applied studies that show the effects of a treatment on changes in nighttime lights.

Chapter 3 gives examples of the mean-reverting error in the most widely used DMSP data and showed impacts on spatial inequality estimates for China and the United States, where these two countries have county-level GDP data as a benchmark. The results show that the inequality indicator is understated if DMSP data are the proxy for regional GDP, and so these spatially mean-reverting measurement errors in DMSP data could mislead policymakers and regional science practice. It suggests that the uncritical use of DMSP data has not been given sufficient attention and may introduce some new misunderstandings about spatial inequality trends. This chapter provides new

evidence that the spatially mean-reverting errors make the DMSP data poorly suited for estimating spatial inequality, even if they may adequately proxy for the average levels of (or growth in) economic activity at the national or aggregated regional level in other research contexts.

Chapter 4 examines how well gridded population estimates can predict the differences in census population between places and changes between time periods. Based on population census data as the benchmark, the four popular sources of gridded population estimates are better at predicting cross-sectional population differences than at predicting the changes in population over time, especially at the most local level. Moreover, gridded population data gives a misleading impression about the trends in spatial inequality, which could lead to inappropriate policy interventions that are based on a misreading of the actual situation. Therefore, using gridded population products in research that depends on temporal variation in the subject under study may be a source of vulnerability. The findings suggest that researchers should be cautious in relying on gridded population estimates as a proxy for the actual temporal changes in population that are occurring at fine spatial scales.

Chapter 5 uses a series of spatial econometric models to examine the effects of converting China's counties to cities, explicitly allowing for spillover effects on local economic activity. The results indicate that the positive direct effects on local economic growth (using GDP and night-time lights as the indicators) of a county being upgraded are amplified by positive indirect effects. The indirect spillovers are especially apparent in the eastern regions of China where economic activity and population are more densely concentrated. The models without spatial lags (which most prior studies used) treated counties as independent of their neighbours and ignored these spillover effects, leading to an understated impact of county upgrading that is just two-fifths to two-thirds as large as the estimated total effects coming from the spatial models. This finding adds to the growing list showing the importance of allowing for spatial spillovers and suggests that the literature studying county upgrading in China might benefit from more widespread use of spatial models that explicitly allow spillover effects to be studied.

Chapter 6 examines the impacts on GDP and luminosity of connecting to the high-speed rail (HSR) network, using a decade-long panel of county-level units. The research framework considers both the spillover effects and (by using an instrumental variables (IV) strategy) the endogenous placement issue that potentially interferes with estimation of HSR impacts on the regional economy. Despite a substantial increase in HSR connectivity in the years studied, the results show that the HSR network may not boost China's local economic growth. These findings echo Qin (2017) and Gao et al. (2020) that connecting to the high-speed rail network might even harm local economic growth through the spatial reallocation of the economically active population from peripheral areas to more central areas, and through the spatial restructuring of industries. It suggests that policy makers may need to do more evaluation of economic impacts on the targeted county-level units and their neighbours before planning and constructing future additions to the high-speed rail network.

Chapter 7 explores China's city size distribution based on population census and city built-up area. The phenomenon shows that there are contrasting changes in the distribution of city size according to land and people, with cities becoming more equal in urbanized area and less equal in resident population. Thus, resident population trends and the expansion of land are not well matched in the process of urbanization. These divergent patterns in land and people suggest that growth in the resident population of the large cities is not being assisted by fast enough built-up area expansion, while the conversion of non-urban land to urban uses for the less populous cities is too fast for their slow growth in resident numbers. These discordant trends may contribute to the misalignment of house prices and to macroeconomic fragility in China.

8.2 Future Work

The findings in this thesis have important implications for studies using nighttime lights to evaluate economic effects of shocks or policy interventions. However, there are still uncertainties and limitations in this thesis that will be further improved in future work.

First, when interpreting how changes in nightlights reflect unmeasured

changes in economic activity, this study mainly focuses on the relationship between nightlights and GDP, and does not consider the correlation with other indicators such as income or income proxies when GDP data are unavailable (Bruederle, and Hodler, 2018). Second, the comparison in chapter 2 shows that the new generation nighttime light data (for example the Black Marble data) have a better performance in predicting county-level GDP. It will require further testing in other settings to see if these apparent advantages of the high-resolution NTL data products hold more widely. Third, the local characteristics of the region and the scale of aggregation may influence the ability of nighttime lights to pick up changes in economic activities. Han et al. (2022) found NTL data more representative of GDP if an administrative region is relatively comprehensive or highly developed, while the representation of GDP becomes weak when regions are unbalanced or undeveloped. This suggests that it may be better to divide the regions according to their economic development level and exploring the optimal NTL data and models within subsets of a sample for better prediction accuracy. It will be helpful to combine more complex data in the process of using NTL data to estimate regional economic development, such as land use data, big data from social media, and other remote sensing data. Moreover, accuracy verification and validation from field surveys could be carried out to improve the estimation accuracy of NTL data. Finally, future work can extend to different countries to test whether the differences in measurement errors are common across regions.

As a final note, more and more attention in the economics literature has been paid to the relationship between remote sensing data (especially nighttime light data) and socio-economic indicators. Armed with this enhanced knowledge, future research can further identify the performance of nighttime light data on other economic research topics. Although this thesis is primarily focused on China, broader implications for the use of nighttime lights and other remote sensing data in research should be apparent.

References

Bluhm, R., and McCord, G. C. (2022). What can we learn from nighttime lights for small geographies? measurement errors and heterogeneous elasticities. *Remote Sensing*, 14(5), 1190.

Bruederle, A., and Hodler, R. (2018). Nighttime lights as a proxy for human development at the local level. *PloS one*, 13(9), e0202231.

Chen, Z., Wei, Y., Shi, K., Zhao, Z., Wang, C., Wu, B., ... and Yu, B. (2022). The potential of nighttime light remote sensing data to evaluate the development of digital economy: A case study of China at the city level. Computers, *Environment and Urban Systems*, 92, 101749.

Cui, Y., Shi, K., Jiang, L., Qiu, L., and Wu, S. (2020). Identifying and evaluating the nighttime economy in China using multisource data. *IEEE Geoscience and Remote Sensing Letters*, 18(11), 1906-1910.

Fan, W., and Hao, Y. (2020). An empirical research on the relationship amongst renewable energy consumption, economic growth and foreign direct investment in China. *Renewable Energy*, 146, 598-609.

Gao, Y., Song, S., Sun, J., and Zang, L. (2020). Does high-speed rail connection really promote local economy? Evidence from China's Yangtze River Delta. *Review of Development Economics*, 24(1), 316-338.

Gibson, J., and Boe-Gibson, G. (2021). Nighttime lights and county-level economic activity in the United States: 2001 to 2019. *Remote Sensing*, 13(14), 2741.

Han, G., Zhou, T., Sun, Y., & Zhu, S. (2022). The relationship between night-time light and socioeconomic factors in China and India. *Plos one*, 17(1), e0262503.

Ji, X., Li, X., He, Y., and Liu, X. (2019). A simple method to improve estimates of county-level economics in China using nighttime light data and GDP growth rate. *ISPRS International Journal of Geo-Information*, 8(9), 419.

Kocornik-Mina, A., McDermott, T. K., Michaels, G., and Rauch, F. (2020). Flooded cities. *American Economic Journal: Applied Economics*, 12(2), 35-66.

Mathen, C. K., Chattopadhyay, S., Sahu, S., and Mukherjee, A. (2024). Which Nighttime Lights Data Better Represent India's Economic Activities and Regional Inequality? *Asian Development Review*, 41(2), 193-217.

Qin, Y. (2017). 'No county left behind?' The distributional impact of high-speed rail upgrades in China. *Journal of Economic Geography*, 17(3), 489-520.

Su, L. (2022). The relationship between the average night light intensify and GDP in Shanghai: Based on the integration data of DMSP-OLS and NPP-VIIRS. *Journal of Computational Methods in Sciences and Engineering*, 22(5), 1729-1735.

Ustaoglu, E., Bovkır, R., and Aydınoglu, A. C. (2021). Spatial distribution of GDP based on integrated NPS-VIIRS nighttime light and MODIS EVI data: A case study of Turkey. *Environment, Development and Sustainability*, 23, 10309-10343.

Xiao, Q. L., Wang, Y., and Zhou, W. X. (2021). Regional economic convergence in China: A comparative study of nighttime light and GDP. *Frontiers in Physics*, 9, 525162.

Zhao, Z., Tang, X., Wang, C., Cheng, G., Ma, C., Wang, H., and Sun, B. (2023). Analysis of the spatial and temporal evolution of the GDP in Henan Province based on nighttime light data. *Remote Sensing*, 15(3), 716.

Zhou, X., Cai, Z., Tan, K. H., Zhang, L., Du, J., and Song, M. (2021). Technological innovation and structural change for economic development in China as an emerging market. *Technological Forecasting and Social Change*, 167, 120671.

Appendix 1: Are Disaster Impact Estimates Distorted by Errors in Popular Night-Time Lights Data?

John Gibson, Yi Jiang, Xiaoxuan Zhang and Geua Boe-Gibson

This paper has been published in the *Economics of Disasters and Climate Change*, (2024), 1-26.

Economics of Disasters and Climate Change
https://doi.org/10.1007/s41885-024-00152-6

RESEARCH

Are Disaster Impact Estimates Distorted by Errors in Popular Night-Time Lights Data?

John Gibson[1] · **Yi Jiang**[2] · **Xiaoxuan Zhang**[1] · **Geua Boe-Gibson**[1]

Received: 31 March 2024 / Accepted: 7 June 2024

Abstract
Satellite-detected night lights data are widely used to evaluate economic impacts of disasters. Growing evidence from elsewhere in applied economics suggests that impact estimates are potentially distorted when popular Defense Meteorological Satellite Program (DMSP) night lights data are used. The low resolution DMSP sensor provides blurred and top-coded images compared to those from the newer and more precise Visible Infrared Imaging Radiometer Suite (VIIRS) images. Despite this, some disaster impact studies continue to use DMSP data, which have also been given a new lease of life through the use of harmonized series linking DMSP and VIIRS data to provide a longer sample. We examine whether use of DMSP data affects evaluations, using expected typhoon damages in the Philippines from 2012–19 as our case study. With DMSP data, negative impacts on economic activity from expected damages at the municipality level appear over 50% larger than when the VIIRS data are used. The DMSP data give the appearance of greater spatial autocorrelation in luminosity and through this channel may tend to spread apparent spatial impacts of shocks. Harmonized data that adjust VIIRS images to be like the DMSP data also have this spurious autocorrelation so researchers should be cautious in using these data for disaster assessments.

Keywords DMSP · Disaster impacts · Night lights · Spatial autocorrelation · Typhoons · VIIRS · Philippines

JEL Classification C21 · Q54

Introduction

Practitioners and researchers can now rapidly evaluate impacts of disasters and monitor the recovery by using satellite-detected earth observation data. Readily accessible night-time lights (NTL) data especially enable this research on disasters (Feeny et al. 2022; Nguyen and Noy 2020). Yet even as this literature expands, questions are raised elsewhere

✉ John Gibson
jkgibson@waikato.ac.nz

[1] Department of Economics, University of Waikato, Private Bag 3105, Hamilton, NZ, New Zealand 3240

[2] Asian Development Bank, Metro Manila, Philippines

about using NTL data as a proxy for changes in local economic activity. These data seem to only weakly relate to changes in traditional measures, such as commune-level employment and household expenditures (Goldblatt et al. 2020) or county-level GDP (Zhang and Gibson 2022), even in countries where economic activity-luminosity relationships seem to strengthen as the data are spatially aggregated (Gibson et al. 2021). The caveats from these studies coming from outside of the disasters literature matter because it is *changes* in economic activity that are especially needed for estimating disaster impacts; hence, the cross-sectional relationships between luminosity and economic activity do not necessarily imply that there are equally useful relationships for measuring the changes in activity (Chen and Nordhaus 2019).

Moreover, it is especially in the low density rural areas of developing countries where relationships between NTL data and local economic activity are weak (Gibson et al. 2021). In such places there is a scarcity of alternative data for evaluating disaster impacts unlike the situation for urban areas. Yet luminosity is especially associated with economic activity in urban areas, and in recognition of this fact some researchers suggest that the NTL data may only be suitable for studying the impacts of (or recovery from) disasters that mainly affect urban areas (Akter 2023). If this caveat is agreed upon more widely it would tend to limit the social usefulness of the rapidly growing literature that estimates impacts of disasters.

One potential reason for why NTL data may be a poor proxy for changes in the local economic activity is that the most popular data are from the Defense Meteorological Satellite Program (DMSP), launched by the US Department of Defense in the 1960s to observe clouds for short-term weather forecasts. Hence, these satellites and sensors were not designed for long-term earth observation, nor for research purposes, and so findings based on these data are somewhat serendipitous. It should be no surprise that it is urban areas whose density and scale allows them to be detected from space even with fairly crude sensors and that the ability to detect changes in local economic activity may be limited. In contrast, newer, more accurate and precise luminosity measures are available since 2012 from the research-oriented Visible Infrared Imaging Radiometer Suite (VIIRS) that detects a far wider range of lighting and thus should provide a better basis for measuring changes in local economic activity.

It might be expected that research on the economic impact of disasters would quickly switch to using these newer VIIRS data. Yet studies on the impacts of disasters continue to be produced using DMSP data, as we show below.[1] Moreover, the DMSP data have been given a new lease of life by so-called 'harmonized' data where VIIRS data and DMSP data are linked to provide one long time-series. This harmonization essentially requires degrading the more precise VIIRS images to make them like the low-resolution DMSP data (Nechaev et al. 2021) and so, indirectly, some of the flaws in the DMSP data still enter into the post-2012 record even when the more precise and accurate VIIRS data are available.

Given the on-going use of the DMSP data in disasters research some evaluation of the sensitivity of estimated impacts to the use of different NTL data may be helpful. We therefore examine whether use of DMSP data affects disaster evaluations, using expected

[1] Overall, economics has seen a persistence in the use of DMSP data not seen in other disciplines. For example, a Scopus search reveals that outside of economics, 12% more articles published in the last two years (2021–22) that mention VIIRS (and 'night' to rule out non-NTL uses) than mention DMSP. Yet for the economics subject area it is the other way around, with 76% more articles mentioning DMSP than VIIRS over the same period.

typhoon damages in the Philippines from 2012–19 as our case study. With the DMSP data, the municipal level (negative) impacts on economic activity appear over 50% larger than when the VIIRS data are used. The potential overstatement of the economic shocks is especially apparent with the spatial models that allow for spillovers; most likely this is because the DMSP data give the appearance of greater spatial autocorrelation in luminosity and through this channel they will tend to spread the apparent spatial impacts of shocks. Harmonized data that degrade VIIRS images to be like the DMSP data also have this spurious autocorrelation so researchers should be cautious in using these data for disaster assessments.

Our focus is on use of NTL data as a proxy for local economic activity in studies of disaster impacts. We pay far more attention to the left-hand side variable in our econometric models than to the key right-hand side variable, which is an expected damage function based on a wind speed model. This same hazard indicator is used in other studies of the economic growth impacts of hurricanes in the United States (Strobl 2011) and Central America (Strobl 2012; Ishizawa et al. 2019), typhoons in China (Elliott et al. 2015) and most pertinently, for typhoon impacts on local economic activity in the Philippines (Strobl 2019). However, there is evidence linking typhoon fatalities in the Philippines to intense rainfall rather wind damage (Yonson et al. 2018), albeit with vulnerability and exposure seeming more important than the destructive characteristics of the hazard itself. Our findings should apply to models that use alternative right-hand side variables, such as rainfall, because the errors we examine are from the left-hand side NTL variable that we (and many other studies) use although it would need to be confirmed with further study using alternative natural hazards indicators.[2]

In the next section we discuss the persistence in the use of DMSP data in the disasters literature. That section also serves as our literature review and highlights some of the main differences between DMSP and VIIRS. Background to our case study is given in Section III, mapping the patterns of luminosity and expected typhoon damages in the Philippines. The data on damages and luminosity, and the econometric methods are discussed in Section IV. The impact estimates are in Section V and Section VI has our conclusions.

Why are we still talking about DMSP in 2024?

There are at least three reasons for ongoing use of DMSP data in the disasters literature even 12 years after newer, more accurate and precise, VIIRS data became available. First, science advances slowly – especially for results published in economics journals – so some disasters with impact estimates now being published were from before VIIRS (pre-2012). The top panel of Table 1 has details on articles and working papers studying disaster impacts and using DMSP data, published from 2019 onwards (7 years after first availability of the VIIRS data). We found 16 papers on disasters using DMSP data that came out

[2] Yonson et al. (2018) use data from 30 rain gauge stations spread across the Philippines, which is approximately one (spatial) data point for every two provinces (which are the spatial units used in their study). In contrast, we use data for over 1600 municipalities and so there would be considerable 'spreading' of the rainfall data to take them down to the municipality level and this is likely to induce some spatial autocorrelation that is a key issue that we wish to study. The question of whether different NTL data distort impact estimates when rainfall-based hazard indicators are used would therefore be best studied in a setting with a higher density of rainfall gauges.

Table 1 a: Recent Papers Studying Disasters using DMSP data

Authors	Year	Journal/Outlet	Disaster	Sample Period
Carballo et al.	2023	Tourism Economics	Hurricane in the Caribbean	2000–2013
Sajid	2023	AAEA conference	Flood in India	1994–2013
Schippers, & Botzen	2023	Nat Hazards and Earth Syst Sci	Hurricane Katrina	2000–2012
Espagne et al.	2022	AFD Research Papers	Typhoons in Vietnam	1992–2013
Felbermayr, et al.	2022	World Development	Weather anomalies, global grid cells	1992–2013
González, London, & Santos	2021	Climate and Development	Disasters in Argentina	1992–2013
Mohan, & Strobl	2021	J of International Development	Tropical storms in the Caribbean	2000–2013
Kocornik-Mina et al.	2020	AEJ: Appl Econ	Flooded cities	2003–2008
Miranda, Ishizawa, & Zhang	2020	Econ Disasters Clim Change	Windstorms in Central America	1992–2013
Qiang, Huang, & Xu	2020	Sustainable Cities and Society	Hurricane Katrina	1992–2013
Strobl, Ouattara,& Kablan	2020	Applied Economics	Hurricanes in the Caribbean	1992–2013
Brei, Mohan, & Strobl	2019	Q Rev Econ Finance	Hurricanes in Dominican Republic	2001–2012
Fabian, Lessmann & Sofke	2019	Environ. & Development Econ	Earthquakes in 1° global grid cells	1992–2013
Ishizawa, Miranda, & Strobl	2019	Int. J. Disaster Risk Science	Tropical cyclones	1992–2013
Li, Liu, Chen, & Meng	2019	J. Appl. Remote Sensing	2010 Haiti earthquake	1992–2013
Strobl	2019	ADB Working Paper	Typhoons in the Philippines	1992–2013
Panel b: Papers using Harmonized Luminosity Data				
Authors	Year	Journal/Outlet	Disaster	Data
Farzanegan & Fischer	2023	cesifo Working Paper	Bam (Iran) earthquake	Li et al.
Joseph	2022	*World Development*	Haiti earthquake	Nechaev et al.
Liu et al.	2021	*IEEE J-STARS*	Wenchuan (Sichuan) earthquake	Li et al.
Miranda et al.	2021	World Bank Working Paper	Hurricanes in the Caribbean	Li et al.

IEEE J-STARS is the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. Li et al.. and Nechaev et al.. are two different sources of VIIRS data being adjusted to be DMSP-like

in the last five years. Some studied particular disasters, such as Hurricane Katrina, that were before availability of VIIRS data. But many studied multiple disasters within a particular sample period, especially the 1992–2013 period that corresponds to availability of the usual DMSP data. Such research designs could instead be applied to, say, the 2012–23 period with VIIRS data.

Moreover, even papers studying pre-2012 disasters could use VIIRS data to examine possible biases from using DMSP data. First, one can examine the proportion of 'false zeros' from places where DMSP annual composites show no luminosity yet VIIRS shows positive luminosity in the same year. False zeros are a non-classical measurement error that can bias impact estimates (Kim et al. 2024). For example, about one-quarter of municipalities in the Philippines show zero luminosity in the DMSP annual composites; almost 3-times higher than what VIIRS shows for the same years. Growth over time in lighting makes true zeros increasingly less likely, so if DMSP lacks sensitivity to pick up newly illuminated but dimly lit places that VIIRS detects, false zeros should be more frequent post-2012. So a researcher studying a pre-2012 disaster could compare post-2012 VIIRS and DMSP data to flag false zeros and form an upper bound for possible bias in their impact estimates from this problem. Likewise, the more spatially precise VIIRS data can also be used to see if DMSP data give a misleading picture of the similarity in luminosity of nearby areas, which is something that we examine below. Our review found no studies with data exercises along these lines.

A second reason for ongoing using of DMSP data is that a researcher may be studying a post-2012 disaster but wants the longest possible time-series. For example, if using a difference-in-differences framework they likely want to test for parallel trends in the pre-disaster period and DMSP data allow one to go back 20 years earlier than VIIRS data allow. Yet, if this was a primary reason for on-going use of DMSP data we would expect to see discussion in these papers of the trade-offs faced when choosing a NTL data source – a more accurate one but with a shorter time-series versus a more error-prone one but with a longer time-series – and we did not find this sort of discussion in the papers we reviewed.

Another reason DMSP data remain relevant to disasters research in 2024 is more subtle, and relates to the growing use of so-called 'harmonized' data where the DMSP data and the VIIRS data are linked to create one long time-series from 1992 to the present. Several studies provide these harmonized data, with Li et al. (2020) perhaps the most widely cited, while Nechaev et al. (2021) is a very thorough example by the authors who are the most familiar with processing both DMSP and VIIRS images. We have listed some of the recent studies of disaster impacts that use these harmonized data in the bottom panel of Table 1. Challenges in trying to make consistent inferences from the inherently different VIIRS and DMSP data include:

- Differences in timing of earth observation alter the quantum and composition of measured luminosity. DMSP nominally observed earth between 8.30-10pm while VIIRS observes at 1.30am.[3] Economic activities after midnight differ from those in the mid-evening.

[3] Unstable orbits saw DMSP observing earlier as satellites aged. This pattern advanced so far that some of the satellites no longer being used for the annual composites (ca. 2013) were observing earth in the late afternoon, and with the 12-h orbit this also provided pre-dawn observations that are used by Ghosh et al. (2021).

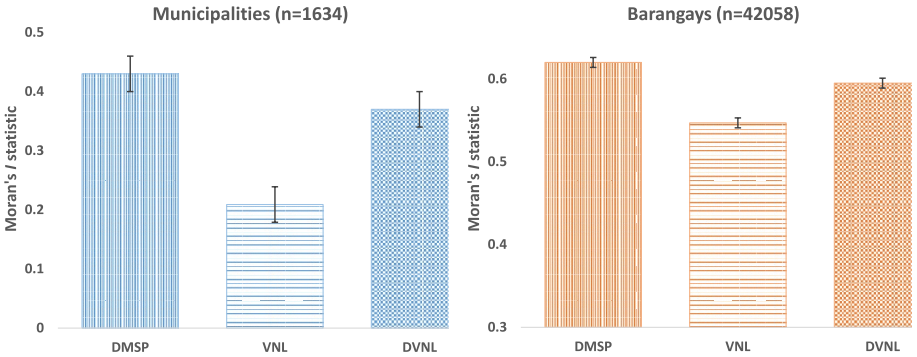


Fig. 1 Moran's I statistics for spatial autocorrelation in night-time lights data from the Philippines (annual composites for 2013) [*Notes*: DMSP is the Defense Meteorological Satellite Program, VNL is VIIRS night lights, DVNL is DMSP-like night time lights derived from VNL. The Moran statistics are calculated using a first-order contiguity spatial weighting matrix and the error bars show 95% confidence intervals.]

- Large differences in spatial precision; at nadir the DMSP ground footprint is 25 km^2 but it is 45-times smaller for VIIRS at 0.55 km^2 (Elvidge et al. 2013).[4] Moreover, the precision advantage for VIIRS is even larger off-nadir because DMSP does not adjust for angular viewing effects, especially blurring images at scan edges (Abrahams et al. 2018).
- Differences in coarseness of quantization; DMSP Digital Numbers (DN) theoretically go from 0-63 but non-zero values below five are rare due to bottom-coding, while top-coding (saturation) may affect pixels with DN values for annual composites as low as 55 (Bluhm and Krause 2022).[5] In contrast to this approximate 50-point scale, the 14-bit quantization of VIIRS allows 16,384 different values and there are no saturation problems.

Given these differences, attempted at harmonization involves degrading the more precise VIIRS images, so some of the DMSP flaws enter into the post-2012 record even with more precise and accurate VIIRS data available. Nechaev et al. (2021) note that while it is possible to smooth (i.e., coarsen) high resolution VIIRS images there is no deterministic way to add detail to the coarse DMSP images. Likewise, it is easy to apply a censor threshold to VIIRS radiances to create a top-coded series but the reverse does not hold; there is nowhere to take brightness values that exceeded the DMSP saturation threshold in order to assign them to equivalent VIIRS values. Hence, it is best to think of the harmonization as

[4] Some studies using DMSP data appear to confuse the output grid (30 arc-seconds, roughly 1 km^2) with the ground footprint, which is far larger. The distinction is discussed by Rogers et al. (2023) and Zhang et al. (2023).

[5] DMSP images reflect unrecorded changes in sensor amplification over the lunar cycle (given the original goal to observe cloud tops rather than to observe earth). Annual composites average over several images per year so Bluhm and Krause (2022) suggest that even annual DN values as low as 55 may be top-coded on some nights.

inherently creating DMSP-like adjusted VIIRS data rather than VIIRS-like adjusted DMSP data.

Figure 1 shows some evidence of the way that the harmonized data can re-introduce DMSP flaws into post-2012 data. The figure has Moran's *I* spatial autocorrelation statistics for night-time lights data for the Philippines in 2013.[6] Intuitively, these statistics show the apparent similarity between an area and adjacent areas, with higher values showing greater similarity (up to a maximum of 1). Spatial autocorrelation is especially important for studying disasters such as typhoons, where the path of storm damage will create correlations between neighbouring spatial units. The left-hand panel has results for municipalities, which have a mean area of 184 km^2, while on the right are results for barangays, whose mean area is 7 km^2. With DMSP data the Moran's *I* statistic is 0.43 for municipalities, slightly over double what VIIRS data show for the same year (the difference is statistically significant at the $p < 0.01$ level). In other words, DMSP overstates similarity of neighbouring areas, which is another perspective on the acknowledged pattern of DMSP data understating spatial inequality due to their spatially mean-reverting errors (Zhang et al. 2023; Mathen et al. 2024).[7]

What happens when VIIRS data are adjusted to be DMSP-like? The third bar in each chart shows the Moran's *I* statistic based on the DVNL harmonized data that are described by Nechaev et al. (2021). The adjustments to VIIRS data to make them DMSP-like increase the Moran's *I* statistic by 77% in the municipality-level results, thereby largely undoing gains in spatial precision achieved by using VIIRS. In other words, the harmonized data re-introduce a distortion that affected the DMSP data. This same pattern is evident in the barangay results, although the baseline level of spatial autocorrelation is higher because a given output of light is being split into smaller areas so there naturally appears to be greater similarity between the adjacent small areas. If we carried on with this exercise all the way down to the (output) pixel scale of 30 arc-seconds used in some studies of disaster impacts (e.g. Strobl 2019) we likely would still see this pattern whereby, for the same set of on-the-ground lights, there is more apparent spatial autocorrelation after the spatially precise VIIRS data have been adjusted (degraded) to be like DMSP.

Why does overstated spatial autocorrelation matter to estimates of disaster impacts? By making adjacent areas look more similar, apparent impacts of a disaster may be spread more widely and this could potentially bias econometric estimates. The patterns in Fig. 1 also make it logically inconsistent to highlight the greatly improved spatial precision of VIIRS while using the DMSP-like VIIRS data (as Joseph (2022) does in a study of the Haiti earthquake) because the adjustments to create DMSP-like VIIRS data largely reverse these gains in spatial precision.

[6] This examines the strength of the relationship between one observation and the spatially weighted average of neighboring observations. We use 2013 as it is the last full year with the usual DMSP data (for satellite F18) and the first full year with VIIRS data (as VIIRS night lights data are only available from April 2012).

[7] We assert that DMSP data overstate similarity, not VNL data understating it, because this is consistent with known flaws in DMSP data, including blurring and top-coding that attenuate apparent differences between areas (Abrahams et al. 2018; Gibson 2021, Kim et al. 2024). Also, evidence from China with local (county) GDP as a benchmark shows Moran's *I* statistics for DMSP luminosity data are biased upwards (Zhang et al. 2024).

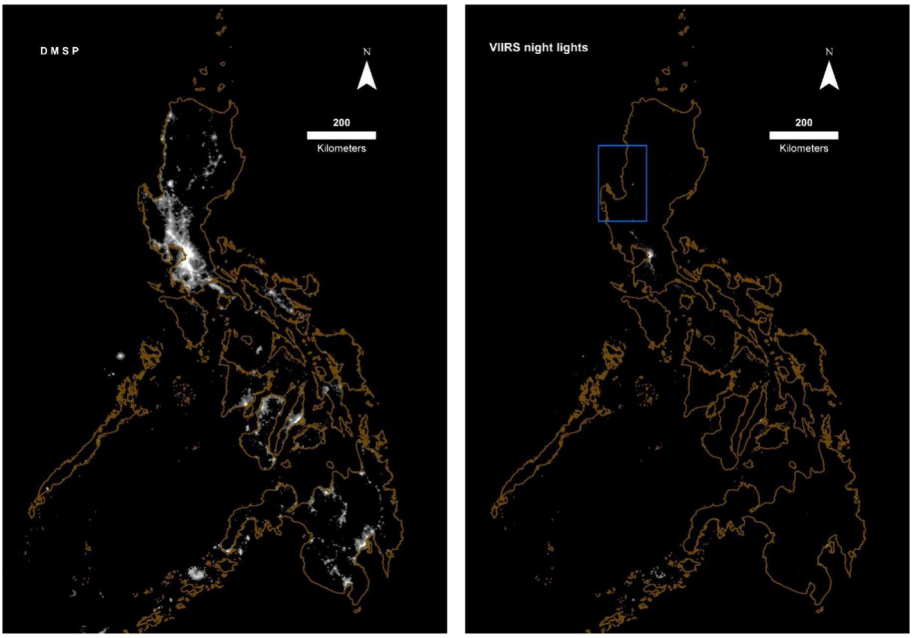


Fig. 2 Night-time lights in the Philippines (annual composites for 2013) according to DMSP (left) and VIIRS (right) [*Notes*: The DMSP annual composite is from satellite F18. The VIIRS night lights are the average masked series. The box in the image at right is the area mapped in Fig. 3.]

Patterns of Luminosity and of Typhoon Damages in the Philippines

We start with luminosity maps to highlight some key differences between DMSP and VIIRS. The map in the left panel of Fig. 2 shows the apparent location of night-time lights in the Philippines in 2013, according to DMSP annual composites. These same data are mapped as Fig. 1 in Strobl (2019), who repeats the commonly made claim that DMSP has a resolution of approximately one km near the equator. As noted above, this claim is ambiguous because it conflates sensor resolution, whose ground footprint is, at best, 25 km^2 at the nadir, with the scale of the grid on to which the data are allocated (Zhang et al. 2023). The map is dominated by a large illuminated area centred on the National Capital Region, with lit area that seems to extend north through much of Central Luzon and into the Ilocos region and south almost to the island of Mindoro. There are also large patches of lights in Central Visayas, centred on Cebu City, in Western Visayas, centred on Iloilo and Bacolod City, and in several parts of Mindanao, especially around Davao City, Cagayan de Oro, and Zamboanga City.

The broad picture provided by DMSP is quite different to what is seen in the VIIRS annual composite for 2013, in the map on the right of Fig. 2. According to the VIIRS data, lights are much more concentrated within the National Capital Region, within Metro Cebu but not spreading over the rest of the island, within Davao City, and in one or two other places. In other words, the DMSP images diffuse the concentrated sources of light to give an exaggerated sense of the scale of brightly lit areas. The remote sensing literature has known for over two decades that the DMSP images may exaggerate city boundaries by a

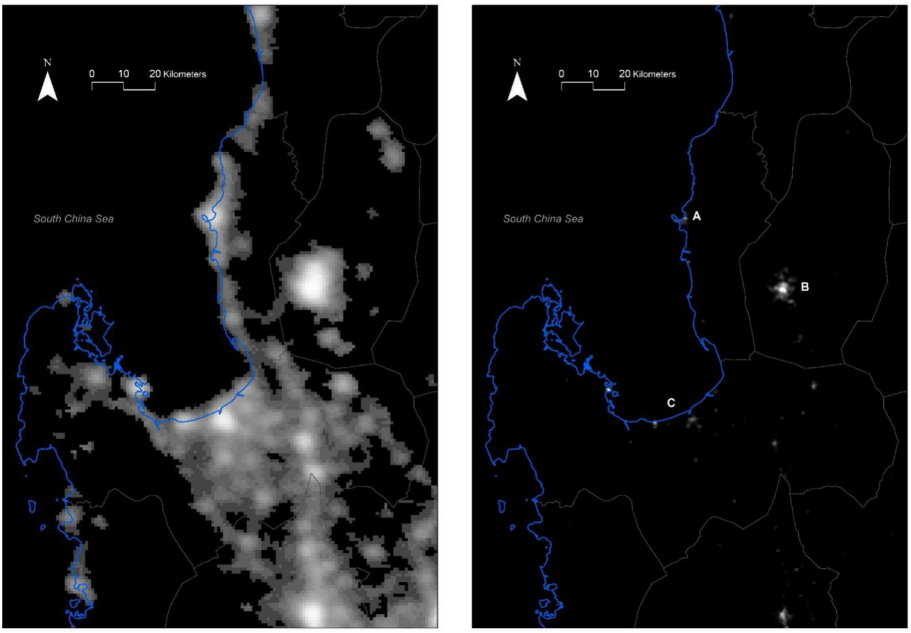


Fig. 3 Night-time lights in a selected part of northern Luzon (annual composites for 2013) according to DMSP (left) and VIIRS (right) [*Notes*: The area mapped is the box in the image at right of Fig. 2. The A, B, and C in the figure at right refer to points of interest noted in the text. For other notes, please see Fig. 2.]

factor of up to ten (Henderson et al. 2003), in a pattern often described as 'blooming' but which is more accurately termed as 'blurring' (Abrahams et al. 2018). This blurring is due to inherent flaws in the DMSP sensor and data management, including pixel aggregation (smoothing) that is done to conserve memory, geolocation errors, and uncorrected angular viewing effects where the expanded ground footprint far away from the nadir overstates the light coming from particular pixels. These features of DMSP have previously been shown to exaggerate the illuminated area of towns and cities in other countries (Gibson 2021; Zhang et al. 2023).[8]

The Philippines-wide maps in Fig. 2 do not allow a full discussion of some of the flaws in DMSP because the scale is too large. So in Fig. 3 we zoom into part of the Ilocos and Cordillera regions of northern Luzon, covering about six percent of the Philippines land area.[9] Once again, the maps have the DMSP image on the left and the VIIRS image on the right, for annual composites for 2013. There are three points of interest (labelled as A, B, and C on the VIIRS map) that help show the potential distortions that may affect the DMSP data. The specific distortions we examine are examples of the general distortions

[8] At the same time, in a pattern that is not apparent to the naked eye, the DMSP data also overstate the extent of completely unlit areas. In the DMSP annual composites for 2013, 24% of municipalities had zero luminosity yet the VIIRS annual composites for the same year show only 12% were unlit. In other words, the DMSP data are simultaneously diffusing luminosity from brightly lit areas, while also bottom coding dimly lit areas.

[9] This area, in the box in the map on the right of Fig. 2, has dimensions of approximately 130 × 200 km.

that underlie the exaggerated spatial autocorrelation in DMSP data that was shown in Fig. 1.

The first point of interest (A) is San Fernando, the capital city of La Union province. The administrative boundaries of this city enclose an area of 100 km^2 although much of that is forested hillsides, with only about one-third as built-up area. The VIIRS image reflects this pattern, with clearly illuminated pixels covering about 30 km^2 and a clear demarcation from the surrounding hinterland. In contrast, the DMSP image makes the lit area appear far larger, extending about 15 km inland and about 50 km north and south along the coast. The second point of interest (B) is the mountain city of Baguio, that was extensively damaged in the 1990 Luzon earthquake. This city with a population of about 370,000 has a highly urbanized area of about 60 km^2, and VIIRS images are consistent with this; a rectangle of roughly 10 km in the north–south direction and 8 km in the east–west direction would cover not just lights from the city but also parts of the adjacent municipalities of La Trinidad and Tublay. In contrast, in the blurred DMSP image Baguio seems to have a highly illuminated area of about 500 km^2 and so city area is overstated by a factor greater than six. The third point of interest is in the Lingayen Gulf (C), just offshore from Binmaley Beach where DMSP images make it seem that lights extend for about five kilometres over the water – while it is true that in parts of the Philippines houses are over water on stilts, photographs of this area show long sandy beaches with no houses. Instead it is due to the distortion from the DMSP blurring that makes it seem that lights are coming from over the water.

Why are these features of DMSP data relevant to the estimation of disaster impacts? Consider a shock that affects a city, such as a typhoon hitting San Fernando or an earthquake damaging Baguio. The direct effect of the shock may reduce luminosity because of damage to electricity generation and transmission infrastructure, or damage to buildings causing a fall in demand for lighting. Indirect effects may then result from harm to livelihoods, creating an economic shock that also leads to lower demand for lighting. If these effects last for a period of weeks or months the annual composites would be likely to show a reduction in luminosity. However there may be uncertainty about the spatial extent of this shock, especially for the indirect effects. Hence, if lights emanating from a concentrated urban area are diffused in the blurred DMSP images, any dimming of city lights in the annual composite will also show up as a dimming in the adjacent areas. Of course these areas may truly be affected by any of the economic channels that link them to the city (potentially positively if activity displaces to undamaged hinterland or negatively if there is a fall in the demand for hinterland labour and output), so with blurred DMSP data it may be quite difficult to get an accurate estimate of the impacts. In contrast, the more spatially precise VIIRS data should allow these impacts to be observed just where they occur and not in adjacent areas if they truly are unaffected.

It will require formal econometric evidence, in Section V below, to ascertain whether features of DMSP data illustrated in Figs. 2 and 3 cause detectable bias when the data are aggregated to the municipality level. There is evidence from elsewhere at similar spatial scale that these flaws in DMSP data do matter. Specifically, Kim et al. (2024) study the closure of an industrial zone in North Korea located in a municipality whose area is 180 km^2 (so similar to the mean area of municipalities in the Philippines) divided into two districts. The district with the industrial zone was at the 99th percentile of North Korea's luminosity distribution in both VIIRS and DMSP (prior to the closure) but there was a big gap between where the two data sources ranked the other district; at the 29th percentile in VIIRS and the 79th percentile in DMSP. In other words, the luminosity gap between the two districts was greatly attenuated in DMSP data (seeming to be only 20 percentiles

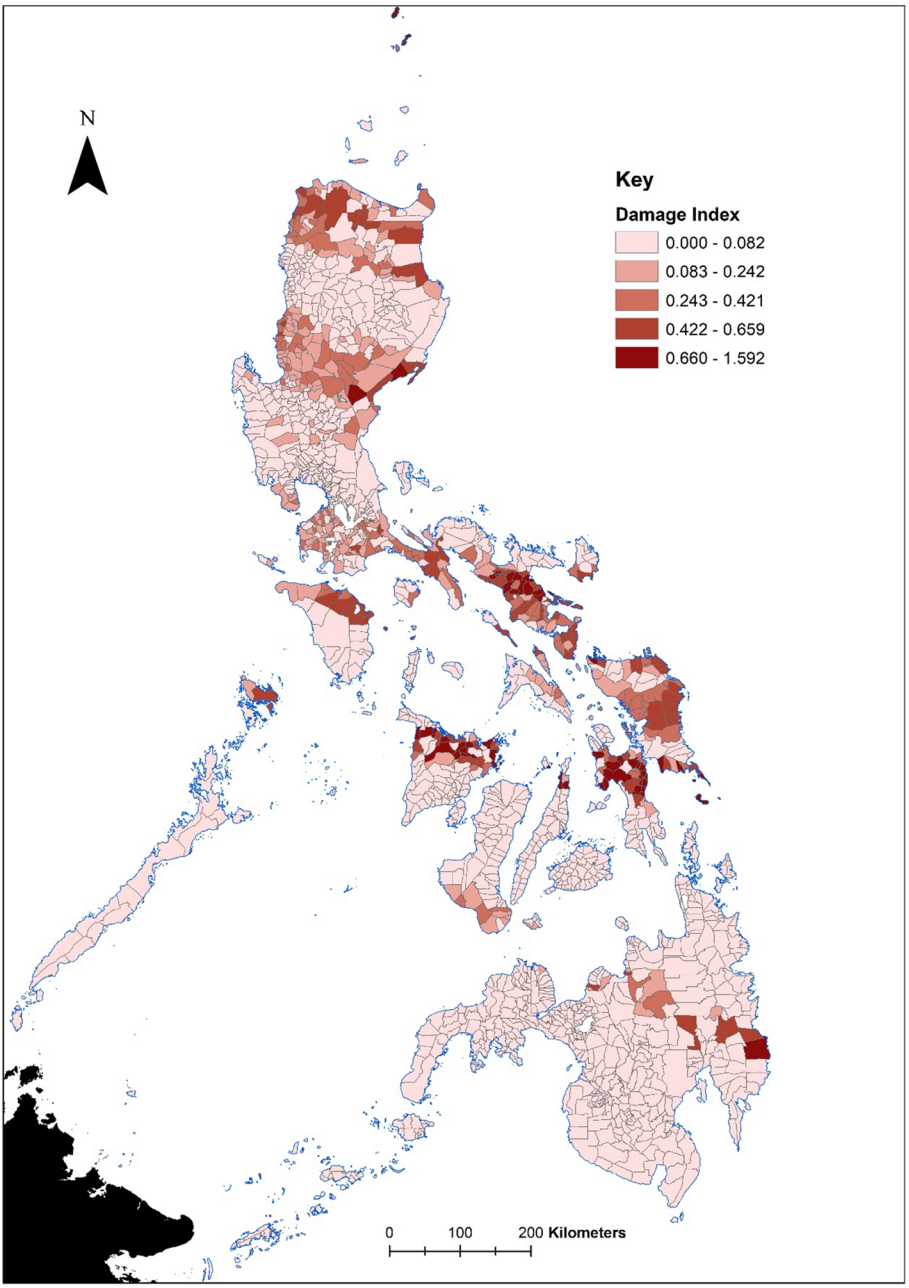


Fig. 4 Damage index derived from typhoon paths, 2012–2019

rather than 70 percentiles) because the blurred DMSP images attributed some lights from the industrial zone to the adjacent district. This flaw in these data carried over into difference-in-differences evaluation of the impact of the closure; VIIRS data yielded a precisely

estimated 50% fall in luminosity after closure of the industrial zone (using the other district as the control group) while DMSP data implausibly implied that closing the industrial zone caused luminosity to increase due to some of the fall in luminosity from the treated district being wrongly attributed to the comparison district.

An important aspect of studying these potential econometric biases is to use models that allow for the spatial autocorrelation that is a likely feature of natural hazards (whereas economic shocks, such as closing an industrial zone, can be highly localized). In Fig. 4 we show a damage index calculated at the municipality level from expected damage estimates for each year from 2012 to 2019, where these are derived from the storm track of typhoons.[10] There are clear bands of expected damage, lying on an east-to-west orientation with a slight tilt towards the northwest; these bands are especially apparent in northern Luzon, but also in the Eastern Visayas through the Bicol Peninsula and up towards Manila. While Mindanao is typically less affected, there is one damage band going across Mindano and into the southern end of Negros island. With these clear spatial patterns, the usual assumption that regression errors for each observation are independent of errors for neighbouring observations may not be expected to hold. Moreover, in addition to the unobserved components (i.e. the errors) of changes in economic activity being spatially correlated there are also likely to be spillovers from the fact that municipalities are engaged in economic interactions with their neighbours. For these reasons, our modelling strategy starts with the most general spatial econometrics model that can allow for both local and global spillovers, and then imposes restrictions to consider various special cases, including the usual aspatial panel data model.

Data and Methods

In order to examine the sensitivity of estimated disaster impacts to the use of different NTL data we need at least two sources of data on night-time lights and also estimates of exposure to disasters. We use expected typhoon damages in the Philippines from 2012 to 2019 as our case study and primarily rely upon two sources of annual NTL data: DMSP stable lights and VIIRS night lights (VNL) version 2.1 masked average radiances.[11] Key references on how these annual composites of the night-time lights data are formed are Baugh et al. (2010) and Ghosh et al. (2021) for DMSP and Elvidge et al. (2017, 2021) for VIIRS.

The DMSP data are 6-bit digital numbers (DN) ranging from 0 to 63 (lower values for less brightly lit areas), for each 30 arc-second output grid (roughly 0.9×0.9 km at the latitude of the Philippines). In the annual composites, 'stable' simply means that ephemeral lights, from sources like fires and gas flaring, are removed before the annual composite is built from nightly images. An alternative meaning of 'stable', in terms of temporal

[10] We provide more details on the construction of this index in Section VI. Briefly, it is based on *ex post* best estimates of the storm track of each typhoon, from a post-season re-analysis of data pooled from sources such as satellites, ships, aviation and surface measurements. The estimates of the wind speed from each typhoon in each location then allow the expected degree of damage to be estimated.

[11] The DMSP data are available from: https://eogdata.mines.edu/products/dmsp/#download and the VIIRS data from https://eogdata.mines.edu/nighttime_light/annual/v21/. The DMSP data are from satellite F18 for 2012–13 and from F15 from 2014 onwards. Comparisons between various VIIRS data products described by Elvidge et al. (2021) show the masked average radiance annual composite used here is the best proxy for economic activity, using a panel of United States county-level and state-level GDP as a benchmark (Gibson and Boe-Gibson 2021).

consistency, does not apply as there is no in-built calibration of the DMSP data to adjust for the constant changes in sensor amplification (that are made without any record kept, where amplification changes are so cloud-tops can be viewed with similar brightness across light and dark parts of the lunar cycle). Consequently, it is best to think of the DN value as a relative measure of brightness, because the same DN value in different years could correspond to different radiance values (Doll 2008). While there are radiance-calibrated DMSP data (Elvidge et al. 1999), these are not available since 2010, and so cannot be used for comparing with the VIIRS data.

The VNL data are version 2.1 annual composites from 2012 to 2019, from Elvidge et al. (2021), derived from monthly cloud-free radiance averages from the Suomi/NPP satellite. These data undergo an initial filtering to remove extraneous features such as fires and aurora before the resulting rough annual composites are subjected to further outlier removal procedures. The lit grid cells are isolated from background noise using thresholds that apply across years, making these data better for change detection than earlier vintages of the VIIRS annual composites (that used year-specific thresholds). The data are in units of nano Watts per square centimeter per steradian ($nW/cm^2/sr$) presented on a 15 arc-second output grid (which is equivalent to 0.47×0.45 km at the latitude of the Philippines).

Our expected damage index from typhoons is for all municipalities in the Philippines, over the period from 2012 to 2019. We use the same approach as described in Strobl (2019), except that rather than calculating at the municipality level Strobl used DMSP output grid cells (of 30 arc-seconds) as the spatial units and 1992–2013 as the study period. Our concern with a grid-cell level analysis is that DMSP sensor resolution is far coarser than the output grid (at best, having a ground footprint of 25 km^2 at nadir) and so the combination of a coarse sensor and a finer (resampled) output grid is likely to create spurious spatial autocorrelation in the luminosity estimates for grid cells. In contrast, with a mean area for the municipalities of 184 km^2 (which is far larger than the sensor resolution) this concern is mitigated.

The original source of data used to calculate the expected damage index is the Best Track Archive for Climate Stewardship (IBTrACS) provided by the National Oceanic and Atmospheric Administration.[12] It has details on the geospatial position, maximum sustained wind speed, radius of maximum winds, and other characteristics at 6-h frequency for each storm. The calculation involves two steps. First, a wind field model developed in Holland (1980) and updated in Boose et al. (2004) is used to estimate the wind speed experienced in a location at a particular time. It is a function of maximum sustained wind velocity anywhere in the typhoon, the radius of maximum winds, the distance from the centre of the storm to the point of interest, the forward velocity of the storm, the clockwise angle between the forward path of the storm and a radial line from the storm centre to the point of interest, and the gust factor.

Second, we apply the damage function that is proposed in Emanuel (2011), which links the maximum wind speed from a storm experienced at point of interest, above a certain threshold, to the expected damage. Following Strobl (2019), we assume that maximum wind speed below 92 km/hr would cause no damage to the location. This step yields a damage index between 0 and 1. For a municipality that is hit by several storms in a single year, we take the maximum of the multiple estimates of expected damages. For more

[12] The data are available from: https://www.ncei.noaa.gov/products/international-best-track-archive

details on the formula, data sources and parameters used in our calculations, please refer to Appendix A.

Estimation Framework

Spatial econometric methods allow for non-randomness in space of phenomena and account for possible spillovers. A key element making these models feasible is that possible interactions between spatial units are summarized with a $N \times N$ spatial weights matrix, W. In this study we use a row normalized contiguity weights matrix with ones for neighbours and zero otherwise, with a diagonal of zeros because a municipality cannot neighbour itself. The average municipality in the Philippines has five neighbours.

In what follows, luminosity in municipality i in year t is denoted as L_{it}, the expected damages are D_{it}, the μ_i are time-invariant fixed effects for each municipality, the ϑ_t are year fixed effects, and e_{it} is a random error. By using the spatial weights matrix we can allow for spatial lags, which are averages of these variables over the neighbouring units. Thus, with the addition of a few extra variables (the spatial lags) it becomes possible to summarize effects of many complicated interactions (if each potential spillover was to be separately estimated the number of parameters would exceed the number of observations).

Our starting point is the very general spatial autoregressive model with spatial autoregressive errors (SARAR).[13] This model allows for spatially lagged dependent variables, spatially lagged independent variables, and spatially lagged errors:

$$L_{it} = \lambda W L_{it} + \beta_1 D_{it} + \beta_2 W D_{it} + \mu_i + \vartheta_t + \rho W u_{it} + e_{it} \tag{1}$$

The SARAR model allows for changes in an outcome variable in a given area to have effects on contemporaneous outcomes in other areas (via the autoregressive spatial lag of the dependent variable, if $\lambda \neq 0$). It also allows changes in independent variables to affect not only own-area outcomes but also outcomes in neighbouring areas (if $\beta_2 \neq 0$). The $\rho W u_{it}$ term allows for spatial autocorrelation, where errors for a given municipality correlate (ρ) with a weighted average of the errors from surrounding municipalities. Equation (1) nests a spatial Durbin model if $\rho = 0$, a spatial auto-regressive model (*aka* spatial lag model) where only the dependent variable is spatially lagged if $\beta_2 = \rho = 0$, a spatial error model where only the errors are spatially lagged (if $\lambda = \beta_2 = 0$), and the most restrictive of all, which is an aspatial model with no spatial lags (if $\lambda = \beta_2 = \rho = 0$). Another feature of Eq. (1) is that including municipality fixed effects provides a 'within' estimator interpretation, which is the appropriate framework for considering how *changes* in luminosity relate to the expected damages index (Zhang and Gibson 2022).

An important feature of spatial econometric models is that lags of either the outcome variable or of independent variables (but not of errors) mean that total effects of changes in an independent variable—e.g. higher expected damages—may be quite different to what the usual regression coefficient shows. While $\widehat{\beta_1}$ is the object of interest in the typical model without spatial lags, in the spatial models when either the spatial lags of outcomes or the spatial lags of independent variables are non-zero then $\widehat{\beta_1}$ does not capture the total effect of a change in the independent variable. A useful decomposition of the more complex spatial relationships that occur relies on rewriting Eq. (1) in matrix notation (for simplicity,

[13] This is estimated by quasi-maximum likelihood (Lee and Yu 2010), using the Drukker et al. (2013) estimator.

Table 2 Descriptive statistics for expected damages and log luminosity

	All municipalities		Municipalities ever damaged		Municipalities never damaged	
	Mean	Std Deviation	Mean	Std Deviation	Mean	Std Deviation
Expected damages index	0.014	0.077	0.030	0.110	n.a	n.a
log (sum DMSP DN values)	4.218	2.759	4.561	2.592	3.918	2.864
log (sum VIIRS radiances)	3.943	2.084	4.184	1.948	3.732	2.175

Notes: There are 12,448 observations, with 5816 for the municipalities that ever had expected typhoon damages greater than zero (over the 2012–19 period) and 6636 for the municipalities that never had expected damages

subscripts are dropped and fixed effects and error terms combined in v because the errors do not affect this decomposition) as:

$$L = (I - \lambda W)^{-1}\left(D\beta_1 + WD\beta_2\right) + (I - \lambda W)^{-1}v \tag{2}$$

Following Elhorst (2012), the $N \times N$ matrix of partial derivatives can be written (noting that diagonal elements of W are zero) as:

$$\frac{\partial L}{\partial D_k} = (I - \lambda W)^{-1}\left(\beta_{1k} I_N + \beta_{2k} W\right) \tag{3}$$

where D_k is the damage index in municipality k. The total marginal effect on luminosity from expected damages has two components, a direct one and an indirect one, that may both vary over space. The estimator that we use follows LeSage and Pace (2009) in reporting a single direct effect, that averages the diagonal elements of the matrix in (3) and a single indirect effect that averages the row sums of the non-diagonal elements of that matrix. Indirect effects arise not just from adjacent area units, if $\beta_{2k} \neq 0$, but also from (potentially) all areas through the spatial autoregressive effect if $\lambda \neq 0$. Thus, there can be both local and global spillovers and when these are accounted for, averages from the matrix of derivatives $\partial L/\partial D_k$ may be quite different to the estimated direct impact effect, $\widehat{\beta}_1$.

Results

In order to the use the data described in Section IV some transformations are needed because the units of DMSP (digital numbers) and of VNL (radiances) are not comparable. We take the logarithm of the sum of lights in each municipality in each year, using the inverse-hyperbolic sine (IHS) transformation to allow for the inclusion of any zeros.[14] The impact estimates can then be interpreted as approximate percentage changes in economic activity (to the extent that this change is proxied by changes in luminosity) for the damages index changing from 0 to 1. The descriptive statistics in Table 2 show that such a movement is more than ten standard deviations, if all municipalities are considered, and

[14] Prior studies with NTL data have shown that for non-zero observations the IHS transformation gives the same results as using logarithms (Gibson et al. 2017), and that ad hoc transformations such as adding a small constant before taking the logarithm do not adequately account for the zeros (Kim et al. 2024).

Table 3 Typhoon damage effects on luminosity (log DMSP): Municipalities in the Philippines, 2012 to 2019

	(1)	(2)	(3)	(4)	(5)
	Spatial lag of errors, covariate and outcome	Spatial lag of the covariate and outcome	Spatial lag of the outcome	Spatial lag of the errors	Standard panel model analysis
Damage	-0.094 (0.177)	-0.102 (0.174)	-0.298*** (0.115)	-0.303** (0.144)	-0.448*** (0.123)
Average impacts:					
Direct	-0.148 (0.163)	-0.138 (0.165)	-0.309*** (0.119)	-0.303** (0.144)	n.a
Indirect	-0.578* (0.312)	-0.525** (0.250)	-0.165** (0.064)	n.a	n.a
Total	-0.726** (0.286)	-0.663*** (0.222)	-0.474*** (0.183)	-0.303** (0.144)	n.a
County fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Spatial lag: upgraded	Yes	Yes	No	No	No
Spatial lag: output	Yes	Yes	Yes	No	No
Spatial lag: errors	Yes	No	No	Yes	No
All covariates = 0	$\chi^2 = 8304$***	$\chi^2 = 2389$***	$\chi^2 = 2387$***	$\chi^2 = 428$***	$\chi^2 = 969$***
Nesting restrictions	n.a	$\chi^2 = 739$***	$\chi^2 = 740$***	$\chi^2 = 4788$***	$\chi^2 = 6411$***

The sample period is 2012–2019, for 1556 municipality-level units, giving an estimation sample of n = 12448. Coefficients for the fixed effects and the spatial lags are not reported. The decomposition of average impacts into direct, indirect and total components is based on LeSage and Pace (2009). The nesting restrictions are imposed on the SARAR model in column (1) to derive the models in columns (2) to (5) Standard errors are in (), with statistical significance at the 1%, 5% and 10% level denoted by ***, **, *

Table 4 Typhoon damage effects on luminosity (log VIIRS night lights): Municipalities in the Philippines, 2012 to 2019

	(1)	(2)	(3)	(4)	(5)
	Spatial lag of errors, covariate and outcome	Spatial lag of the covariate and outcome	Spatial lag of the outcome	Spatial lag of the errors	Standard panel model analysis
Damage	-0.098 (0.067)	-0.026 (0.076)	-0.213*** (0.050)	-0.197** (0.063)	-0.336*** (0.053)
Average impacts:					
Direct	-0.053 (0.070)	-0.057 (0.072)	-0.221*** (0.052)	-0.197** (0.063)	n.a
Indirect	-0.379*** (0.096)	-0.458*** (0.108)	-0.116*** (0.027)	n.a	n.a
Total	-0.431*** (0.093)	-0.516*** (0.095)	-0.337*** (0.078)	-0.197** (0.063)	n.a
County fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Spatial lag: upgraded	Yes	Yes	No	No	No
Spatial lag: output	Yes	Yes	Yes	No	No
Spatial lag: errors	Yes	No	No	Yes	No
All covariates = 0	χ^2 = 2213***	χ^2 = 11,735***	χ^2 = 11,716***	χ^2 = 4213***	χ^2 = 9206***
Nesting restrictions	n.a	χ^2 = 2147***	χ^2 = 2161***	χ^2 = 301***	χ^2 = 3803***

The sample period is 2012–2019, for 1556 municipality-level units, giving an estimation sample of n = 12,448. Coefficients for the fixed effects and the spatial lags are not reported. The decomposition of average impacts into direct, indirect and total components is based on LeSage and Pace (2009). The nesting restrictions are imposed on the SARAR model in column (1) to derive the models in columns (2) to (5) Standard errors are in (), with statistical significance at the 1%, 5% and 10% level denoted by ***, **, *

nine standard deviations if attention is restricted just to the municipalities (47% of the total) where expected damages were greater than zero over 2012–19. Hence, a more useful interpretation may be to consider one-tenth of the estimated impact, corresponding to an approximately one standard deviation change in the damage index. The descriptive statistics also show that the municipalities that ever faced typhoon damage are more economically active, with about 15% higher luminosity than the never damaged municipalities.

The results of estimating Eq. (1) and then imposing the various parameter restrictions and estimating the nested models are given in Table 3 (for DMSP) and Table 4 (for VNL). The nesting restrictions are rejected in all cases, so that the SARAR models appear to be the most data-acceptable models in both cases. In other words, the spatial Durbin model, the spatial lag model, the spatial error model, and the aspatial model that does not have any lags all rely on restrictions that are rejected by the data. The discussion therefore concentrates mostly on the results in column (1) for the SARAR model, which allows for spatial lags of the outcome (luminosity as a proxy for economic activity), of the independent treatment (the expected typhoon damages), and of the errors. Some attention is also paid to the column (5) results, for the typical panel data model with area and time period fixed effects that does not have any spatial lags because models of this type are frequently found in the literature using NTL data to estimate disaster impacts.

Given the statistical relevance of the spatial lags of the outcomes and of the treatment, it is important to note that the usual regression coefficients do not tell the full story and so the matrices of marginal effects based on Eq. (3) need to be used. The results of these marginal effects calculations are reported in the "average impacts" rows of Tables 3 and 4, using the decomposition due to LeSage and Pace (2009). As a preliminary comparison, the "total" row of Table 3 that is based on using the DMSP NTL data is almost 70% larger than the corresponding value in Table 4 using VNL data, for the SARAR model that is the most data acceptable of the five models. This same pattern, of larger negative average total impacts when using DMSP data, holds across all five models and so even if a researcher approached these data without using the most general spatial econometric models they would find that the DMSP data was yielding apparently larger negative impacts from the typhoon damages.

This difference in the apparent impacts is more plausibly attributed to overstatement when using the DMSP data rather than understatement when using VNL data, although we cannot be too firm in this claim due to the lack an independent benchmark estimate of actual changes in economic activity.[15] Nevertheless, the VIIRS sensors have better properties and the prior evidence (including in settings with benchmarks such as county-level GDP) also is consistent with the claim that the VNL data are more accurate and precise and give results that are closer to the truth (Zhang and Gibson 2022; Zhang et al. 2023). The exaggerated spatial autocorrelation shown by the DMSP data in Fig. 1 provides one mechanism for the spatial impacts of a shock to apparently spread more widely. In this regard, it is especially notable that the indirect impacts exceed the direct impacts, except when constrained models (as in column (3)) are used that a priori rule out including spatial lags of the damages index. The map of the damages index patterns in Fig. 4 suggests that these spatial lags are likely to be important, given that they are based on the typhoon tracks, and so for disasters with this type of damages pattern, the exaggerated spatial autocorrelation

[15] The imprecision in the estimates using DMSP data also hinders comparisons, and this has been observed in other exercises that compare impact estimates derived from DMSP and VNL data (Kim et al. 2024).

in luminosity seen with the DMSP data may prove to be especially distorting to the impact estimates.

In terms of the magnitude of the estimated impacts, using the results in Table 4, for a standard deviation increase in the expected damages index, VNL luminosity will decline by about four percent according to the SARAR model. If one used the aspatial model, the fall in luminosity would be about three percent. While the first of these estimates allows for spatial spillovers neither of them allow for temporal lags. In the previous results for the Philippines reported by Strobl (2019) the temporal lags were all imprecisely estimated and only the contemporaneous effects on luminosity in the same year as the expected damages showed up as statistically significant.[16] We therefore have concentrated on spatial effects, which are especially salient in light of the exaggerated spatial autocorrelation shown with the DMSP data (and with VIIRS data that have been adjusted to be like DMSP).

Discussion and Conclusions

In this study we have provided some evidence on the sensitivity of estimated disaster impacts to the use of different night-time lights data. Although the newer, more accurate and precise, VIIRS data have been available for over a decade, there is still some ongoing use of the older and less suitable DMSP data in the disasters literature. Moreover, the attempt to create longer time-series with various 'harmonized' data that splice together DMSP data and VIIRS data also makes some of the features (or flaws) of DMSP data relevant in the current era because these harmonized data inherently involved degrading the more precise VIIRS data. We show one aspect of this by using spatial autocorrelation statistics where the harmonized VIIRS data are very different than the original VNL data (with a Moran's *I* statistic that is 77% higher), and are much more like DMSP data in showing a high degree of similarity of adjacent areas. In other words, the harmonized data give the appearance of greater spatial autocorrelation.

In our case study of typhoons in the Philippines over the 2012–19 period we found that the estimated negative impacts on luminosity were larger when using DMSP data than when using VNL data for the same period. These damages were especially expressed through local spillovers from spatial lags of the covariates, which is consistent with damages that are derived from the storm path of typhoons. In this case, the tendency of DMSP data to overstate the degree of spatial autocorrelation, which is the flipside of these data understating local inequality, seems to cause estimated impacts to spread. We also use imagery from selected areas of the Philippines which provides examples of this blurring of lights from the highly concentrated sources such as cities, and these blurring patterns have also been shown in other countries. While more evidence is needed, to the extent that it is possible to study disasters in other settings for years where there is overlapping availability of different sources of night-time lights data, it would be useful to repeat the comparisons carried out here.

We focussed on the blurring and spatial autocorrelation aspects of night-time lights data because damages from natural hazards like typhoons are often spatially autocorrelated. Other aspects of different types of night-time lights data, including overpass time,

[16] Annual composites may not be the best indicator if lights can be re-established just hours or weeks following a disaster. In that case, higher frequency (e.g. nightly) luminosity data when used in an event study framework might be a better way to identify the effects of natural hazards.

also likely affect usefulness for impact evaluations. While DMSP is usually described as observing earth between 8.30-10pm, when much consumption occurs, unstable satellite orbits saw overpass times become progressively earlier as satellites aged (even as early as late afternoon). Hence, to extend the DMSP time-series, Ghosh et al. (2021) resorted to using data from overpasses that were 12 h earlier, effectively giving a reading at about 3.30am—so the same activity that is observed by VIIRS at about 1.30am should be seen. Thus, our comparisons should not be affected by time-of-observation issues. However, a future study might use data from new satellites, such as China's SDGSAT-1 (https://sdg.casearth.cn/en/datas/SDGSAT) that have a more favourable overpass time of ca. 9pm and have spatial resolution that is about 20-times finer than for VIIRS.

Of course our comparisons are less informative than is ideal because we do not have a true measure of on-the-ground damages. We also lack conventional subnational economic data, such as GDP, at municipality level in the Philippines. Two options for future study of these issues would be to either focus on a highly insured setting, where individual claims data might provide a measure of actual damages (but behaviour might differ from the less-insured settings of interest in developing countries), or on a setting with benchmark economic data at a sufficiently disaggregated level. For example, China has GDP estimates for county-level units (the third sub-national level), and for a province like Guangdong that is frequently hit by typhoons, third-level units are less than an order of magnitude larger than municipalities in the Philippines. However, these economic data are only available annually and in places that are able to recover rapidly from natural hazard damages an annual timescale may not be fine enough, and higher frequency daily or weekly data might be better.

Appendix A: Details on the Damage Index Calculations (Data and Parameters)

1. Data Definitions

Variable Name	Units	Description
Sid		Storm identifier
Year	year	
Number		Cardinal number of the system for that season (year)
Basin		Western North Pacific
Name		Name provided by the agency
iso_time		Format is YYYY-MM-DD. Most points are in 6-h intervals and some are in 3 h
Nature		Combined storm type This is assigned based on all available storm types. They include: DS—Disturbance TS—Tropical ET—Extratropical SS—Subtropical NR—Not reported MX—Mixture (contradicting nature reports from different agencies)
Lat	Deg north	Latitude of storm center
Lon	Deg east	Longitude of storm center

Variable Name	Units	Description
dist2land	Km	Distance to land from the current position
Landfall	Km	Nearest location to land within next 6 h This can be thought of a landfall flag: =0 – Landfall within 6 h >0 – No landfall within next 6 h *Calculations are based on storm center (columns 9,10). Values less than 60 nmile likely are impacted by the system even though the center of the system is not over land
usa_wind	knots	Maximum sustained wind speed in knots: 0—300 kts
usa_pres	mb	Minimum sea level pressure, 850—1050 mb
usa_sshs		Saffir-Simpson Hurricane Scale information based on the wind speed provided by the US agency wind speed (US agencies provide 1-min wind speeds) -5 = Unknown -4 = Post-tropical -3 = Miscellaneous disturbances -2 = Subtropical Tropical systems classified based on wind speeds -1 = Tropical depression (W < 34) 0 = Tropical storm [34 < W < 64] 1 = Category 1 [64 < = W < 83] 2 = Category 2 [83 < = W < 96] 3 = Category 3 [96 < = W < 113] 4 = Category 4 [113 < = W < 137] 5 = Category 5 [W > = 137]
usa_poci	mb	pressure in millibars of the last closed isobar, 900—1050 mb (not reanalyzed or not Best Tracked)
usa_roci	N. miles	radius of the last closed isobar, 0—999 n miles (not reanalyzed or not Best Tracked)
usa_rmw	N. miles	radius of max winds, 0—999 n miles (not reanalyzed or not Best Tracked)
tokyo_r50_long	N. miles	The longest radius of 50kt winds or greater
tokyo_r50_short	N. miles	The shortest radius of 50kt winds or greater
tokyo_r30_long	N. miles	The longest radius of 30kt winds or greater
tokyo_r30_short	N. miles	The shortest radius of 30kt winds or greater
storm_speed	knots	Translation speed of the system as calculated from the positions in LAT and LON
storm_dir	degrees	Translation direction of the system as calculated from the positions in LAT and LON. Direction is moving toward the vector pointing in degrees east of north [range = 0–360 deg]

2. Wind Field Model

To estimate the local winds experienced by a municipality we estimate V_{ijt} which is the wind experienced in point i (municipality) amidst typhoon j, at time t. The wind field model takes the following form:

$$V_{ijt} = GF\left[Vm_{jt} - S\left(1 - sin\left(T_{ijt}\right)\right)\frac{Vh_{jt}}{2}\right]\left[\left(\frac{Rm_{jt}}{R_{it}}\right)^{B_{jt}} exp\left(1 - \left[\frac{Rm_{jt}}{R_{it}}\right]^{B_{jt}}\right)\right]^{1/2}$$

Our data fills in some of the necessary parameters. However, there are some parameters that will be assumed (based on previous literature) or estimated.

Data var	Parameter	Description
	V_{ijt}	wind experienced at time t due to typhoon j at any point (P = i)
usa_wind	Vm	maximum wind speed
storm_dir*	T	clockwise angle between the typhoon's forward path and radial line between typhoon center and point of interest (P = i)
storm_speed	Vh	forward velocity
usa_rmw	Rm	radius of maximum winds
estimated	R	radial distance from typhoon center to point (P = i)
Assumed	G	gust factor
Assumed	F	surface friction
Assumed	S	asymmetry due to forward motion of the storm
Estimated	B	shape of wind profile curve

How each factor will be estimated or assumed:

- R and T is calculated factoring in the location of municipality centroid or P = i
- G (gust) is assumed to be 1.5 (Paulsen and Schroeder 2005)
- F (surface friction) linearly decreases further inland

Vickery et al.. (2009) suggest a reduction factor of 0.7, thus a 14% reduction on the coast and 28% reduction 50km inland.

- S (asymmetry) is assumed to be 1 (Boose et al.. 2004)
- B (wind profile) is estimated using approximation method by Holland (2008);

$$B = b_s\left(\frac{v_{mg}}{v_m}\right)^2 \approx 1.5b_s$$

$\frac{V_{mg}}{V_m}$ is a conversion factor from gradient to surface wind

$b_s = -4.4x10^{-5}\Delta p^2 + 0.01\Delta p + 0.03\frac{\partial p_c}{\partial t} - 0.014\psi + 0.15V_T^x + 1.0$

$x = 0.6\left(1 - \frac{\Delta p}{215}\right)$

Δp is the pressure drop from the typhoon center.

$\frac{\partial p_c}{\partial t}$ is the intensity change

ψ is the absolute value of latitude

V_t is the typhoon transition speed

- Rm is given by variable usa_rmw[17] but can also be estimated with a parametric model Xiao et al.. (2009)

$$lnR_m = 5.3259 + (-0.0249)\Delta p - 0.0161\psi$$

3. Damage Function

After estimating local wind experience, V_{ijt}, we can estimate the damage function. Typhoon damage comes in three forms: (1) wind destruction, (2) flooding, and (3) storm surge. Since these factors are all correlated, wind speed can be used as a general proxy for potential damage.

To estimate potential damage, index by Emanuel (2011) is used.[18]

$$f_{ijt} = \frac{v_{n,ijt}^3}{1 + v_{n,ijt}^3}$$

where,

$$v_{n,ijt} = \frac{MAX\left[\left(V_{ijt} - V_{thresh}\right), 0\right]}{V_{half} - V_{thresh}}$$

On top of the local wind estimation, we need to establish wind thresholds, details of each factor outlined below:

- v_{ijt} is wind experienced at time t due to typhoon j at any point (P = i)
- V_{thresh} is the threshold below which point i does not incur damage.

  Emanuel (2011) uses 50kts

- V_{half} is the threshold where half the property in point i is damaged.

  Emanuel (2011) uses 150kts

Acknowledgements Gibson, Zhang, and Boe-Gibson are all with the University of Waikato, Yi is with the Asian Development Bank. We are grateful for helpful comments from two anonymous reviewers and from participants at the ADB/IRIDeS Conference on Big Data for Disaster Response and Management in Asia and the Pacific.

Author Contributions JG developed the study and wrote the main manuscript and handled the revisions. YJ calculated the damages estimates and reviewed the manuscript. XZ and GBG handled the GIS and Remote Sensing analyses. GBG prepared the figures. XZ prepared the tables. All authors reviewed the manuscript.

[17] The variable usa_rmw is not 'best tracked' or reanalyzed. The parametric model may be used as a robustness check.

[18] In earlier research the expected damage proxy was estimated as the cubic power of wind speed. However, this does not consider that there is likely a wind speed level below which no property damage occurs. To deal with this issue Emanuel (2011) creates a threshold-based index.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. Funding was received from the Asian Development Bank.

Data Availability The datasets generated during and/or analysed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Declarations

Competing Interests All authors declare that they have no relevant financial or non-financial interests to disclose.

References

Abrahams A, Oram C, Lozano-Gracia N (2018) Deblurring DMSP nighttime lights: A new method using Gaussian filters and frequencies of illumination. Remote Sens Environ 210:242–258

Akter S (2023) Australia's Black Summer wildfires recovery: A difference-in-differences analysis using nightlights. Glob Environ Chang 83:102743

Baugh K, Elvidge C, Ghosh T, Ziskin D (2010) Development of a 2009 stable lights product using DMSP-OLS data. Proceed Asia-Pacific Adv Network 30:114

Bluhm R, Krause M (2022) Top lights: Bright cities and their contribution to economic development. J Dev Econ 157:102880

Boose E, Serrano M, Foster D (2004) Landscape and regional impacts of hurricanes in Puerto Rico. Ecol Monogr 74(2):335–352

Brei M, Mohan P, Strobl E (2019) The impact of natural disasters on the banking sector: Evidence from hurricane strikes in the Caribbean. Q Rev Econ Finance 72:232–239

Carballo Chanfón P, Mohan P, Strobl E, Tveit T (2023) The impact of hurricane strikes on cruise ship and airplane tourist arrivals in the Caribbean. Tour Econ 29(1):68–91

Chen X, Nordhaus W (2019) VIIRS nighttime lights in the estimation of cross-sectional and time-series GDP. Remote Sens 11(9):1057

Doll C (2008) CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications. Columbia University, New York, Center for International Earth Science Information Network

Drukker D, Prucha I, Raciborski R (2013) Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. Stand Genomic Sci 13(2):221–241

Elhorst JP (2012) Dynamic spatial panels: models, methods, and inferences. J Geogr Syst 14(1):5–28

Elliott R, Strobl E, Sun P (2015) The local impact of typhoons on economic activity in China: A view from outer space. J Urban Econ 88:50–66

Elvidge C, Baugh K, Dietz J, Bland T, Sutton P, Kroehl H (1999) Radiance calibration of DMSP-OLS low-light imaging data of human settlements. Remote Sens Environ 68(1):77–88

Elvidge C, Baugh K, Zhizhin M, Hsu FC (2013) Why VIIRS data are superior to DMSP for mapping night-time lights. Proceed Asia-Pacific Adv Network 35:62

Elvidge C, Baugh K, Zhizhin M, Hsu FC, Ghosh T (2017) VIIRS night-time lights. Int J Remote Sens 38(21):5860–5879

Elvidge C, Zhizhin M, Ghosh T, Hsu F, Taneja J (2021) Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. Remote Sens 13(5):922

Emanuel K (2011) Global warming effects on US hurricane damage. Weather, Climate, Soc 3(4):261–268

Espagne E, Ha YB, Houngbedji K, Ngo-Duc T (2022) Effect of typhoons on economic activities in Vietnam: evidence using satellite imagery. AFD Research Papers (263):1–18

Fabian M, Lessmann C, Sofke T (2019) Natural disasters and regional development–the case of earthquakes. Environ Dev Econ 24(5):479–505
Farzanegan M, Fischer S (2023) The impact of a large-scale natural disaster on local economic activity: evidence from the 2003 bam earthquake in Iran. Munich, Center for Economic Studies and ifo Institute (CESifo). Working Paper No. 10502. https://ssrn.com/abstract=4477990
Feeny S, Trinh T, De Silva A (2022) Detecting disasters and disaster recovery in Southeast Asia: Findings from space. Nat Hazard Rev 23(2):04021065
Felbermayr G, Gröschl J, Sanders M, Schippers V, Steinwachs T (2022) The economic impact of weather anomalies. World Dev 151:105745
Ghosh T, Baugh K, Elvidge C, Zhizhin M, Poyda A, Hsu FC (2021) Extending the DMSP nighttime lights time series beyond 2013. Remote Sens 13(24):5004
Gibson J (2021) Better night lights data, for longer. Oxford Bull Econ Stat 83(3):770–791
Gibson J, Boe-Gibson G (2021) Nighttime lights and county-level economic activity in the United States: 2001 to 2019. Remote Sens 13(14):2741
Gibson J, Datt G, Murgai R, Ravallion M (2017) For India's rural poor, growing towns matter more than growing cities. World Dev 98:413–429
Gibson J, Olivia S, Boe-Gibson G, Li C (2021) Which night lights data should we use in economics, and where? J Dev Econ 149:102602
Goldblatt R, Heilmann K, Vaizman Y (2020) Can medium-resolution satellite imagery measure economic activity at small geographies? Evidence from Landsat in Vietnam. World Bank Econ Rev 34(3):635–653
González FAI, London S, Santos ME (2021) Disasters and economic growth: evidence for Argentina. Climate Dev 13(10):932–943
Henderson M, Yeh E, Gong P, Elvidge C, Baugh K (2003) Validation of urban boundaries derived from global night-time satellite imagery. Int J Remote Sens 24(3):595–609
Holland G (1980) An analytic model of the wind and pressure profiles in hurricanes. Mon Weather Rev 108(8):1212–1218
Holland G (2008) A revised hurricane pressure–wind model. Mon Weather Rev 136(9):3432–3445
Ishizawa O, Miranda J, Strobl E (2019) The impact of hurricane strikes on short-term local economic activity: Evidence from nightlight images in the Dominican Republic. Int J Disaster Risk Sci 10:362–370
Joseph IL (2022) The effect of natural disaster on economic growth: Evidence from a major earthquake in Haiti. World Dev 159:106053
Kim B, Gibson J, Boe-Gibson B (2024) Measurement errors in popular night lights data may bias estimated impacts of economic sanctions: Evidence from closing the Kaesong Industrial Zone. Econ Inq 62(1):375–389
Kocornik-Mina A, McDermott T, Michaels G, Rauch F (2020) Flooded cities. Am Econ J Appl Econ 12(2):35–66
Lee LF, Yu J (2010) Estimation of spatial autoregressive panel data models with fixed effects. J Econom 154(2):165–185
LeSage J, Pace RK (2009) Introduction to spatial econometrics. Chapman and hall/CRC. http://journals.openedition.org/rei/3887
Li X, Zhou Y, Zhao M, Zhao X (2020) A harmonized global nighttime light dataset 1992–2018. Sci Data 7(1):168
Li X, Liu Z, Chen X, Meng Q (2019) Assessment of the impact of the 2010 Haiti earthquake on human activity based on DMSP/OLS time series nighttime light data. J Appl Remote Sens 13(4):044515–044515
Liu Z, Zhang J, Li X, Chen X (2021) Long-term resilience curve analysis of Wenchuan earthquake-affected counties using DMSP-OLS nighttime light images. IEEE J Sel Top Appl Earth Observations Remote Sensing 14:10854–10874
Mathen C, Chattopadhyay S, Sahu S, Mukherjee A (2024) Which nighttime lights data better represent India's economic activities and regional inequality? https://ssrn.com/abstract=4590715
Miranda J, Ishizawa O, Zhang H (2020) Understanding the impact dynamics of windstorms on short-term economic activity from night lights in Central America. Econ Dis Cli Cha 4:657–698
Miranda J, Butron L, Pantoja C, Gunasekera R (2021) Mangroves as coastal protection for local economic activities from hurricanes in the Caribbean. Policy Research Working Paper No. 9863, The World Bank. https://ideas.repec.org/p/wbk/wbrwps/9863.html
Mohan P, Strobl E (2021) The impact of tropical storms on tax revenue. J Int Dev 33(3):472–489
Nechaev D, Zhizhin M, Poyda A, Ghosh T, Hsu FC, Elvidge C (2021) Cross-sensor nighttime lights image calibration for DMSP/OLS and SNPP/VIIRS with residual U-net. Remote Sens 13(24):5026
Nguyen CN, Noy I (2020) Measuring the impact of insurance on urban earthquake recovery using nightlights. J Econ Geogr 20(3):857–877

Paulsen B, Schroeder J (2005) An examination of tropical and extratropical gust factors and the associated wind speed histograms. J Appl Meteorol Climatol 44(2):270–280
Qiang Y, Huang Q, Xu J (2020) Observing community resilience from space: Using nighttime lights to model economic disturbance and recovery pattern in natural disaster. Sustain Cities Soc 57:102115
Rogers G, Koper P, Ruktanonchai C, Ruktanonchai N, Utazi E, Woods D, Cunningham A, Tatem AJ, Steele J, Lai S, Sorichetta A (2023) Exploring the relationship between temporal fluctuations in satellite nightlight imagery and human mobility across Africa. Remote Sens 15(17):4252. https://www.mdpi.com/2072-4292/15/17/4252
Sajid O (2023) Economic and demographic effects of increased flood susceptibility: evidence from rural India. In: 2023 Annual meeting, July 23–25, Washington DC (No. 335442). Agric Appl Econ Assoc. https://ideas.repec.org/p/ags/aaea22/335442.html
Schippers V, Botzen W (2023) Uncovering the veil of night light changes in times of catastrophe. Nat Hazard 23(1):179–204
Strobl E (2011) The economic growth impact of hurricanes: Evidence from US coastal counties. Rev Econ Stat 93(2):575–589
Strobl E (2012) The economic growth impact of natural disasters in developing countries: Evidence from hurricane strikes in the Central American and Caribbean regions. J Dev Econ 97(1):130–141
Strobl E (2019) The impact of typhoons on economic activity in the Philippines: evidence from nightlight intensity. Asian Development Bank Economics Working Paper Series, (589). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3590202
Strobl E, Ouattara B, Kablan SA (2020) Impact of hurricanes strikes on international reserves in the Caribbean. Appl Econ 52(38):4175–4185
Vickery P, Masters F, Powell M, Wadhera D (2009) Hurricane hazard modelling: The past, present, and future. J Wind Eng Ind Aerodyn 97(7–8):392–405
Xiao Y, Xiao Y, Duan Z (2009) The typhoon wind hazard analysis in Hong Kong of China with the new formula for Holland B parameter and the CE wind field model. In: Proc. Seventh Asia-Pacific Conf. on Wind Engineering. November 8–12, 2009, Taipei. https://iawe.org/Proceedings/7APCWE/M2B_4.pdf
Yonson R, Noy I, Gaillard JC (2018) The measurement of disaster risk: An example from tropical cyclones in the Philippines. Rev Dev Econ 22(2):736–765
Zhang X, Gibson J (2022) Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. Remote Sens 14(5):1282
Zhang X, Gibson J, Deng X (2023) Remotely too equal: Popular DMSP night-time lights data understate spatial inequality. Reg Sci Policy Pract 15(9):2106–2125
Zhang X, Gibson J, Li C (2024) The role of spillovers when evaluating regional development interventions: Evidence from administrative upgrading in China. Lett Spat Resour Sci 17(1):9

Appendix 2: Remotely measuring rural economic activity and poverty: Do we just need better sensors?

John Gibson, Xiaoxuan Zhang, Albert Park, Yi Jiang and Xi Li

Abstract

It is difficult and expensive to measure rural economic activity in developing countries. The usual survey-based approach is less informative than often realized due to combined effects of the clustered samples dictated by survey logistics and the spatial autocorrelation in rural livelihoods. Administrative data, like sub-national Gross Domestic Product for lower level spatial units, are often unavailable and informality and seasonality of many rural activities raises doubts about accuracy of such measures. A recent literature argues that high-resolution satellite imagery can overcome these barriers to the measurement of rural economic activity and rural living standards and poverty. Potential advantages of satellite data include greater comparability between countries irrespective of their varying levels of statistical capacity, cheaper and more timely data availability, and the possibility of extending estimates to spatial units below the level at which GDP data or survey data are reported. While there are many types of remote sensing data, economists have particularly seized upon satellite-detected nighttime lights (NTL) as a proxy for local economic activity. Yet there are growing doubts about the universal usefulness of this proxy, with recent evidence suggesting that NTL data are a poor proxy in low-density rural areas of developing countries. This study examines performance in predicting rural sector economic activity and poverty in China with different types of satellite-detected NTL data that come from sensors of varying resolution. We include the most popular NTL source in economics, the Defense Meteorological Satellite Program data, whose resolution is, at best, 2.7 km, two data sources from the Visible Infrared Imaging Radiometer Suite (VIIRS) on the Suomi/NPP satellite which have spatial resolution of 0.74 km, and data from the Luojia-01 satellite which is even more spatially precise, with a resolution of 0.13 km. The sensors also vary in ability to detect feeble light and in the time of night that they observe the earth. With this variation we can ascertain whether better sensors lead to better predictions. We supplement this statistical assessment with a set of ground-truthing exercises. Overall, our study may help to inform decisions about future data directions for studying rural economic activity in developing countries.

I. Introduction

It is costly and difficult to measure rural economic activity in developing countries. Survey-based approaches are less informative than often realized; rural livelihoods have a high degree of spatial autocorrelation (Gibson et al, 2011) and the costs of getting interviewers into rural areas makes clustered sampling the most practical approach. Thus rural economic survey data have more uncertainty than would be found in a simple random sample of similar size that was fielded on a population whose livelihoods correlate less strongly with their neighbors (Gibson, 2019). In addition to imprecision, survey-based approaches often have a lack of consistency in methods over time and space (DeWeerdt et al, 2020) and measurement errors can be large, especially in (self-)reported agricultural production data (Jerven and Johnston, 2015). These survey errors can spill over into national accounts data on rural sector activity, which also face other measurement constraints in developing countries (Angrist et al, 2021).

Recently it has been argued that high-resolution satellite imagery can help to overcome difficulties in measuring rural economic activity, living standards and poverty (Jean et al, 2016; Watmough et al, 2019; Lobell et al, 2020; Yeh et al, 2020). Potential advantages of remote sensing approaches include greater comparability between countries irrespective of their varying levels of statistical capacity, lower cost, the availability of higher frequency data that arrive sooner, and the possibility of building up from the pixel level to form estimates for spatial units below the level at which national accounts data or survey data are typically reported.[1] For example, Burke et al (2022) suggest that a single satellite image might be able to tell the story of a village's economic health.

There are many types of remote sensing data but economists have especially focused on satellite-detected night-time lights (NTL) as an indicator of economic

[1] Survey-to-census imputation methods (e.g. Elbers et al, 2003) can facilitate construction of small-area welfare estimates (which have extended beyond expenditures and poverty to include things like food security indicators, as in Hossain et al, 2020) but the temporal coverage is usually limited by census infrequency.

activity (Chen and Nordhaus, 2011; Henderson et al, 2012; Gibson et al, 2020). For example, Burke et al (2022) suggest that researchers can use satellite-detected data on a village's use of lights at night to draw inferences about village-level economic productivity; this is a far more granular level of analysis than would be traditionally possible. Yet there are emerging doubts about the universal usefulness of NTL data as a proxy. A study of variation in economic activity over space (but not over time) found that NTL data are a poor proxy in low-density rural areas of developing countries (Gibson et al., 2021). Relatedly, a cross-country panel study found that changes in NTL data did not positively correlate with changes in GDP in countries where agriculture and forestry were large components of GDP (Keola et al, 2015). A failure of luminosity changes to predict changes in economic activity indicators, such as employment and household spending, is also seen at local levels in developing countries (Goldblatt et al, 2020; Asher et al, 2021).

These patterns may reflect the fact that rural economic activity relies less on night-time lights than does urban sector activity. If so, finding that NTL data are less useful as a proxy for rural economic activity should hold throughout the development spectrum. Indeed, Gibson and Boe-Gibson (2021) find no relationship between changes in county-level agricultural GDP and changes in NTL data in the United States, even as activity changes in other sectors significantly correlate with changes in lights. Likewise, the cross-sectional lights-GDP relationship for the primary sector is less than one-fifth as strong as for other sectors of the US economy. The GDP data for the United States should be quite reliable so these differences in the strength of the lights-GDP relationships suggest that something about agriculture makes it less amenable to being remotely measured with NTL data.[2]

[2] Khachiyan et al (2022) use daytime imagery (from Landsat) to predict income and population for micro-grids with far greater accuracy than they achieve when using DMSP NTL data. However, their study is restricted to urbanized US census blocks and does not use modern NTL data that are more accurate (Gibson et al, 2020) and so it remains an open question as to whether rural economic activity is also hard to proxy with daytime imagery.

An alternative hypothesis is that because the most popular NTL data are from sensors originally designed to detect clouds for Air Force weather forecasts, failure to accurately detect dimly-lit and dispersed rural economic activity should be no surprise, even if brightly-lit cities are serendipitously detected. Along these lines, Chen and Nordhaus (2015) show that data from the more accurate research-focused Visible Infrared Imaging Radiometer Suite (VIIRS) sensor reveal activity that is not detected by the widely used Defense Meteorological Satellite Program (DMSP) data; amongst 600 cells (of 1°×1°) in Africa whose populations are below 10,000, all showed light using VIIRS but 72% showed no light according to DMSP. In other words, better sensors do a better job. Thus, the failure to remotely detect rural economic activity may just be because studies have used data that rely on remote sensors that were not designed for this task.

Determining whether better sensors are all that is needed for remotely measuring rural activity and poverty, rather than it being the case that such phenomena are inherently ill-suited to remote measurement, matters to the statistics investment decisions made by developing countries and donors. Most low- and middle-income countries fund less than one-half of their national statistical plans, and rely on donors for about US$1 billion annually in funding of statistical systems (World Bank, 2021). The development of machine-learning algorithms that successfully harness ‘big data’ is gaining considerable interest amongst development agencies even if these algorithms may yield little improvement in value-added over more traditional statistical approaches (Mahler et al, 2022). If monitoring rural economic activity and poverty is not very amenable to remote measurement approached but such approaches get favored by new investments in statistical systems just because of their current popularity it may exacerbate an existing bias. Urban areas are relatively data-rich, even in developing countries, and NTL data are quite suitable for measuring many urban phenomena so skewing statistics investments toward remote measuring approaches may make rural areas even more relatively data poor.

The current study provides some evidence on these questions. We use satellite-detected NTL data from sensors of varying precision to predict rural economic activity and poverty in China. We use the most popular NTL data source in economics, DMSP, with spatial resolution of 2.7 km at best, two data sources from VIIRS which have spatial resolution of 0.74 km, and data from the Luojia-01 satellite whose spatial resolution is 0.13 km[3]. The ground footprints for these various sensors show the dramatic improvements in spatial precision achieved with new generations of remote sensing systems; VIIRS is a product of the 2000s and has a ground footprint that is 45-time smaller than that of DMSP, which is a product of the 1970s (Elvidge et al, 2013). In turn, Luojia-01, which launched in 2018, has a ground footprint that is 28-times smaller than that of VIIRS (and over 1500-times smaller than that of DMSP).

Another key way in which these remote sensing systems have improved is in detecting feeble lights. The Day-Night Band (DNB) sensor on VIIRS can detect lights over a range of seven orders of magnitude of radiance while DMSP detects over only two orders of magnitude so it is either top-coded in brightly lit areas (Bluhm and Krause, 2022) or it misses many dimly lit places (Chen and Nordhaus, 2015). The Luojia-01 sensor also has superior ability to detect feeble lights compared to other NTL data sources (Li et al, 2019), and this should make it very suitable for studying rural economic activity and poverty. Overall, we expect that the variation in spatial precision and in the range of luminosity conditions that can be detected by the three different remote sensing systems that we use should help to ascertain whether better sensors are sufficient to yield better predictions for rural areas.

In the next section we describe properties of the three types of NTL sensors that we use and summarize prior literature showing the superiority of Luojia-01 over VIIRS and of VIIRS over DMSP. We also report on some ground-truthing exercises we used to establish first-hand that Luojia can detect features in rural areas of China

[3] Evidence for the ongoing popularity of DMSP data in economics comes from a Scopus search (on 27/11/23) of journal articles: outside of economics, 12% more articles published in the last two years mention VIIRS (and 'night' to rule out non-NTL uses) than mention DMSP. Yet for the economics subject area it is the other way around, with 76% more articles mentioning DMSP than VIIRS over the same two-year period.

that are less well detected by the other two sensors. In Section III we describe our China-wide sample of 1460 rural counties and the indicators we use when comparing the predictive ability data from the NTL sensors. Our main results are in Section IV, using primary sector GDP, poverty status and fiscal revenue (all at county-level), as economic activity and welfare indicators for assessing predictions from each remote sensing system. Section V has a discussion of the broader implications of our results.

II. The Competing Sensors

Our evaluations cover four main sources of nighttime lights (NTL) data: DSMP, two flavours of VIIRS, and Luojia-01. Table 1 has a summary of relevant attributes for the four data sources, with the top panel dealing with the satellite and sensor (just three columns are needed because the information is the same for both flavours of VIIRS data) and the bottom panel dealing with the data products.

We use the DMSP annual composites for 2018 from satellites F15 and F16.[4] These data exclude ephemeral lights from fires and flaring but there may be other forms of ephemeral light still captured in the composite. Any pixel-nights affected by clouds, moonlight, sunlight and other glare are removed at the processing stage. The observation time is ca. 3.30am, compared to the usually reported 8.30pm observation time for the DMSP time-series that ended in 2013.[5] The data are 6-bit digital numbers (DN) ranging from 0–63 (higher for more light), presented on a 30 arc-second output grid, which is roughly 0.9×0.7 km at China's latitude (but the underlying resolution of the sensor is far coarser, as discussed below).

We use VIIRS data for 2018 that are processed by two different groups in two different ways. The Earth Observation Group (EOG) at the Colorado School of Mines will be familiar to economists because the same group worked extensively on DMSP. They provide VIIRS Night Light (VNL 2.1) annual composites from monthly cloud-free radiance averages from the Suomi/NPP satellite. An initial filtering removes

[4] These data are available from: https://eogdata.mines.edu/products/dmsp/#download

[5] Unstable orbits for DMSP satellites meant they observed the earth earlier as they aged, so Ghosh et al (2021) extended the time series past 2013 by switching from late afternoon observations, which were not useful for measuring luminosity, to the corresponding pre-dawn observations (given the 12 hour revisit time).

extraneous features such as fires and aurora before the resulting rough composites have further outlier removal procedures applied (Elvidge et al, 2021). The data are in units of nano Watts per square centimetre per steradian (nW/cm^2/sr) on a 15 arc-second output grid (ca. 0.5×0.4 km at China's latitude). In addition to the masking of ephemeral lights there is also a stray-light correction. The second set of VIIRS data are from NASA's Black Marble collection (Román et al, 2018). The data are corrected for atmospheric, terrain, vegetation, snow, lunar, and stray light effects on radiance values, which are calibrated across time and are also validated against ground measurements. The Black Marble (BM) data differ from VNL data in four ways: reporting is with 16-bit precision (n=65,536 values) rather than 14-bit precision (n=16,384 values), the user has some control over the angle of detection, with near-nadir, off-nadir, and all-angles composites, data are given separately for snow-free and snow-covered nights, and users have to build their own composites from tiled data.[6]

The Luojia-01 satellite was developed by Wuhan University as China's first NTL sensing satellite, and was launched in June 2018. The spatial resolution is finer than any prior NTL sensor, at 130 metres, and it covers a wider range of radiance values (Li et al, 2019; Wu et al, 2021). The 14-bit digital numbers are on a 4.44 arc-second output grid (ca. 0.1×0.1 km at China's latitude). Luojia-01 better detects feeble lights than do earlier generation sensors, such as the VIIRS DNB, (Li et al., 2019a), which is likely to matter for studying rural areas. By May 2019, Luojia-01 imagery had fully covered China and some parts of Southeast Asia (Levin et al., 2020). In terms of observing economic activity, the local observation time is ca. 9.30pm, as opposed to 1.30am (BM and VNL) and 3.30am (DMSP) for the other sensors.[7]

A key attribute of these various remote sensing systems is their pixel "footprint" that is projected onto the Earth as the oscillating sensor scans a swath, of approximately 3000 km for DMSP, VNL and BM, and 250 km for Luojia-01 (Table 1).

[6] The VNL data are available from: https://eogdata.mines.edu/nighttime_light/annual/v21/ and the Black Marble data from: https://ladsweb.modaps.eosdis.nasa.gov/achive/allData/5000/VNP46A4/. For both BM and VNL the earth observation time is ca. 1.30am.

[7] The data are available to download from the High-Resolution Earth Observation System of the Hubei Data and Application Center (http://59.175.109.173:8888/app/login.html).

At nadir (the point on earth below the satellite) the lights are captured from a circular footprint but away from the nadir the angular viewing effect means that the lights are from a larger ellipse. The DMSP footprint at the nadir is approximately 25 km^2 (Elvidge et al, 2013); this reflects aggregation of the original ('fine') pixels into 5×5 blocks in order to conserve data storage prior to transmitting the signal to earth plus effects of geo-location errors that displace the signal by about 3 km from where the light sources are located (Tuttle et al, 2013). The footprint expands away from the nadir due to the angular viewing effect, so only images from the inner 1500 km part of the swath are used due to excessive blurring at scan extremities (Abrahams et al, 2018). Even at the half-scan point, the footprint will be about 60 km^2 (which is far coarser than the output grid).

In contrast to DMSP, VIIRS maintains a constant 0.55 km^2 footprint because the sensor compensates for effects of viewing the earth at an angle by turning off some detector elements; the spatial precision of VIIRS is also due to there being no need to aggregate pixels (there is more than adequate onboard data storage) and no geolocation errors. The Luojia-01 sensor is even more spatially precise, with a ground footprint of just 0.02 km^2 which is 28 times smaller than for VIIRS and 1500 times smaller than for DMSP. The relative sizes of the footprints are shown in Figure 1, which draws inspiration from a similar figure in Elvidge et al (2013) that shows why VIIRS data are superior to DMSP data for mapping nightlights.

2.1 Prior Literature on Performance of the Sensors

The properties of the sensors outlined in Table 1, (especially their varying spatial precision, as shown in Figure 1) lead us to expect that the data from VIIRS should more accurately predict economic activity and poverty than does the data coming from DMSP, and likewise data from Luojia-01 should perform better than data from VIIRS. This ranking is generally borne out by the literature. For example, an early cross-sectional study found that VIIRS data are better predictors of county-level GDP in China than are DMSP data (Shi et al, 2014) and the same pattern was found at provincial

and prefectural level (Dai et al, 2017). Similar comparisons have been made (cross-sectionally) for a wider set of socio-economic indicators; VIIRS data were generally a better proxy although the later observation time (compared to the pre-2014 DMSP data that observed earth in the early evening) lowered the correlations for some indicators (Jing et al, 2015).

One reason DMSP data may be predictively less accurate is outliers due to ephemeral and stray lights (Table 1). Gibson and Boe-Gibson (2021) evaluate DMSP data as a predictor of county-level GDP in the United States by comparing against two types of VIIRS data, one with outliers masked and one without masking; DMSP results are most like the ones from the unmasked VIIRS data, which suggests that the DMSP data also are affected by outliers. A feature of the DMSP measurement errors is that they are spatially mean-reverting (Gibson, 2021); hence these errors become more apparent when DMSP data are used to proxy for the economic activity of smaller, lower-level, spatial units. In contrast, aggregation to larger spatial units, such as provinces, tends to disguise the effects of the errors, because aggregation is inherently mean-reverting. Another factor that may matter is population density; China's county-level GDP is more strongly related to VIIRS data (either BM or VNL) in more densely areas but the DMSP data show no similar gradient (Zhang and Gibson, 2022). The saturation effect (top-coding) in DMSP data may be a factor here given that saturation causes densely populated areas but the DMSP data show no similar gradient (Zhang and Gibson, 2022). One reason for this lack of gradient may be from the saturation effect (top-coding) in DMSP data which make densely populated cities (which tend to be more brightly-lit) seem to be no brighter than far smaller places.

The comparisons of Luojia-01 to earlier generation sensors mostly use VIIRS data as the DMSP data for 2018 are only recently available (Ghosh et al, 2021), and 2018 is the sole year with Luojia-01 data. For example, when modelling socio-economic parameters in eastern and central regions of China, Zhang et al (2019) found that Luojia-01 data out-performed the equivalent VIIRS data. Liu et al (2020) created a 20-factor multi-dimensional economic development index for counties in three

provinces (Hubei, Hunan and Jiangxi) and found that the Luojia-01 data had better fit with this index in a Random Forest model than did VIIRS data. Zhang et al. (2020) showed that Luojia-01 images could detect changes in lights during the 2019 Spring Festival in six cities in western China and that these changes correlate with Baidu-sourced data on short-term (holiday) migration. Lin et al (2022) found that there was a slight improvement over VIIRS when using Luojia-01 data in a random forest model for estimating the poverty status of 126 counties (and districts) in Chongqing and Guizhou in China.

2.2 Ground-truthing Exercises in Rural China

To establish first-hand that Luojia-01 data capture features in rural areas of China that are less well detected by earlier generation sensors we conducted ground-truthing in three provinces (Henan, Shanxi, and Shaanxi)[8]. In each province, five counties were selected for the fieldwork and two types of analyses were conducted—village transects and luminosity curves along the development spectrum.

For the transects, photographs were taken at the main crossroads in a village and then at 1 km intervals out to a 3 km perimeter. We then compared satellite-detected luminosity data for these points with what was seen on the ground. For example, in Xiyang village of Mianchi county, Henan Province, by the 3 km mark the land was devoid of buildings and was covered in crops (and an unlit pathway), at the 2km mark there were sparsely scattered building, while from the 1 km mark inwards it was all paved surfaces and built-up area, including some two storey buildings (Figure 2). The Luojia-01 data reflected this transect very closely; the highest luminosity was at the crossroads, at the 1 km mark it was 86% of that value, at the 2 km mark it was 35% of luminosity at the crossroads, and by the 3 km mark it had fallen to seven percent. The DMSP data showed no similar sensitivity; at the 3 km mark the values were still two-thirds of light reported for the crossroads while at the 1 km and 2 km marks they were

[8] To check if these provinces are representative we estimated elasticities of primary sector GDP with respect to luminosity from Luojia-01 using data for rural counties, province-by-province. The elasticities for these three provinces were not statistically significantly different to the elasticities for the other 24 provinces (Appendix A).

88% and 80% of the value for the crossroads. Thus, DMSP data hardly distinguished the village centre from surrounding farmland. Black Marble data had a similarly flat profile, even though the VIIRS sensors has better spatial resolution than DMSP. Another example comes from Dajianbei village, Pinglu county, Shanxi Province where inability of earlier generation sensors to detect village features was even more pronounced; by 2 km from the crossroads, where there were hardly any buildings, Luojia-01 showed that luminosity had declined to just 7% of the brightest point in the village while DMSP (Black Marble) was still at 87% (79%) of the brightest value (Figure 2c, Appendix A).

For the luminosity curves we ranked each county in a province by primary sector GDP (total rather than per capita to establish where rural economies are overall more active). We then plotted GDP against total luminosity (from the various sensors) and selected five counties per province, spanning the range from low GDP to high, and visited villages in these counties to photograph the level of development seen on the ground. For example, in Taiqian county in northeast Henan we observed very sparse buildings and unpaved roads, while in Gushi, whose primary sector GDP is over seven-times higher, there were modern two-storey buildings, paved streets and vigorous activity (Figure 3). While Luojia-01 data reflected these differences, with luminosity from Gushi over twice that of Taiqian, the DMSP data had it the other way around; the poor county seemed to emit more total light than the far richer county. A likely explanation is that inherent blurring in the DMSP images seems to spread the light from brightly-lit places far and wide; Tiaqian is near big cities in Shandong (Jining and Liaocheng) and some light they emit may be wrongly attributed to Tiaqian in DMSP data.[9] Across all rural counties in Henan, the DMSP data provide no ability to distinguish between the poor areas and prosperous ones, with a flat trend line in Figure 3, while the Luojia-01 data do generally reflect differences seen on the ground. The

[9] Likewise, Zhongmu county in Figure 3 has a middling GDP level, at just below CNY 4 billion yet the DMSP luminosity is almost twice as high as any other rural county in Henan (yet Luojia-01 shows nothing remarkable). Zhongmu is midway between the provincial capital Zhengzhou, with an urban population of over 10 million, and Kaifeng, with an urban population of two million, and the blurred DMSP images attribute some of the lights from these two big cities to this largely rural county.

differences were less marked for the two VIIRS-based data sources, which show recorded luminosity rising with primary sector GDP (Figure 1b, Appendix A) and were less marked in the other two provinces where ground-truthing fieldwork was conducted.

III. Estimation Sample, Selected Indicators and Comparison Procedures

3.1 Sample

We selected all rural counties in China to use for testing if the NTL data coming from better sensors provide more accurate predictions of rural economic activity and poverty. From all third-level sub-national units in China's 2020 census (*n*=2848) we first removed districts (these are the urban core of cities). Next we removed county-level cities; the primary sector share of GDP in such places averages just 1.3% (and is no higher in county-level cities than in districts). This left us with a sample of n=1460. This sample is predominantly counties (n=1415) but in the parts of China that are not administered as prefectures (as the second-level unit) the other third-level units may be banners and special areas (n=45).[10] We refer to all of these as counties for simplicity. In this sample, the share of GDP from the primary sector averages 37%. The sample selection flowchart is in Appendix B, while the map showing the areas we cover is in Figure 4. The map also highlights the provinces where the ground-truthing exercises described in Section 2.2 took place.

3.2 Indicators of Rural Economic Activity and Welfare

We use primary sector GDP (in billions of Yuan, CNY) for each county in 2018, from the 2019 edition of the *China Statistical Yearbook* (county-level), known in Chinese as *Zhongguo Chengshi Tongji Nianjian*.[11] The same yearbook provides our second

[10] For non-county and non-city areas (e.g. banners), if there were doubts about their suitability for inclusion the primary sector share of GDP was used as a criteria. The excluded areas were especially coal mining areas from Inner Mongolia.

[11] Such disaggregated GDP data, at the third sub-national level, are rarely available; while the United States also reports county-level GDP, other large countries like India and Indonesia only report at the second sub-national (district) level, and Eurostat only reports down to the NUTS2 level, which is aggregations of counties in some countries (e.g., the UK) and provinces in others.

indicator of economic activity (and of ability to fund local public goods that may affect household welfare), which is the fiscal revenue generated in each county. Our third indicator is a binary variable for the poverty status of each county, which is determined by the State Council Leading Group Office of Poverty Alleviation and Development. The basis of this classification was that a group of *n*=832 rural counties had been defined as poverty-stricken in 2013 and thereafter the National Bureau of Statistics used an annual National Rural Poverty Monitoring survey to update the progress of these counties in moving out of poverty.[12] We use the situation as it was at the beginning of 2018, when *n*=679 counties were still defined as poverty stricken across all of China, with *n*=578 of these counties in our sample.

3.3 Luminosity Indicators

We use data from the competing sensors to form three sets of luminosity variables, all of which are found in applied economics studies using NTL data. The first is the sum of lights by county, which is the product of lit area and average brightness within lit areas; previous studies relating luminosity to poverty suggest this is a sufficient statistic for the relevant variation (Gibson et al, 2017; Gibson et al, 2023) and non-nested tests show that the sum of lights outperforms other luminosity variables for predicting China's county-level GDP (Zhang and Gibson, 2022). Next we use the sum of radiance divided by county area, as the way that several studies use NTL data (Henderson et al, 2012; Castelló-Climent et al, 2018) even if GDP is rarely normalized by area. The third approach is to use the proportion of pixels in a county that are illuminated; the inverse of this measure, or relatedly the share of the population living in unlit rural areas, is used in studies of rural poverty by Smith and Wills (2018) and Maldonado (2023).

[12] Earlier classifications of counties as poverty stricken were found to have targeting errors with respect to per capita income (Park et al, 2002). A switch to village-level targeting (the Integrated Village Development Program) was made in 2001. This intervening regime, and the fact that the number of poverty counties in 2013 exceeded the number of poor counties from the pre-2001 regime, means that the earlier criticism of targeting errors for the pre-2001 poor county designation does not automatically apply to the post-2013 classification. In support of this claim, we show below that the current designation reflects environmental influences (such as elevation and temperature) that were previously neglected in the poor county designation (Olivia et al, 2011). Hence it is reasonable to consider this new poor county classification as a break from the past.

The machine learning models used in the literature to predict rural poverty often draw upon a large set of predictors and so we want to evaluate the NTL data both as unconditional predictors and also when we include covariates. For each county we have the average elevation, precipitation, and temperature, and five land cover categories—cultivated, forested, urbanized, village settlements, and industrial and infrastructural—from daytime Landsat images (Song and Deng, 2017) that we use as covariates.

For the land cover categories that refer to built-up area, there are varying relationships with luminosity that are shown by all four NTL data sources (Table 2). In particular, village settlement areas have a far smaller effect on the county sum of lights than from similarly sized urban area or industrial and infrastructural area. Averaging across the four NTL data sources, for each 1 km^2 of village built-up area the sum of lights is (unconditionally) 0.7 percent higher, while it is 3.8 (1.6) percent higher per 1 km^2 of urban (industrial and infrastructure) area. If we include all three types of built-up area in the same regression, county lights are 0.3 percent higher per km^2 of village area, on average and holding the other types of built-up area constant; 2.5% higher per km^2 of urban area and 1.4% higher per km^2 of industrial and infrastructural area. For all four sources of NTL data, we would reject the hypothesis that village built-up area has the same effect on the county luminosity totals as do the other two types of built-up area. The results in Table 2 emphasize that NTL data are far more responsive to differences in urban area and in industrial and infrastructural area than they are to variations in rural areas, such as in the extent of village built-up area.

3.4 Empirical Specification

We have three economic activity and welfare indicators, and these are related to the three luminosity indicators by the following regressions:

$$\ln(activity\ or\ welfare\ indicator)_i = \alpha + \beta \ln(luminosity\ indicator)_i + \varepsilon_i \quad (1)$$

where we also include control variables in some of these regressions. For one outcome (county poverty status) and for one luminosity indicator (the share of illuminated

pixels) we cannot take logarithms because of the possible presence of zeros and so results in those cases use standardized variables. We place no causal interpretation on equation (1), which is just one way to find the best predictors amongst the set of NTL variables.

IV. Results

The results of estimating equation (1) for our three outcome measures are shown in Figure 5, using the sum of lights as the luminosity indicator. Irrespective of whether we control for land cover and environmental factors, or simply look at the unconditional predictions, there is no gradient whereby the data from the more precise sensors lead to either larger elasticities of the activity or welfare indicators with respect to luminosity, or to higher R^2 values for the prediction equations. In other words, it does not seem that the data from newer and more precise sensors do a better job of predicting agricultural activity, poverty or fiscal revenue for these counties in rural China.[13]

In terms of each of the outcome measures, fiscal revenue is the one that is best predicted by NTL data and has the highest elasticities with respect to luminosity. While this indicator seems to show the beginning of a gradient, in the sense that the R^2 values and the elasticities have a ranking *BM* > *VNL* > *DMSP* that pattern is then broken by the values for Luojia-01 being lower than for all of the other NTL data sources.[14] The predictions for primary sector GDP have the next highest R^2 values and elasticities but with little semblance of a gradient either with or without the control variables included in the specifications. For example, the R^2 values when using the lowest

[13] In order to facilitate comparison of results across these three diverse indicators, the variables are standardized. The figures show one value for each source of NTL data (so four values), but overall we have results for seven sets of NTL data; from DMSP satellites F15 and F16 as well as their average, from VNL, from Luojia-01, and from the snow-free and weighted-average Black Marble data (BMwa). The tables report all seven results but to avoid clutter figures do not show component results, such as for DMSP F15 and F16 or for the Black Marble data that is only composed from readings on snow-free nights.

[14] Confidence intervals are not shown on the charts, to reduce clutter, but are reported in the appendix tables. For fiscal revenue and primary sector GDP the standard errors are about 0.02 and for poverty status are 0.03 in unconditional regressions and 0.04 with control variables. To provide a sense about the statistical significance of differences in Figure 5, one example is the unconditional elasticity of fiscal revenue with respect to luminosity; when using BM data it is four standard errors larger than the elasticity estimated using DMSP data.

resolution DMSP data to make the predictions exceed those from two of the models using data from the more modern and spatially precise sensors (in the results with control variables included). The inability of the NTL data to predict county-level poverty status is especially clear, with none of the adjusted R^2 values for poverty predictions from any specification or from any NTL sensor, either with or without control variables, exceeding 0.2. In contrast, in the equations for fiscal revenue and primary sector GDP some of the R^2 values were approximately 0.7.

In the full results of the regressions that underpin Figure 5 (as reported in Appendix C) the lack of gradient, whereby better sensors are expected to yield data that better predict, is also seen if lights are normalized by area or if the percentage of illuminated pixels is used. For the regressions for primary sector GDP and fiscal revenue, these alternative specifications of the NTL variables are almost never as good as using the sum of lights; for example, for just two out of the 84 results in Tables 1 and 3 of Appendix C are the specifications that use variables other than the sum of lights giving the best fit. The regressions for county poverty status showed more mixed results; if land cover and the environmental controls are included then the sum of lights is generally the best fitting specification whereas if no control variables are used then lights normalized by area was generally the best fitting specification. Nevertheless, it needs to be emphasized that even with detailed earth observation data from nights (across multiple sensors) and days (using Landsat) we could not predict more than 20% of the variation in whether a county in rural China is classified as poor or not.

4.1 Machine Learning Results

The results reported thus far have used traditional regression approaches but much of the recent literature using satellite imagery for poverty predictions is based on machine learning (Hall et al, 2023). To ensure that our comparisons of the predictive power of data derived from different generations of satellite remote sensing systems are salient to this recent literature we also used a random forests (RF) algorithm (which is a form of supervised machine learning) to predict poverty. This is an ensemble learning

method that combines multiple decision trees for solving classification and regression problems (Breiman, 2001). Recent studies using luminosity data in Asia find that the RF method has the highest predictive performance (Puttanapong et al, 2023; Fenz et al, 2024). A key feature for our application is that the RF method is well suited to the task of assessing variable importance, when compared to other frequently used machine learning methods (Grömping, 2015).

To obtain the RF predictions we used the randomForest package in R which allows us to have enough decision trees to generate them with randomness when building the forest. Each of the individual trees is built independently with a subset of the entire training dataset, where this subset is drawn using bootstrap sampling with replacement. Two-thirds of the instances in the bootstrap samples were used to train the individual trees, while the rest were used as testing data to evaluate the final RF classifier. The instances used for training the individual trees are referred to as in-bag instances, and the instances for testing are out-of-bag instances. In order to achieve robust results, the processing of building and selecting the final RF model is repeated 1000 times. This ensured that the RF model can include different sets of predictor variables in the training and testing samples, so that stable and reliable results are produced.

We first used just the four NTL predictors (the log of the sum of lights from DMSP, VNL, BM and Luojia-01) and then subsequently included the eight environmental factors and land cover variables as features in the ensembles of decision trees used for classifying counties as either poor or not. We use the index "%IncMSE" (the percentage increase in the mean square error) to measure the relative importance of each predictor variable. This index is constructed by randomly assigning a value to a predictor variable; for a predictor that is relatively important there will be a larger rise in the MSE (in other words, a penalty) if an actual value is replaced with a random value, compared to the case of replacing values for other variables with random values. As shown in Figure 6, the BM and DMSP data showed more importance than the other two NTL variables in the RF model without environmental variables (panel a). In other

words, even using a machine learning approach, there is no appearance of a gradient whereby the data from newer and precise NTL sensors such as Luojia-01 seem to provide better predictions of rural poverty. This lack of gradient is also apparent when environmental variables are included as predictors (panel b), and the inclusion of these variables also shows that the most important variables for predicting county-level poverty in rural China appear to be elevation and temperature, rather than luminosity.

4.2 Heterogeneity Analysis

The rural counties that we study vary widely in terms of their population density; the 10th percentile is under 10 persons per km2 and the 90th percentile is 567 per km2 the (median density is 131). Previous evidence for China shows that it is for the more densely populated areas that VIIRS data (from both VNL and BM) are most closely related to total GDP at county level (this evidence included urban districts, that are at an equivalent level in the administrative hierarchy), whereas the DMSP data had no such gradient with respect to density (Zhang and Gibson, 2022). To check if this same pattern holds here with our exclusively rural sample, we split the sample at the median density, with results of the prediction equations for the low density and high density counties reported in Table 3. In comparison to Figure 5 (and to the tables in Appendix C), the outcome measure now is total GDP rather than primary sector GDP and so R^2 values are higher just from this change (from 21% higher for VNL to 47% higher for BM; the average rise is 36%). This is just another manifestation of the fact that agricultural activity is not very well proxied by NTL data; thus, if an outcome measure is predominantly agricultural—as is the case for primary sector GDP—there are less accurate relationships with NTL data than when the outcome measure depends less on agricultural activity, as is the case for total GDP.

For three of the four sources of NTL data—VNL is the exception—the R^2 values for the high density counties are higher, by an average of almost 50%, compared to the low density counties, in the regressions without control variables. In the same regressions, elasticities are also about 45% larger, on average, in the high density

counties reflecting the closer fit between luminosity and economic activity in densely populated areas. This result corroborates patterns previously seen elsewhere (e.g. Gibson (2021); Gibson et al (2021)). However, the introduction of control variables changes these patterns, with generally little difference in explanatory power or in elasticities between the high density and low density counties once the controls for land cover and environmental factors are included. The correlation between population density and land cover, especially the area of a county that is urban built-up area, particularly contributes to the reduced importance of variation in the sources of the NTL data.

In terms of our main focus, there is still no apparent gradient whereby the modern, more precise, sensors are yielding NTL data that are more accurate predictors of county-level GDP, in either high density or low density areas, with or without control variables. Thus the existing finding that NTL data are a poor proxy for economic activity and welfare in low density rural areas (e.g. Gibson et al, 2021) does not appear to be an artefact from researchers relying on data from outdated, low resolution, remote sensing systems such as DMSP. Instead, the lack of improvement in the predictions, when using newer and far more precise sensors, is consistent with the hypothesis that rural activity is poorly suited to remote measuring approaches using NTL data, even when these measurement efforts are in conjunction with the use of daytime Landsat images (such as in the Table 3 results that include control variables).

V. Discussion and Conclusions

In our analysis of rural counties drawn from all parts of China, where we use three generations of light-detecting satellite remote sensing systems, we do not find any gradient whereby more accurate predictions of rural economic activity and poverty come from the newer, more precise, systems. Compared to the DMSP sensor, the newer systems have sensors that should better detect dim lights, which should be an advantage when studying rural areas. The newest of the sensors, Luojia-01, also has an advantage of observing in the evening rather than after midnight and so it is timed to coincide with

a greater range of rural economic activities. The lack of gradient is notable because our own ground-truthing exercises showed that data from the newest and most precise sensor, Luojia-01, did capture features of the built environment in villages in these rural counties that were less well detected by the older and coarser resolution sensors. Our findings also go against the general thrust of the literature that suggests that newer sensors do a better job of predicting economic activity and poverty.

The question motivating our study is what accounts for poorer predictive performance of NTL data when used for low density rural areas, as found by Gibson et al (2021) and Zhang and Gibson (2022). A prior macro-level study also found night-time lights were a poor proxy for economic activity if agriculture and forestry are a large share of GDP (Keola et al, 2015). Are these results just arising because these studies used data from sources such as DMSP that were not designed to observe sparsely spread and dimly lit lights on earth, and so it should be no surprise that data from these older sensors are poor predictors in such settings, or is it, instead, due to the fact that most rural economic activity is inherently ill-suited to these remote measurement approaches because use of night-time lights is limited. Our results support the second hypothesis; the predominant rural activity is agriculture that rarely needs concentrated sources of light (Bluhm and McCord, 2022), so activity in this sector and the economic welfare of sector participants are ill-suited to remote measurement using light-sensing satellites. In other words, the payoffs to using newer and better NTL sensors to measure and predict rural outcomes are fairly muted (and may be zero) even if there may be a reward to using these newer sensors for studying urban areas.

In terms of differences between these findings and the published literature, the current evaluation is far more comprehensive, both over space and in terms of indicators. Prior studies for China typically compare just two sources of luminosity data, such as VIIRS and DMSP or Luojia-01 and VIIRS, and are usually for just a few provinces. In terms of the difference from our own ground-truthing exercises, which showed that Luojia-01 could pick up village features not picked up by the other sensors, the scale and spatial autocorrelation of county-level data is likely to matter. The median

area of the counties we study is 2200 km2 and only 12% of them are below 1000 km2. At this scale, the fact that the ground footprint of Luojia-01 is 0.02 km2 whereas for DMSP it is at least 25 km^2 provides less advantage than it might for smaller spatial units. Furthermore, there is considerable spatial autocorrelation in rural China, with statistically significant Moran's I statistics for primary sector GDP and poverty of 0.49 and 0.36.[15] Thus, the fact that the older and lower resolution DMSP system yields blurred data (Abrahams et al, 2018), with some lights attributed to particular counties coming from elsewhere (as discussed in Figure 3), may matter less than it would if these same data were used to study more spatially variegated indicators at a local scale. Hence, it would be useful to repeat our research in a setting where there are village-level indicators available as a benchmark, which may provide a testing ground that shows a greater payoff to spatial precision

Some support for our main finding of a lack of gradient comes from a recent review of studies that use machine learning and earth observation satellite imagery to predict poverty (Hall et al, 2023). The meta-analysis based on 60 studies found that the spatial resolution of the satellite imagery was unrelated to the accuracy of the subsequent predictions made about poverty. In other words, studies using data from better sensors did not seem to do a better job of predicting poverty, which is the same pattern that we find for rural China.

A potential criticism of our findings is that some high profile remote sensing studies of poverty in developing areas concentrate on using daytime images rather than NTL data, because it is argued that very poor rural areas lack sufficient variation in night-time lights (Yeh et al, 2020; Hall et al, 2023). Yet some of these high profile studies have used transfer learning approaches that combine daytime and night-time images (e.g. Jean et al, 2016) and so they still utilize NTL data. Furthermore, some studies in economics claim to directly measure poverty with night-time lights data (or

[15] Moran's I is equivalent to the slope coefficient in a linear regression of Wz on z where W is a spatial weights matrix (Anselin, 1988). In other words, it examines the strength of the relationship between one observation and the spatially weighted average of its neighboring observations. The NTL data also show spatial autocorrelation, with Moran's I values of 0.41 (DMSP), 0.43 (VNL), 0.27 (BM) and 0.41 (Luojia-01). All of the I statistics are statistically significant at the $p<0.01$ level.

with the absence of lights), such as Smith and Wills (2018) and Maldonado (2023). This prior direct use of night-time lights data to study poverty helps make our findings salient. Moreover, given that we also use daytime Landsat images to create land cover variables we believe that our findings can be related to some of the high profile studies that used both night-time and daytime images. We therefore believe that our findings should be considered when investment decisions are made about future data directions for studying rural economic activity and poverty in developing areas.

In arguing that newer is not necessarily better, in terms of the types of data used to study poverty in rural areas of developing countries, we are not defending the status quo methods. While rural livelihoods will remain spatially correlated, given the use of environmental inputs, which therefore means that rural samples will be less informative than similarly-size samples22 in less correlated places, there is still scope to improve the surveys.[16] For example, many rural surveys are highly clustered, with a dozen or more households surveyed per enumeration area (EA). Less clustered samples would help to increase our confidence in survey results. Such a change could be timely because recent FAO and World Bank (2018) guidelines for food data collection recommend switching to 7-day recall surveys. With this survey design, interviewers can spend less time in each EA compared to when they were required to monitor households who were filling out expenditure diaries for 14-days.[17] Evidence for this time-saving comes from a recent experiment where the same teams of interviewers switched between diary and recall formats when assigned a new workload in each survey round; with recall they had ample free time but if fielding diaries they struggled to complete their workload (Sharp et al, 2022). Hence, survey experiments where fewer households per EA are surveyed but more EAs are sampled could be fruitful, just as

[16] Many of the SDGs are tail-based measures, such as proportions of the population either hungry or poor, and so they depend not only on measures of means and totals (what many household economic surveys historically focused on) but also on variances (Gibson, 2020). Hence, sample efficiency matters to SDG monitoring.

[17] At the very least, interviewers for diary surveys would return each week to pick up the completed diary and distribute a new diary for the coming week. In situations with illiterate households the interviewers often needed to revisit every second day (Beegle et al, 2012).

experiments like those of Beegle et al (2012) helped to change other aspects of survey design in developing countries.

Yet interest in such experiments may be limited by apparent success of the high profile machine learning studies using satellite remote sensing to predict rural poverty. Donors may reason that there is less need for surveys if we can rely on satellites. Our results provide grounds for caution. If rural economic activity and poverty are poorly suited to remote study, as we find, but current fashion leads to more use of satellite-based approaches, it may exacerbate an existing data inequality. Specifically, rural areas are already data poor. We showed here (in Table 2) that night-time lights are far more sensitive to urban activity than rural activity, and so further investments into remote sensing approaches may inadvertently be pro-urban even if motivated by a desire to modernize the study of rural poverty. Indeed, to the extent that future funding for remote measurement approaches to rural poverty is at the expense of improvements in the traditional survey approaches, these investments may impair our understanding of rural poverty in the future.

References

Abrahams, A., Oram, C., and Lozano-Gracia, N. (2018). Deblurring DMSP nighttime lights: A new method using Gaussian filters and frequencies of illumination. *Remote Sensing of Environment, 210*, 242-258.

Angrist, N., Goldberg, P. K., and Jolliffe, D. (2021). Why is growth in developing countries so hard to measure? *Journal of Economic Perspectives*, *35*(3), 215-42.

Anselin, L. (1988). *Spatial Econometrics*: Methods and Models, Dordrecht: Kluwer Academic Publishers.

Asher, S., Lunt, T., Matsuura, R., and Novosad, P. (2021). Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the SHRUG open data platform. *The World Bank Economic Review*, 35(4), 845-871.

Beegle, K., De Weerdt, J., Friedman, J., and Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, 98(1), 3-18.

Bluhm, R., and Krause, K. (2022). Top lights: Bright cities and their contribution to economic development. *Journal of Development Economics*, *157*(1), 102880.

Bluhm, R., and McCord, G. (2022). What can we learn from nighttime lights for small geographies? Measurement errors and heterogeneous elasticities. *Remote Sensing*, 14(5), 1190.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

Burke, M., Driscoll, A., Lobell, D., and Ermon, S. (2022). Using satellite imagery to understand and promote sustainable development. *HAI Policy Brief*, Stanford Institute for HumanCentered Artificial Intelligence. https://hai.stanford.edu/policy-brief-using satelliteimagery-understand-and-promote-sustainable-development.

Chen, X., and Nordhaus, W. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences, 108*(21), 8589-8594.

Chen, X., and Nordhaus, W. (2015). A test of the new VIIRS lights data set: Population and economic output in Africa. *Remote Sensing*, *7*(4), 4937-4947.

Castelló-Climent, A., Chaudhary, L., and Mukhopadhyay, A. (2018). Higher education and prosperity: From Catholic missionaries to luminosity in India. *The Economic Journal*, *128*(616), 3039-3075.

Dai, Z., Hu, Y., and Zhao, G. (2017). The suitability of different nighttime light data for GDP estimation at different spatial scales and regional levels. *Sustainability*, *9*(2), 305.

De Weerdt, J., Gibson, J., and Beegle, K. (2020). What can we learn from experimenting with survey methods? *Annual Review of Resource Economics*, *12*, 431-447.

Elbers, C., Lanjouw, J., and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, *71*(1), 355-364.

Elvidge, C., Baugh, K., Zhizhin, M., and Hsu, F-C. (2013). Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proceedings of the Asia-Pacific Advanced Network*, *35*(0), 62.

Elvidge, C., Zhizhin, M., Ghosh, T., Hsu, F-C., and Taneja, J. (2021). Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, *13*(5), 922.

Fenz, K., Mitterling, T., Martinez, A., Bulan, J., Durante, R., Martillan, M., Addawe, M., and Roitner-Fransecky, I., (2024). Compiling granular population data using geospatial information. *Asian Development Review*, 41(1), 1-39.

Food and Agriculture Organization of the United Nations (FAO) and the World Bank. (2018). *Food Data Collection in Household Consumption and Expenditure Surveys: Guidelines for Low-and Middle-Income Countries*. Accessed on 2 June, 2022 from: https://openknowledge.worldbank.org/handle/10986/32503. Rome and Washington D.C.: Food and Agriculture Organization (FAO) and the World Bank.

Ghosh, T., Baugh, K., Elvidge, C., Zhizhin, M., Poyda, A., and Hsu, F-C. (2021). Extending the DMSP nighttime lights time series beyond 2013. *Remote Sensing*, *13*(24), 5004.

Gibson, J. (2019). Are you estimating the right thing? An editor reflects. *Applied Economic Perspectives and Policy*, *41*(3), 329-350.

Gibson, J. (2020). Measuring chronic hunger from diet snapshots. *Economic Development and Cultural Change*, 68(3), 813-838

Gibson, J. (2021). Better night lights data, for longer. *Oxford Bulletin of Economics and Statistics*, *83*(3), 770-791.

Gibson, J., and Boe-Gibson, G. (2021). Nighttime lights and county-level economic activity in the United States: 2001 to 2019. *Remote Sensing*, *13*(14), 2741.

Gibson, J., Datt, G., Murgai, R., and Ravallion, M. (2017). For India's rural poor, growing towns matter more than growing cities. *World Development*, *98*, 413-429.

Gibson, J., Jiang, Y., and Susantono, B. (2023). Revisiting the role of secondary towns: How different types of urban growth relate to poverty in Indonesia. *World Development*, *169*, 106281.

Gibson, J., Kim, B., and Olivia, S. (2011). Spatial correlation in household choices in rural Indonesia. *Asian Economic Journal*, *25*(3), 271-289.

Gibson, J., Olivia, S., and Boe-Gibson, G. (2020). Night lights in economics: Sources and uses. *Journal of Economic Surveys*, *34*(5), 955-980.

Gibson, J., Olivia, S., Boe-Gibson, G., and Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, *149*, 102602.

Grömping., U. (2015). Variable importance in regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7, 137-152.

Hall, O., Dompae, F., Wahab, I., and Dzanku, F. M. (2023). A review of machine learning and satellite imagery for poverty prediction: Implications for development research and applications. *Journal of International Development*. https://doi.org/10.1002/jid.3751

Henderson, V., Storeygard, A., and Weil, D. (2012). Measuring economic growth from outer space. *American Economic Review*, *102*(2), 994-1028.

Hossain, M., Das, S., Chandra, H., and Islam, M. (2020). Disaggregate level estimates and spatial mapping of food insecurity in Bangladesh by linking survey and census data. *PloS one*, *15*(4), e0230906.

Jean, N., Burke, M., Xie, M., Davis, W., Lobell, D., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790-794.

Jerven, M., and Johnston, D. (2015). Statistical tragedy in Africa? Evaluating the data base for African economic development. *The Journal of Development Studies*, *51*(2), 111-115.

Jing, X., Shao, X., Cao, C., Fu, X., and Yan, L. (2015). Comparison between the Suomi-NPP Day-Night Band and DMSP-OLS for correlating socio-economic variables at the provincial level in China. *Remote Sensing*, *8*(1), 17.

Keola, S., Andersson, M., and Hall, O. (2015). Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. *World Development*, *66*, 322-334.

Khachiyan, A., Thomas, A., Zhou, H., Hanson, G., Cloninger, A., Rosing, T., and Khandelwal, A. (2022). Using neural networks to predict micro-spatial economic growth. *American Economic Review: Insights*, *4*(4), 491-506.

Levin, N., Kyba, C. C. M., Zhang, Q., Sánchez de Miguel, A., Román, M. O., Li, X., Portnov, B. A., Molthan, A. L., Jechow, A., Miller, S. D., Wang, Z., Shrestha, R. M., and Elvidge, C. D. (2020). Remote sensing of night lights: A review and an outlook for the future. *Remote Sensing of Environment*, 237, 111443.

Li, X., Li, X., Li, D., He, X., and Jendryke, M. (2019). A preliminary investigation of Luojia-1 night-time light imagery. *Remote Sensing Letters*, *10*(6), 526-535.

Li, X., Liu, Z., Chen, X., and Sun, J. (2019). Assessing the ability of Luojia 1-01 imagery to detect feeble nighttime lights. *Sensors*, *19*(17), 3708.

Lin, J., Luo, S., and Huang, Y. (2022). Poverty estimation at the county level by combining LuoJia1-01 nighttime light data and points of interest. *Geocarto International*, *37*(12), 3590-3606.

Liu, H., Luo, N., and Hu, C. (2020). Detection of county economic development using LJ1-01 nighttime light imagery: A comparison with NPP-VIIRS data. *Sensors*, *20*(22), 6633.

Lobell, D., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., and Murray, S. (2020). Eyes in the sky, boots on the ground: Assessing satellite-and ground-based approaches to crop yield measurement and analysis. *American Journal of Agricultural Economics*, *102*(1), 202-219.

Mahler, D., Castañeda Aguilar, R., and Newhouse, D. (2022). Nowcasting global poverty. *World Bank Economic Review*, *36*(4), 835-856.

Maldonado, L. (2023). Living in darkness: rural poverty in Venezuela. *Journal of Applied Economics*, *26*(1), 2168464.

Olivia, S., Gibson, J., Rozelle, S., Huang, J., and Deng, X. (2011). Mapping poverty in rural China: how much does the environment matter? *Environment and Development Economics*, 16(2), 129-153.

Park, A., Wang, S., and Wu, G. (2002). Regional poverty targeting in China. *Journal of Public Economics*, 86(1), 123-153.

Puttanapong, N., Prasertsoong, N., and Peechapat, W. (2023). Predicting provincial Gross Domestic Product using satellite data and machine learning methods: A case study of Thailand. *Asian Development Review*, 40(2), 39-85.

Román, M., Wang, Z., Sun, Q., Kalb, V., Miller, S., Molthan, A., Schultz, L., Bell, J., Stokes, E., Pandey, B., and Seto, K. (2018). NASA's Black Marble nighttime lights product suite. *Remote Sensing of the Environment***,** *210*, 113-43.

Sharp, M. K., Buffière, B., Himelein, K., Troubat, N., and Gibson, J. (2022). Effects of data collection methods on estimated household consumption and survey costs. *Policy Research Working Paper* No. 10029, The World Bank.

Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., Chen, Z., Chen, L., and Wu, J. (2014). Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross

domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data. *Remote Sensing*, *6*(2), 1705-1724.

Smith, B., and Wills, S. (2018). Left in the dark? Oil and rural poverty. *Journal of the Association of Environmental and Resource Economists*, *5*(4), 865-904.

Song, W., and Deng, X. (2017). Land-use/land-cover change and ecosystem service provision in China. *Science of the Total Environment*, *576*, 705-719.

Tuttle, B., Anderson, S., Sutton, P., Elvidge, C., and Baugh, K. (2013). It used to be dark here. *Photogrammetric Engineering and Remote Sensing*, *79*(3), 287-297.

Watmough, G., Marcinko, C., Sullivan, C., Tschirhart, K., Mutuo, P., Palm, C., and Svenning, J. (2019). Socio-ecologically informed use of remote sensing data to predict rural household poverty. *Proceedings of the National Academy of Sciences*, *116*(4), 1213-1218.

World Bank. (2021). *World Development Report 2021: Data for Better Lives.* World Bank Group, Washington DC.

Wu, J., Zhang, Z., Yang, X., and Li, X. (2021). Analyzing pixel-level relationships between Luojia 1-01 nighttime light and urban surface features by separating the pixel blooming effect. *Remote Sensing*, *13*(23), 4838.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., and Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, *11*(1), 2583.

Zhang, G., Guo, X., Li, D., and Jiang, B. (2019). Evaluating the potential of LJ1-01 nighttime light data for modelling socio-economic parameters. *Sensors*, *19*(6), 1465.

Zhang, C., Pei, Y., Li, J., Qin, Q., and Yue, J. (2020). Application of Luojia 1-01 nighttime images for detecting the light changes for the 2019 spring festival in western cities, China. *Remote Sensing*, *12*(9), 1416.

Zhang, X., and Gibson, J. (2022). Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. *Remote Sensing*, *14*(5), 1282.

Table 1. Attributes of the various NTL data sources and data products

	DMSP	VNL	Black Marble	LJ1-01
Satellite/Sensor Attributes				
Operator	US DoD	NASA/NOAA		Wuhan University
Sensor	OLS	VIIRS		CMOS
Available years	1992-2019	2012-present		June 2018- March 2019
Wavelength range	500-900 nm	500-900 nm		460-980 nm
Orbit type and height	Polar, 850 km	Polar, 827 km		Polar, 645 km
Spatial resolution at nadir	2.7 km	742 m		130 m
Swath	3000 km	3000 km		250 km
Revisit time	12 hr	12 hr		3-5 days
Pixel saturation	Saturated	Not saturated		Not saturated
On-board calibration	No	Yes		Yes
Data Products				
Creator of annual composites	EOG	EOG	NASA	Wuhan University
Tiled	No	No	Yes, 648 tiles	Yes
Quantization	6-bit (n=64)	14 bit (n=16,384)	16 bit (n=65,536)	Digital number (DN) values
Masking of ephemeral light sources	No	Yes	Yes	Yes
Stray-light correction	No	Yes, from 2014	Yes	Yes
User control over angle of detection	No	No	Yes	No
Treatment of snow	No	No	Yes	No

Note: DoD is Department of Defense, OLS is Operational Linescan System, VIIRS is Visible Infrared Imaging Radiometer Suite, EOG is Earth Observation Group

Table 2. Relationships between Landsat detected built-up area and luminosity (sum of lights)

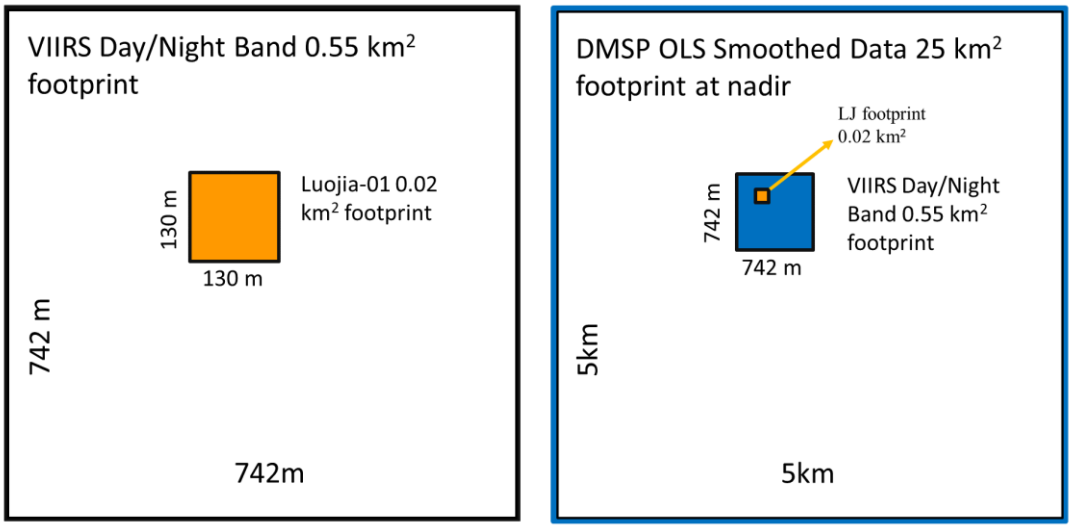
	(1)	(2)	(3)	(4)
	---------- DMSP ----------			
Urban (towns and cities)	0.045***			0.028***
	(0.004)			(0.002)
Village residential area		0.008***		0.004***
		(0.000)		(0.000)
Industrial/infrastructural			0.019***	0.015***
			(0.004)	(0.001)
Adjusted R^2	**0.303**	**0.256**	**0.105**	**0.416**
	---------- VNL ----------			
Urban (towns and cities)	0.037***			0.021***
	(0.003)			(0.003)
Village residential area		0.007***		0.004***
		(0.000)		(0.000)
Industrial/infrastructural			0.015***	0.013***
			(0.003)	(0.003)
Adjusted R^2	**0.274**	**0.251**	**0.100**	**0.391**
	---------- Black Marble ----------			
Urban (towns and cities)	0.041***			0.030***
	(0.004)			(0.004)
Village residential area		0.006***		0.003***
		(0.000)		(0.000)
Industrial/infrastructural			0.018***	0.015***
			(0.004)	(0.003)
Adjusted R^2	**0.276**	**0.177**	**0.114**	**0.366**
	---------- Luojia-01 ----------			
Urban (towns and cities)	0.028***			0.019***
	(0.002)			(0.002)
Village residential area		0.005***		0.002***
		(0.000)		(0.000)
Industrial/infrastructural			0.013***	0.011***
			(0.003)	(0.002)
Adjusted R^2	**0.194**	**0.134**	**0.088**	**0.269**

Note: The dependent variable is county-level luminosity. Intercepts are not reported to save space. Built-up area is measured in km2 and the sum of lights in logs, so coefficients are (approximately) percentage differences in luminosity per 1 km^2 of each type of built up area. The land use for industrial and infrastructural land includes factories, mines and quarries, airports and roads. Based on n=1460 counties, with robust standard errors in ( ) and ***, **, and * denoting statistica significance at 1%, 5% and 10%.

Table 3. Heterogeneity analysis: comparing predictive power of night-time lights data for total GDP in low and high density areas of China

	No control variables			Controls for land cover and environmental factors		
	All regions	Low density	High density	All regions	Low density	High density
DMSP	0.700***	0.392***	0.586***	0.343***	0.365***	0.275***
	(0.019)	(0.024)	(0.032)	(0.021)	(0.026)	(0.043)
Adjusted R^2	**0.490**	**0.265**	**0.310**	**0.745**	**0.577**	**0.527**
VIIRS Night Lights	0.714***	0.409***	0.485***	0.347***	0.354***	0.165***
	(0.018)	(0.023)	(0.034)	(0.020)	(0.024)	(0.036)
Adjusted R^2	**0.510**	**0.301**	**0.217**	**0.752**	**0.590**	**0.513**
Black Marble	0.759***	0.437***	0.645***	0.374***	0.318***	0.347***
	(0.017)	(0.024)	(0.029)	(0.020)	(0.024)	(0.037)
Adjusted R^2	**0.577**	**0.313**	**0.399**	**0.759**	**0.571**	**0.552**
Luojia-01	0.431***	0.255***	0.419***	0.150***	0.145***	0.124***
	(0.024)	(0.024)	(0.026)	(0.016)	(0.021)	(0.027)
Adjusted R^2	**0.185**	**0.133**	**0.260**	**0.716**	**0.497**	**0.513**

Notes: The dependent variable is the logarithm of GDP, and the independent variable is the logarithm of the sum of light. The sample is split amongst 1460 counties into the above median and below median population density sub-samples. The low-density and high density subsamples have 731, and 729 observations, respectively. Robust standard errors in ( ), ***, **, and * denote statistical significance at 1%, 5% and 10% levels. The control variables in columns (4) to (6) are elevation, precipitation, temperature, and five land cover classes (three types of urban land, cultivated land, and forest).



(a) Luojia-01 (LJ) versus VIIRS (VNL and BM) (b) LJ versus VIIRS versus DMSP

Figure 1. The DMSP-OLS nighttime visible band are collected with a 5 km×5 km footprint (after on-board averaging), VIIRS DNB data, are collected with a 742 m×742 m pixel footprint from nadir out to the edge of scan, and Luojia-01 (LJ) nighttime data has a 130 m×130 m footprint. (After: Elvidge et al (2013) "Why VIIRS data are superior to DMSP for mapping night time lights")

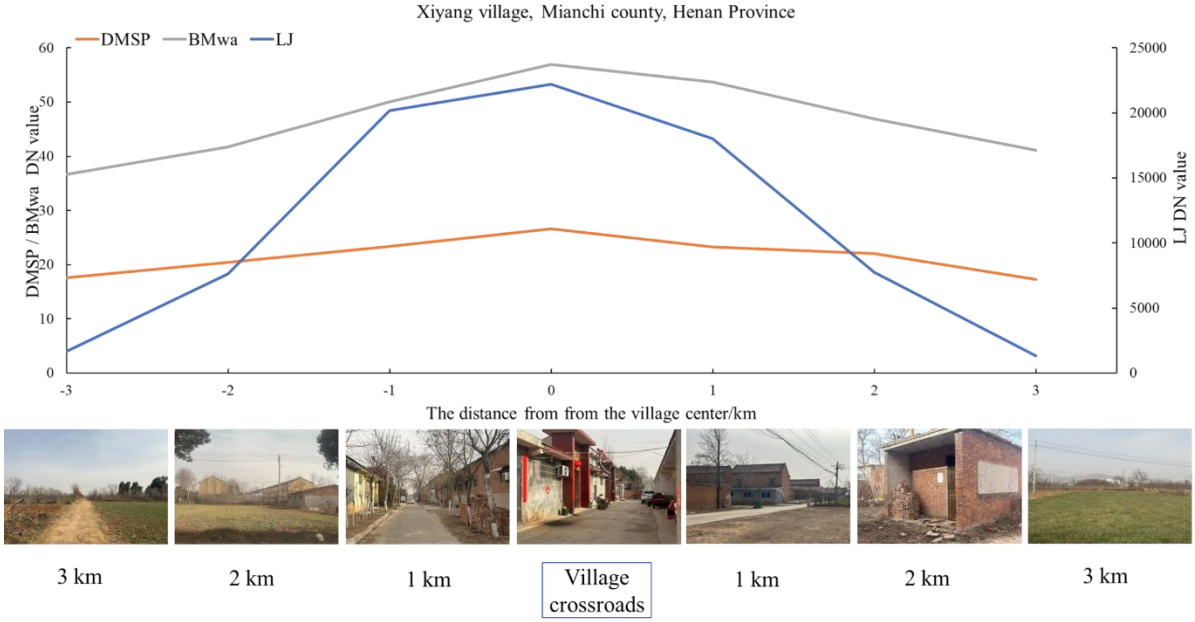


Figure 2. Example of a ground-truthing transect: Xiyang village, Mianchi county, Henan Province

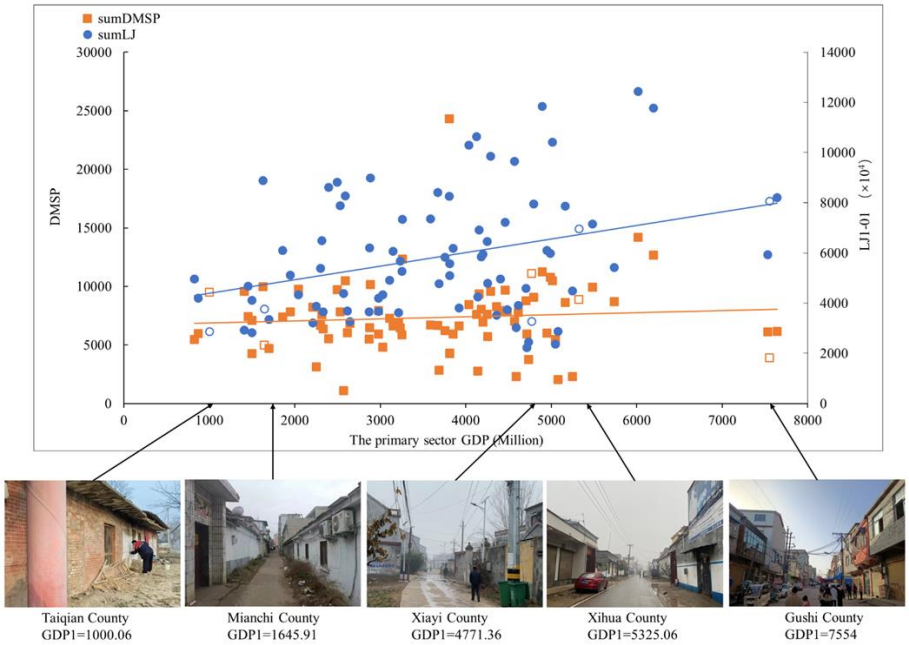


Figure 3. Example of a Primary sector GDP—luminosity curve: Henan Province (hollow symbols are counties where the ground-truthing photographs are taken)

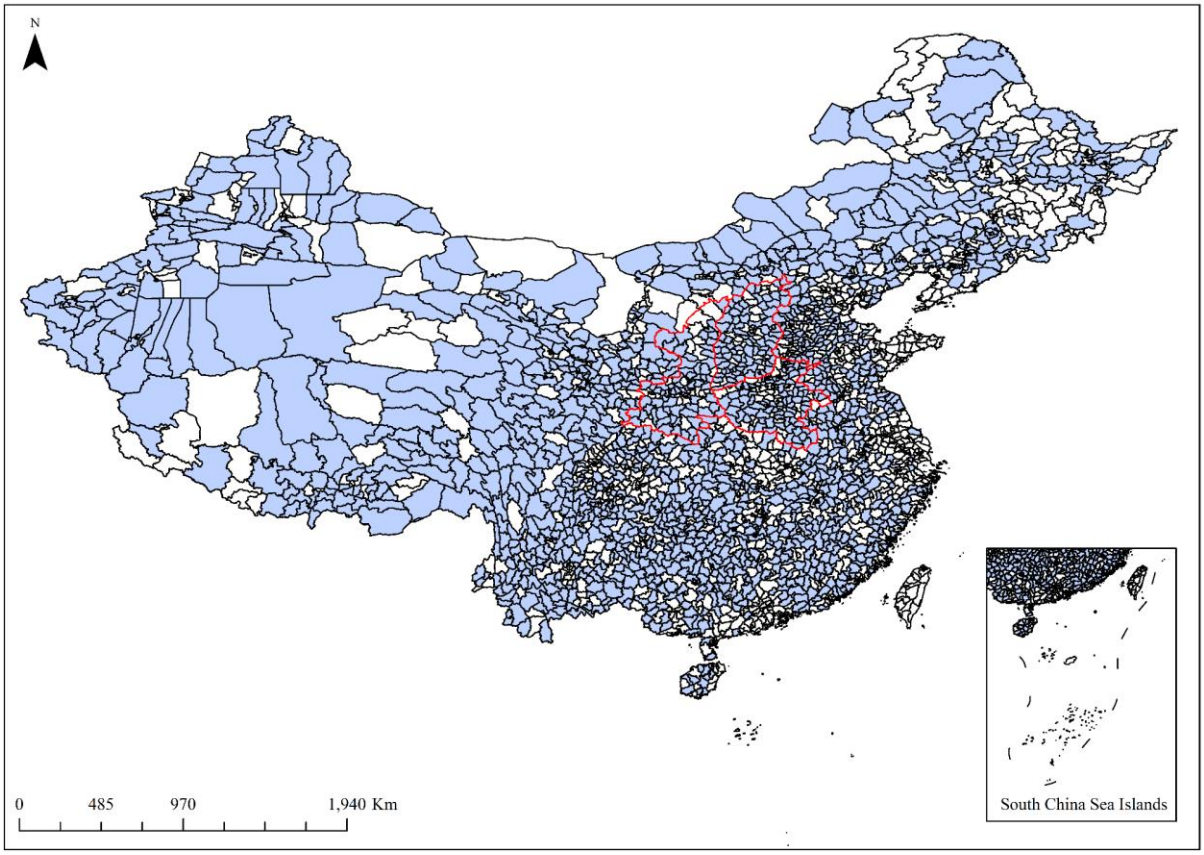


Figure 4. Location map for the *n*=1460 rural counties (provinces with red borders had ground-truthing exercises)

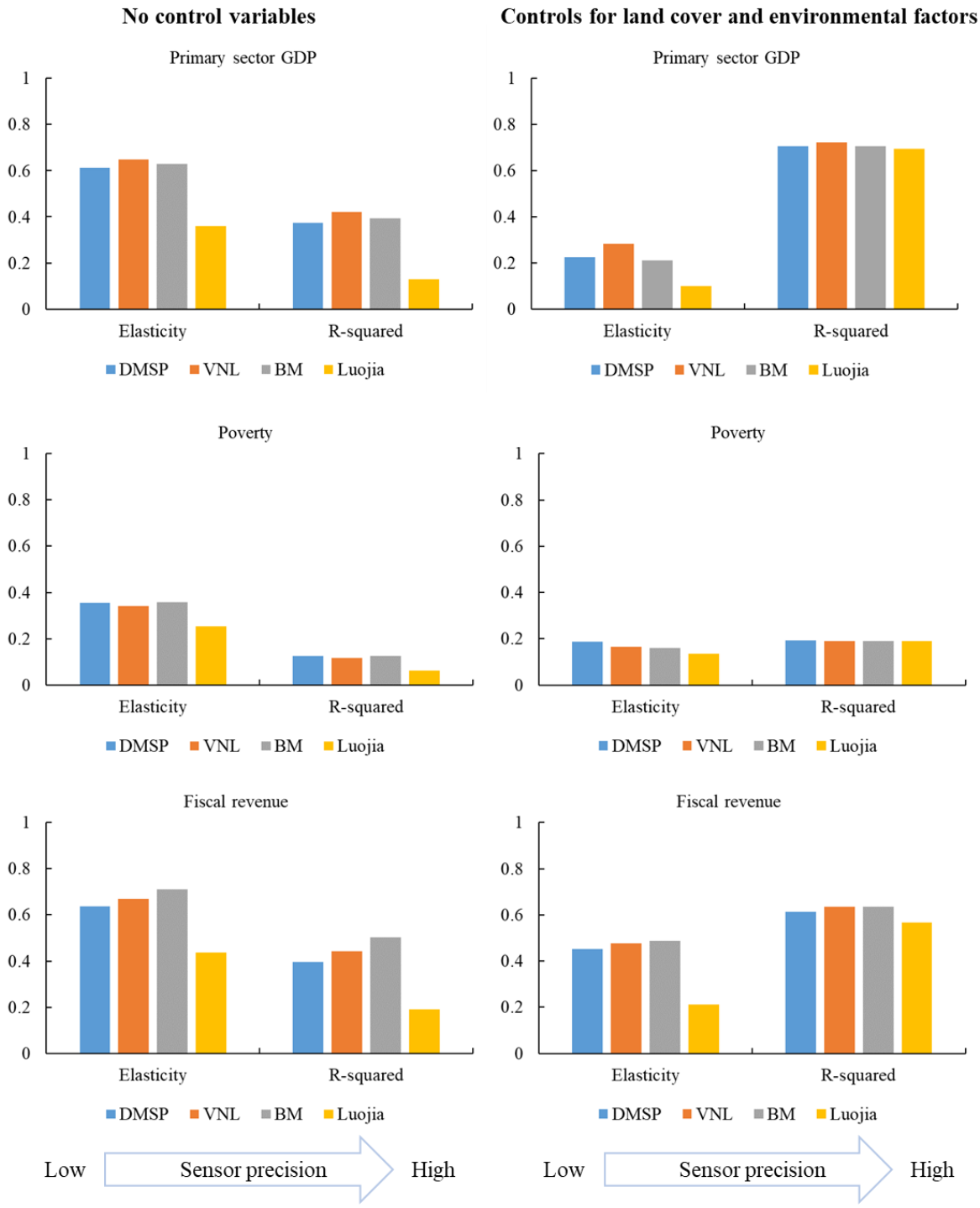
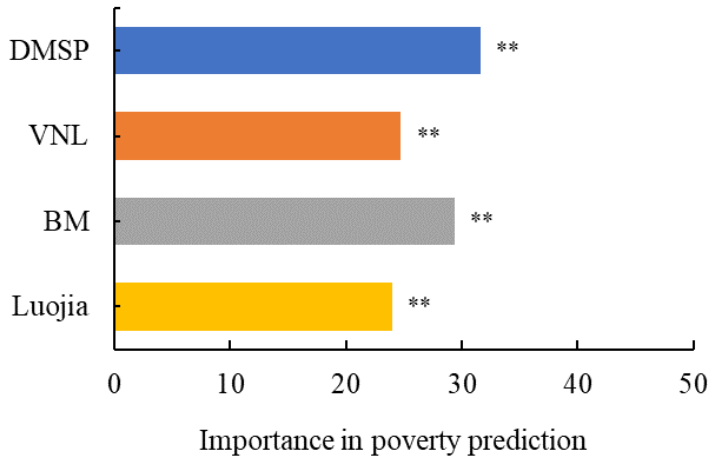


Figure 5. Variation in the Predictive Performance of NTL data from Sensors of Varying Precision Notes: Based on *n*=1460 counties, with full results reported in Tables 1-3 of Appendix C. The sign of the elasticities of poverty with respect to luminosity are reversed for display purposes. The luminosity indicator used is the sum of lights. The R-squared reported is the adjusted R-squared to account for the different number of covariates.

(a)



(b)

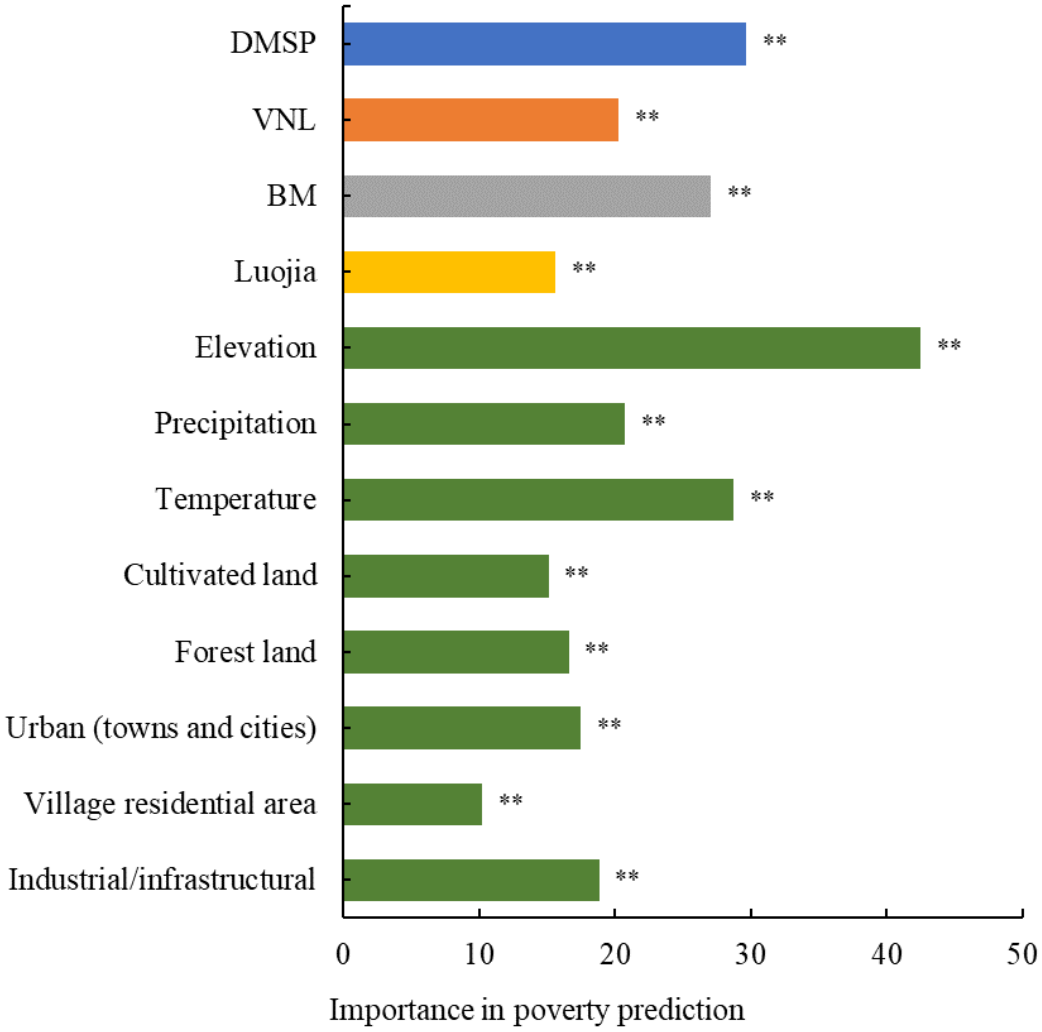


Figure 5. Average contributions of each NTL variable to poverty predictions. Notes: (a) only includes NTL variables, (b) also includes environmental variables.

Appendix A: Details on the Ground-Truthing Exercises

We selected sites in three provinces (Henan, Shanxi, and Shaanxi) for ground-truthing exercises that aimed to provide first-hand insight into the ability of Luojia, BM, VNL and DMSP to detect small scale rural development. We chose these three provinces because of their advantages of adjacency and centrality (for the last several decades both the population-weighted and the GDP-weighted center of China have been located in Henan). As a check on whether these provinces had significant deviations from China's 'typical' relationships between rural economic activity and luminosity we first estimated province-by-province elasticities of primary sector GDP with respect to luminosity (using double-log specifications). The results are in Table 1. For Luojia-01, which is our main sensor of interest, there was no difference between the GDP-luminosity elasticities for these three provinces and for the other provinces (a mean difference of -0.08, $p<0.54$). Likewise, the R^2 of the prediction equations was not significantly lower for these three provinces ($p<0.34$). Hence, we expect that our ground-truthing exercises should be broadly informative.

Within these three provinces, we chose ground-truthing sites that took into account varying wealth levels and varying levels of economic activity, while also achieving a spatial spread of locations in each province (Figures 1a, 2a, and 3a shows these locations). We took a field trip to each location, and selected villages for more detailed study by taking transects from the village centers. We then plot radiance along the same transect. For example, in Figures 1c, 2c, and 3c it can be seen that Luojia-01 radiance is very low/zero outside the village and then spikes in the village center and goes back to being very low. In contrast, DMSP tends to not show much variation along the transect, most likely because of blurring problems. Even though the resolution of the Black Marble and VNL data is much better than for DMSP, these sources also tended to obscure the differences seen on the ground along the village transects. In other words, at a very fine scale the Luojia data seem to have more sensitivity to reflect economic activities in China's rural areas that are less well detected by the other sensors.

Moving from the village to the county level, we plotted primary sector GDP against luminosity and illustrated at various points the difference in on-the-ground conditions (Figures 1b, 2b, 3b). For example, in Henan province there is no relationship between luminosity and county primary sector GDP when using DMSP data, while the Luojia data are positively correlated with GDP (and more closely related to GDP than are VNL or BM data). The scatter plots for the other two provinces do not show such marked contrasts between the various remote sensing data sources.

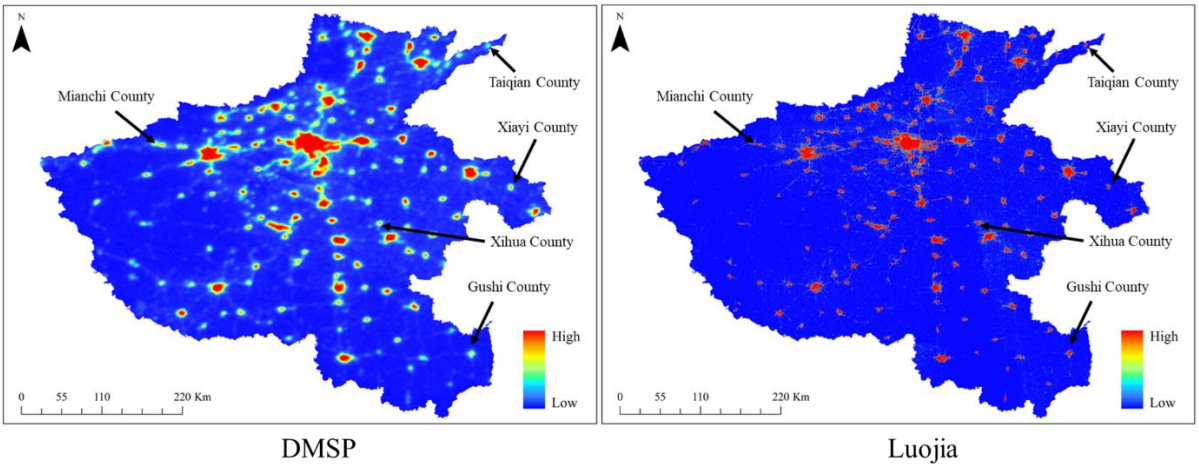
Table 1. Province-by-province elasticities of primary sector GDP with respect to luminosity

	DMSP	VNL	Black Marble	LuoJia-01
Ground-truthing provinces				
Shanxi	0.463***	0.364**	0.396***	0.275*
	(0.115)	(0.122)	(0.099)	(0.118)
Hunan	0.697***	0.794***	0.596***	0.738***
	(0.116)	(0.105)	(0.098)	(0.121)
Shaanxi	0.325***	0.433***	0.314***	0.339***
	(0.072)	(0.087)	(0.069)	(0.099)
Other provinces				
Hebei	0.343***	0.320**	0.297***	0.233**
	(0.098)	(0.104)	(0.088)	(0.072)
Inner Mongolia	0.505***	0.627***	0.474***	0.251*
	(0.106)	(0.108)	(0.097)	(0.130)
Liaoning	-0.016	0.179	0.109	0.185
	(0.104)	(0.120)	(0.107)	(0.166)
Jilin	0.815**	1.114***	0.812**	0.575*
	(0.287)	(0.268)	(0.262)	(0.228)
Heilongjiang	0.565***	0.620***	0.467***	-0.026
	(0.094)	(0.106)	(0.090)	(0.178)
Jiangsu	0.439*	0.515***	0.391*	0.463*
	(0.177)	(0.127)	(0.156)	(0.175)
Zhejiang	0.432***	0.329*	0.551***	0.445*
	(0.102)	(0.165)	(0.104)	(0.190)
Anhui	0.468***	0.557***	0.502***	0.697***
	(0.045)	(0.058)	(0.051)	(0.124)
Fujian	0.379***	0.533***	0.354***	0.421***
	(0.087)	(0.106)	(0.083)	(0.116)
Jiangxi	0.540***	0.562***	0.487***	0.586***
	(0.071)	(0.092)	(0.078)	(0.086)
Shandong	0.094	0.272*	0.304**	0.278*
	(0.128)	(0.116)	(0.110)	(0.135)
Hubei	0.494***	0.450***	0.389**	0.502***
	(0.099)	(0.127)	(0.128)	(0.103)
Hunan	0.697***	0.794***	0.596***	0.738***
	(0.116)	(0.105)	(0.098)	(0.121)
Guangdong	0.565***	0.658***	0.547***	0.596***
	(0.116)	(0.141)	(0.116)	(0.132)
Guangxi	0.660***	0.697***	0.714***	0.604***
	(0.083)	(0.065)	(0.077)	(0.146)
Hainan	0.713**	0.658***	0.574**	0.755***

	(0.202)	(0.098)	(0.153)	(0.132)
Sichuan	0.808***	0.859***	0.839***	0.308**
	(0.073)	(0.077)	(0.058)	(0.116)
Guizhou	0.531***	0.531***	0.488***	0.146*
	(0.106)	(0.096)	(0.111)	(0.077)
Yunnan	0.544***	0.535***	0.524***	0.205**
	(0.063)	(0.062)	(0.062)	(0.068)
Tibet	0.083	0.549***	-0.056	0.540***
	(0.086)	(0.136)	(0.129)	(0.132)
Gansu	0.270**	0.513***	0.196*	0.048
	(0.093)	(0.099)	(0.082)	(0.082)
Qinghai	0.361***	0.435**	0.468***	0.200
	(0.096)	(0.142)	(0.123)	(0.258)
Ningxia	0.415	0.494*	0.250	0.266*
	(0.225)	(0.215)	(0.191)	(0.098)
Xinjiang	0.650***	0.726***	0.407***	0.151
	(0.120)	(0.096)	(0.121)	(0.123)

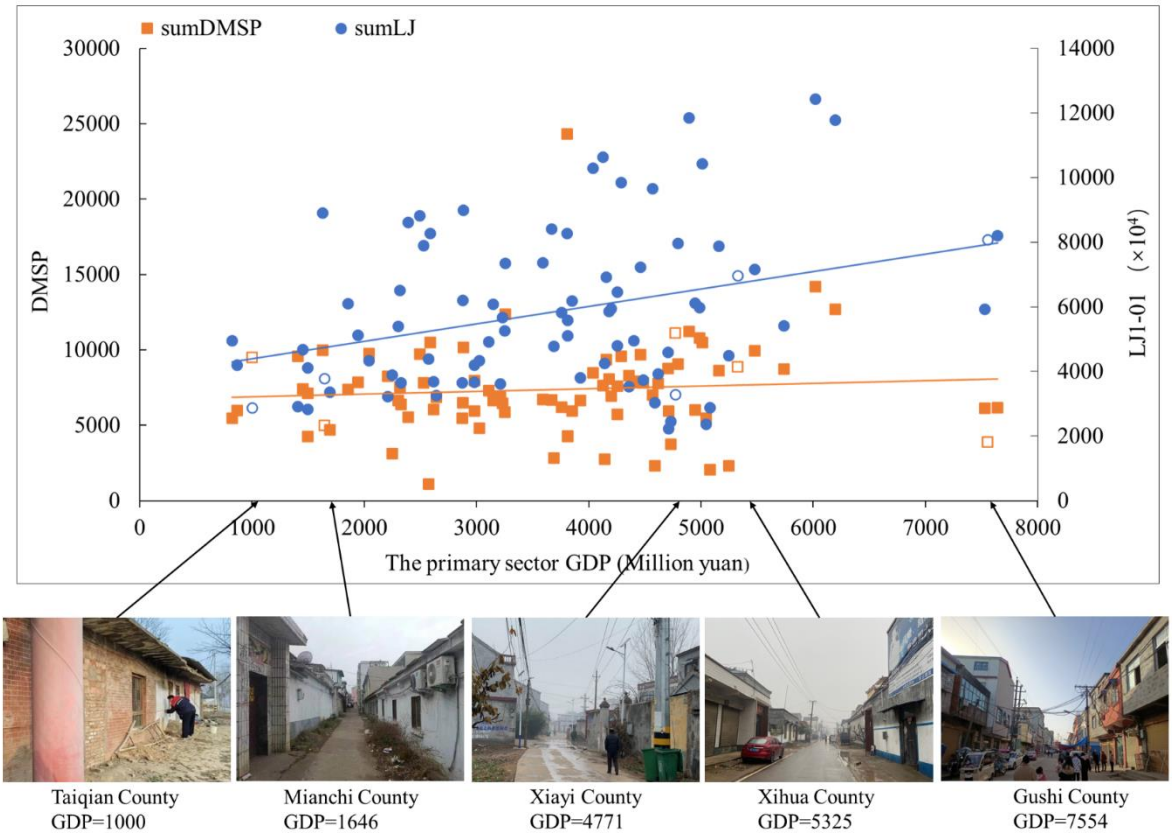
Notes: Standard errors in (), with statistical significance at 1%, 5%, or 10% denoted by ***, **, or *.

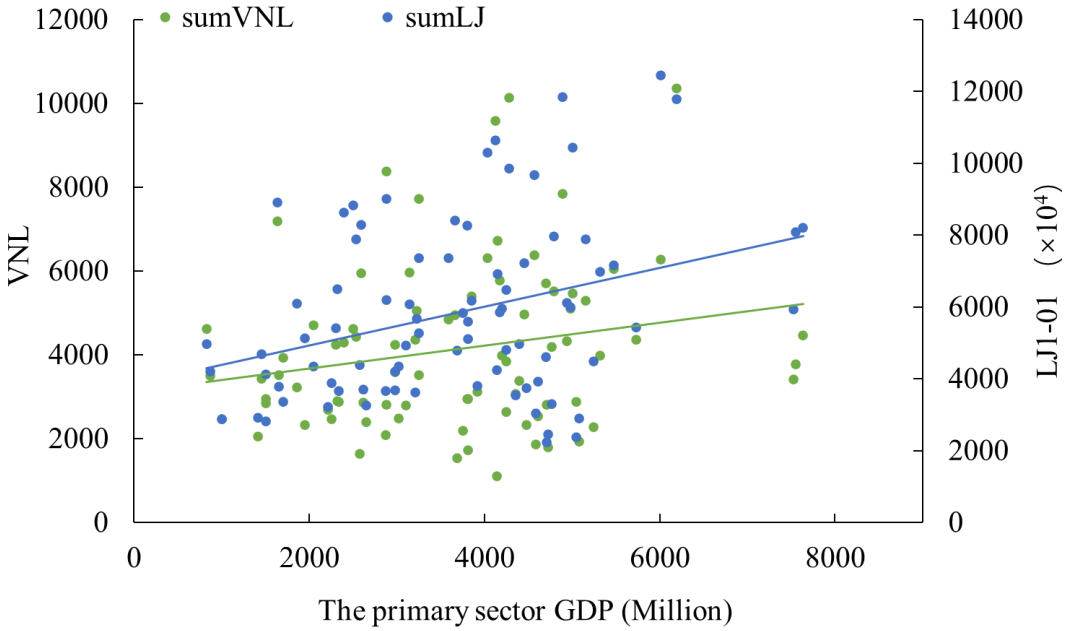
Figure 1a. Ground-truthing Map for Henan Province.



Note: The selected counties for ground-truthing are named in the image.

Figure 1b. Scatter plot of Henan Province.




sumVNL
sumLJ
VNL
LJ1-01 (×10⁴)
The primary sector GDP (Million)
0
2000
4000
6000
8000
10000
12000
14000

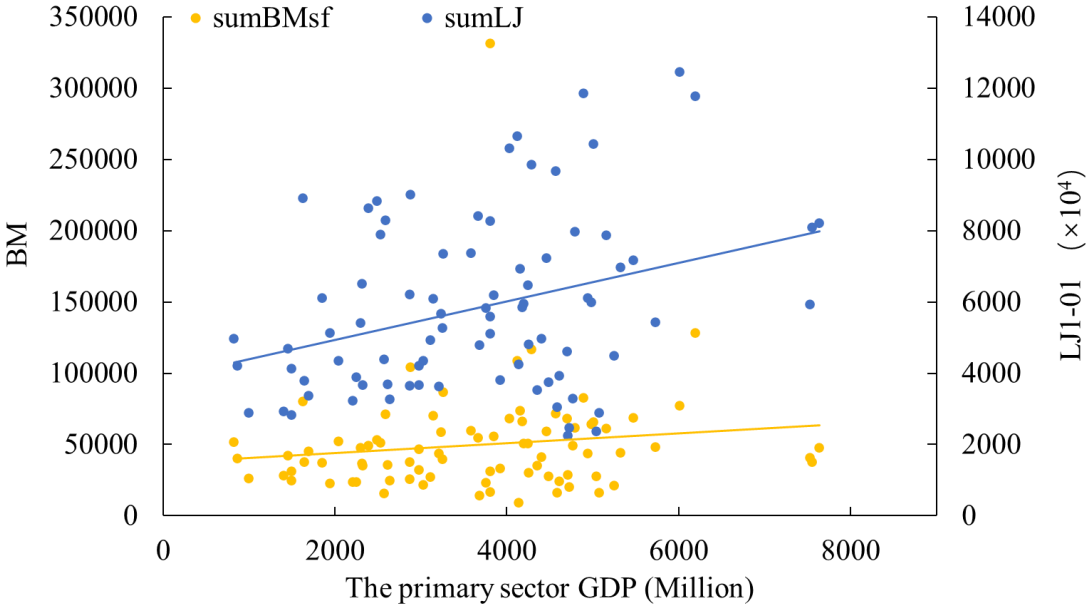

sumBMsf
sumLJ
BM
LJ1-01 (×10⁴)
The primary sector GDP (Million)
0
50000
100000
150000
200000
250000
300000
350000
2000
4000
6000
8000
10000
12000
14000

Figure 1c. The transects map for Xiyang village, Mianchi county, Henan Province.

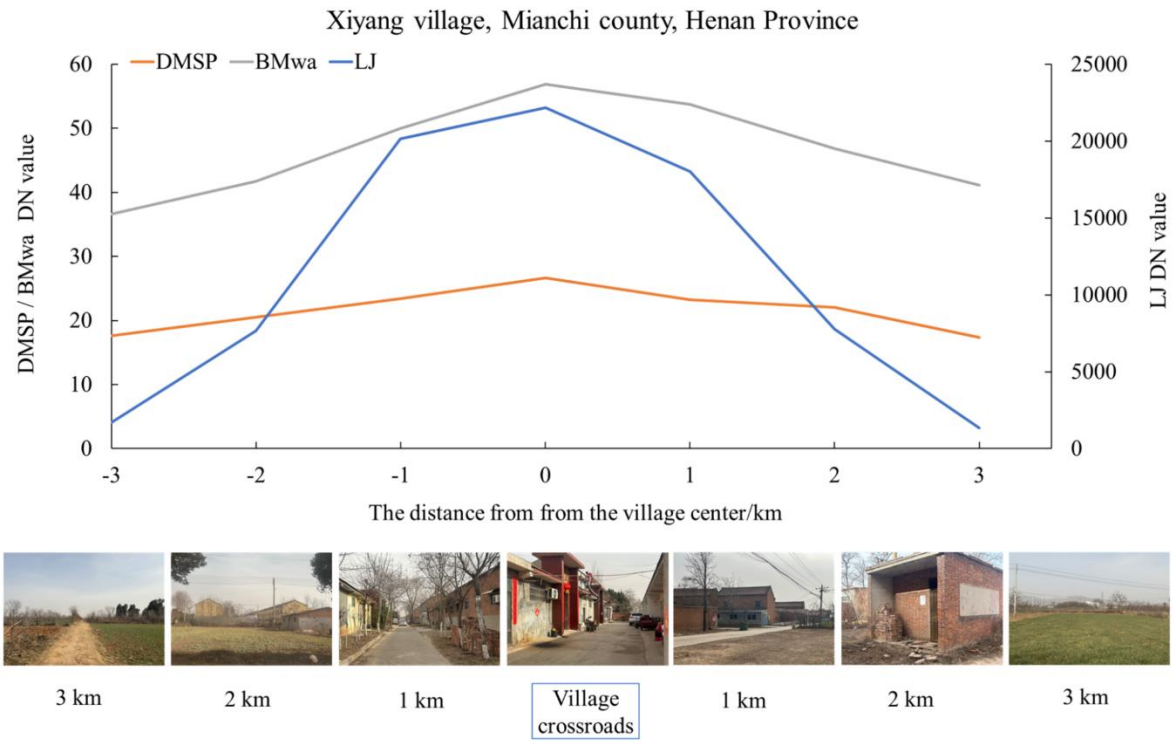
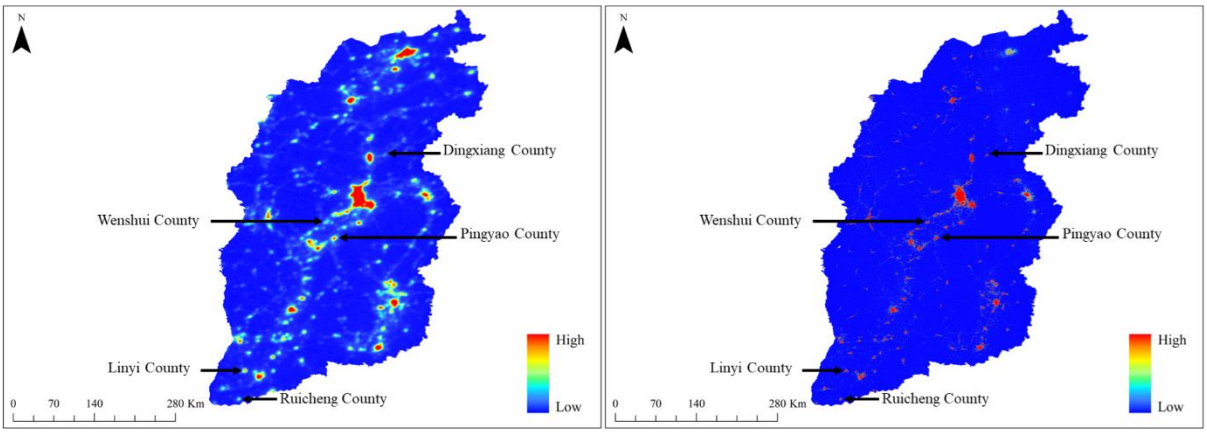
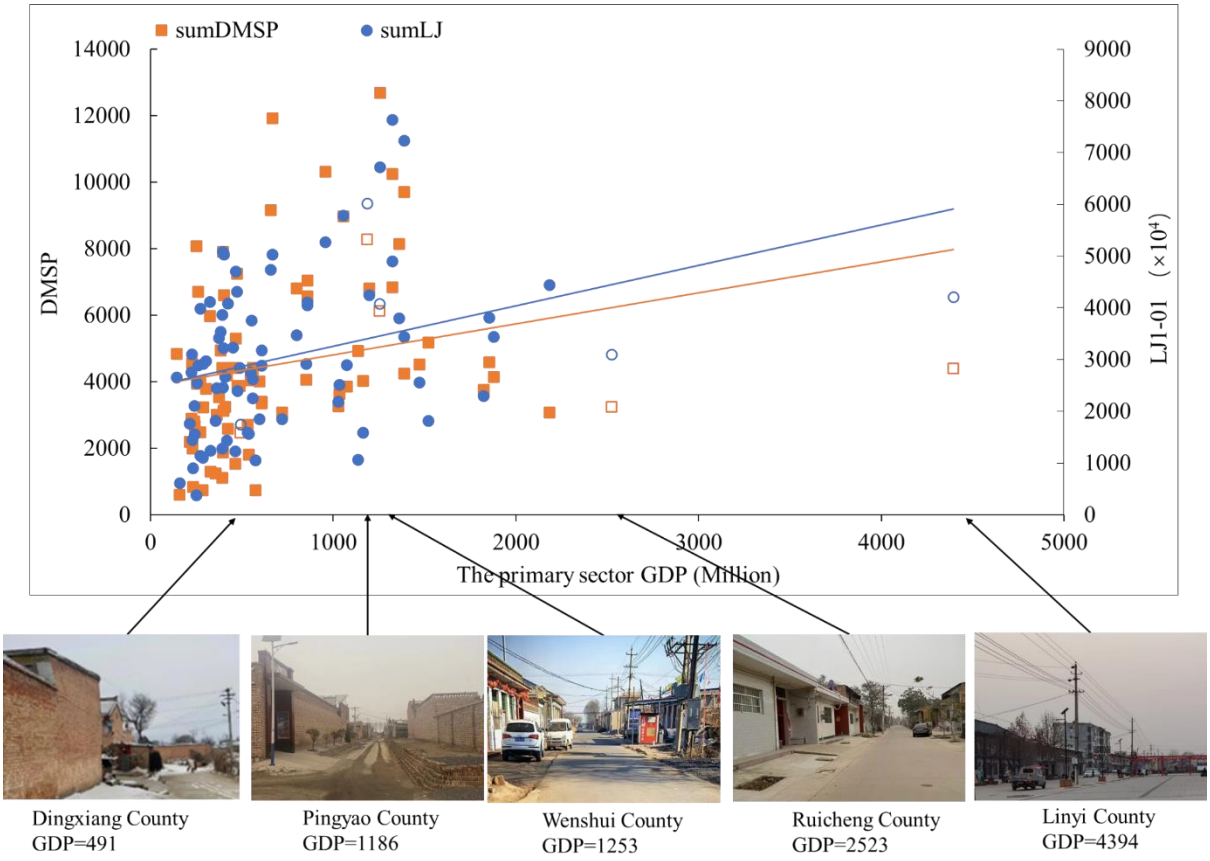


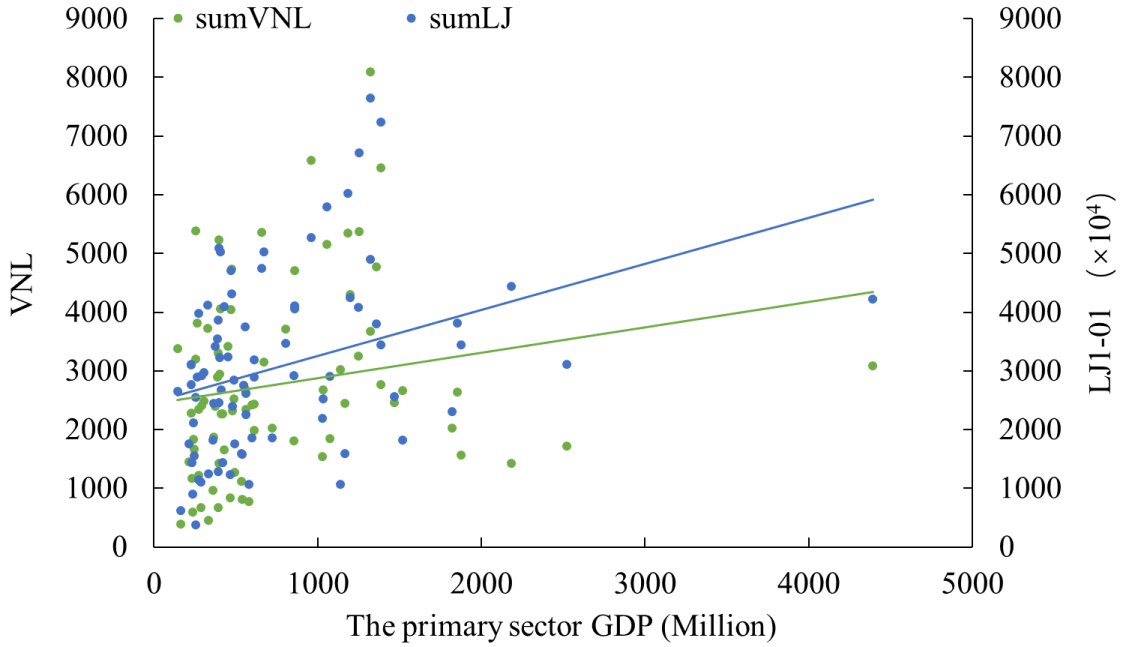
Figure 2a. Ground-truthing Map for Shanxi Province.



Note: The selected counties for ground-truthing are named in the image.

Figure 2b. Scatter plot of Shanxi Province.




sumVNL
sumLJ
VNL
LJ1-01 (×10⁴)
The primary sector GDP (Million)
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

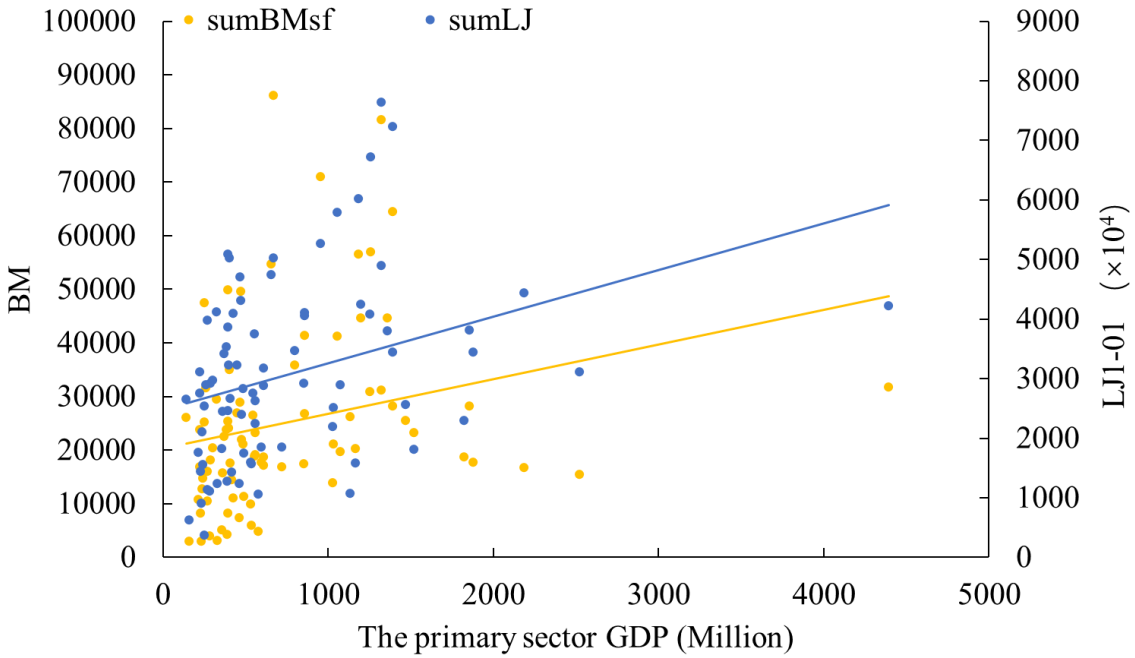

sumBMsf
sumLJ
BM
LJ1-01 (×10⁴)
The primary sector GDP (Million)
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
1000
2000
3000
4000
5000
6000
7000
8000
9000

Figure 2c. The transects map for Dajianbei village, Pinglu county, Shanxi Province.

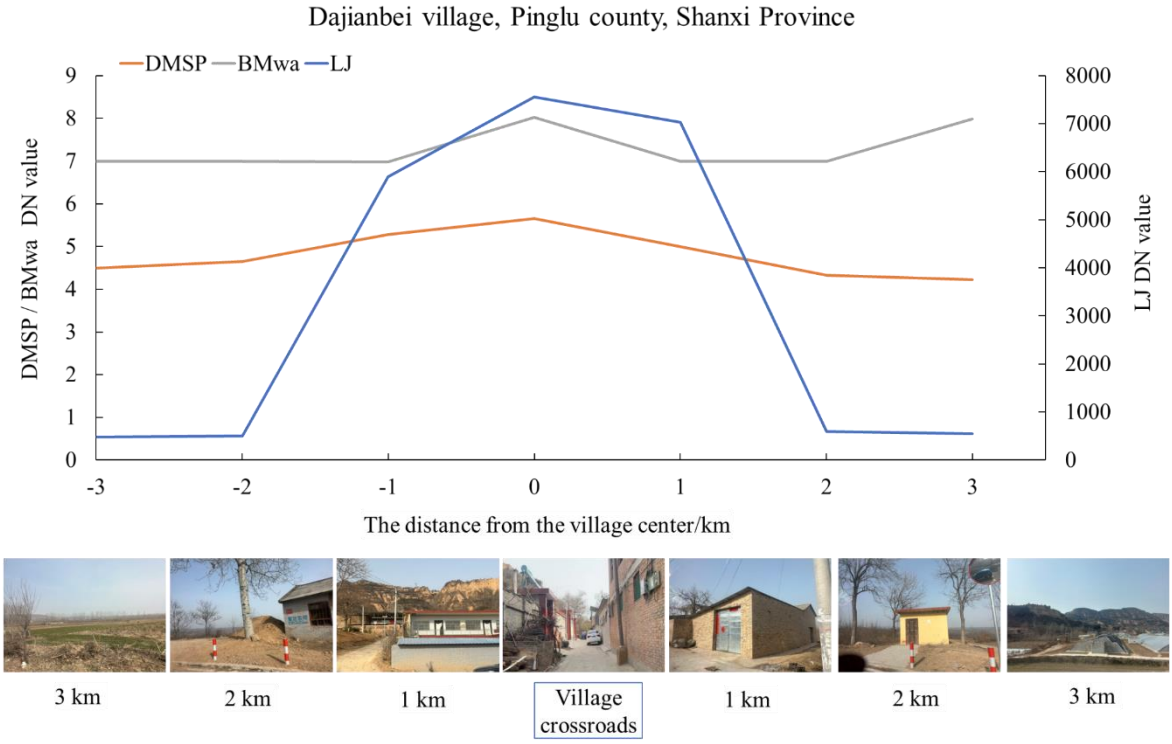
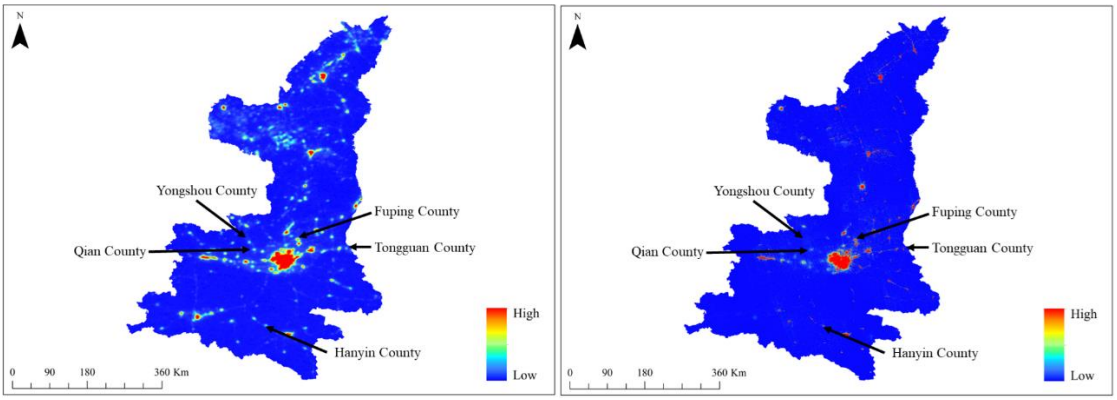
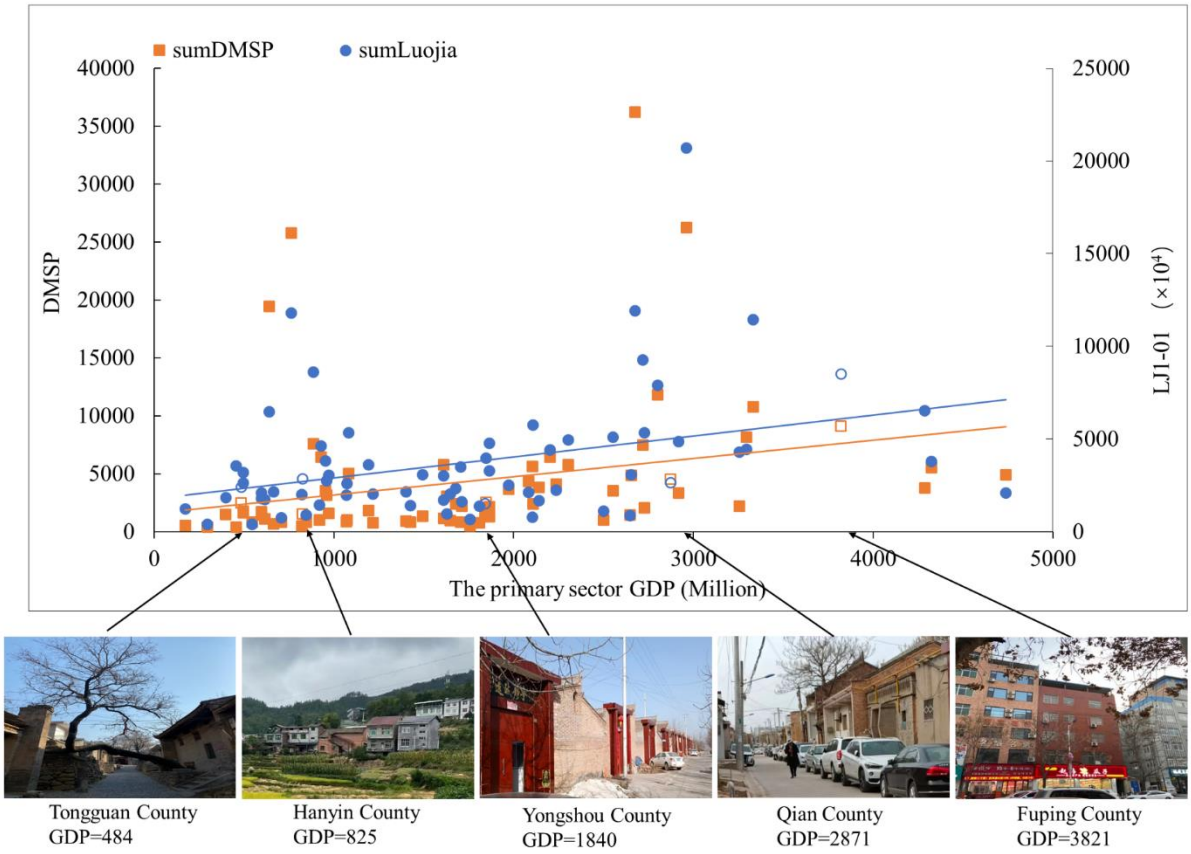


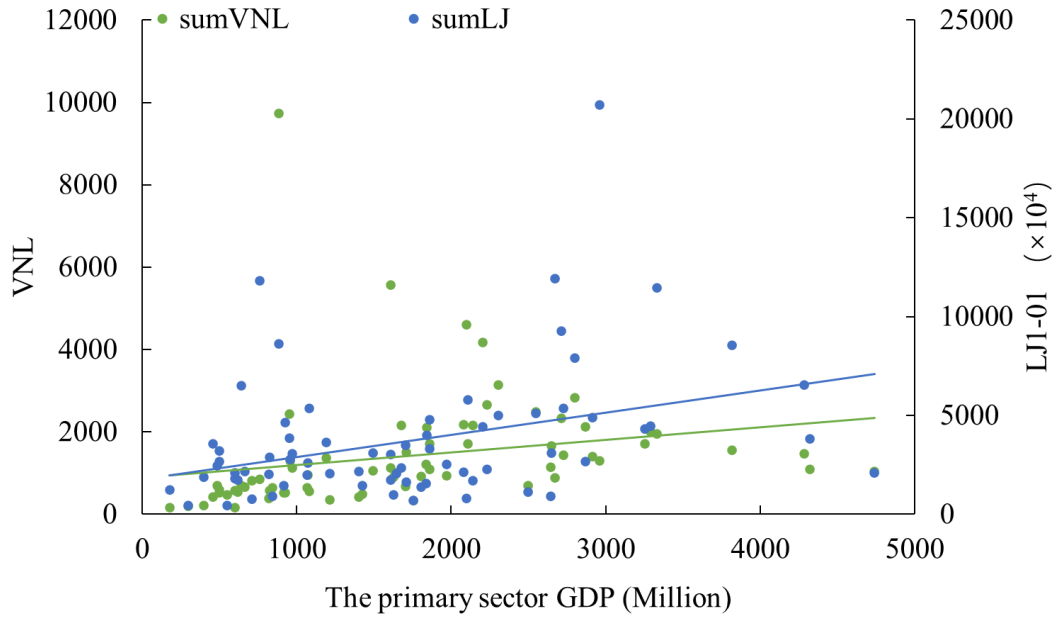
Figure 3a. Ground-truthing Map for Shaanxi Province.



Note: The selected counties for ground-truthing are named in the image.

Figure 3b. Scatter plot of Shaanxi Province.




sumVNL
sumLJ
VNL
LJ1-01 (×10⁴)
12000
10000
8000
6000
4000
2000
0
25000
20000
15000
10000
5000
0
0
1000
2000
3000
4000
5000
The primary sector GDP (Million)

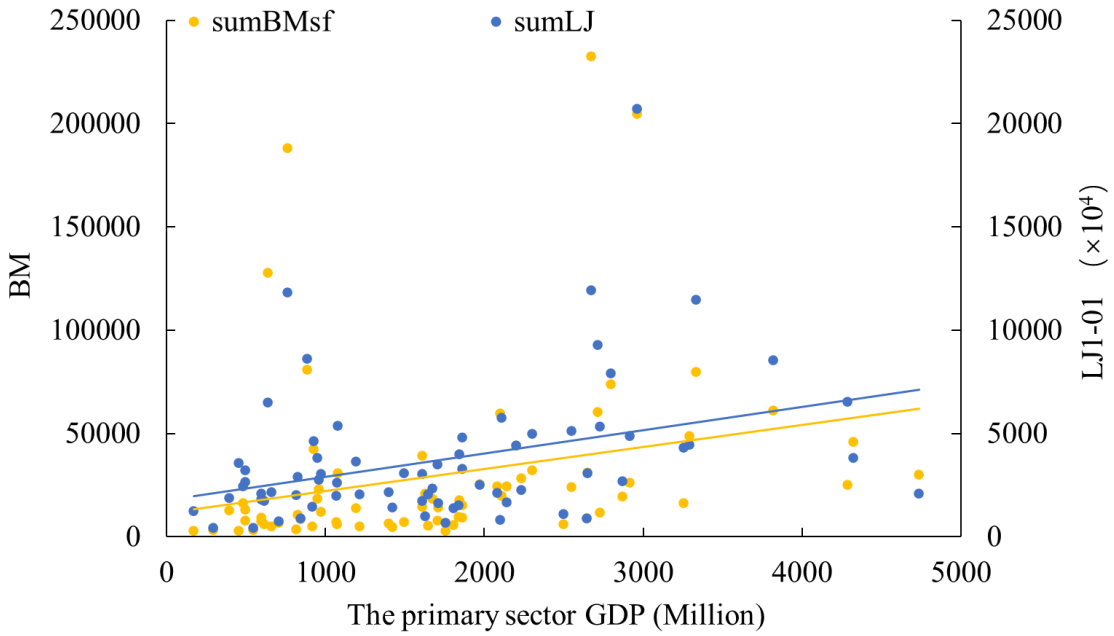
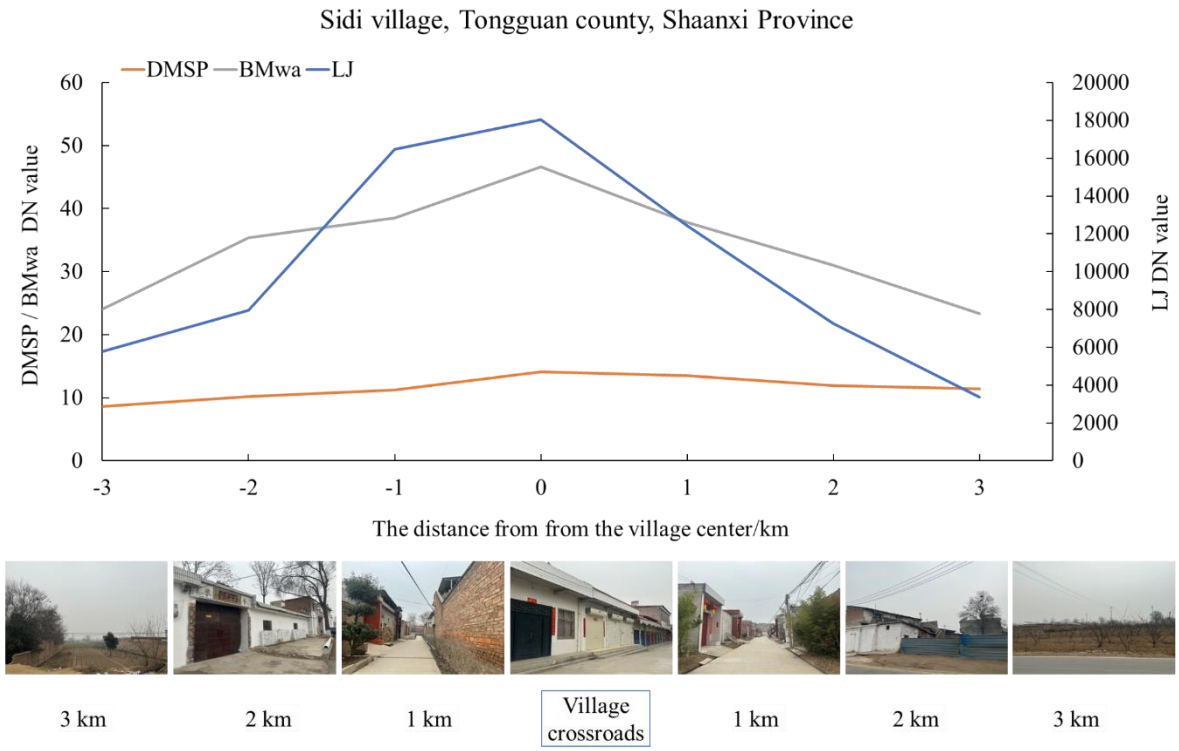
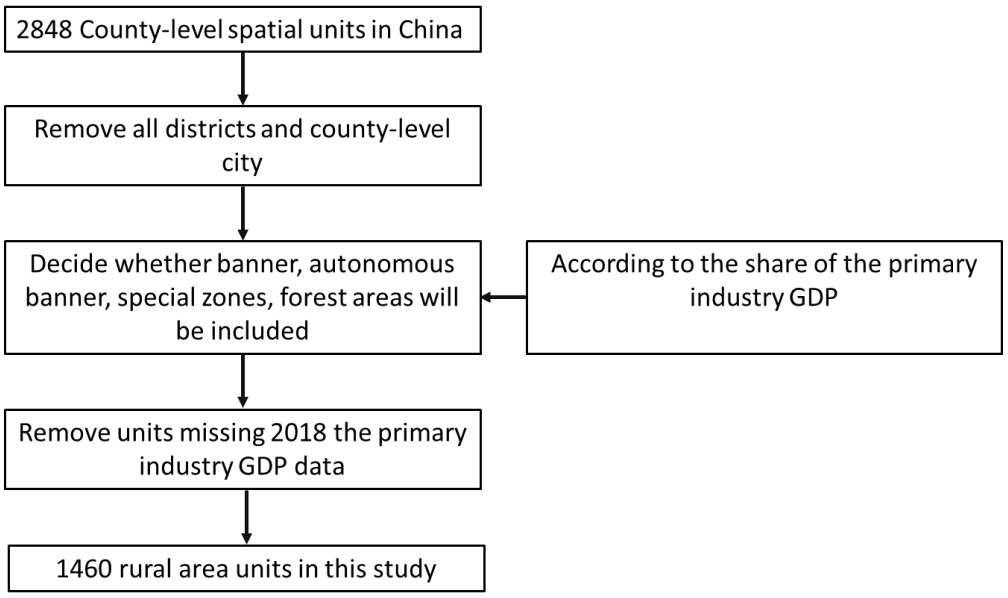

sumBMsf
sumLJ
BM
LJ1-01 (×10⁴)
250000
200000
150000
100000
50000
0
25000
20000
15000
10000
5000
0
0
1000
2000
3000
4000
5000
The primary sector GDP (Million)

Figure 3c. The transects map for Sidi village, Tongguan county, Shaanxi Province.



Appendix B: Sample Selection Flowchart



Co-authorship Forms

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 2
Zhang, X., and Gibson, J. (2022). Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. Remote Sensing, 14(5), 1282.

Nature of contribution by PhD candidate	Conceptualizing and designing the study, designing the empirical strategy, data cleaning, empirical analysis and writing of initial draft
Extent of contribution by PhD candidate (%)	75

CO-AUTHORS

Name	Nature of Contribution
John Gibson	Guidance, critical feedback, revising and dealing with the journal submission process

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
John Gibson	John Gibson	23/9/24

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work. Chapter 3 Zhang, X., Gibson, J., and Deng, X. (2023). Remotely too equal: Popular DMSP night time lights data understate spatial inequality. Regional Science Policy & Practice, 15(9), 2106-2126.	
Nature of contribution by PhD candidate	Conceptualizing and designing the study, designing the empirical strategy, data cleaning, empirical analysis and writing of initial draft
Extent of contribution by PhD candidate (%)	75

CO-AUTHORS

Name	Nature of Contribution
John Gibson	Guidance, critical feedback and assisting with the journal revision process
Xiangzheng Deng	Data provision, guidance and resourcing

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
John Gibson	John Gibson	23/9/24
Xiangzheng Deng	Xiangzheng Deng	2024.09.26

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4
Zhang, X., and Gibson, J., (2024). How well do gridded population estimates proxy for actual population changes? Evidence from four gridded data products and three population censuses for China.

Nature of contribution by PhD candidate	Conceptualizing and designing the study, designing the empirical strategy, data cleaning, empirical analysis and writing of initial draft
Extent of contribution by PhD candidate (%)	80

CO-AUTHORS

Name	Nature of Contribution
John Gibson	Guidance, critical feedback and assisting with the journal submission process

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
John Gibson	John Gibson	23/9/24

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 5
Zhang, X., Li, C., and Gibson, J. (2024). The role of spillovers when evaluating regional development interventions: evidence from administrative upgrading in China. Letters in Spatial and Resource Sciences, 17(1), 9.

Nature of contribution by PhD candidate	Conceptualizing and designing the study, designing the empirical strategy, data cleaning, empirical analysis and writing of initial draft
Extent of contribution by PhD candidate (%)	70

CO-AUTHORS

Name	Nature of Contribution
Chao Li	Data provision and data cleaning, guidance on the empirical analysis
John Gibson	Guidance, critical feedback and assisting with the journal revision process

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
Chao Li	Chao Li	24/09/2024
John Gibson	John Gibson	23/9/24

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 6
Zhang, X., and Gibson, J., (2024). Local economic effects of connecting to China's high-speed rail network: Evidence from spatial econometric models

Nature of contribution by PhD candidate	Conceptualizing and designing the study, designing the empirical strategy, data cleaning, empirical analysis and writing of initial draft
Extent of contribution by PhD candidate (%)	80

CO-AUTHORS

Name	Nature of Contribution
John Gibson	Guidance, critical feedback and assisting with the journal submission process

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
John Gibson	John Gibson	23/9/24

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 7
Zhang, X., Li, C., and Gibson, J., (2024). China's City Size Distribution: Diverging in terms of people and converging in terms of area

Nature of contribution by PhD candidate	Conceptualizing and designing the study, designing the empirical strategy, data cleaning, empirical analysis and writing of initial draft
Extent of contribution by PhD candidate (%)	70

CO-AUTHORS

Name	Nature of Contribution
Chao Li	Data provision and guidance on the empirical analysis
John Gibson	Guidance, critical feedback and assisting with the journal submission process

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
Chao Li	Chao Li	24/09/2024
John Gibson	John Gibson	23/9/24

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Appendix 1
Gibson, J., Jiang, Y., Zhang, X., and Boe-Gibson, G. (2024). Are disaster impact estimates distorted by errors in popular night-time lights data? Economics of Disasters and Climate Change, 1-26.

Nature of contribution by PhD candidate	Data cleaning, drafting text, and empirical analysis
Extent of contribution by PhD candidate (%)	30

CO-AUTHORS

Name	Nature of Contribution
John Gibson	Conceptualization, drafting and revisions
Yi Jiang	Conceptualization, data construction, drafting
Geua Boe-Gibson	Data acquisition, visualization

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
John Gibson	John Gibson	23/9/24
Yi Jiang	Yi Jiang	Sept. 23, 2024
Geua Boe-Gibson	GBoe	23/9/24

THE UNIVERSITY OF WAIKATO

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email SGR@waikato.ac.nz
Website http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Appendix 2
Gibson, J., Zhang, X., Park, A., Jiang, Y., and Li, X. (2024). Remotely measuring rural economic activity and poverty: Do we just need better sensors?

Nature of contribution by PhD candidate	Data cleaning, empirical analysis and writing of initial draft
Extent of contribution by PhD candidate (%)	25

CO-AUTHORS

Name	Nature of Contribution
John Gibson	Conceptualization, drafting and revisions
Albert Park	Resourcing, reviewing and guidance
Yi Jiang	Data provision, resourcing, drafting and revisions
Xi Li	Data provision and drafting

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
John Gibson	John Gibson	23/9/24
Albert Park	Albert Park	Sept 25, 2024
Yi Jiang	Yi Jiang	Sept. 23, 2024
Xi Li	Xi Li	25/9/2024