UNIVERSITY OF WAIKATO

Hamilton
New Zealand

Rising Regional Income Inequality in China: Fact or Artefact?

Chao Li and John Gibson

Department of Economics

Working Paper in Economics 09/12

July 2012

Corresponding Author

John Gibson

Economics Department
University of Waikato
Private Bag 3105
Hamilton, New Zealand

Tel: +64 (7) 838 4045

Email: jkgibson@waikato.ac.nz

Chao Li

Economics Department
University of Waikato
Private Bag 3105
Hamilton, New Zealand

Email: cl95@waikato.ac.nz

Abstract

China's local populations can be counted in two ways; by how many people have *hukou* household registration from each place and by how many people actually reside in each place. The counts differ by the non-*hukou* migrants – people that move from their place of registration – who have grown from fewer than five million when reform began in 1978 to over 200 million by 2010. For most of the first three decades of the reform era, the *hukou* count was used to produce *per capita* GDP figures. In coastal provinces the resident count is many millions more than the *hukou* count, while for migrant-sending provinces it is the reverse, creating a systematic and time-varying distortion in provincial GDP *per capita*. Moreover, a sharp discontinuity occurred when provinces recently switched from the *hukou* count to the resident count when reporting GDP *per capita*. A double-count also resulted because some provinces switched before others and initial resident counts were incomplete. This paper describes the changing definition of provincial populations in China and their impact on inequality in provincial GDP *per capita*. We show that much of the apparent increase in inter-provincial inequality disappears once a consistent series of GDP per resident is used.

Keywords

inequality
growth
regional development
population
China

JEL Classification

O15

1. Introduction

Nobel Prize winner Amartya Sen recounts how, as a nine-year-old boy, he saw unbelievably emaciated people dying in the streets during the 1943 Bengal famine (Sen, 1995). These were the rural destitute, who had trekked from the countryside into the cities in search of food; by one estimate they numbered at least 100,000 in Calcutta alone (Sen, 1981). Yet just 15 years later in the Great Leap famine in China, which killed at least 30 million people,[1] there was no similar trek by desperately hungry villagers seeking food in the cities. Instead, the strict controls on mobility in China under the *hukou* household registration system prevented the mass movement of people in search of food. So limited was movement that the excess mortality rate in one province – Anhui – was 17 times higher than that of a neighboring province, adding up to a difference of over six million excess deaths in Anhui between 1958 and 1961.[2]

Fast forward two decades to the beginning of China's economic reform and the *hukou* system was largely intact. Every person in China in 1958 had been given either "agricultural" (*nongye*) or "non-agricultural" (*fei nongye*) *hukou*, to determine entitlements to state-provided goods and services (regardless of location), which thereafter passed from mothers to children. In addition to *hukou* type, each person was categorized by their place of *hukou* registration (*hukou suozaidi*), which was their official and permanent residence in the eyes of the state (Chan, 2009). In the pre-reform era there was no freedom of movement, with all internal migration subject to approval from authorities at the destination, so the *hukou* operated much like an internal passport system (Chan and Zhang, 1999). But China's population was maladjusted for the needs of a market economy, with too many people in the interior, and too few in the soon-to-be-booming coastal cities. So the non-*hukou* migrants emerged, people who moved to live and work somewhere other than their place of *hukou* registration – often hundreds of miles away.

China's statistical system had no way to deal with the non-*hukou* migrants, so for most of the first three decades of the reform era, local GDP was divided by *registered* population rather than by *resident* population whenever *per capita* figures were reported. In other words, official data on provincial GDP *per capita* (and also for prefectures, counties and cities) were not based on the average number of residents but instead, on the household registered population. In coastal provinces the household registered population is many millions less than the resident population, while the reverse is true for migrant-sending inland provinces, so a systematic and time-varying distortion in provincial GDP *per capita* data was created. For example, at the time of the 2000 census, Guangdong province had a registered population of 75 million but residents numbered 86 million, so using registered population to

[1] Smil (1999) estimates that 30 million starved to death between 1959 and 1961 and about the same number of births were lost or postponed.

[2] Specifically, from 1958 to 1961 there were 6.3 million excess deaths in Anhui, equivalent to 18.4 percent of the population, but just 0.2 million excess deaths (one percent of the population) in neighboring Jiangxi (Yixin, 2010), even though these provinces share a 260 mile border. Brown (2010) reports evidence that urban residents even smuggled food *out* of cities to their nearby but immobilized (by inflexible *hukou*) rural friends and relatives.

calculate GDP *per capita* gives a 15 percent overstatement. At finer spatial scales, such as for individual counties and cities, the error is much larger. The city of Shenzhen provides an outstanding example; its registered population was just over one million by the time of the 2000 census but its residents numbered seven million, so *per capita* GDP was overstated by almost 600 percent in the official data of the time (Chan, 2009a).

Every year an avalanche of data comes out of China's statistical system, with yearbooks for each province and for topics such as population, urbanization and the economy. The fact that these yearbooks regularly reported economic output of each province or sub-provincial area divided by a population concept that had no bearing to the actual number of people living in those areas has been routinely ignored by economists. As noted a decade ago by Rawski and Xiao (2001), few economists hesitated to utilize these standard yearbook data for their research. Instead, most debate by economists about economic statistics concerns the possible falsification of China's GDP growth figures (Rawski, 2001, Chow, 2006) and the threat to time series comparisons from varying rules for constructing consumption data (Holz, 2004). Much less attention is paid to the simpler issue of how to count China's local populations, despite the systematic distortion of *per capita* data created by using the wrong population figures.

In this paper we document these problems with China's population data and highlight some implications, especially for the literature on inter-provincial and inter-regional (coastal-inland) inequality. A general finding is that inequality rose sharply in the reform era (e.g. Fan, 1995; Chen and Fleisher, 1996; Huang et al., 2003; Fan and Sun, 2008; Li and Wei, 2010) but this may be a statistical artefact caused by the unusual way that China has calculated GDP *per capita*. We also explore two problems related to the issue of the wrong population being used for *per capita* statistics. First, a sharp discontinuity occurred when some provinces recently switched from reporting GDP per registered population to GDP per resident. We show that this discontinuity causes a spurious trend change for measures of inter-provincial inequality, which coincides with initiatives to reduce inequality such as the West China Development Project (Fan and Sun, 2008), the Northeast China Revitalization Campaign (Zhang, 2008) and the Rise of Central China Plan (Lai, 2007). Second, there was a double-counting problem, since some provinces switched to reporting output per resident several years before others. Hence, the same person may have been in the denominator of GDP *per capita* for two different places at the same time; as a resident of one province and in the registered population of a province that was slow to switch to reporting output per resident. We estimate that up to 26 million people were included in this double-count.

These population errors and their impact on GDP *per capita* figures have been discussed in the geography literature for some time (Chan, 2007). But economists have largely ignored them, with just a single working paper devoted to this issue (Hoshino, 2011). In contrast to that previous study, we explore all of the problems with China's local population data while Hoshinojust examines the problem of using GDP per registered

population.[3] Furthermore, a common thread of much recent applied research on China has been to use data from sub-provincial levels, such as prefectures (Roberts et al, 2012) or counties (Banerjee, Duflo and Qian, 2012). Yet the problem of using the wrong population in *per capita* GDP statistics is much worse for smaller jurisdictions like counties, because many more people are defined as non-*hukou* migrants when working with smaller geographic units (Chan, 2012). Moreover, intercensal population figures for lower-level units only refer to the *hukou* registered population (Scharping, 2001). Consequently, there are no annual estimates of the *resident* population of small units like counties, so *per capita* GDP estimates for those small areas continue to be based on the wrong concept of local population.

2. Impacts on GDP *per capita* from Using the Wrong Population Concept

The discrepancy between registered and resident population caused by the non-*hukou* migrants can only be clearly observed in years with either a microcensus or a full population census. The intercensal population data rely on the annual National Sample Survey on Population Changes, which appears to have understated non-*hukou* migrants, as we show below. In terms of the years with more reliable data, China has carried out a national census in 1982, 1990, 2000 and 2010, with a one percent microcensus in each of 1987, 1995 and 2005. But even the time-trend from census and microcensus data is not fully consistent since the duration required to be defined as a resident changed (Scharping, 2001). The 1982 and 1990 census counted non-*hukou* migrants as residents at their place of in-migration if they had lived there for more than one year, but each microcensus and the census in 2000 and 2010 used a six-month stay for someone to be defined as a *de facto* resident.[4]

This inconsistency means that the number of non-*hukou* migrants will be understated in earlier years and the gap between GDP per resident and GDP per registered population will be smaller than if a consistent definition had been used. However our main point is not much undermined by this inconsistency, since the non-*hukou* migrants were a much smaller group in the 1980s, increasing by just 15 million (from 6.6 million in 1982 to 21.6 million in 1990 – both figures on a one year residency definition). In contrast, Chan (2012) reports that the number of non-*hukou* migrants increased by almost 100 million between the 1995 microcensus and the 2000 census (both on a six month definition). So it is in the 1990s, and subsequently, that the trends in GDP *per capita* are most affected by the discrepancy between registered and resident population.

---

[3] Moreover, the procedure used by Hoshino to develop an alternative to the officially reported GDP per registered population can be questioned because the way that the 1990 and 2000 census populations are spread over intercensal years suffers from the double-counting problem.

[4] Tsui (2007) is an example of a study that (wrongly) ignores this change in the length of stay required to be defined as a resident. Tsui projects age-specific population cohorts forward from the 1982 and 1990 census and uses the gap between the actual figure in the next census and the projected figure for that year to estimate net in-migration and then spreads these evenly over the intercensal years. The six-month rule in the 2000 census gives more residents than the 12-month rule used in 1990, so this method will wrongly overstate the number of in-migrants in the 1990s.

In Figure 1 we illustrate the trends in GDP *per capita* for six provinces, three migrant-senders in the left column and three migrant-receivers in the right.[5] The solid line is the ratio of provincial GDP (in 2010 prices) to the population with *hukou* for that province. The dashed line shows GDP divided by the local resident population, as counted in the census or microcensus (six-month residency criteria).[6] The GDP per resident of the migrant-sending provinces is about seven percent higher than GDP per registered population in 2005, and about 15 percent higher in 2010. The reason for the gap is that the permanent residents of the migrant-sending provinces are far fewer than the number of people with *hukou* registration for these provinces. In the case of Sichuan (including Chongqing) the difference in the two population counts is almost 15 million people by 2010, which is the number who left that province to live elsewhere – typically in the coastal cities. Yet for most of the economic reform period the officially reported GDP *per capita* figures for these inland provinces have included these missing people in their denominator.

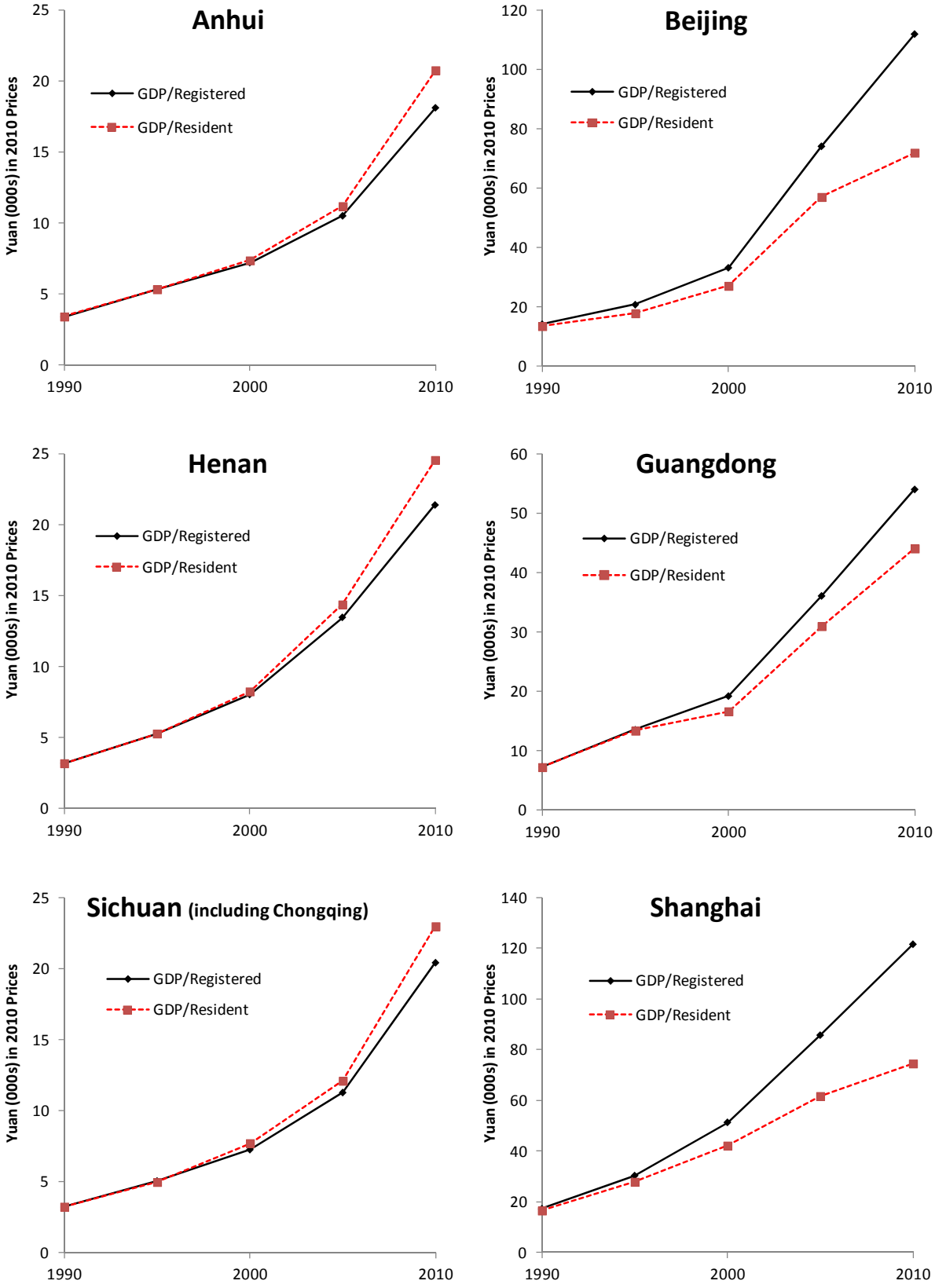
The discrepancy between GDP per resident and GDP per registered population is much larger for the three migrant-receiving provinces shown in the right column of Figure 1. This reflects the funneling of migrants from many source provinces into just a few destinations, along the coast between Shanghai and Guangdong, and into the Beijing-Tianjin conurbation. For example, the registered and resident populations for Shanghai were 13.2 million and 16.7 million in the 2000 census, so dividing GDP just by the registered population results in a 27 percent upward bias in the *per capita* estimates. The gaps are even larger in 2010, at 55 percent for Beijing, 23 percent for Guangdong and 63 percent for Shanghai. In terms of annual trend growth rates in GDP *per capita*, wrongly using the GDP per registered population would overstate growth rates for Beijing and Shanghai by about two percentage points per year. Since there is an understated growth rate for all the migrant sending provinces, the scope for distorted findings from studies that use uncorrected provincial GDP *per capita* figures should be apparent.

---

[5] Beijing and Shanghai are province-level municipalities, along with Tianjin and (since 1997) Chongqing. Each one is included as a separate entity in province-level statistics. Each of these municipalities also includes agricultural populations. For example, Chongqing municipality had a registered population of 30.9 million in the 2000 census but the non-agricultural population in its city districts was just 3.8 million (Chan, 2007).

[6] The resident population for 2000 and 2005 is based on revised figures for those years produced by the National Bureau of Statistics (NBS) after the 2010 census. The revisions raised the estimated resident population of Shanghai in 2005 by six percent and lowered the estimate for 2000 by four percent, while for the other provinces shown in the figure there were almost no changes from the initial counts made in 2000 and 2005.

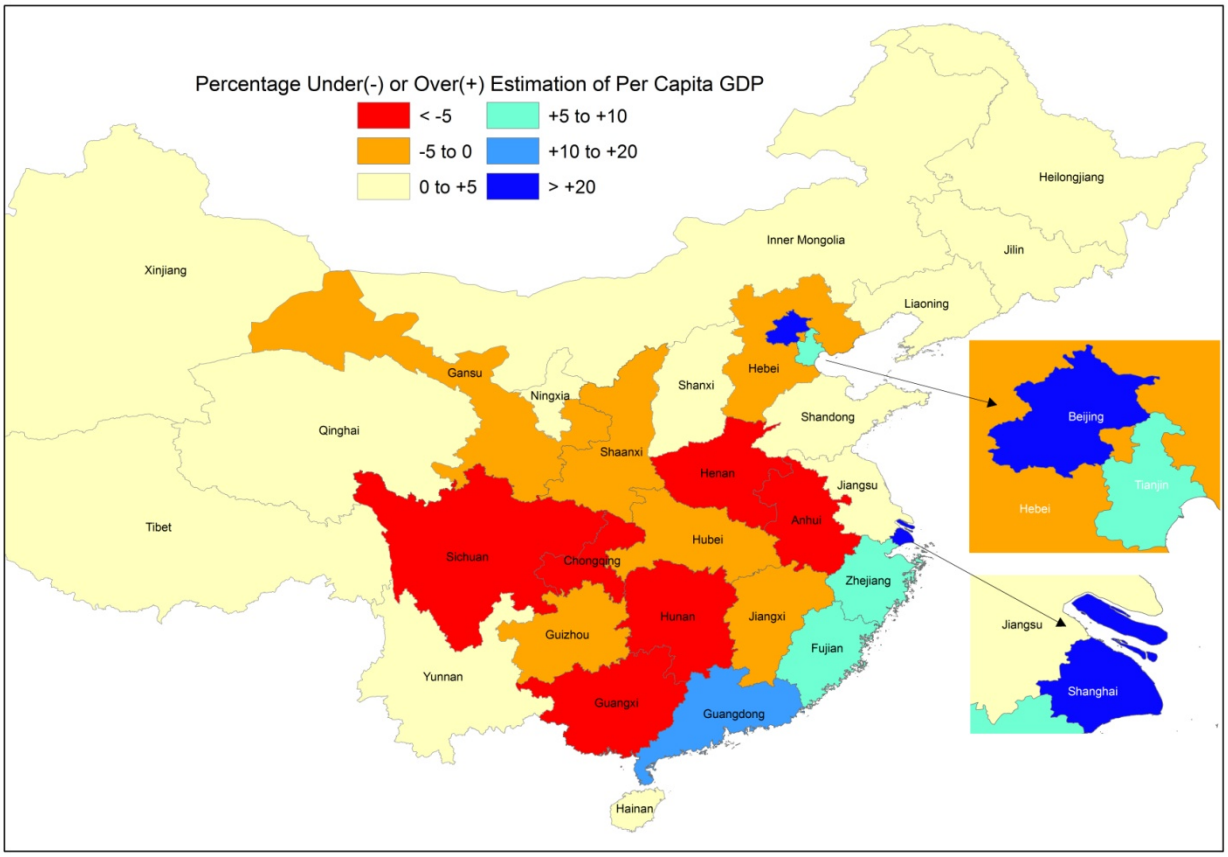
Figure 1: Trends in GDP per Registered and per Resident Population: Selected Provinces



Notes: Chongqing included in Sichuan since they were one province until 1997. Values are in 2010 prices using the GDP deflator for China provided by the World Bank.

To illustrate the situation more generally across all provinces, the percentage understatement or overstatement of *per capita* GDP at the time of the 2005 microcensus is mapped in Figure 2. For the six provinces in the central interior of China with the largest understatement, on average GDP *per capita* was seven percent lower using registered rather than resident population. A group of six adjacent (but still interior) provinces had lower rates of understatement, averaging two percent. The next group of 13 provinces forms a band around the western, northern and northeastern periphery where a very slight upward bias, averaging just two percent, results from dividing GDP by the registered population. These provinces are neither important sources nor destinations for non-*hukou* migrants. The final six provinces are all main destinations for non-*hukou* migrants, with *per capita* GDP overstated by at least five percent (and averaging 15 percent). This final group of provinces includes those on the southeast coast (Shanghai, Zhejiang, Fujian and Guangdong) and the Beijing-Tianjin conurbation.

Figure 2: Percentage Difference in GDP per Registered and per Resident Population: 2005



If the map in Figure 2 was redrawn for 2010 the patterns would be the same but the rates of over- and under-statement would be larger. Such a map could be considered misleading, however, because by 2009 all provinces in China had, finally, switched to reporting *per capita* GDP on a resident basis rather than using the misleading *hukou*-registered population as the denominator. But the way in which this switch was implemented introduced a further set of problems with China's local population data, which we describe in the next section.

3. Discontinuities and Double-Counts

The population denominator used for *per capita* economic statistics of many provinces switched abruptly from a registered population basis to a resident population basis. This change in statistical procedures occurred around 2005 for the largest migrant-senders and receivers, but the switchover date varied for other provinces. This switch in the denominator for GDP *per capita* statistics raises the apparent living standards in inland provinces while lowering it for coastal provinces, since it involves (correctly) accounting for the inter-provincial non-*hukou* migrants. But the sudden switch made by the provinces with the largest migrant in-flows and out-flows creates a discontinuity in estimates of inter-provincial and coastal-inland inequality, at about the time that many government-led initiatives to reduce regional inequality were at their peak. It is likely that at least part of the claimed success from investment in central and western China in reducing the coastal-inland gap (Fan et al, 2011) was simply due to the timing of the change in the way that China's economic statistics counted local populations.

In order to illustrate the sudden switch in the denominator of official GDP *per capita* estimates we chart time series of populations for the six provinces used in Figure 1. Two other features of China's provincial population data are also highlighted by this exercise; first, the reports of the resident population in the *Statistical Yearbooks* appear unreliable for most years, and second, the most plausible time series of resident population for each province have only been available since 2011. Consequently, research that uses provincial population estimates that were published before 2011 and which relies on a concept of *resident* population – as most *per capita* economic statistics should – is likely to be distorted.

To explain each of these points fully, four time series are needed for each province:

- GDP implied population - the population values underlying reported GDP *per capita* figures. We derived these annual average population estimates by dividing current reported provincial GDP by current reported provincial GDP *per capita*.[7]
- Registered population - the year-end registered population (*hukou suozaidi*) for each province, as reported in *China Population Statistical Yearbook* 1991-2006, *China Population and Employment Statistical Yearbook* 2007-2010, and *China Health Statistical Yearbook* 2011.
- Resident (initial) population - the estimated year-end resident population, reported each year in *China Population Statistical Yearbook* 1991-2006, *China Population and Employment Statistical Yearbook* 2007-2010, and *China Statistical Yearbook* 2011. In years without a census or microcensus, this resident count comes from the annual National Sample Survey on Population Changes and from the census data otherwise.
- Resident (revised) population - the annual average of the resident population in previous years back to 1990 that was released by NBS in 2011in the statistical yearbooks for each

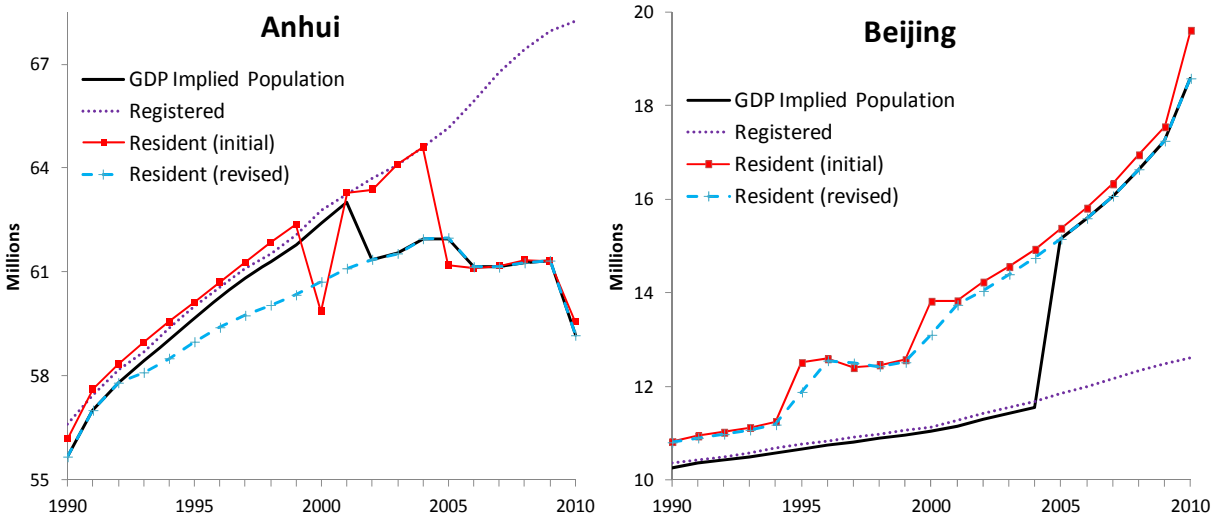
[7] Holz (2004) uses the same approach to derive the urban and rural populations that are implied by household consumption statistics. Our particular sources are the *Comprehensive Statistical Data and Materials on 50 Years of New China*, the *China Statistical Yearbook for Regional Economy 2001-2011*, and the historical sections of the 2011 edition of the *Statistical Yearbook* for each province.

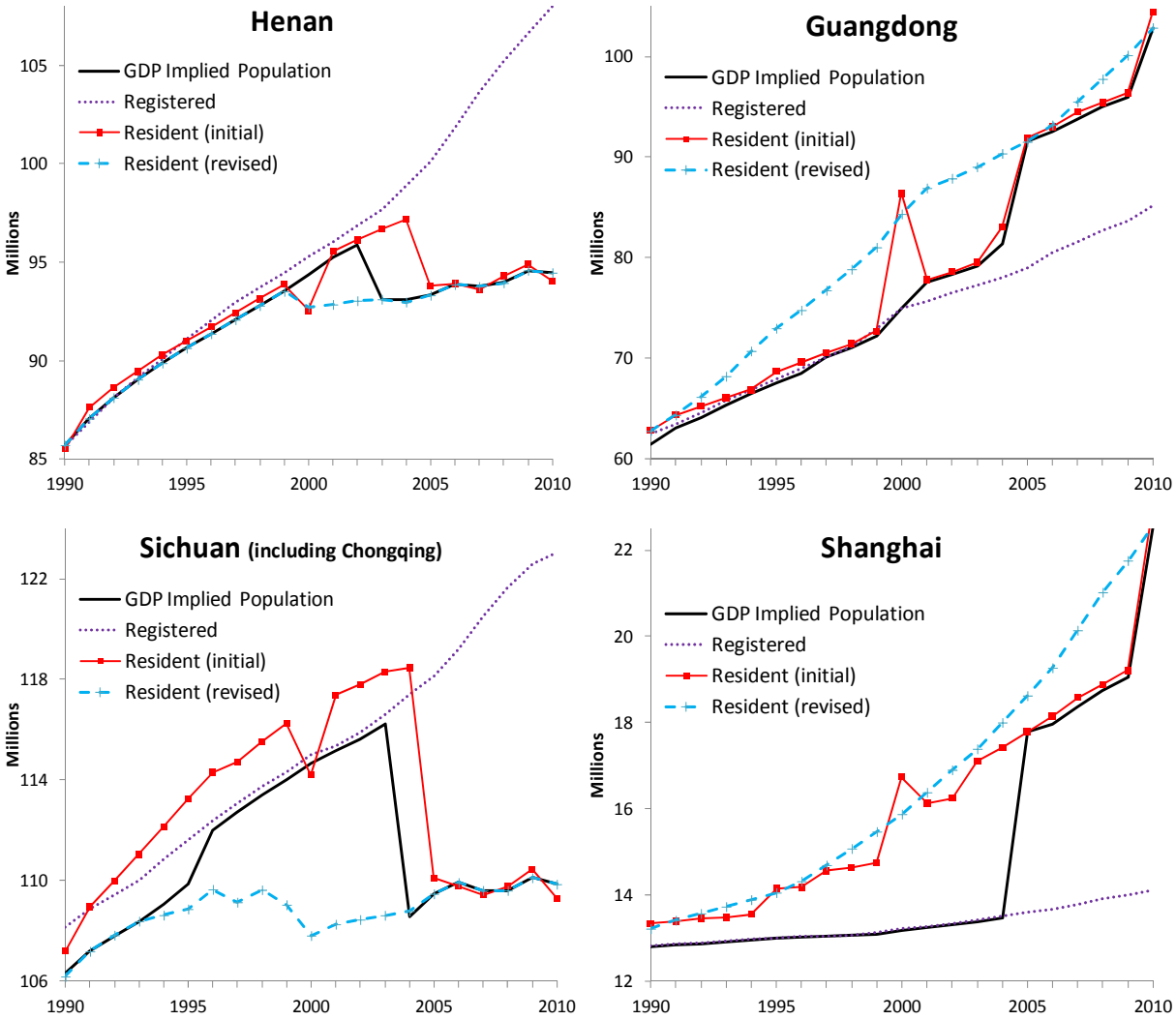
province, derived by dividing current reported provincial GDP by current reported provincial GDP *per capita*. This revision uses the results from the 2010 census and attempts to smooth out discrepancies in the time series of resident population.

For each of the six provinces shown in Figure 3, the GDP implied population tracks very closely with the registered population for the province from 1990 to 2004 (Beijing and Shanghai) or to slightly earlier (Guangdong and Sichuan – 2003; Henan – 2002; Anhui – 2001). There is then a sharp upward jump in the implied population in Beijing and Shanghai and a sharp downward drop in the implied population for the three migrant-sending provinces. For Beijing and Shanghai the population jump is equivalent to over 30 percent of the previous year population, while for the three provinces shedding population the decline is equivalent to four percent of the previous year population. After this abrupt change the implied population then closely tracks either or both of the two resident population time series. The only province where the switch from registered to resident population time series takes more than one year is Guangdong, where it is spread over the years from 2001 to 2005.

The next feature of the provincial population data shown in Figure 3 is the volatility of the time series of initial estimates of the resident population. The three migrant-sending provinces all show sharp dips in the time series in 2000 and then sharp declines in 2005 that are not reversed. For the migrant-receivers, the estimated number of residents spikes in 2000 and shows rapid rises again in 2010, and also in 2005 for Guangdong. For each province the year(s) that do not fit with the trend for the surrounding years, or that mark a shift to a different time trend all correspond with the timing of either a census or microcensus. For the intercensal years the resident population count comes from the National Sample Survey on Population Changes, which is just a 0.1 percent sample. It appears that this sample is not able to adequately cover non-*hukou* migrants, compared with the more extensive efforts of the census and microcensus.

Figure 3: Trends in Implied, Registered and Resident Population: Selected Provinces





Source: Author's calculations from the sources described in text. For other notes see Figure 1.

In order to test our contention that the initial estimates of resident population in the *Statistical Yearbooks* understated non-*hukou* migrants, we compared the 2000 census counts with what the resident population would have been estimated as in that year from the trend for the two surrounding years. For example, in Anhui the census count of residents was 59.9 million but the trend line joining the 1999 and 2001 *Yearbook* figures implies that there would have been 62.9 million residents in 2000, if estimated in the same way as in the surrounding years. We then regressed this imputed yearbook population for 2000 on the census count of the resident population in 2000:

$$\text{yearbook} = \beta_0 + \beta_1 \text{census} + \varepsilon$$

separately for migrant-sending provinces and for migrant-receiving provinces.

The results were that $\beta_1^{\text{senders}}=1.023$ while $\beta_1^{\text{receivers}}=0.953$ and the null hypothesis that these coefficients are equal is clearly rejected (t=2.20). In other words, the method used to form the estimates of resident population that were reported in the *Statistical Yearbooks* for

non-census years overstates the population of migrant-sending provinces and understates it for receiving provinces. This pattern could only result from underestimating the number of non-*hukou* migrants. Thus even if economists had been concerned that the official GDP *per capita* figures relied on a conceptually incorrect denominator – the registered population – and had attempted to adjust them by dividing by the reported resident population, the data available would not have allowed them to make a correct adjustment, except in census years.

The third feature of the provincial population data shown in Figure 3 is that the most plausible time series of resident population for each province have only been available since 2011. These data are shown by the dashed line labeled "Resident (revised)" and they represent the attempt by the National Bureau of Statistics (NBS) to use the latest census counts to correct the initial estimates of the resident population in each province.[8] In contrast to the initial estimates, reported year-by-year in each province's *Statistical Yearbook*, which are quite volatile between census and non-census years, the revised estimates smooth the time series. Nevertheless, we must stress that these are only estimates and it will never be known for certain what the annual resident population of China's provinces was for most of the first three decades of economic reform.

The final idiosyncratic feature of China's provincial population data, which is not apparent from Figure 3, is that a few provinces had been reporting output per resident rather than output per registered population from as far back as 1990. The same person can be counted as a resident of one province and in the registered population of another province, so poor coordination between provinces in when they switched to reporting output per resident creates scope for a double-counting problem. The year in which each province switched to reporting GDP per resident is shown in Table 1. We determine this date using charts like those in Figure 1, but for all provinces, since there is no explicit discussion of the (changing) population basis for GDP *per capita* calculations in contemporary statistical documents. The registered and resident population for each province, as counted in the 2010 census is also reported in Table 1, along with the number of non-*hukou* migrants (positive for in-migrants and negative for out-migrants) and this number as a percentage of the resident population. In unreported regressions we examined whether the switch date for each province was related to whether they were a migrant-sender or receiver, to the number of their non-*hukou* migrants and to the size of their total population. There were no clear relationships to explain why the statistical authorities in some provinces switched to reporting GDP per resident much earlier than did the authorities in other provinces.

[8] The 2010 population census (and the Second Economic Census, conducted in 2008) provides corrected endpoints for recalculating trends in GDP and GDP *per capita*. This "new trend" can then be used to revise previous annual estimates. Specifically, annual deviations from a "historical trend" (based on the previous *Yearbook* data that did not benefit from the latest census results) are applied to the new trend. Wu (2007) provides details for GDP revisions.

Table 1: Year of First Using Resident Population in Official GDP *per capita* Estimates

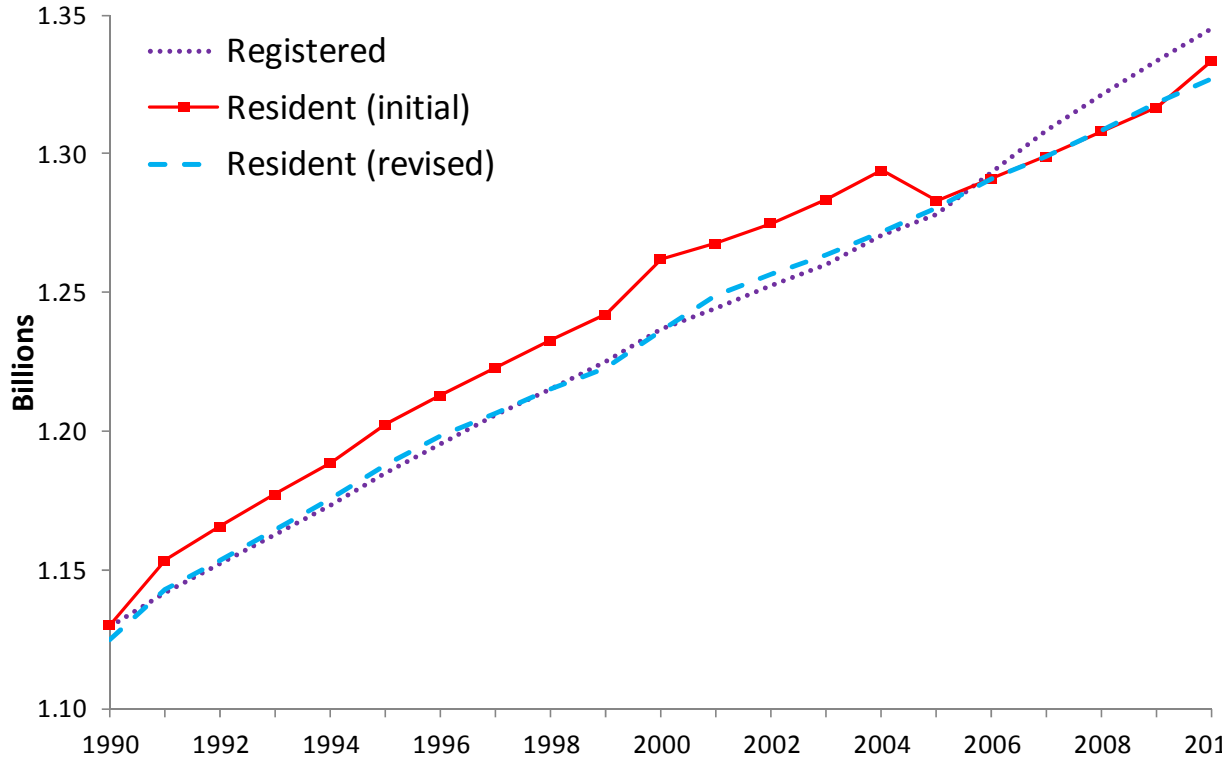
Province	Registered Population (2010 mil)	Resident Population (2010 mil)	Non-*hukou* migrants (2010 mil)	Migrants as % of residents	Year of GDP switch to residents
Shanghai	14.1	23.0	8.9	39%	2005
Beijing	12.6	19.6	7.0	36%	2005
Tianjin	9.9	13.0	3.1	24%	2005
Guangdong	85.2	104.4	19.2	18%	2001
Zhejiang	47.5	54.5	7.0	13%	2001
Jiangsu	74.7	78.7	4.0	5%	1990
Fujian	35.3	36.9	1.6	4%	2000
Shanxi	34.7	35.7	1.0	3%	1991
Liaoning	42.5	43.7	1.2	3%	1990
Qinghai	5.5	5.6	0.1	2%	1990
Tibet	2.9	3.0	0.1	2%	2000
Yunnan	45.3	46.0	0.7	2%	1991
Xinjiang	21.6	21.9	0.2	1%	1990
Jilin	27.2	27.5	0.2	1%	1997
Inner Mongolia	24.5	24.7	0.2	1%	1991
Shandong	95.4	95.9	0.5	1%	1997
Heilongjiang	38.4	38.3	-0.1	0%	1990
Hebei	73.0	71.9	-1.0	-1%	1991
Ningxia	6.4	6.3	-0.1	-2%	2001
Hainan	9.0	8.7	-0.3	-3%	2003
Shaanxi	38.7	37.4	-1.4	-4%	1992
Jiangxi	46.9	44.6	-2.3	-5%	1990
Gansu	27.1	25.6	-1.5	-6%	1995
Hubei	61.5	57.3	-4.2	-7%	2005
Hunan	70.7	65.7	-5.0	-8%	2001
Sichuan	90.0	80.4	-9.6	-12%	2004
Chongqing	33.0	28.8	-4.2	-15%	2004
Anhui	68.3	59.6	-8.7	-15%	2002
Henan	108.0	94.1	-13.9	-15%	2003
Guangxi	53.3	46.1	-7.2	-16%	2003
Guizhou	41.9	34.8	-7.1	-20%	2008

Notes: Year of GDP switch to residents is the year in which official average population became resident population (before which it was registered population). Provinces are sorted according to the migrant percentage.

In Figure 4 we show the scope of this double-counting problem, by adding up the population of each province for the 1990-2010 years. We note in advance that the national sum of the resident population should be less than the national sum of the *hukou* registered population, since the People's Liberation Army and any Chinese with foreign permanent residency are not counted as residents but do have household registration status. Yet between

1990 and 2005 the resident population obtained by summing across all provinces (shown by the solid line) exceeded the registered population (shown by the dotted line). The excess number of people counted as residents was especially apparent between 2000 and 2004, when it averaged 24 million people (with a maximum of 26 million). The source of this excess is that what was called 'resident' in many provinces was, in fact, registered population, which exceeded their true number of residents. In other words, some people were counted as residents in the province of their *hukou* registration and were counted again, as residents in the provinces that they actually lived in. Consequently it would have been no simple matter for economists to abandon the officially reported GDP per registered population and calculate their own estimates of GDP per resident, since the data on the resident population were so distorted.

Figure 4: The Double-Count in Resident Population (Summation over all Provinces)



Source: Author's calculations from the sources described in text.

The final population time series shown in Figure 4 is for the sum of each province's revised resident population, shown by the dashed line. These revised counts use the results from the 2010 census to correct the initial estimates of resident population in each province. Specifically, the trend-deviation interpolation approach (described in footnote 8) was used by the NBS in 2011 to revise previous annual estimates of GDP and GDP *per capita*, giving an implied annual average of the resident population for each province. For most provinces this revision went back as far as 1990, but for nine provinces only a shorter time series was

revised.[9] Therefore, for those nine provinces we use the same approach as the NBS to backdate the revised counts of the resident population to 1990. It is the sum across provinces of these revised counts which are shown in Figure 4 and they almost completely eliminate the double-count problem. Whereas the double-count in the year-by-year initial estimates of resident population averaged 17 million over 1990 to 2005, in the revised series the average double-count is just one million. Still, even the revised series only obeys the expected pattern of being less than the sum of the registered counts since 2006, so it is only for the most recent years that there can be any great confidence in the estimates of the resident population of China's provinces.

4. Impacts of Population Errors on Trends in Inter-provincial and Inter-regional Inequality

Despite the shaky foundations created by China's official data using a population concept that had no bearing to the actual number of people living in each province, a large literature that relies on reports of provincial GDP *per capita* has already emerged. In this section we show how the population errors described above may have distorted some of the conclusions reached by one strand of this literature – that related to inter-provincial and inter-regional (coastal-inland) inequality. In Table 2 we survey studies published in the last 15 years that use provincial GDP *per capita* as their economic indicator for analyzing trends in regional inequality. Only four of the 14 studies make any adjustment to the data from the *Statistical Yearbooks* to account for the distortions from using registered population as the denominator.[10] Findings from the studies that rely on unadjusted data are potentially misleading, at least for the period since 1990 when the number of non-*hukou* migrants greatly increased. Even when the potential bias from using GDP per registered population is recognized, the adjustment procedures can be questioned, not least because they all lacked the results from the latest 2010 census that support the most plausible revisions of resident population estimates.[11]

To illustrate the potential bias in previous studies we calculate three of the most commonly used statistics of inter-provincial inequality – the Theil index, the coefficient of

[9] The provinces and their earliest year with revised data are: Zhejiang (2000), Jiangxi (2000), Henan (2003), Hubei (2000), Hunan (2002), Guangxi (1998), Sichuan, including Chongqing (2000), Guizhou (2006) and Gansu (2006).

[10] China occasionally releases compilations of data from the *Statistical Yearbooks* covering two or more decades, typically with titles of "20 years …", "40 years …" but these do not show any consistent adjustment to the population data initially published year-by-year so studies using these compilations are counted in the group using unadjusted data.

[11] Specific concerns with the adjustments of Hoshino (2011) and Tsui (2007) are listed in footnotes 3 and 4. The Fan and Sun (2008) adjustment is to replace census estimates of resident population with the surrounding year estimates from the *Yearbooks*, while the discussion of Figures 3 and 4 above suggests that the opposite procedure should be used. The Chan and Wang (2008) paper does not adjust for the double-count issue shown in Figure 4.

variation (CoV), and the Gini coefficient.[12] We compare the time series for each statistic using three different population denominators to calculate GDP *per capita* for each province:

- GDP implied population
- Registered population
- Resident (revised) population

Each of these three population time series has been previously described in the discussion of Figures 3 and 4.[13] The revisions to the resident population estimates only go back as far as 1990 (by the NBS for 21 provinces and by us for the remaining nine) so this comparison is limited to the 21 years from 1990 to 2010. To put this starting date into context, from the beginning of economic reform in 1978 until 1990, inter-provincial inequality in China fell rapidly, with trend annual rates of decline of 4.6 percent (Theil), 3.9 percent (CoV) and 0.5 percent (Gini).[14] This falling inequality resulted from the early decollectivization of agriculture, which saw incomes in rural provinces close some of the gap with the three richest urban provinces (Shanghai, Beijing and Tianjin). Also, at the start of the reform period, the soon-to-be-booming southeast coastal provinces of Guangdong, Fujian and Zhejiang were not even in the top one-third of provinces (ranking 10th, 23rd and 16th according to GDP *per capita* in 1978), so the tremendous growth that they soon experienced had an equalizing impact.[15] Therefore in our illustration of the impact of population errors on regional inequality trends, the starting point happens to be the most equal year in the reform era.

For all three of the inequality statistics shown in Figure 5, the 1990 to 2000 trend using the GDP implied population (shown by the solid line) is almost identical to that using registered population (shown by the dotted line). This is unsurprising since most large provinces were still reporting GDP *per capita* on a registered population basis during these years (Table 1). Yet, while inequality in the officially reported GDP *per capita* was rapidly rising in the decade from 1990 (at a three percent annual rate in the case of the Theil index) it was largely divorced from the (correct) trend shown when GDP *per capita* is calculated using the revised resident population estimates (the dashed line in Figure 5). In fact, the apparent

---

[12] The Theil index is: $T_w = \sum_{j=1}^{m}(p_j/P)(y_{wj}/\mu)\ln(y_{wj}/\mu)$ where m=30 provinces (Sichuan and Chongqing merged), p_j is the population of the j^{th} province, P is overall population, y_{wj} is the GDP *per capita* of the j^{th} province, and μ is the overall population-weighted mean of GDP *per capita* for all provinces. The (weighted) coefficient of variation is: $CoV = \sqrt{\sum_{j=1}^{m}(p_j/P)(y_{wj}-\mu)^2}\Big/\mu$. The Gini coefficient is: $G = \left(\sum_{i=1}^{m}\sum_{j=1}^{m} p_i p_j \left|y_{wi}-y_{wj}\right|\right)\Big/ 2\sum p_i^{\,2}\mu$.

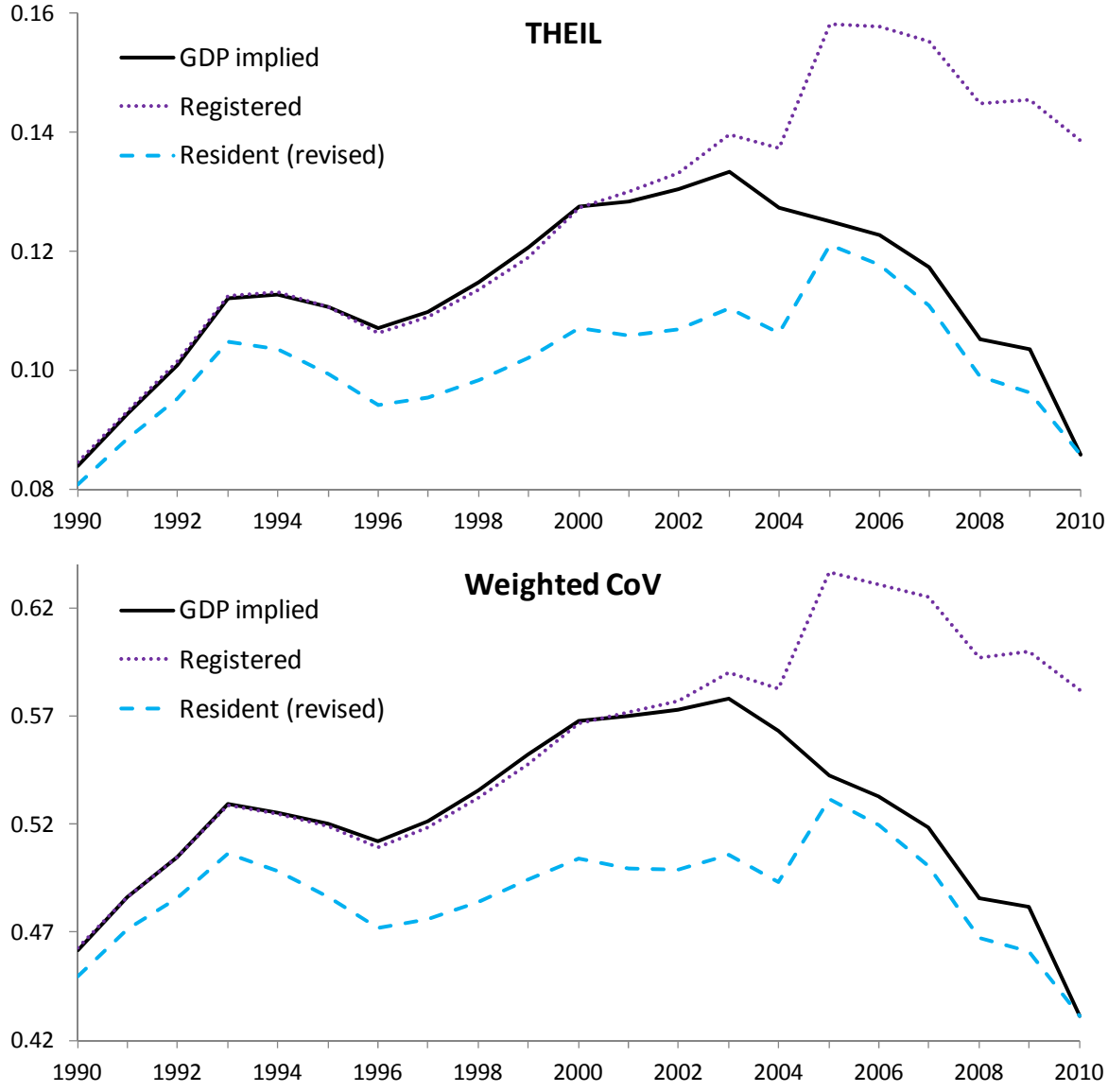
[13] We do not need to use the "Resident (initial) population" as an alternative denominator for GDP *per capita* in this illustration because we have already shown that these initial estimates of the resident population were unreliable (Figure 4).

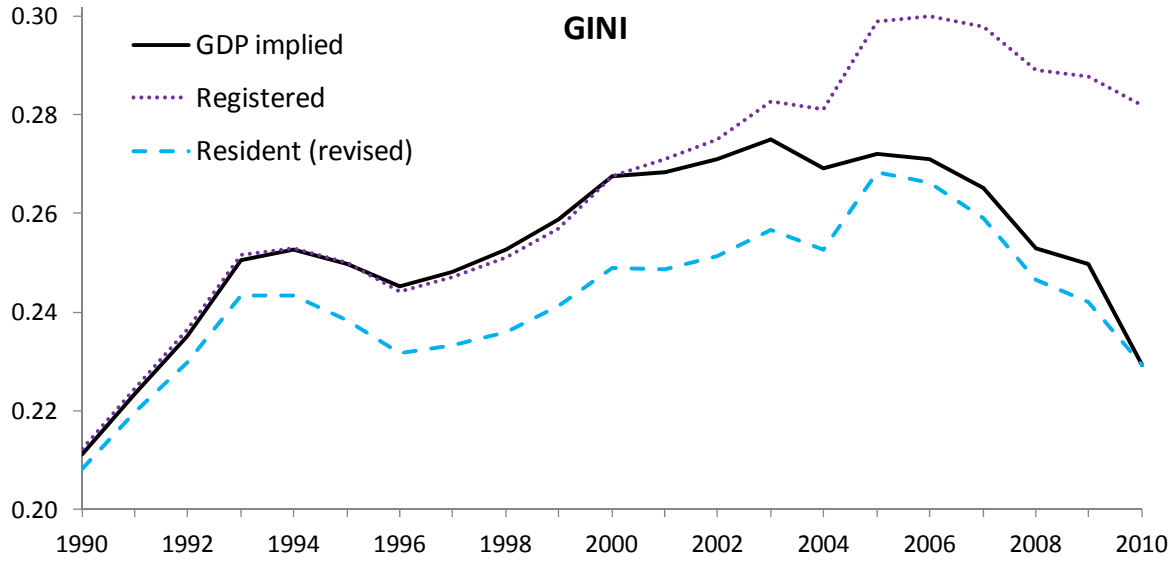
[14] Trend rates are estimated from a semi-logarithmic regression on time, and are statistically significant at the one percent level (Theil and CoV) and the eight percent level (Gini) using Newey-West heteroscedasticity and autocorrelation consistent (HAC) standard errors with a single lag. The provincial GDP *per capita* data underlying the inequality statistics are all based on registered population from 1978 to 1990.

[15] China had only 29 provinces until 1988.

rise in inter-provincial inequality between 1990 and 2000, when using the Theil index calculated from the official data of the time, is 66 percent greater than the actual rise, based on the revised resident series. For the CoV and the Gini, the overstatement created by using uncorrected *Yearbook* data for GDP *per capita* is 94 percent and 38 percent, as of the year 2000. If there had been no change in the basis of GDP *per capita* calculations, the apparent trend in inter-provincial inequality would have carried on rising strongly (for example, there is a 1.4 percent per annum trend rise in the Theil index when using registered population over 2001-10). But, in fact, inter-provincial inequality returns almost exactly to its starting point by 2010 (for the Theil index; the CoV is slightly lower and the Gini slightly higher than their 1990 values) in the time series based on the corrected data. Hence, if GDP per registered population had continued to be used as the official data, *all* of the apparent rise in inequality would have been measurement error.

Figure 5: Inter-Provincial Inequality in GDP *per capita* With Various Denominators



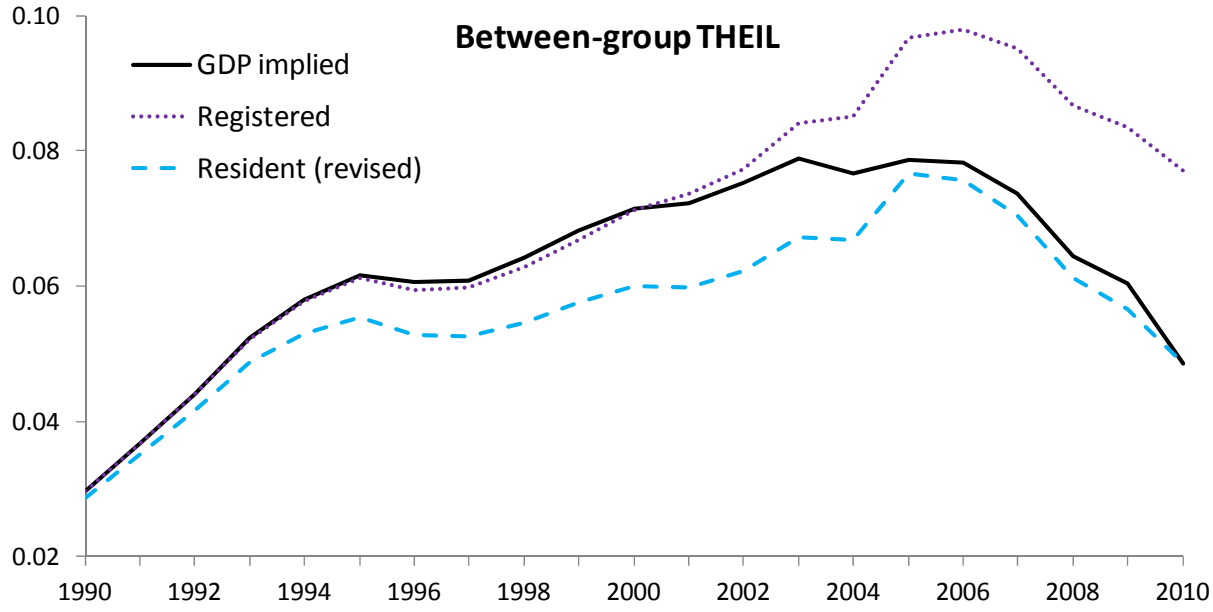


The other feature illustrated by Figure 5 is the apparent change in the trend in inter-provincial inequality, when using the GDP implied population series. From 1990 to 2003 the trend annual rate of increase was 2.9 percent (for the Theil index; 1.5 percent for the CoV and 1.6 percent for the Gini), but this then sharply reversed, to decline at a rate of 5.6 percent per year (3.8 percent using CoV and 2.3 percent for the Gini) over 2003-2010. But rather than reflecting any major change in the true situation in 2003, this is an artefact resulting from the switch to reporting *per capita* GDP on a resident basis. Since inequality in GDP per resident is considerably lower than inequality in GDP per registered population, this switch in statistical procedure automatically reduces measured inequality in the officially reported GDP *per capita* without any underlying change necessarily occurring in the economy in that year. In fact, in both the registered-based and resident-based series in Figure 5, the turning point in the inequality trend does not occur until 2005, which happens to be the year that the coastal provinces reached their maximum share of GDP. Whether the reduction in inequality thereafter is from lagged effects of the investments into western and inland China (Fan et al, 2011) rather than due to the contraction in the export-oriented coastal provinces as a result of the Global Financial Crisis is an open question.

A reasonable summary of the patterns in the inter-provincial inequality time series, based on the most plausible revised population estimates, is that there was a largely stable pattern of inter-provincial inequality between 1993 and 2004. Over those years, the average annual deviation was just four percent of the 1993-04 mean (for the Theil index). The only two occasions of notable change in inter-provincial inequality were from 1990-93, when there was a sharp increase in inequality; and from 2005-2010, when that initial increase was more than fully reversed. Recalling that the 12 years prior to 1990 had witnessed an almost continuous decline in inter-provincial inequality (reducing the Theil index from 0.16 to 0.08), the solitary three year episode of rising inequality from 1990-93 – which lifted the Theil index back to just over 0.10 – hardly qualifies as strong support for the claim from the literature that regional inequality has risen in the reform era.

Our final illustration of how apparent trends may be affected by the switch from reporting GDP per registered population to GDP per resident is for a regional decomposition of inequality. Several papers in the literature group China's provinces into regions and examine the between-region inequality trends. We carry out a similar exercise for the coastal-inland comparison, using the Theil index (which decomposes exactly, unlike the Gini coefficient). In Figure 6 we show the time series for the between-group Theil index and it is apparent that inequality between the coast and inland provinces rises, by about 0.04 points, from 1990 to around 2005.[16] But from 2000 to 2005, while the time series of inequality using the GDP implied denominator is detaching from the registered population series and moving toward the resident-based series, this movement makes it seem that there is a largely unchanged between-group component. Specifically, the increase from 2000 to 2005 in the GDP implied time series in Figure 6 is just ten percent, versus a 36 (27) percent increase when inequality is measured using the registered (resident) population.

Figure 6: Coastal-Inland Inequality With Various Denominators for GDP *per capita*



Source: Decomposition of Theil index into within-group and between-group components.

In terms of policy evaluation, the most active period of state intervention to close the regional gaps had already run its course by 2005, by which time more than 1 trillion Yuan (US$180 billion) of state-led infrastructural investment had been placed in western China (Yao, 2009). Some authors have suggested that a leveling off of the coastal-inland gap in this period reflects the good effects of these interventions (Fan et al, 2011). But an alternative

[16] The identical time series, but rescaled, would be shown if we examined the ratio of mean GDP *per capita* in the coastal provinces to mean GDP *per capita* in the inland provinces. This ratio starts at just over 1.6 in 1990, peaks at 2.2 in 2005 and was back below 1.9 by 2010.We group Hainan, Guangdong, Fujian, Zhejiang, Shanghai, Jiangsu, Shandong, Hebei, Tainjin and Beijing together as the coastal provinces.

explanation is that the apparent leveling off over those years is just an artefact created as the population series used to denominate GDP *per capita* progressively switched from one based on registered population to one based on resident population which involved switching from a higher trending inequality time series to a lower trending one, creating an apparently static period.

5. Conclusions and Implications

We have shown how changes over time in the population denominator used for China's provincial GDP *per capita* statistics may distort understanding of recent trends in regional inequality. Much of the apparent trend increase, and change in trend in 2003, disappear once proper account is taken of non-*hukou* migrants, so that GDP per resident is calculated correctly for each province of China over 1990-2010. Putting the corrected data for the last two decades into the context of the entire reform era, the changing pattern of inter-provincial inequality can be thought of as composing four episodes, only one of which involves rising inequality. From 1978 until 1990 inter-provincial inequality declined almost continuously; about one-third of this decline was reversed over the next three years; then a decade of little change in inequality was ended by a sharp one-year rise in 2005 but even with that rise, inequality had returned to only two-thirds of its starting value in 1978; finally, inter-provincial inequality fell sharply after 2005, so that by 2010 it was right back at the low levels previously seen in 1990. Thus the only sustained episode of rising inequality in the reform era was from 1990 to 1993, representing just three out of 33 years since liberalization began.

The three errors in China's local population data described here – using registered population as a denominator, the abrupt and uncoordinated switch to using a resident denominator, and the double-count introduced by the partial and inconsistent way that residents were counted by different provinces in any given year – are also likely to bias econometric research using the published data on provincial GDP *per capita*. The errors are correlated with the levels and growth rates of GDP per capita, since the non-*hukou* migrants were moving into richer, faster growing provinces, and the errors are increasing over time (at least until each province switches to reporting GDP per resident). In contrast to random (classical) errors, which cause no bias when just in the dependent variable and which always attenuate the coefficient on a single error-ridden explanatory variable, the effects of these errors will be more complex and pernicious. Moreover, the main correction used by economists for measurement error bias – instrumental variables (IV) – is inconsistent when errors are correlated with true values (Black, Berger and Scott, 2000), while bounding estimates based on reverse regression are unlikely to be effective in practice (Gibson and Kim, 2010). The recent trend in applied research on China to use data from smaller spatial units is especially vulnerable to the threat posed by these errors.

Finally, we acknowledge that we have only considered errors in the population denominator for GDP *per capita*, while leaving possible errors in the numerator unchallenged. In fact, in 2006 and 2009, the National Bureau of Statistics retrospectively adjusted provincial GDP for 1993-2003 and 2005-2008 according to the results from China's

First and Second Economic Census in 2004 and 2009. In order to cleanly describe the effects of the population errors, we chose to work with the originally published GDP data rather than use the recently revised series. However, the patterns we find, of mostly falling inequality that returns to levels seen in 1990, are also apparent if we had used the revised GDP data for the numerator. For the same reason of wanting to cleanly identify the effect of population errors while not introducing auxiliary variations, we did not adjust for spatial differences in prices despite richer provinces in China being more expensive places to live, especially in urban areas (Brandt and Holz, 2006)**.** In other words, there would be less apparent inter-provincial inequality if we were to recast our analysis using spatially real prices; however, spatial deflators are not available for years since 2004. The task of generating reliable data for local areas of China that reflect appropriate concepts in terms of populations, prices and nominal variables will be an on-going task for the research community.

References

Banerjee, A., Duflo, E., and Qian, N. (2012). On The Road: Access to Transportation Infrastructure And Economic Growth In China. *Working Paper* No. 17897, National Bureau of Economic Research, Cambridge MA.

Black, D., Berger, M., and Scott, F. (2000). Bounding Parameter Estimates with Nonclassical Measurement Error. *Journal of the American Statistical Association* 95(451): 739-748.

Brown, J. (2010). Great Leap City: Surviving the Famine in Tianjin. In *Eating Bitterness: New Perspectives on China's Great Leap Forward and Famine*, K. Manning and F. Wemheuer (eds), Vancouver: UBC Press, pp. 226-250.

Brandt, L., and Holz, C. A. (2006). Spatial Price Differences in China: Estimates and Implications. *Economic Development & Cultural Change,* 55(1), pp. 43-86.

Cai, F., Wang, D., and Du, Y. (2001). Impact of Labor Market Distortion on Regional Disparities *(Laodongli shichang niuqu dui quyu chaju de yingxiang). Zhongguo shehui kexue (China Social Science)*, 2, 4–14.

Chan, K. W., and Zhang, L. (1999). The *Hukou* System and Rural-Urban Migration: Processes and Changes. *The China Quarterly* 160:818–855.

Chan, K. W. (2007). Misconceptions and Complexities in the Study of China's Cities: Definitions, Statistics, and Implications. *Eurasian Geography and Economics*, 48(4):383-412.

Chan, K. W., and Wang, M. (2008). Remapping China's Regional Inequalities, 1990-2006: A New Assessment of de Facto and de Jure Population Data. *Eurasian Geography and Economics*, 49(1), pp.21-56.

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

Chan, K.W. (2009a). Measuring the Urban Millions. *China Economic Quarterly*, March, 21-26.

Chan, K.W. (2012). Internal Labor Migration in China: Trends, Geography and Policies. In *Population Distribution, Urbanization, Internal Migration and Development: An International Perspective*, United Nations Population Division, New York: pp.81-102.

Chen, J., and Fleisher, B. (1996). Regional Income Inequality and Economic Growth in China. *Journal of Comparative Economics* 22(2): 141-164.

Chow, G. (2006). Are Chinese Official Statistics Reliable? *CESifo Economic Studies* 52(2): 396-414.

Duncan, R., and Tian, X. (1999). China's Inter-Provincial Disparities: An Explanation. *Communist and Post-communist Studies*, 32, 2:211–224.

Fan, C. C. (1995). Of Belts and Ladders: State Policy and Uneven Regional Development in post-Mao China. *Annals of the Association of American Geographers,* 85(3), pp. 421-449.

Fan, C. C. & Sun, M. (2008). Regional Inequality in China, 1978-2006. *Eurasian Geography and Economics,* 49(1), pp. 1-17.

Fan, S., Kanbur, R., & Zhang, X. (2011). China's Regional Disparities: Experience and Policy. *Review of Development Finance*, 1(1): 47-56.

Fujita, M., and Hu, D. (2001). Regional Disparity in China, 1985–1994: The Effects of Globalization and Economic Liberalization. *The Annals of Regional Science*, 35, 3:3–37.

Gibson, J. and Kim, B. (2010). Non-classical Measurement Error in Long-term Retrospective Surveys. *Oxford Bulletin of Economics and Statistic*s, 72(5): 687-695.

Holz, C. (2004). Deconstructing China's GDP statistics. *China Economic Review* 15(2): 164-202.

Hoshino, M. (2011). Measurement of GDP per capita and regional disparities in China, 1979-2009. *Discussion Paper* DP2011-17, Research Institute for Economics and Business Administration, Kobe University, Kobe, Japan.

Hsueh T-T., Li, Q., and Liu, S. (Eds). (1993). China's provincial statistics 1949-1989. Boulder CO: Westview Press.

Huang, J-T., Kuo, C-C., and Kao, A-P. (2003). The Inequality of Regional Economic Development in China Between 1991 and 2001. *Journal of Chinese Economic and Business Studies* 1(3): 273-285.

Kanbur, R., & Zhang, X. (2005). Fifty Years of Regional Inequality in China: A Journey through Central Planning, Reform and Openness. *Review of Development Economics,* 9(1), 87-106.

Lai, H. Y. (2002). China's Western Development Program: Its Rationale, Implementation, and Prospects. *Modern China,* 28(4), pp. 432-453.

Lai, H. Y. (2007). Developing Central China: A New Regional Programme. *China: An International Journal,* 5, pp. 109-128.

Li, Y. R. & Wei, Y. H. D. (2010). The Spatial-Temporal Hierarchy of Regional Inequality of China*. Applied Geography,* 30, pp. 303-316.

Liu, S., and Zhang, X. (2007). The features of China's sustained high growth and the decline of regional economic disparity. *Jingji Yanjiu (Economic Research)*, 10, pp.17-31.

Lu, M., and Wang, E. (2002). Forging Ahead and Falling Behind: Changing Regional Inequalities in Post-Reform China. *Growth and Change*, 33, 1:42–71.

National Bureau of Statistics Bureau (NBS). (1998). The twenty years of spectacular achievements. Beijing: Statistics Press of China.

National Bureau of Statistics Bureau (NBS). (1990). Historical Statistical Materials for Provinces, Autonomous Regions and Municipalities 1949-1989 *(Quanguo Gesheng Zizhiqu Zhixiashi Lishi Tongji Ziliao Huibian)*. Beijing: China Statistics Press.

National Bureau of Statistics Bureau (NBS). (1999). Comprehensive Statistical Data and Materials on 50 Years of New China *(Xin Zhongguo Wushinian Tongji Ziliao Huibian)*. Beijing: China Statistics Press.

National Bureau of Statistics Bureau (NBS). (2000). Comprehensive Agricultural Statistical Data and Materials on 50 Years of New China *(Xin Zhongguo Wushinian Nongye Tongji Ziliao Huibian)*. Beijing: China Statistics Press.

National Bureau of Statistics Bureau (NBS). (2005). China Compendium of Statistics 1949-2004 *(Xin Zhongguo Wushiwunian Tongji Ziliao Huibian)*. Beijing: China Statistics Press.

National Bureau of Statistics Bureau (NBS). (2011). China Health Statistical Yearbook 2011 *(Zhongguo Weisheng Tongji Nianjian)*. Beijing: Peking Union Medical College Press.

National Bureau of Statistics Bureau (NBS). (1991-2006). China Population Statistical Yearbook 1991-2006 *(Zhongguo Renkou Tongji Nianjian)*. Beijing: China Statistics Press.

National Bureau of Statistics Bureau (NBS). (2007-2010). China Population and Employment Statistical Yearbook 2007-2010 *(Zhongguo Renkou he Jiuye Tongji Nianjian)*. Beijing: China Statistics Press.

National Bureau of Statistics Bureau (NBS). (2011). China Statistical Yearbook 2011. Beijing: China Statistics Press.

National Bureau of Statistics Bureau (NBS). (2011). Provincial Statistical Yearbook 2011 (individual issue for each of the 31 provinces). Beijing: China Statistics Press.Rawski, T. (2001). What is Happening to China's GDP Statistics. *China Economic Review* 12(4): 347-354.

Rawski, T. and Xiao, W. (2001). Roundtable on Chinese Economic Statistics: Introduction. *China Economic Review* 12(4): 298-302.

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): 580-94.

Scharping, T. (2001). Hide and Seek: China's Elusive Population Data. *China Economic Review* 12(4): 323-332.

Sen, A. (1981). *Poverty and Famines: an Essay on Entitlement and Deprivation* Oxford: Clarendon Press.

Sen, A. (1995). Nobody need starve. *Granta* 52(1): 213-220.

Smil, V. (1999). China's great famine: 40 years later. *British Medical Journal* 319(7225): 1619-1621.

Tsui, K-Y. (2007). Forces Shaping China's Interprovincial Inequality. *Review of Income and Wealth* 53(1): 60-92.

Wu, H. (2007). The Chinese GDP Growth Rate Puzzle: How Fast Has the Chinese Economy Grown? *Asian Economic Papers* 6(1): 1-23.

Xu, Z. Y., & Li, S.T. (2006). Analysis of the Recent Changes of Regional Disparity in China *(Jinnianlai zhongguo diqu chaju de bianhua qushi)*. *Jingji Yanjiu (Economic Research)*, 7:106–116.

Yao,Y. (2009). The Political Economy of Government Policies Toward Regional Inequality In China. In: Y. Huang, A. Bocchi (Eds.), *Reshaping Economic Geography in East Asia*.World Bank,Washington, DC, pp. 218–240.

Ying, L. (1999). China's Changing Regional Disparities during the Reform Period, *Economic Geography*, 75, 1:59–70.

Yixin, C. (2010). Under the same Maoist sky: accounting for death rate discrepancies in Anhui and Jiangxi. In *Eating Bitterness: New Perspectives on China's Great Leap Forward and Famine*, K. Manning and F. Wemheuer (eds), Vancouver: UBC Press, pp. 197-225.

Zhang, P. Y. (2008). Revitalizing Old Industrial Base of Northeast China: Process, Policy and Challenge. *Chinese Geographical Science,* 18(2), pp. 109-118.

Table 2. Selected Recent Studies of Inter-Provincial and Inter-Regional Inequality in China that Use Provincial GDP *per capita* as an Indicator

Authors	Period	Statistic(s)	Scale	Key patterns in inequality (in reform era)	Population data sources used in GDP *per capita* denominator
Duncan and Tian (1999)*	1952–95	CV	InterP, IntraP	Rose in the first half of the 1990s	China Statistical Yearbook, Provincial Statistical Yearbook
Ying (1999)	1978–94	Theil	InterP	Declined until 1990; then increased	40 Years (a); China Statistical Yearbook, 1985-94
Fujita and Hu (2001)	1985–94	CV, Theil	InterP	Declined in the 1980s; increased in 1990s	China Statistical Yearbook
Cai et al. (2001)	1978–99	CV, Theil	InterP	Declined between 1978 and 1990 but increased since around 1990	50 Years
Lu and Wang (2002)*	1978–98	CV, GINI, Theil	InterP, 3 regions, R/U	Declined from 1978 to 1990 and then increased steadily	China Statistical Yearbook, 40 Years (a) and 20 Years
Huang et al. (2003)	1991–01	GINI	7 regions	Increased	China Statistical Yearbook
Kanbur and Zhang (2005)	1952–00	GINI	InterP, R/U coastal-inland	Increased sharply and steadily since 1984	1952-1978: 40 Years (b); 1978-2000: 50 Years, Agricultural 50 Years, China Statistical Yearbook, China Rural Statistical Yearbook
Xu and Li (2006)	1978-04	GINI	InterP	A slight decline beginning around 2003	Provincial Statistical Yearbooks
Tsui (2007)	1952–99	Theil	InterP	Upward trend in first half of 1990s, results based on official data show sharp increase since 1995 while author's adjusted data show no change	1952-1981: 50 Years and Provincial Statistical Yearbooks; 1982-1999: extrapolation from 1982, 1990 and 2000 population censuses
Liu and Zhang (2007)	1952-06	CV, GINI	InterP	Declined in 2000-2001 and 2003-2006	1952-1998: 50 Years; 1999-2006: China Statistical Yearbook
Fan and Sun (2008)	1978–06	CV, Theil, GINI	InterP, 3 regions, InterR	Declined in 1980s; increased in1990s, then stabilized	1978-2004: 55 Years; 2005-2006 China Statistical Yearbook; extrapolation: replacing 1990, 1995, 2000, 2005 census and sample survey population by averaging prior and subsequent year annual yearbook population
Chan and Wang (2008)	1990-06	CV	InterP	Rose in the first half of the 1990s; levelled off and stable since 1995 due to impact of long-distance migration	1990-2000: Benchmark of population censuses and sample survey in 1990, 2000 and 1995; 2000-2006: China Statistical Abstract
Li and Wei (2010)	1978-07	CV, Theil, GINI	InterP, 3 regions, InterR	Declined in 1980s; increased since 1990; declined since 2004	China Data Online
Hoshino (2011)	1979-09	CV, Theil, GINI, MLD, A(e)	InterP, 3 regions, IntraR	Declined in 1980s; increased in 1990s, especially if using registered population as denominator; stable or declining since 2004, depending on choice of population denominator	1978-1981: China Population Statistical Yearbook 1990; 1982-2005: extrapolation from 1982, 1990, 2000 population censuses and 2005 population sample survey; 2006-2009: China Statistical Abstract 2010

Notes: *=also uses indicators other than *per capita* GDP. CV = coefficient of variation, MLD = mean logrithimic deviation, A(e)=Atkinson inequality index. InterP = interprovincial; IntraP = intraprovincial; InterR = interregional, IntraR = intraregional; R/U = rural/urban.

Details for Population Source Data:

20 Years = National Bureau of Statistics (NBS). 1998. The twenty years of spectacular achievements. Beijing: Statistics Press of China.

40 Years (a) = Hsueh T-T., Li, Q. and Liu, S. (Eds). 1993. China's provincial statistics 1949-1989. Boulder CO: Westview Press.

40 Years (b) = National Bureau of Statistics (NBS), Historical Statistical Materials for Provinces, Autonomous Regions and Municipalities 1949-1989 (*Quanguo Gesheng Zizhiqu Zhixiashi Lishi Tongji Ziliao Huibian*). Beijing: China Statistical Publishing House (1990).

50 Years = National Bureau of Statistics (NBS), Comprehensive Statistical Data and Materials on 50 Years of New China (*Xin Zhongguo Wushinian Tongji Ziliao Huibian*). Beijing: China Statistical Publishing (1999).

Agricultural 50 Years = National Bureau of Statistics (NBS), Comprehensive Agricultural Statistical Data and Materials on 50 Years of New China (*Xin Zhongguo Wushinian Nongye Tongji Ziliao Huibian*). Beijing: China Statistical Publishing House (2000).

55 Years = National Bureau of Statistics (NBS), China Compendium of Statistics 1949-2004 (*Xin Zhongguo Wushiwunian Tongji Ziliao Huibian*). Beijing: Statistics Press of China, 2005.