http://researchcommons.waikato.ac.nz/

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

Evaluation of estimators for ill-posed statistical problems subject to multicollinearity

A thesis

submitted in fulfilment

of the requirements for the Degree

of

Master of Management Studies

in Economics

at

The University of Waikato

by

Luke M. Holland

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2014

Abstract

Multicollinearity is a significant problem in economic analysis and occurs in any situation where at least two of the explanatory variables in a model are related to one another. The presence of multicollinearity is problematic, as changes in the dependent variable cannot be accurately attributed to individual explanatory variables. It can cause estimated coefficients to be unstable and have high variances, and thus be potentially inaccurate and inappropriate to guide management or policy. Due to this problem, many alternative estimators have been developed for the analysis of multicollinear data.

The primary objective of this thesis is to compare and contrast the performance of some of these common estimators, as well as a number of new estimators, and test their prediction accuracy and precision under various circumstances. Through the use of non-trivial Monte Carlo experiments, the estimators are tested under 10 different levels of multicollinearity, with regressors and errors drawn from different distributions (normal, $t_{(3)}$, $\chi^2_{(5)}$, and in the case of errors, mixed Gaussian). Insights are gained through response surface analysis, which is conducted to help summarise the output of these simulations.

A number of key findings are identified. The highest levels of mean square error (MSE) are generally given by a Generalised Maximum Entropy estimator with narrow support bounds defined for its coefficients (GMEN) and the One-Step Data Driven Entropy (DDE1) model. Yet, none of the estimators evaluated produced sufficiently high levels of MSE to suggest that they were inappropriate for prediction. The most accurate predictions, regardless of the distributions tested

or multicollinearity, were given by Ordinary Least Squares (OLS). The Leuven-2 estimator appeared relatively robust in terms of MSE, being reasonably invariant to changes in condition number, and error distribution. However, it was unstable due to variability in error estimation arising from the arbitrary way that probabilities are converted to coefficient values in this framework. In comparison, MSE values for Leuven-1 were low and far more stable than those reported for Leuven-2.

The estimators that produced the least precision risk, as measured through mean square error loss (MSEL), were the GMEN and Leuven-1 estimators. However, the GMEN model requires exogenous information and, as such, is much more problematic to accurately apply in different contexts. In contrast, two models had very poor precision in the presence of multicollinear data, the Two-Step Data Driven Entropy (DDE2) model and OLS, rendering them inappropriate for estimation in such circumstances.

Overall, these results highlight that the Leuven-1 estimator is the most appropriate if a practitioner wishes to achieve high prediction accuracy and precision in the presence of multicollinearity. Nevertheless, it is critical that more attention is paid to the theoretical basis of the Leuven-1 estimator, as relating estimated probabilities to coefficients using concepts drawn from the theory of light appears highly subjective. This is illustrated through the differences in empirical results obtained for the Leuven-1 and Leuven-2 estimators.

Acknowledgements

I would like to thank my supervisor Dr Graeme Doole for many things. This thesis would not have been possible without his help with everything from the topic itself, through to the final editing. I could not be more grateful for everything he has done. Without his support and direction this never would have happened.

A huge thanks goes to Dr Steven Tucker, for loaning me his office to work in and thus saving me from having to fight first years for space in the library.

My parents, Murray and Christine Holland, thank you so much for the support and help throughout this. For everything from sending me newspaper articles which could potentially, somehow have been relevant to what I was researching, through to ringing me in the middle of the day and giving me a small distraction from my thesis when I was going slowly insane. I never would have gotten here without their help.

And finally, if I were to write out how much thanks she deserves this would be as long as the thesis itself. So I will just leave it at eternal thanks for Kristi Bennett. I never would have managed to finish this without her. All the times I was going insane and she helped me. All the times she proofread various sections even if she had no idea what I was talking about, I don't think she has any idea how much she has helped me over the course of this thesis. Even if seeing the insanity as she finished hers terrified me, I knew that no matter what, she would help me get through mine, and she has. She gets more thanks than I could ever explain.

Table of Contents

Section **Page Number**

Abstract ii

Acknowledgements iv

Table of Contents v

List of Tables.....ix

List of Figures xiii

List of Appendices xiv

Chapter 1: Introduction 1

Chapter 2: Review of the Relevant Literature 8

2.1 Multicollinearity.....9

2.2 Effects of multicollinearity 11

2.3 Detection of multicollinearity 15

2.4 General methods of dealing with multicollinearity.....26

2.5 Generalised Maximum Entropy regression to overcome multicollinearity 32

2.6 Leuven regression to overcome multicollinearity.....37

2.7 Conclusions.....40

Chapter 3: Methods 41

3.1 Numerical Simulations.....42

3.2 OLS & RLS estimators 45

3.3 Ridge-1 & ridge-2 estimators.....46

3.4 GME estimator.....48

3.5 Leuven-1 & Leuven-2 estimators.....50

3.6 DDE2 estimator.....53

3.7 DDE1 estimator.....58

3.8 Estimator Comparison.....62

3.9 Response surface analysis.....63

3.10 Summary.....64

Chapter 4: Results and Discussion.....65

4.1 Comparison of mean squared error for each estimator.....66

4.1.1 Mean squared error for all models when condition number is 1.....66

4.1.2 Mean squared error for all models when condition number is 40.....69

4.1.3 Mean squared error for all models when condition number is 60.....70

4.1.4 Mean squared error for all models when condition number is 100.....72

4.1.5 Mean squared error for all models when condition number is 500.....74

4.1.6 Mean squared error: Comparison of means for condition numbers: 1, 40, 60, 100, 500.....75

4.2 Comparison of mean squared error loss for each estimator.....78

4.2.1 Mean squared error loss for all models when condition number is 1.....79

4.2.2 Mean squared error loss for all models when condition number is 40.....80

4.2.3 Mean squared error loss for all models when condition number is 60.....82

4.2.4 Mean squared error loss for all models when condition number is 100.....84

4.2.5 Mean squared error loss for all models when condition number is 500 86

4.2.6 Mean squared error loss: Comparison of means for condition numbers: 1, 40, 60, 100, 500 88

4.3 Response Surface Results 92

4.3.1 Response Surface Results for mean squared error 92

4.3.2 Response Surface Results for mean squared error loss 99

Chapter 5: Conclusion 106

5.1 Motivation and Method 107

5.2 Key findings 108

5.3 Limitations and further research 112

5.4 Final word 113

Literature Cited 115

Appendix 1: GAMS code 122

A1.1 Importing of the data from Microsoft Excel. 122

A1.2 DDE ST50 Programming for Monte Carlo simulations 123

Appendix 2: Extended Results 146

A2.1 Comparison of mean squared error for each estimator 146

A2.1.1 Mean squared error for all models when condition number is 10 146

A2.1.2 Mean squared error for all models when condition number is 20 148

A2.1.3 Mean squared error for all models when condition number is 80 149

A2.1.4 Mean squared error for all models when condition number is 250 151

A2.1.5 Mean squared error for all models when condition number is 1000 152

A2.2 Comparison of mean squared error loss for each estimator 154

A2.2.1Mean squared error loss for all models when condition number is 10 154

A2.2.2 Mean squared error loss for all models when condition number is 20 155

A2.2.3 Mean squared error loss for all models when condition number is 80 157

A2.2.4Mean squared error loss for all models when condition number is 250 158

A2.2.5 Mean squared error loss for all models when condition number is 1000 159

A2.3 Average tables for each condition number and estimator 161

A2.3.1 Average MSE values for each estimator over all 10 studied condition numbers 161

A2.3.2 Average MSEL values for each estimator over all 10 studied condition numbers 162

List of Tables

Table **Page Number**

Table 2.1: Annual data for the Canadian Economy from 1975–1994 16

Table 2.2: Correlations between explanatory variables 17

Table 2.3: Eigenvectors of sample data 20

Table 2.4: Variance inflation factors for the data set used in this review 23

Table 4.1: Mean squared error for estimators with a condition number of 1 67

Table 4.2: Mean squared error for estimators with a condition number of 40 69

Table 4.3: Mean squared error for estimators with a condition number of 60 71

Table 4.4: Mean squared error for estimators with a condition number of 100 72

Table 4.5: Mean squared error for estimators with a condition number of 500 74

Table 4.6: Mean squared error average values for each estimator, and condition numbers: 1, 40, 60, 100, 500 75

Table 4.7: Mean squared error loss for estimators with a condition number of 1 79

Table 4.8: Mean squared error loss for estimators with a condition number of 40 81

Table 4.9: Mean squared error loss for estimators with a condition number of 60 82

Table 4.10: Mean squared error loss for estimators with a condition number of 100 84

Table 4.11: Mean squared error loss for estimators with a condition number of 500 86

Table 4.12: Mean squared error loss average values for each estimator, and condition numbers: 1, 40, 60, 100, 500 88

Table 4.13: Summary Statistics for the MSE (Dependent Variable) used in the Response Surface Analysis for all Estimators.....93

Table 4.14: Response surface results for MSE 96

Table 4.15: Summary Statistics for the MSEL (Dependent Variable) used in the Response Surface Analysis for all Estimators.....99

Table 4.16: Response surface results for MSEL 101

Table A2.1: Mean squared error for estimators with a condition number of 10 147

Table A2.2: Mean squared error for estimators with a condition number of 20 148

Table A2.3: Mean squared error for estimators with a condition number of 80 150

Table A2.4: Mean squared error for estimators with a condition number of 250 151

Table A2.5: Mean squared error for estimators with a condition number of 1000 153

Table A2.6: Mean squared error loss for estimators with a condition number of 10 155

Table A2.7: Mean squared error loss for estimators with a condition number of 20 156

Table A2.8: Mean squared error loss for estimators with a condition number of 80 157

Table A2.9: Mean squared error loss for estimators with a condition number of 250 158

Table A2.10: Mean squared error loss for estimators with a condition number of 1000 160

Table A2.11: Mean squared error average values for each condition number and estimator.....161

Table A2.12: Mean squared error loss average values for each condition number and estimator 162

List of Figures

Figure **Page Number**

Figure 2.1: Scatterplot of relationships between personal disposable income per person and CPI for chicken from the Canadian economy data set (Table 2.1) 18

List of Appendices

Appendix **Page Number**

Appendix 1: GAMS code used in this thesis 122

Appendix 2: Extended results 146

Chapter 1

Introduction

Stochastic linear inverse problems are prevalent in Economics. These problems involve the general linear model $y=X\beta+e$, where y is a T-vector of observations, X is a $T \times K$ data matrix reporting values for T observations across K parameters, β is a vector of unknown explanatory variables to be estimated, and e is an unobservable vector of T length. An ill-conditioned data matrix causes estimated coefficients to have unstable magnitudes and high variances, and also makes inversion potentially impossible (Belsley, 1991). There are numerous causes for ill-conditioned regression problems, particularly in Economics, as data is often passively generated or obtained from surveys rather than generated from formulated experiments. Some of these causes include using non-experimental data in analysis, data transformations, or using an ill-defined model where there are more parameters than there are observations ($K > T$) (Doole, 2013).

While ill-conditioning describes any effect in a data matrix which causes large changes in the regression estimates due to a small change in the data (Belsley, 1991), multicollinearity is the primary cause of this type of behaviour in regression models, and thus is broadly referred to as ill-conditioning, particularly in econometrics (Heckelei, 2002). Multicollinearity occurs in a regression when at least two of the explanatory variables in the model being looked at have a relationship with each other, as well as with the dependent variable (Alauddin & Nghiem, 2010; Alin, 2010; Belsley, 1991; Murray, 2006; Yamagata & Orme, 2005). Multicollinearity is generally split into two types: perfect and imperfect multicollinearity. Perfect multicollinearity occurs when at least two of the regressors move exactly in step with each other (Murray, 2006). This causes Ordinary Least Squares (OLS) procedures to break down due to the creation of

estimates which are not unique (Stewart, 1987). However, such instances are relatively rare, unless an error is made in specification.

Imperfect multicollinearity is much more difficult to deal with, as well as being a lot more prevalent than perfect multicollinearity. Imperfect multicollinearity[1] means that at least two of the explanatory variables are related to each other; however, this relationship need not be perfect. Imperfect multicollinearity tends to occur when there is deficient data, which is often seen in observational studies such as those conducted in Economics (Alin, 2010). Sometimes this is an inherent flaw in the data set due to uncontrollable aspects of data gathering (Belsley, 1991). In the case of Economics, data is produced by the functioning economic system, and the multicollinearity in the data reflects that system (Johnston, 1984). Multicollinearity causes problems when interpreting regression results, as the parameters cannot be interpreted on an individual basis, given their interdependency (Alin, 2010; Murray, 2006). Multicollinearity also makes estimates and model parameters less reliable through creating large standard errors and inflated variances (Farrar & Glauber, 1967; Greenberg & Parks, 1997; Kiers & Smilde, 2007; Lauridsen & Mur, 2006; Meloun et al., 2002; Stewart, 1987; Thomas, 1993). However, regressions containing multicollinear variables are often still useful for prediction, provided that the multicollinearity is expected to continue in the future (Greenberg & Parks, 1997).

There are many forms of regression that attempt to deal with multicollinearity. The primary objective of this thesis is to compare and contrast the performance of

[1]When multicollinearity is referred to in this thesis, it will be referencing imperfect multicollinearity.

the main estimators used within Economics to deal with multicollinearity, and a number of new estimators that have been proposed but have received little testing, relative to these established techniques. Ten estimators are assessed. These are Ordinary Least Squares (Campbell & Hill, 2005), ridge regression with the ridge parameter computed in two different ways (Hoerl & Kennard, 1970; Mittelhammer et al., 2000), Restricted Least Squares (Akdeniz et al., 2011), Generalised Maximum Entropy (Golan et al., 1996) with both narrow and wide support bounds, a Two-Step Data Driven Entropy estimator (Doole, 2013), a One-Step Data Driven Entropy estimator, a Leuven-1 regression (Paris, 2001, 2004), and a Leuven-2 regression (Paris, 2004). A particular focus is placed on the Two-Step Data Driven Entropy and the One-Step Data Driven Entropy methods that have been recently developed at the University of Waikato (Doole, 2013), but have not yet been empirically tested relative to a broad range of alternatives.

This set of estimators is compared through a series of non-trivial, multi-factorial Monte Carlo experiments that allow comparisons between the methods across 10 significantly different levels of multicollinearity. Regressors and errors are drawn from different distributions to test the robustness of the methods across different data-generating scenarios. Insights are drawn from the interpretation of output of the simulations, including the use of response surface analysis (Campbell & Hill, 2005) to summarise the effects of alternative estimators. Response surface analysis involves the use of regression methods to summarise the output of a high number of simulation experiments for a wide range of econometric estimators.

Ordinary Least Squares (OLS) and Restricted Least Squares (RLS) methods are very similar, except for one important aspect. Both work to estimate the unknown

parameters in a regression via the minimisation of the squared deviation of the error. However, while OLS employs unrestricted coefficients, RLS involves setting a restriction on the coefficient values. In comparison, ridge regression works through the addition of a small value *k* (the ridge parameter) to the diagonal of the correlation matrix (Fourgeaud et al., 1984). The addition of this parameter allows ridge regression to be flexible enough to reduce coefficient variances from their OLS values, when multicollinearity is present (Li et al., 2010). Nevertheless, the optimal value of *k* is difficult to estimate; therefore, two alternative methods are tested in this thesis.

Another estimator that deals with ill-posed regressions is the Generalised Maximum Entropy (GME) estimator (Golan et al., 1996). It has been developed to provide robust solutions for ill-posed stochastic inverse problems such as those described above, and involves using the maximum entropy principle (Jaynes, 1957) to recover the unknown probability distributions of coefficients and error terms for ill-posed problems (Ciavolino & Al-Nasser, 2009). Maximisation of entropy conservatively forces the probability terms for each coefficient/error to be the most uniform (most uncertain), while still satisfying the data constraints. This allows a practitioner to identify the probabilities that could have been generated in the most number of ways consistent with the data (Jaynes, 1957). This is the most appropriate approach, as to choose a solution with lower entropy requires extrapolation, while choosing a solution with higher entropy violates existing data (Golan, 2006). GME is able to cope with small sample sizes, avoid specific assumptions regarding the distribution of errors, and incorporate prior information through appropriate support definition (Golan et al., 1996). However, there are problems with this estimator, primarily the fact that it is susceptible to high

variability in estimated outcomes as the proposed ranges (i.e. supports) of the coefficients and error terms vary (Paris, 2001, 2004).

The Leuven-1 and Leuven-2 estimators have been designed to overcome multicollinearity. These estimators were put forward in Paris (2001), and Paris (2004) respectively. Based on the theory of light, they aim to create a maximum entropy estimator which does not require additional information. As such, these estimators do not experience variability in outcomes with different bounds placed on the coefficients, as the GME estimator does (Macedo et al., 2010; Mishra, 2004; Paris, 2001), but are still robust to multicollinearity (Macedo et al., 2010; Paris, 2004).

The remaining two estimators examined in this thesis are also based on the GME estimator, but are designed to be robust to variability in the ranges of coefficients and error terms. Like the Leuven estimators, these estimators endogenously estimate the ranges of these terms. However, whereas the Leuven estimators base their bounds on the theory of light, these approaches estimate the bounds from the data itself. The Two-Step Data Driven Entropy (DDE2) estimator and One-Step Data Driven Entropy (DDE1) estimator both use interval-valued estimation to identify endogenous support bounds. These bounds assume that no information on the shape of a given distribution is available, thus a uniform distribution is adopted, aligning with the maximum entropy principle. The Least Absolute Deviations (LAD) estimator is used, following transcription to an interval-valued programming context using the transcription of Wu (2008). The results from this estimation are then used to give informative bounds for the GME estimator, thus removing the need to formulate exogenous support bounds by using appropriate

estimates of these from the data. The DDE2 estimator involves separate estimation of the support bounds and the regression, while the DDE1 estimator involves their integration.

The thesis has the following structure. Chapter 1 is the introduction, followed in Chapter 2 by a review of the relevant literature. Chapter 3 will provide an overview of the methods used in this thesis, Chapter 4 will contain the results and discussion from the experiments, and Chapter 5 will contain conclusions.

Chapter 2

Review of the Relevant Literature

2.1 Multicollinearity

Multicollinearity is a condition that occurs in a regression when at least two of the explanatory variables in the model being looked at have a relationship between each other (Alauddin & Nghiem, 2010; Alin, 2010; Belsley, 1991; Murray, 2006; Yamagata & Orme, 2005). It is often seen as a symptom of poor experimental design (Farrar & Glauber, 1967; Haitovsky, 1969).

Distinctions are made between different types of multicollinearity, although each type often has different names in different studies. The common distinctions are those between perfect (exact) multicollinearity and imperfect multicollinearity. Moreover, the terms ill-conditioning and collinearity are also sometimes used interchangeably, though ill-conditioning describes any effect in a data matrix that causes large changes in the regression estimates, due to a small change in the data, so does not involve multicollinearity alone (Belsley, 1991). However, multicollinearity is the primary cause of such behaviour in regression models; thus, it is also broadly called ill-conditioning, particularly in econometrics (e.g. Heckelei, 2002).

Multicollinearity can cause problems regardless of whether it is exact or not. It tends to occur when there is deficient data, which often occurs in observational studies, such as those conducted in Economics (Alin, 2010; Murray, 2006). Regression problems involving perfect multicollinearity can usually be fixed easily through appropriate diagnosis. However, in contrast, imperfect multicollinearity is much harder to deal with. Perfect multicollinearity occurs when there are explanatory variables in the regression that move exactly in step with each other (Murray, 2006). Perfect multicollinearity is caused by the $[X'X]^{-1}$

matrix not being invertible for the linear-squares estimator (Thomas, 1993). This effect leads to a situation where there is an infinite number of coefficients from the regression that minimise the sum of squared residuals (Stewart, 1987). Thus, perfect multicollinearity creates a state of complete uncertainty as to the value of the unknown regression parameters (Thomas, 1993). However, perfect multicollinearity occurs very rarely in reality (Thomas, 1993). The most common instance in empirical work is where practitioners include binary variables for all options represented by a dummy variable. This "dummy variable trap" causes perfect multicollinearity between the dummy variable and the constant term in the regression (Stock & Watson, 2007).

If the relationship between the explanatory variables is close, but not exact, then these variables are referred to as collinear or multicollinear (Murray, 2006), as opposed to perfectly multicollinear. When data is collinear (but not perfectly collinear), the linear combination of the relevant columns is small, however not zero (as with perfect multicollinearity) (Stewart, 1987), which is revealed by a large partial correlation coefficient between two of the explanatory variables (Kumar, 1975). An example is where age and the number of years that a person has spent in a profession are strongly correlated. Collinearity is used to describe the case when the inclusion of a variable makes the total collinearity present within a regression high enough to be classified as harmful (Lauridsen & Mur, 2006).

Some authors define multicollinearity as the presence of a relationship between more than two of the explanatory variables, inherently defining collinearity as describing a single relationship (Alauddin & Nghiem, 2010). However, while the

distinction is mentioned here for the sake of completeness, it is rarely maintained in practice and the terms are generally used interchangeably (Alauddin & Nghiem, 2010). Multicollinearity is also often confused in this way with correlation, which exists when there is a linear relationship between only two variables (as in the case of collinearity above), making it a special case of multicollinearity. As such, high correlation implies the existence of multicollinearity—though, this is not true in reverse. There can be multicollinearity between explanatory variables without there being high correlation between pairs of these variables (Belsley, 1991; Chennamaneni et al., 2008; Thomas, 1993). This occurs when a relationship exists between sets of explanatory variables. This can occur in any situation with at least three independent variables, where none of the three have a high correlation with the others independently. As such, the correlation matrices would not reflect this due to low levels of correlation between any two variables (Belsley, 1991; Chennamaneni et al., 2008). Accordingly, multicollinear relationships between more than two explanatory variables are often identified using auxiliary regressions (Belsley, 1991).

2.2 Effects of multicollinearity

One of the key implications of multicollinearity is that it causes issues when interpreting multidimensional evidence. While information can be obtained as to the linear combinations of the regression coefficients, regression will not show how to allocate the effect to each individual explanatory variable. This means that results cannot be interpreted parameter by parameter (Leamer, 1973), as multicollinearity confuses the relative importance of each coefficient. This makes

it hard to identify the individual effect of changes in any one explanatory variable on the dependent variable (Alin, 2010; Murray, 2006).

Another key problem that multicollinearity creates is a decrease in the reliability of estimates and model parameters through the creation of large standard errors, ill-conditioned results, improper specification, and inflated variances of the coefficient estimates (Alin, 2010; Farrar & Glauber, 1967; Greenberg & Parks, 1997; Kiers & Smilde, 2007; Lauridsen & Mur, 2006; Meloun et al., 2002; Murray, 2006; Stewart, 1987; Thomas, 1993). This can potentially harm estimation, hypothesis testing, and forecasting. High standard errors found in the presence of multicollinearity demonstrate the lack of precision it causes (Thomas, 1993), reducing the information content of the data, relative to a data set of an equivalent size in which such a pattern is not evident (Farrar & Glauber, 1967). Least squares estimators have the smallest variance when the least squares assumptions are met[2] (Stock & Watson, 2007). However, when there is multicollinearity, even this variance may be larger than reasonable, and thus least squares may not be as efficient as other methods in this case (Meloun et al., 2002). Estimation difficulties caused by multicollinearity—including the reduction of the reliability of least squares and maximum likelihood estimators—become more pronounced the more closely related the explanatory variables are (Lauridsen & Mur, 2006; Thomas, 1993).

Multicollinearity is only a problem if it is present in the variables which are crucial in the analysis (Farrar & Glauber, 1967). This is because multicollinearity

[2]The least squares assumptions being that: For a given value of X, the distribution of the errors has a mean of 0, the sample is independently and identically distributed, and large outliers are unlikely.

only affects the variances of the collinear explanatory variables, not the entire model (Murray, 2006). Another aspect of note is that when there is severe multicollinearity in a regression, the parameter estimates that are identified are very sensitive to changes in the model specifications, such as adding or deleting variables that are statistically insignificant (Alin, 2010; Farrar & Glauber, 1967).

Multicollinearity will not affect predictions, as long as there is multicollinearity in the period for which the predictions are being made, as well as in the sample period (Greenberg & Parks, 1997). Likewise, Kiers & Smilde (2007) point out that multicollinearity does not cause prediction problems if it is present in the population the sample was taken from, as well as the sample itself.

It is possible for all individual variables to appear insignificant in the presence of multicollinearity, though the model as a whole could be significant (Alin, 2010). This is because instability in the regression coefficients does not affect the proportion of the variance of the dependent variable which is explained by movements in the explanatory variables (as indicated by the R^2 measure) (Rockwell, 1975).

In a misspecified model, multicollinearity can potentially be an asset, as opposed to a liability (Haitovsky, 1969). While this opinion is seldom put forward, it has been noted that multicollinearity is not always harmful, as in some cases its effects can be negated through sufficiently small error variances, i.e. in a situation where enough information or data is present to reduce the error variances (Alin, 2010; Thomas, 1993). It is also possible for multicollinearity to not be a problem, even for least squares estimators, provided that the model is properly specified

and feedback is absent[3] (Beaver, 1974): however, it is acknowledged that the presence of multicollinearity makes this unlikely (Farrar & Glauber, 1967). The harmfulness of multicollinearity, of course, depends upon both its degree and the practitioner's goals. If the goal is simply to predict a dependent variable from a set of *x* variables using OLS, then the predictions will still be accurate (Paul, 2008). It has also been suggested that a high R^2 and large sample size can offset the problems caused by multicollinearity (Grewal et al., 2004). Multicollinearity does not by itself lead to biased or inconsistent estimates; however, interactions with variables such as measurement or rounding errors can lead to large inconsistencies (Green & Kiernan, 1989; Stewart, 1987). Some practitioners act as if there is no problem with multicollinearity, primarily because least squares estimators should still in theory work when there is multicollinearity. However, they also admit that, through its presence, multicollinearity creates bias towards incorrectly-specified models (Farrar & Glauber, 1967). Some of these problems can be seen through the fact that with multicollinearity the Wald test for sample selection bias rejects too often, and Lagrange multiplier test statistics are often negative (Yamagata & Orme, 2005).

The key problem posed by multicollinearity in regression stems from there not being enough information in the sample to be able to facilitate appropriate model creation and individual parameter estimation (Farrar & Glauber, 1967; Feldstein, 1973). An important aspect of the multicollinearity problem comes from prior information. It is claimed that if our prior knowledge of parameter values was either completely certain or completely uncertain, then the most concerning

[3]The term "feedback" is a recursive condition which refers to the case where a change in *x* affects *y*, and this change in *y* affects *x*.

aspects of multicollinearity would disappear. Indeed, if we have total prior information it is easier to work out the effects of variables. In contrast, if there is no prior information, we are prevented from letting our estimates for one parameter depend on our uncertain prior information about another parameter, thus exacerbating the instability of coefficient estimates (Leamer, 1973).

2.3 Detection of multicollinearity

There are numerous empirical applications in which multicollinearity is observed in Economics (e.g. Ahn et al., 2013; Alauddin & Nghiem, 2010; Bowles & Levin, 1967; Agudo & Gimeno, 2005; Kalatzis et al., 2011; Mahajan et al., 1977; Mittelhammer et al., 1980; Salinas & Hillmer, 1987; Wheeler & Tiefelsdorf, 2005). Accordingly, the detection of multicollinearity is important in applied studies. To help illustrate some of these methods for detecting multicollinearity, we will use a data set containing annual data from the Canadian economy from 1975 to 1994 that has been retrieved from the CANSIM Statistics Canada database. This data set is presented in Table 2.1 below.

Table 2.1: Annual data for the Canadian Economy from 1975–1994.

Year	Per capita use of beef - retail basis (in pounds)	Fresh or frozen beef - Consumer price index (1986=100)	Fresh or frozen chicken - Consumer price index (1986=100)	Personal disposable income per person (in dollars)
1975	79.2	42	51.3	4,883
1976	84.2	39.3	51.3	5,453
1977	80.1	41.8	51.6	5,941
1978	75	61.1	59.9	6,634
1979	63.7	80.3	66	7,408
1980	63.1	87.2	71	8,281
1981	64.8	89.5	84.5	9,545
1982	64.2	88.9	86.8	10,430
1983	63.9	89.6	89.9	10,843
1984	60.9	95.5	95.8	11,686
1985	61.7	97.9	91.8	12,387
1986	61.2	100	100	12,902
1987	58.2	109.1	106.1	13,613
1988	58.2	110.8	107.6	14,658
1989	56.6	113.2	120	15,783
1990	54.5	117.6	126.1	16,263
1991	53.4	118.3	123.5	16,570
1992	51.7	116.8	123	16,753
1993	49.6	123.1	125.6	16,874
1994	50.6	124.1	119.3	17,003

When detecting multicollinearity, one rule of thumb defines multicollinearity as too severe to work with when the Pearson correlation coefficient values computed for the relationships between explanatory values are higher than $r=0.8$ (Farrar & Glauber, 1967; Grewal et al., 2004; Kumar, 1975). Table 2.2 indicates that the correlation coefficients for the data set in Table 2.1 are very high, indicating the presence of strong multicollinearity.

Table 2.2: Correlations between explanatory variables

	Fresh or frozen beef - Consumer price index (1986=100)	Fresh or frozen chicken - Consumer price index (1986=100)	Personal disposable income per person (in dollars)
Fresh or frozen beef - Consumer price index (1986=100)	1	0.96	0.96
Fresh or frozen chicken - Consumer price index (1986=100)	0.96	1	0.99
Personal disposable income per person (in dollars)	0.96	0.99	1

A second rule of thumb is that multicollinearity is not a problem unless it is high when looked at relative to the overall degree of correlation in the regression. In other words, it is not a problem if the correlation between two independent variables is equal to or higher than the level of multiple correlation between the dependent and independent variables (Farrar & Glauber, 1967).

A third rule of thumb is that multicollinearity can be detected from scatter plots of the relationships between variables. However, multicollinearity in these figures can be hidden by other effects, such as leverage points or influential points in the data which may cause distortion (Meloun et al., 2002). For example, in Figure 2.1 it seems that there is a correlation between two explanatory variables from Table 2.1. This can also be seen in Table 2.2, where the explanatory variables from Table 2.1 seem to have very high correlations between each other (all of the correlations being above 0.95); this implies the existence of strong multicollinearity in the data. However, this test is not always accurate, as there can still be multicollinearity even if the correlations between variables are all low. This occurs when the multicollinearity is caused by linear combinations of

variables, as opposed to being between variables themselves (as discussed at the end of Section 2.1) (Alin, 2010; Belsley, 1991; Chennamaneni et al., 2008).

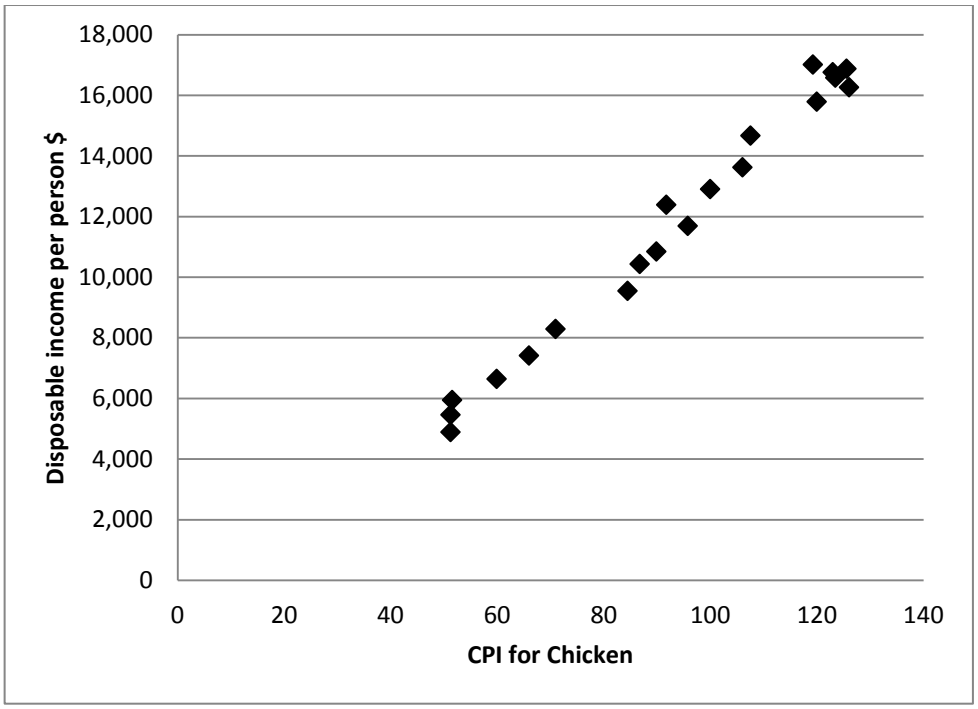

Figure 2.1: Scatterplot of relationships between personal disposable income per person and CPI for chicken from the Canadian economy data set (Table 2.1).

Another relatively informal test for multicollinearity is looking for the effects which multicollinearity tends to create in a regression model. If there are large standard errors, ill conditioned results, improper specification, inflated variances, and yet a good fit of the model (as measured through the R^2 measure), then there is a good chance that multicollinearity is present (Alin, 2010; Farrar & Glauber, 1967; Greenberg & Parks, 1997; Kiers & Smilde, 2007; Lauridsen & Mur, 2006; Meloun et al., 2002; Stewart, 1987; Thomas, 1993).

Multicollinearity can also be identified through the use of various numerical and statistical tools (Meloun et al., 2002). Some of the most common of these being

eigenvalues and eigenvectors (Alin, 2010; Grewal et al., 2004; Meloun et al., 2002), condition numbers (Alin, 2010; Grewal et al., 2004; Meloun et al., 2002; Stewart, 1987), determinant of *R* (Alin, 2010; Grewal et al., 2004), variance inflation factors (VIF) (Alauddin & Nghiem, 2010; Alin, 2010; Grewal et al., 2004; Meloun et al., 2002; Stewart, 1987), and collinearity indices (Stewart, 1987).

The method of examining eigenvalues and eigenvectors in an effort to identify multicollinearity is often called the principal component approach and reveals the number of linear dependencies in the model, as well as their structure (Alin, 2010; Mahajan et al., 1977; Paul, 2008). However, the problem with this method is attempting to clarify what values are significant enough to signify potentially harmful multicollinearity, i.e. when the multicollinearity can be expected to cause errors to any inferences made from the model (Grewal et al., 2004). Eigenvectors of X^tX or the matrix of closely-related correlations (R) are the set of nonzero vectors ζ that follow the rule $X^tX\zeta=\lambda\zeta$. In other words, eigenvectors are vectors which are turned into a scalar of themselves having been multiplied by λ when multiplied by X^tX (Belsley, 1991). The eigenvectors for the data set used in this review (Table 2.1) can be seen below in Table 2.3.

The value of λ is sometimes termed the "stretch factor" and is the corresponding eigenvalue to a given eigenvector. If an eigenvector has an eigenvalue of zero then this means that $X\zeta=0$, meaning that there is an exact linear dependence between the columns of the matrix X; in other words, perfect multicollinearity exists (Belsley, 1991). It has been suggested that small eigenvalues can signal the presence of multicollinearity (Silvey, 1969). This is logical as very small

eigenvalues would lead to $X\zeta$ being approximately zero, and so signalling multicollinearity. However, in practicality, the usefulness of this statement is limited as there is no definition of what is classed as significantly small (Belsley, 1991). Indeed, while Table 2.2 clearly indicates the presence of multicollinearity, Table 2.3 does not, because whether the values can be classified as small or not is not entirely obvious.

Table 2.3: **Eigenvectors of sample data.**

	F1	**F2**	**F3**
Fresh or frozen beef - Consumer price index (1986=100)	0.57	0.82	-0.05
Fresh or frozen chicken - Consumer price index (1986=100)	0.58	-0.36	0.73
Personal disposable income per person (in dollars)	0.58	-0.45	-0.68

Looking at the condition number is another way of diagnosing and measuring the overall level of multicollinearity (Belsley, 1991). These condition numbers can be found through the formula:

$$k(A) = \frac{\mu_{max}}{\mu_{min}} , \qquad (2.1)$$

where *k(A)* is the condition number of the matrix *A*, μ_{max} is $||A||$, and the inversion of matrix *A* yields $1/\mu_{min}$. The condition number of a matrix with orthogonal columns is 1, thus showing that the condition number of a matrix is at its lowest in this situation, which is consistent with no multicollinearity. The higher the condition number, the more ill-conditioned the data matrix is, and the higher the degree of multicollinearity. Equation 2.1 can also be described as the square root

of the ratio between the maximum and minimum eigenvalues (Belsley, 1991). The degree of this ill conditioning is dependent on how small the minimum singular value (μ_{min}) is in comparison to the maximum singular value, thus what is classified as small depends on the data. The *k*th condition index can be defined in the formula:

$$\eta_k \equiv \frac{\mu_{max}}{\mu_k}, \qquad (2.2)$$

where $k=[1,\ldots,p]$ (Belsley et al., 1980). Generally, a condition index of 5 – 10 is considered weak multicollinearity, while 30 – 100 is considered moderate to severe multicollinearity (Alin, 2010; Belsley, 1991; Friendly & Kwan, 2012). Condition numbers are often used in numerical analysis (e.g. Paris, 2004); however, one problem is that while the condition number can be very useful as a multicollinearity indicator, it may not be specific enough for statistical applications since it distils a large amount of information into a single number (Stewart, 1987).

Matrix *R* is the correlation matrix of the explanatory variables. The determinant of *R* (*detR*) method indicates the potential severity of multicollinearity (Rockwell, 1975). The determinant of a perfectly collinear (singular) matrix is 0, whereas being close to one implies near orthogonality. However, the gradient between these two values is not very well defined, making it difficult to indicate the true level of multicollinearity, instead only providing good information regarding its potential existence (Farrar & Glauber, 1967). Also, there are times when the determinant may imply a lack of multicollinearity when it does in fact exist, particularly if there are pair wise correlations in the regression. Also, while it

identifies multicollinearity, it does not identify the number of, or the structure of, the relationships between the variables (Alin, 2010).

A very common multicollinearity diagnostic is the variance inflation factor (VIF) (Alauddin & Nghiem, 2010; Stewart, 1987). A variance inflation factor is computed through:

$$VIF_i = \frac{1}{1-R_i^2}, \qquad (2.3)$$

(Alin, 2010; Meloun et al., 2002; O'Brien, 2007), where R_i^2 is the R^2 in the regression of x_k on all other independent variables (Alauddin & Nghiem, 2010). The statistic produced denotes the increase in the variance of the parameters due to multicollinearity, relative to the level of variance which would exist if there was no multicollinearity. If there is no multicollinearity at all then R_i^2 will be near 0, and so the VIF will approach 1 (Alauddin & Nghiem, 2010). The larger the VIF produced, the more the variance has been inflated due to the presence of multicollinearity. This large value indicates a linear dependency, however, it does not tell us which variables are involved, which is a key limitation of this method, as such information is useful to guide improvement of the situation (Alin, 2010; Meloun et al., 2002). A term which is sometimes used interchangeably with variance inflation factors is tolerance (TOL). This is the inverse of the VIF, and as such if a VIF is high then the TOL would be low (Alauddin & Nghiem, 2010).

A general rule of thumb tells us that multicollinearity is serious if the VIF exceeds 10 (Alauddin & Nghiem, 2010; Alin, 2010; O'Brien, 2007) (i.e. TOL<0.1); however, this should always be looked at in context, as there are cases where even a very high VIF (i.e. 40) does not require corrective action (Alauddin & Nghiem,

2010). It can be seen from Table 2.4 below that the variance inflation factors for the CPI on beef, the CPI on chicken, and personal disposable income are all over 10 in the example from Table 2.1, with these being 12.98, 77.65 and 68.72 respectively and that the TOL results are all less than 0.1. This supports the results from Table 2.1 and Table 2.2, in telling us that multicollinearity is a definite problem in this data set.

Table 2.4: Variance inflation factors for the data set used in this review.

Statistic	Fresh or frozen beef - Consumer price index (1986=100)	Fresh or frozen chicken - Consumer price index (1986=100)	Personal disposable income per person (in dollars)
VIF	12.98	77.65	68.72
TOL	0.08	0.01	0.01

Variance decomposition proportions are multicollinearity diagnostics that have also been developed based on the concept of multicollinearity inflating variances; these can be used in conjunction with VIF (Alin, 2010). If there are collinear relationships between columns of the data matrix X, and then X is used in a regression, the variances will be inflated (Belsley, 1991). The least squares estimator can be written as: $b= (X^TX)^{-1}X^ty$, with its variance matrix being given by $\sigma^2(X^tX)^{-1}$, where σ^2 is the variance (Johnston, 1984). If we use the decomposition techniques described in Belsley (1991) where $X=UDV^T$, the variance-covariance matrix is written as $V(b) = \sigma^2(X^TX)^{-1} = \sigma^2VD^{-2}V^T$ (Belsley, 1991).

This means that the variance of the *k*th regression coefficient can be stated:

$$var(b_k) = \sigma^2 \sum_j \frac{v_{kj}^2}{\mu_j^2}, \quad (2.4)$$

with μ_j representing a diagonal matrix of singular values or eigenvalues of X (Belsley, 1991; Ettarid, 1996; Liao & Valliant, 2012). In this case, a small singular value of μ_j can lead to a large variance due to its position in the denominator. This can tell us that an unusually high proportion of variance is associated with a particular value of μ_j and that there is a near dependency or collinearity associated with that value and thus the associated regression coefficients (Belsley, 1991). We define the variance decomposition proportion for the point *(k,j)* as the proportion of the variance found in the *k*th regression coefficient, which is associated with the *j*th component of its decomposition (Belsley, 1991). To obtain the variance decomposition proportions, we use the equations:

$$\phi_{kj} \equiv \frac{v_{kj}^2}{\mu_j^2}, \quad (2.5)$$

$$\phi_k \equiv \sum_{j=1}^{p} \phi_{kj}, \quad (2.6)$$

where *k=[1,…,p].* This is then used to compute the actual variance decomposition proportions through:

$$\pi_{jk} \equiv \frac{\phi_{kj}}{\phi_k}, \quad (2.7)$$

where *k,j=[1,…,p]*. These proportions are often displayed in tables where each row corresponds to one of the values of μ_j (Belsley, 1991). It has been suggested to show both the matrix of these proportions and the condition indices of X in a

table, and if at least two elements in a row are relatively large, and the corresponding condition index is also large, then it is likely that collinearity is a factor (Liao & Valliant, 2012).

Collinearity indices provide a number for each variable, and are always larger than 1. These indices are designed to be a scale-invariant measure of multicollinearity and are useful in the sense that they are easy to interpret. While there are various ways to calculate them, the simplest of these is to take the square root of the variance inflation factors for each coefficient (VIF) (Stewart, 1987).

These indices can indicate three things (Stewart, 1987):

1. How near the regression matrix is to being exactly collinear.
2. The negative effects of errors in the variables.
3. Likelihood of declaring an important variable insignificant.

The primary advantages of these indices lie in their scale invariance and because they are easier to interpret than VIF. However, these indices are not appropriate for use in situations (like polynomial regression) where error grows from variable to variable due to functional relationships (Stewart, 1987). They are also not widely used, given that use of VIF is established and both techniques yield similar results.

2.4 General methods of dealing with multicollinearity

There are many different ways of dealing with multicollinearity.

One method that has always been relatively widely used is eliminating the variables believed to be causing multicollinearity and which seem to be contributing the least to the model (Alauddin & Nghiem, 2010; Feldstein, 1973; Grewal et al., 2004; Hill & Judge, 1987; Leamer, 1973). This is justified by the fact that multicollinearity does not affect prediction accuracy, when the entire model is used for forecasting, and the deleted variable is, at least partly, represented by its collinear counterpart that is retained in the model. However, overall, this practice is not advisable (Feldstein, 1973; Grewal et al., 2004; Leamer, 1973), primarily due to the potential for creating misspecification errors (Grewal et al., 2004) and, because of the nature of multicollinearity, removing a variable can drastically affect the magnitudes of the other estimates (Leamer, 1973). Despite these problems, the removal of a multicollinear variable is still widely used to deal with multicollinearity, mainly due to its simplicity.

A common recommendation to resolve multicollinearity is to obtain data which does not suffer from multicollinearity (Alin, 2010; Fabrycy, 1973; Farrar & Glauber, 1967; Grewal et al., 2004). However, this is often impractical (Grewal et al., 2004). It is often said that the effects of multicollinearity can be removed through this additional data (Alin, 2010; Johnston, 1984; Murray, 2006; Thomas, 1993), which could be furnished through various methods such as extra primary data collection, using parameter estimates from secondary data sources, the introduction of subjective information, or through Bayesian estimation procedures (Farrar & Glauber, 1967). However, even after any multicollinearity between the

determinants has been removed (i.e. through using prior information to turn two prices into a price ratio), it is possible for multicollinearity to persist (Fabrycy, 1973). A related idea is that of artificial orthogonalization, which reduces the information requirements of the model to the level of information in the data at hand through the decomposition of the data matrix into a set of statistically significant orthogonal common factors and residuals. However, sometimes this can lead to even worse parameter estimates than those identified with multicollinearity (Farrar & Glauber, 1967).

Principal component regression (PCR) can also be used to remove the effects of multicollinearity (Alin, 2010; Kiers & Smilde, 2007; Meloun et al., 2002). In PCR, the explanatory variables are summarised via the means of a set of "principal components", and these principal components are used as explanatory variables in the regression (Kiers & Smilde, 2007). This regression yields biased estimates (Meloun et al., 2002), but decreased standard errors, when compared to OLS (Butler & McNertney, 1991).

Generalised principal component regression (GPCR) regresses the dependent variable on the important principal components in the regression, and then splits the effect of these variables among the full original set of variables. This is a similar concept to removing variables from the regression, as it aims to remove from consideration those variables which cause the multicollinearity. However, the factor dropped is not a single variable, but rather a linear combination of variables. GPCR reduces the variances in the model, but does add some bias (Meloun et al., 2002).

A similar, yet distinct, method for potentially dealing with the multicollinearity problem is that of principal covariates regression (PCovR). PCovR aims to reduce some of the potential problems caused by decreasing some of the errors which are caused in PCR when the information about the variables is reduced. It does so by trying to ensure that not only are the predictor variables summarised, but that the components used to do so also perform well at predicting the criterion variables. It does so through the inclusion of a parameter α which is used to choose the focus point of importance for the regression in a weighted sum. If this parameter is set at 1 the entire focus is on X, hence giving us a principal components regression. However, if it is set at 0, the focus is entirely on Y, thus yielding a reduced rank regression (as discussed below). However, problems often accompany trying to discern the correct level of α to use, thus making it a difficult tool to use (Kiers & Smilde, 2007).

Another interesting, though not widely used method, is reduced rank regression or RRR. This focuses entirely on optimising the explained variance of Y and not the explained variance of X. The regression weights of RRR have the same decomposed form as those for PCR; however, obviously with different values (Kiers & Smilde, 2007).

Another biased regression technique which can be used to account for multicollinearity is partial least squares (Alin, 2010; Ciavolino & Al-Nasser, 2009; Diamantopoulos & Winklhofer, 2001; Grewal et al., 2004; Kiers & Smilde, 2007). Partial least squares aims to summarise the explanatory variables of the data set, and then performs the regression analysis upon these summaries instead of the explanatory variables themselves (Kiers & Smilde, 2007). It has been

shown that in the presence of multicollinearity the predictive ability of partial least squares does not suffer anywhere near as much as does its ability to accurately recover the values of regression variables. This result is similar for principal component regression, principal covariates regression, and reduced rank regression (Kiers & Smilde, 2007).

Ridge regression is a very common treatment for multicollinearity (Alauddin & Nghiem, 2010; Alin, 2010; Grewal et al., 2004; Kiers & Smilde, 2007; Meloun et al., 2002), as it makes full use of the data and does not require the addition or removal of explanatory variables for its use (Li et al., 2010). It was introduced by Hoerl & Kennard (1970). It is also a biased technique, with variances reduced in return for an introduction of some bias (Alin, 2010; Grewal et al., 2004; Mason & Brown, 1975). Ridge regression works through the addition of a small value k (which is a symmetric positive matrix) (Fourgeaud et al., 1984) to the correlation matrix of the variables in all of the diagonal elements (thus creating the ridge which gives the regression its name). Where a normal regression estimation is based upon a standard $X'X$ matrix, ridge regression uses an estimator defined as $b=[X'X+K]^{-1}X'Y$ (Hoerl & Kennard, 1970; Wan, 2002), where $K=\text{diag}(k_1,k_2,\ldots,k_p)$, where the k_i's are biasing parameters (Wan, 2002). It has been shown that defining $k_i = \sigma^2/\gamma_i^2$ (where γ_i is the ith element of γ) will minimise the mean squared error (Wan, 2002). This addition of the variable k allows ridge regression to have enough flexibility to reduce the inflated variances of OLS coefficients that arise from multicollinearity (Li et al., 2010), and thus increase the reliability of point estimates (Butler & McNertney, 1991).

There have been various criteria put forward for enabling the most appropriate choice of *k*, such as using a variance inflation factor, generalised cross-validation, and a ridge trace (Hoerl & Kennard, 1970; Li et al., 2010). Hoerl & Kennard (1970) acknowledged the difficulty in choosing an appropriate level of *k*, and provide some hints as to how to choose it. Their general recommendation is to find values of *k* at which the system stabilises, where coefficients do not have unreasonable values, coefficients which have incorrect signs when *k=0* (when *k=0* this regression is the same as OLS) now have the correct signs, and the residual sum of squares will not be unreasonably large. One potential method of choosing a ridge parameter is through selecting a value of *k* which when viewing the ridge trace[4] appears to cause the parameters to stabilize (Meloun et al., 2002). Another method for choosing the ridge parameter using the ridge trace is called generalised cross-validation, which is a minimiser equation involving the ridge trace. The advantage of this method is that it does not require an estimate of the variance (Golub et al., 1979). The introduction of a ridge parameter introduces bias into the model, but also improves efficiency (Grewal et al., 2004). It has been shown that the larger the value of *k* used, the smaller the variance will become (Hoerl & Kennard, 1970). However, this also causes more bias to arise, thus creating a situation where a balance must be found to choose the appropriate level of *k* (Li et al., 2010).

Ridge regression has been used in many areas, such as nuclear science, environmental science, biology and social science, with it being generally found

[4]A ridge trace is a plot of the values of the estimated ridge regression coefficients for different values of *k.* The value where the lines merge horizontally can tell us the most appropriate level of *k* (Schroeder et al., 1990).

that ridge regression may be useful in obtaining improved point estimates of model parameters (Butler & McNertney, 1991). However, one issue with its use is that due to the biasing of the estimator, it is no longer certain that the ratio of the parameter estimate to its standard error gives the appropriate *t*-statistic, thus making statistical inferences harder to perform (Butler & McNertney, 1991).

Other models which have been used to deal with multicollinearity (although rarely) include factor analysis regression (Lauridsen & Mur, 2006) and the combining of the collinear variables into a single index (Alauddin & Nghiem, 2010; O'Brien, 2007; Schroeder et al., 1990). The latter can be difficult to perform without compromising the validity of the model (Schroeder et al., 1990). The combination of the collinear variables into a single index differs from approaches such as PCR and GPCR, which also attempt to aggregate multicollinear variables in some manner, in that it involves combining only the collinear variables into an index. In contrast, PCR involves the aggregation of all explanatory variables into a set of combined variables, and GPCR removes a linear combination of variables from the model and then splits the aggregate effect between the original variable set (Kiers & Smilde, 2007; Meloun et al., 2002).

In regressions involving dummy variables, it is common to find multicollinearity between the dummy variable and other independent variables. One way to combat this is through measuring the independent variables by how different they are from the group means, and then weighting this by the value of the dummy variable (Buck & Hakim, 1981). This has, however, been shown to be unsuccessful (Kennedy, 1983).

Differencing of time series data to reduce the level of multicollinearity in the data set is also sometimes suggested as a solution to multicollinear data (Burt, 1987; Maccini, 1981). However, this differencing has been shown to be poorly conceived when it comes to removing multicollinearity, as differencing the data will also difference the disturbance term (Burt, 1987). It has even been put forward that differencing in this situation involves so much guesswork that it should certainly not be used (Cassidy, 1981).

2.5 Generalised Maximum Entropy regression to overcome multicollinearity

Generalised Maximum Entropy (GME) regression (Golan et al., 1996) was developed to provide robust solutions for ill-posed stochastic inverse problems. A problem is classed as ill-posed due to non-stationarity of data, or because the number of parameters is larger than the number of data points (Akdeniz et al., 2011). GME is based on the classic maximum entropy principle put forward by Jaynes (1957), which states that the probabilities associated with generation of the data that are the most uniform (most uncertain) should be chosen, while satisfying constraints specifying the data (Jaynes, 1957; Leon et al., 1999; Wu, 2009). This identifies the probabilities that could have been generated in the most number of ways consistent with the data (Jaynes, 1957). This is justified as the selection of a distribution with lower entropy would require extrapolation, while selection of a distribution with higher entropy would violate existing data (Golan, 2006).

The GME estimator is an expansion of the classical maximum entropy estimator, which has been reformulated for stochastic moments to allow for more flexibility and more stable estimates (Golan, 2006). This generalisation was created to allow

for the broad analysis of economic data (which is often ill-posed) and to allow the inclusion of prior information (Golan, 2006). Solution of GME problems requires reformulating each coefficient and error term as a discrete probability distribution (Jaynes, 1957). Coefficient β_k in GME is reformulated as:

$$\beta_k = \sum_{c=1}^{C} P_{k,c} z_{k,c} \ , \tag{2.8}$$

where $P_{k,c}$ are probabilities and $z_{k,c}$ for $c = [1,2,...,C]$ are fixed supports that are determined exogenously and defined over the closed interval $\left[z_{k,1}, z_{k,C}\right]$. Additionally, estimates of the disturbance e_t are computed through the equation:

$$e_t = \sum_{d=1}^{D} W_{t,d} v_{t,d} \ , \tag{2.9}$$

where $W_{t,d}$ are probabilities and $v_{t,d}$ for $d = \left[1,2,...,D\right]$ are fixed supports that are determined exogenously and defined over the closed interval $\left[v_{k,1}, v_{k,C}\right]$.

The probabilities $P_{k,c}$ and $W_{t,d}$ are decision variables that are determined through nonlinear optimisation with an objective to maximise entropy. As such, they must be non-negative and all values for a given coefficient or error term must add to one. The objective function is stated:

$$MaxJ = -\sum_{k=1}^{K}\sum_{c=1}^{C} P_{k,c} \ln(P_{k,c}) - \sum_{t=1}^{T}\sum_{d=1}^{D} W_{t,d} \ln(W_{t,d}) \, , \tag{2.10}$$

where *J* is the level of entropy. Although it is also sometimes multiplied by -1 to become a minimisation problem (see Section 3.4). This GME function is based on

the maximum entropy approach of Jaynes (1957), and Shannon's (1948) entropy information measure which is used to determine the probability distribution of undetermined problems. Through reflecting the level of uncertainty relating to a set of events, Jaynes (1957) proposed the maximisation of entropy (subject to fitting with the observed data) as a way of determining these unknown probabilities. GME builds on this by taking into account individual observations, thus generalising the maximum entropy problem (Golan et al., 1996; Golan & Perloff, 2002).

The structure of the GME estimator provides an inherent capacity to deal with ill-conditioned problems. It is also able to cope with small sample sizes, avoid specific assumptions regarding the distribution of errors, and incorporate prior information through appropriate support definition (Golan et al., 1996). Prior data can also be introduced in a structured way through the use of generalised cross entropy, which is closely related to Bayesian regression in that a regression will be weighted toward an existing set of priors (Golan et al., 1996). Moreover, the broad applicability of GME is assisted by an expanding theoretical literature generalising its application to different econometric formulations, including panel (Golan et al., 1996), discrete dependent variable (Golan et al., 1997), and weighted least squares (Wu, 2009) contexts.

Extensive numerical experiments have established the value of GME, relative to alternative approaches, in the context of ill-posed stochastic inverse problems. Golan et al. (1996) compared the GME estimator to OLS, RLS, and ridge regression. These authors found that GME had an improved ability to identify the true coefficients within a sampling experiment, relative to these other methods.

Moreover, GME predictions were robust to escalating degrees of ill-conditioning, whereas the other methods performed poorly as the simulated degree of multicollinearity increased. Campbell and Hill (2005) extended these findings to show the superiority of GME over OLS and RLS for small sample sizes and high error variances. Ciavolino & Al-Nasser (2009) also identified that GME is superior to partial least squares, in terms of minimising mean squared error, when data is contaminated with missing data, outliers, and multicollinearity. GME has also been shown to offer a reasonable alternative to maximum likelihood and Least Absolute Deviation procedures when attempting to recover estimates of parameters from censored and ordered multinomial data (Golan et al., 1997).

Nevertheless, a growing body of literature highlights the insufficiency of the GME estimator for general application. Golan et al. (1996) employed the wrong notion of a condition number in their simulations in which they show the GME estimator to outperform OLS, RLS, and ridge regression (Paris, 2001). Their condition number was the ratio of maximum to minimum eigenvalues, not the square root of this ratio which is what was defined in Belsley et al. (1980). Thus, while they purported to study condition numbers up to 100, they essentially only tested the equivalent of condition numbers up to 10 (Paris, 2001). Accordingly, none of their testing involved a harmful level of multicollinearity as defined by Belsley et al. (1980). Additionally, numerous applied studies that employ a GME estimator have demonstrated how sensitive parameter estimates are to the exogenous support values (e.g. Akdeniz et al., 2011; Golan et al., 1996; Leon et al., 1999; Paris & Caputo, 2001; Paris & Howitt, 1998). These support values specify the bounds of the estimates and must be provided prior to the estimation, for the coefficient and error terms specified in the entropy model. This is

problematic, as appropriate support values are difficult to identify. In an ideal world, we would choose these supports based on prior information or the economic theory underlying the issue at hand. However, economic theory often does not provide us with enough information to select the values appropriately (Campbell & Hill, 2005).

In line with this, several ways have been proposed that attempt to make bound creation more objective. The most common is using the "three-sigma rule" for establishing supports for the error terms. This rule defines the bounds through $v = \pm 3\hat{\sigma}_y$, where $\hat{\sigma}_y$ is the standard deviation of the dependent variable y in the sample (Akdeniz et al., 2011). This three-sigma rule ensures that at least 88.8% of unknown errors would fall within these bounds (Campbell & Hill, 2005), although it has also been found that a four-sigma rule could also be appropriate (Akdeniz et al., 2011), which would include at least 93.75% of unknown errors. Interestingly, using the standard deviation of the sample, instead of the sample deviation of the population (which is commonly unknown by the researcher), will often result in slightly larger bounds, as the sample deviation is commonly larger than the true deviation. Thus, the three-sigma rule based on the sample deviation will often correspond closely to the true value of the four-sigma value (Campbell & Hill, 2005). It is generally expected that wider bounds here would lead to a lower GME variance, however, with a subsequent increase in bias (Campbell & Hill, 2005). In contrast, no rules such as the three-sigma formalism exist for regressors. This is a major deficiency since often prior data is poor and regression results change greatly, depending upon the supports used (Caputo & Paris, 2008). Potentially due to these problems, the GME estimator is now seldom used in the Economics

community, although it has received some attention (Akdeniz et al., 2011; Howitt & Msangi, 2002, 2006; Leon et al., 1999; Paris & Howitt, 1998).

2.6 Leuven regression to overcome multicollinearity

Another set of estimators that have been put forward to attempt to deal with the problem of multicollinearity are the Leuven estimators. The first one of these was put forward by Paris (2001) and was based on the theory of light (Macedo et al., 2010). This estimator is generally referred to as the Maximum Entropy Leuven estimator (MEL)[5]. The GME estimator was introduced in an attempt to combat ill-posed problems, and has been claimed to also help deal with multicollinearity. However, it is susceptible to variability in predicted outcomes arising from the arbitrary selection of support bounds, allowing cases where different researchers with the same data could reach different results. The Leuven-1 estimator was the attempt by Paris (2001) to solve this problem and create a maximum entropy estimator which does not fall into the trap of requiring additional subjective information (Macedo et al., 2010; Mishra, 2004; Paris, 2001), while still appropriately handling multicollinear data (Macedo et al., 2010; Paris, 2001, 2004). The MEL estimators also use the information available in the sample more efficiently than OLS (Mishra, 2004).

When there is noise in an economic model, it is of course impossible to perfectly measure the parameters involved in sample data generation, as each parameter is

[5]I shall refer to it here as the Leuven-1 estimator, as it is called in Paris (2004), to remove any confusion by the introduction of the Leuven-2 estimator later in this review.

dependent upon all others in the model. This estimator follows the theory that it is possible to estimate the probability based upon the revealed image of the parameter (based on the theory of light). This revealed image of a parameter is the estimable dimensionality that depends on the sample information which is available (Paris, 2004). In the Leuven-1 estimator, the probability of a parameter is defined as the square of its normalised dimensionality or "amplitude". This probability can be shown as:

$$p_{\beta_k} = \beta_k^2 / \sum_k \beta_k^2 , \qquad (2.11)$$

where β is the vector of parameters to estimate, and where $p_{\beta_k} \geq 0$, $k = [1,\ldots,K]$, $t = [1,\ldots,T]$, and the amplitude of the parameter β_k is defined by $\beta_k/\sqrt{L_\beta}$. The solution of the Leuven-1 estimator requires nonlinear programming (Paris, 2004). This is a factor that the Leuven-1 estimator has in common with GME, however, both methods were found to solve in similar timeframes to OLS in Paris (2001). It was also found by this author that in situations where the data is well conditioned, the Leuven-1 estimator performs as well as the OLS estimator, and that when multicollinearity is considered, it performs better than the OLS estimator (Paris, 2001). The Leuven-1 estimator showed low levels of mean squared error loss (MSEL)[6] in the presence of more multicollinearity, whereas OLS experienced a dramatic increase in MSEL in the same circumstances. When GME was used on the same data, if narrow support intervals were used, the MSEL was comparable with the Leuven-1 estimator, whereas with larger support intervals it was less

[6]The mean squared error loss is a measure of empirical precision, and shows the ability of an estimator to reliably estimate true coefficients. A higher value of MSEL shows that a given estimator was less able to estimate these true parameters than an estimator with a lower value of MSEL.

effective, thus highlighting the potential advantage of the Leuven-1 estimator over standard GME (Paris, 2001). Despite the non-linearity and complexity of the parameters of the Leuven-1 model, Paris (2004) found its solution to be quick and efficient, relative to OLS. However, when small samples were used, the Leuven-1 estimator was not as efficient as OLS when the data was well-conditioned. However, in the presence of multicollinearity, the Leuven-1 estimator was far more stable than both OLS or the Leuven-2 estimator (Paris, 2004).

The Leuven-2 estimator is very similar to the Leuven-1 estimator, except that it introduces probabilities as part of the error term, as well. The probabilities for the error term are shown by:

$$p_{u_t} = u_t^2 / \sum_t u_t^2 \ , \qquad (2.12)$$

where u_t is the vector of the error terms to estimate.

The Leuven-2 estimator seems to be slightly less stable than the Leuven-1 estimator, although for high levels of multicollinearity it appears to converge to the same level of MSEL (Paris, 2001). Despite being less stable than the Leuven-1 estimator, the Leuven-2 estimator still uniformly outperforms the OLS estimator (Paris, 2004). Both the Leuven-1 and the Leuven-2 estimators are shown to have advantages over the standard GME estimator, with no subjective information having to be included into the analysis (Mishra, 2004; Paris, 2004). They are also shown to be consistent and asymptotically normal, and are scale invariant in the same way that OLS is (Paris, 2004). Nevertheless, the theory of light does not appear entirely convincing as an intuitive basis for an estimator, thus the theoretical justification of this approach has not yet been clearly established. This

is probably, at least partially, the reason that the Leuven-1 and Leuven-2 estimators have received little use.

2.7 Conclusions

There have been many efforts made to reduce the effects of multicollinearity in econometrics. However, it has been shown that there is typically no easy solution, and that generally to remove multicollinearity one needs to look very closely at the original problem (Stewart, 1987), and carefully manage factors which are known to affect it, such as measurement error, from the beginning (Grewal et al., 2004). However, this is difficult because much data in Economics comes from sources and measurements in which an analyst is not explicitly involved. Also, if data does suffer from multicollinearity, it may be inappropriate to disguise this fact, either through the simplification of the economic theory being applied to the data or the use of statistical methods to make it less obvious. Indeed, if multicollinearity is present, it may be best to retain it so that the reader can make their own assessment as to the legitimacy of any conclusions reached with that information (Feldstein, 1973).

Chapter 3

Methods

This section outlines the methodology used to assess the different estimators in the presence of differing degrees of multicollinearity. It describes the numerical simulation process, the estimators, and the metrics used for the comparison of these estimators, including an outline of the response surface analysis procedure adopted. Summation notation is used throughout this document, rather than matrix formulation, as this improves the clarity with which the estimators are described.

3.1 Numerical Simulations

The numerical simulations used here are based upon the work of Doole (2013) and Campbell & Hill (2005). They employ the use of a non-trivial Monte Carlo experiment to test the suitability of the OLS and RLS estimators, two forms of ridge regression (ridge-1 and ridge-2), GME (with wide and narrow support bounds), the Leuven-1 and Leuven-2 estimators, as well as the Two-Step DDE estimator (DDE2), and the One-Step DDE estimator (DDE1).

These estimators are compared in their ability to estimate a given linear model. The linear model is specified as:

$$y_t = \sum_{k=1}^{5} \beta_k X_{k,t} + e_t \;\; \forall t \,, \qquad (3.1)$$

where y_t is the dependent variable, β_k for $k = 1,2,...,K$ are unknown coefficients to be estimated, $X_{k,t}$ are data for each parameter $k = 1,2,...,K$ over $t = 1,2,...,N$ observations, and e_t are a series of unobserved disturbances. The true value of the coefficients used here are: $\beta_1 = -4$, $\beta_2 = 2$, $\beta_3 = 1$, $\beta_4 = -3$, and $\beta_5 = 2$.

Coefficients $\beta_2 - \beta_5$ are the same as defined by Golan et al. (1996), Campbell & Hill (2005), and Doole (2013). An intercept (β_1) has been added to create more consistency with applied econometric problems.

Each simulation run is defined in the interval $s = [1,2,...,100]$.

Belsley (1991) specifies the condition number (κ) as a measure of the degree of multicollinearity. κ is the ratio of the largest and smallest singular values of data matrix *X* once columns have been scaled to a unit length, and is given by:

$$k(A) = \frac{\mu_{max}}{\mu_{min}}, \qquad (3.2)$$

where *k(A)* is the condition number of the matrix *A*, μ_{max} is ||*A*||, and the inversion of matrix *A* yields $1/\mu_{min}$. $\kappa = 1$ is consistent with no multicollinearity, while $\kappa > 30$ indicates that multicollinearity exists to a point where we would expect it to degrade inference. Here, 10 levels of multicollinearity have been simulated through the generation of data matrices consistent with a given condition number, however, for the sake of parsimony, as well as adding focus to the results, five levels are presented in the results (Chapter 4) and five are presented in Appendix 2. The 10 levels are: $\kappa = \{1,10,20,40,60,80,100,250,500,1000\}$. These scenarios include slight multicollinearity, but also a broad range of cases in which very severe multicollinearity could be expected to degrade estimation.

The regressors used were randomly selected from a $N(0,1)$ distribution, a $t_{(3)}$ distribution, and a $\chi^2_{(5)}$ distribution. The $t_{(3)}$ and $\chi^2_{(5)}$ distributions were corrected to incorporate a mean of 0 and variance of unity. The $t_{(3)}$ distribution is selected

because it has thicker tails than a $N(0,1)$ distribution. The $\chi^2_{(5)}$ distribution is selected since it is asymmetric, thus enabling us to test the ability of the estimators under a range of data distributions.

Errors were randomly selected from a $N(0,1)$ distribution, a mixed Gaussian distribution, a $t_{(3)}$ distribution, and a $\chi^2_{(5)}$ distribution. The mixed Gaussian distribution is central to corrupted sampling theory (e.g. Kreider & Pepper, 2008) and is computed through $e_t = 0.9N(0,1) + 0.1N(0,10)$. The $t_{(3)}$ and $\chi^2_{(5)}$ distributions are again corrected to incorporate a mean of 0 and variance of unity. The $t_{(3)}$ distribution is selected because it has thicker tails than a $N(0,1)$ distribution. The $\chi^2_{(5)}$ distribution is selected since it is asymmetric. The above conditions allow the robustness of the estimators to be tested in a broad range of situations.

Each observation of the dependent variable for each simulation is then generated through:

$$y_t = \sum_{k=1}^{5} \beta_k X_{k,t} + e_t \ \forall t . \qquad (3.3)$$

Each observation of the dependent variable in the experiment is present within the interval $y_t = [-73.26, 78.94]$.

3.2 OLS & RLS estimators

The first two estimators tested were Ordinary Least Squares (OLS) and Restricted Least Squares (RLS). OLS and RLS are very similar, except for in one particular aspect, as can be seen below. Both OLS and RLS estimate the unknown parameters in a regression through the minimisation of the squared deviation of the error. However, RLS provides a lower and upper bound on its coefficients, in this case -5 and 5, whereas OLS has no such restriction.

The objective function of both OLS and RLS problems follows:

$$\min O = \sum_{t=1}^{T} e_{ols}{}^{2} \,. \tag{3.4}$$

They also share the same data constraint:

$$y_t = \sum_{k=1}^{K} \beta_k X_{k,t} + e_{ols} \,, \tag{3.5}$$

where β_k is the estimated coefficient for OLS and e_{ols} is the estimated errors for OLS.

RLS employs the same objective function and data constraint as OLS, however it includes the added constraint of $\beta_k = [-5,5]$. The use of these bounds provides better predictive capacity in the presence of multicollinearity, one of the effects of which is the amplification of parameter estimates, giving RLS an advantage over OLS. However, the inclusion of useful constraints requires good prior information to provide meaningful lower and upper bounds.

3.3 Ridge-1 & ridge-2 estimators

The two ridge estimators examined here are very similar, the only difference being in the calculation of the ridge parameter.

The objective function for both ridge estimators is given by:

$$\min R = \sum_{t=1}^{T} v_r^{\,2} + \lambda \sum_{k=1}^{K} \beta_k^{\,2}\,, \tag{3.6}$$

where R is the MSE of the ridge estimator, v_r is the error estimate for the ridge regression, and λ is the ridge parameter. The ridge parameter penalises large values of the coefficients, thus providing some protection against a key effect of multicollinearity (Cule & De Iorio, 2013).

The data constraint for the ridge regressions is given by:

$$y_t = \sum_{k=1}^{K} (\beta_k X_{k,t}) + \mathrm{v}_r\,. \tag{3.7}$$

The first ridge regression model examined here (ridge-1) involves the ridge parameter being selected according to Mittelhammer et al. (2000), where $K = 6$ is the number of variables, and σ_y is the variance of y. The rule is specified:

$$\lambda = \frac{(K-1)\sigma_y}{\sum_{k=1}^{5} \left(B_k\right)^2}\,. \tag{3.8}$$

For the second ridge regression model (ridge-2), the ridge parameter is calculated using the *f* statistic. The *f* statistic in this study is calculated using the explained sum of squares and the residual sum of squares for the ridge regression.

The explained sum of squares is computed through:

$$ess = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 , \tag{3.9}$$

where y_t is the estimate for the dependent variable and $\hat{y}_t$ is the mean of the dependent variable.

The residual sum of squares is calculated through:

$$rss = \sum_{t=1}^{T} e_r{}^2 , \tag{3.10}$$

where e_r is the error estimate.

The f statistic is then calculated using the explained sum of squares and the residual sum of squares. The generalised formula is:

$$fstat = \frac{(\frac{ess}{g-1})}{(\frac{rss}{n-g})}, \tag{3.11}$$

where n is the number of sample observations and g is the number of estimated parameters (Johnston, 1984). The specific equation for this instance is:

$$fstat = \frac{(\frac{ess}{4})}{(\frac{rss}{5})}. \tag{3.12}$$

The ridge parameter for ridge-2 tested here is then:

$$\lambda = \frac{1}{fstat} \ . \tag{3.13}$$

3.4 GME estimator

To use the GME estimator, the unknown coefficients β_k for $k = 1, 2, ..., K$ and the error terms must be transcribed to a form involving probabilities. Coefficient β_k is estimated through:

$$\beta_k = \sum_{c=1}^{C} P_{k,c} z_{k,c} \ \forall k \,, \tag{3.14}$$

where probabilities $P_{k,c}$ are decision variables computed through nonlinear optimisation, fixed supports $z_{k,c}$ are determined exogenously, and $c = [1, 2, ..., C]$ is the index of support points. Probabilities $P_{k,c}$ are subject to the constraints $P_{k,c} \in [0,1]$ and $\sum_{c=1}^{C} P_{k,c} = 1$.

Estimates of the disturbance e_t are computed through:

$$e_t = \sum_{d=1}^{D} W_{t,d} v_{t,d} \ \forall t \,, \tag{3.15}$$

where probabilities $W_{t,d}$ are decision variables computed through nonlinear optimisation, $v_{t,d}$ are fixed supports determined exogenously, and $d = [1, 2, ..., D]$ is the index of support points. Probabilities $W_{t,d}$ are subject to the constraints

$W_{t,d} \in [0,1]$ and $\sum_{d=1}^{D} W_{t,d} = 1$, as a set of probabilities must always sum to 1. Five support points have been used here for both error and coefficient terms, in accordance with other studies (Akdeniz et al., 2011; Golan et al., 1996).

The substitution of equations (3.14) and (3.15) into the general stochastic linear inverse problem defined in equation (3.1) yields the GME data equations as:

$$y_t = \sum_{k=1}^{K} X_{k,t} \sum_{c=1}^{C} P_{k,c} z_{k,c} + \sum_{d=1}^{D} W_{t,d} v_{t,d} \quad \forall t . \tag{3.16}$$

The objective function used in GME regression is the minimisation of the entropy criterion:

$$MinJ = \sum_{k=1}^{K} \sum_{c=1}^{C} P_{k,c} \ln(P_{k,c}) - \sum_{t=1}^{T} \sum_{d=1}^{D} W_{t,d} \ln(W_{t,d}) . \tag{3.17}$$

In applied problems, $\upsilon = 10^{-8}$ is added to the term within each natural logarithm to ensure $ln(0) \approx 0$.

Maximisation of entropy conservatively forces the probability terms for each coefficient/error to be the most uniform (i.e. most uncertain), while satisfying constraints specifying the data. This identifies the probabilities that could have been generated in the most number of ways consistent with the data (Jaynes, 1957). This is justified, as the selection of a distribution with lower entropy would require extrapolation, while selection of a distribution with higher entropy would violate existing data (Golan, 2006) In this thesis the standard GME equation is multiplied by -1 to give a minimisation problem.

Two GME regressions were performed in this analysis, the first with relatively narrow support bounds of $z_{k,c} = \{-5,5\}$ (GMEN) for regressors, and the second with wider support bounds of $z_{k,c} = \{-20,20\}$ (GMEW) for regressors. The narrow support bounds of GMEN are set in a manner which is consistent with knowing broadly the correct range of estimates. It is also in line with the bounds used in the RLS example above and those defined by Golan et al. (1996) in their numerical experiments. However, GME is susceptible to variability based on its support bounds, this being one of the prime disadvantages of its use (Paris, 2004). To reflect this, a second GME regression with broad support bounds (GMEW) was conducted with these support bounds set to be equal to the wide support bounds used in Paris (2004).

The error supports are estimated through the symmetric n-sigma rule, thus giving the symmetric error supports of $v_{t,d} = \left\{-3\sqrt{\sigma_y}, -1.5\sqrt{\sigma_y}, 0, 1.5\sqrt{\sigma_y}, 3\sqrt{\sigma_y}\right\}$ (see Section 2.5).

3.5 Leuven-1 & Leuven-2 estimators

The Leuven-1 and Leuven-2 estimators are very similar to each other; they are formalised in much the same way, and use the same entropy system. The difference is that the Leuven-2 estimator extends the entropy to the error term, whereas the first only defines it for the regressors (Paris, 2004).

The Leuven-1 estimator involves minimising:

$$\min J = \sum_{k=1}^{K} P_{\beta_k} \ln(P_{\beta_k}) + \eta_f \ln(\eta_f) + \sum_{t=1}^{T} e_f{}^2 , \tag{3.18}$$

where P_{β_k} is the probability of coefficient β_k, η_f is a variable used as a base variable for deriving amplitude, and e_f is the error.

This problem is solved subject to the following constraints. First, the data is defined through:

$$y_t = \sum_{k=1}^{K} X_{k,t}\beta_k + e_f \quad \forall t , \tag{3.19}$$

where β_k is the coefficient matrix for the Leuven estimators.

The Leuven variable is used as a basis for generating the probabilities for β_k:

$$\eta_f = \sum_{k=1}^{K} \beta_k{}^2 . \tag{3.20}$$

In the Leuven-1 estimator, the amplitude of parameter β_k is given by:

$$\psi_{\beta_k} = \frac{\beta_k}{\sqrt{\eta_f}} . \tag{3.21}$$

As such, the probability of parameter β_k is given by the square of its amplitude:

$$P_{\beta_k} = \frac{\beta_k{}^2}{\eta_f} . \tag{3.22}$$

The Leuven-1 estimator does not require any subjective prior information, using the statistical linear model and the theory of light to define amplitudes for parameters.

The Leuven-2 estimator is very similar to the Leuven-1 estimator, with only the probability specification of the error term changing. Again, it does not require any subjective exogenous information, as the GME estimator does (Paris, 2004).

The objective function for the Leuven-2 estimator is:

$$\min J = \sum_{k=1}^{K} P_{\beta_k} \ln(P_{\beta_k}) + \eta_f \ln(\eta_f) + \sum_{t=1}^{T} W_d \ln(\mathrm{W}_d) + \vartheta_f \ln(\vartheta_f), \quad (3.23)$$

where W_d is the probabilities for the errors, and ϑ_f is the Leuven variable to help determine the amplitude for the errors.

The Leuven-2 estimator uses the same data constraints as the Leuven-1 estimator, with two additional equations. The first equation states:

$$\vartheta_f = \sum_{t=1}^{T} e_f^{\,2}, \quad (3.24)$$

where e_f is the error term. This equation describes that the base parameter which is used for establishing the probability of the error term is equal to the squared sum of the vector of error terms. This is equivalent to the way that the probabilities of the coefficients are established in the Leuven-1 estimator.

The second equation states:

$$W_d = \frac{e_f^{\,2}}{\vartheta_f}. \tag{3.25}$$

This equation describes the probabilities of the error terms. For both Leuven estimators, probabilities P_{β_k} are subject to the constraints $P_{\beta_k} \in [0,1]$ and $\sum_{k=1}^{K} P_{\beta_k} = 1$. Moreover, for the Leuven-2 estimator, the probabilities W_d are subject to the constraints $W_d \in [0,1]$ and $\sum_{d=1}^{D} W_d = 1$.

The Leuven estimators are shown by Paris (2004) to be consistent and asymptotically normal.

3.6 DDE2 estimator

The DDE2 estimator is a two-step process introduced in Doole (2013). It firstly uses bounded set regression based on interval-valued linear programming and the Least Absolute Deviations (LAD) estimator to endogenously determine support bounds, for both the coefficient and error terms, which are then used in the standard GME model. This provides an empirically-based technique for providing support bounds for GME, as these are endogenously created rather than being selected according to the bias of the researcher.

The LAD estimator is used due to its simplicity, as well as the fact that it can be transcribed for use in interval-valued situations. The linear programming analogue of the LAD estimator is:

$$\min J = \sum_{t=1}^{N} (D_t + E_t), \tag{3.26}$$

subject to:

$$y_t = \sum_{k=1}^{K} \beta_k X_{k,t} - D_t + E_t \text{ for rows } t = [1, 2, ..., N], \tag{3.27}$$

where D_t and E_t are negative and positive deviations, respectively, between the observed data and the estimated values for observation *t*. It should also be noted that $D_t, E_t \geq 0$.

Interval-valued estimation is when the coefficients and errors are defined as closed intervals instead of single values to identify the endogenous support bounds. A closed interval $f = \left[f^L, f^U \right]$ is a set of real numbers including its endpoints. The endpoints consist of a lower bound (f^L) and an upper bound (f^U). This closed interval is used to approximate a structure where the experimenter does not know the shape of a given distribution, so a uniform distribution is used (as the uniform probability distribution is the distribution with the lowest information content). This concept is aligned with the maximum entropy principle, which uses the options implying the least information in the presence of high uncertainty.

Interval-valued estimation is hard to solve due to its inclusion of interval-valued expressions in the equations. This problem can be solved through transcription as in Wu (2008); this transcription can be seen below. A standard interval-valued linear program (i.e. interval-valued estimation) can be defined as:

$$\min J^{IV} = \sum_{p=1}^{P} [c_p^L, c_p^U] x_p \,, \tag{3.28}$$

subject to:

$$\sum_{p=1}^{P} [a_{pq}^L, a_{pq}^U] x_p \geq [b_q^L, b_q^U] \text{ for rows } q = [1, 2, ..., Q], \tag{3.29}$$

where J^{IV} is an interval-valued index of performance, there are p decision variables and q constraints, and $[c_p^L, c_p^U]$ consists of c_p^L which is the lower bound of the cost accruing to decision variable x_p and c_p^U is the upper bound. In the same way, in the set $[a_{pq}^L, a_{pq}^U]$, a_{pq}^L is the lower bound on the matrix of coefficients (a_{pq}) and a_{pq}^U is the upper bound, and set $[b_q^L, b_q^U]$ provides the lower (b_q^L) and upper (b_q^U) bounds of possible values for the lower bound placed upon the constraint q.

As previously mentioned, the interval-valued expressions in these equations make solution difficult as the objective function is not single-valued. However, they can (following Wu (2008), Proposition 5.1) be transcribed to a problem that can be solved using standard mathematical programming. This transcription is defined:

$$\min J = \sum_{p=1}^{P} c_p^L x_p + \sum_{p=1}^{P} c_p^U x_p \,, \tag{3.30}$$

subject to:

$$\sum_{p=1}^{P} a_{pq}^{L} x_p \geq b_q^{L} \text{ for rows } q = [1, 2, ..., Q], \text{ and} \tag{3.31}$$

$$\sum_{p=1}^{P} a_{pq}^{U} x_p \geq b_q^{U} \text{ for rows } q = [1, 2, ..., Q]. \tag{3.32}$$

This transcription removes the interval-valued expressions, as well as turning constraint (3.29) into a constraint for each bound ((3.31) and (3.32)). By placing the lower and upper bounds in the objective function, (3.30) creates a non-dominated solution (meaning that neither bound can be improved without worsening the other). This concept has been taken from multi-objective programming (Doole, 2013) to deal with the solutions not being single-valued.

Robust estimation using the LAD estimator, as is used here, involves the minimisation of the sum of the absolute value of the deviations. When the standard linear programming analogue of the LAD estimator is used, in conjunction with our interval-valued linear model, we get an equation which is difficult to solve due to the interval-valued expressions in the equations discussed above. The LAD estimator for the interval-valued linear statistical model can be stated in linear programming form as:

$$\min J = \sum_{t=1}^{N} [d_t^{L}, d_t^{U}], \tag{3.33}$$

subject to:

$$y_t = \sum_{k=1}^{K} [\beta_k^{L}, \beta_k^{U}] X_{k,t} + [-d_t^{L}, d_t^{U}] \text{ for } t = 1, 2, ..., N, \tag{3.34}$$

where $[\beta_t^L, \beta_t^U]$ are the interval-valued coefficients to be estimated, and where d_t^L is the negative deviation for the lower regression line and d_t^U is the positive deviation for the upper regression line.

This can be reformulated using the transcription in Wu (2008) to the below objective function (3.35), subject to the two data constraints reported below it (equations (3.36) and (3.37)). This system can be solved using linear programming as a bounded set regression and minimises the absolute deviations.

The objective equation for the bounded set regression in the first step of DDE2 is to minimise the absolute deviation through:

$$\min J = \sum_{t=1}^{T} \left(d_t^L + d_t^U \right), \tag{3.35}$$

This objective function is subject to:

$$y_t = \sum_{k=1}^{K} \beta_k^L X_{k,t} - d_t^L \quad \forall t \text{, and} \tag{3.36}$$

$$y_t = \sum_{k=1}^{K} \beta_k^U X_{k,t} + d_t^U \quad \forall t . \tag{3.37}$$

The minimisation of equation (3.35), subject to equations (3.36) and (3.37), is termed a bounded set regression.

A bounded set regression produces a set of lower and upper bounds for the coefficient terms (β_k^L and β_k^U, respectively) as well as sets of lower and upper bounds for the deviation terms (d_t^L and d_t^U, respectively). These bounds are then used directly as support bounds for the coefficient and error terms in the GME

regression discussed previously (Section 3.4). This gives endogenously established support bounds for the GME estimator, instead of its usual exogenously determined bounds. This is implemented through the rules $\left[z_{k,1}, z_{k,C}\right] = \left[\beta_k^L, \beta_k^U\right]$ and $\left[v_{t,1}, v_{t,D}\right] = \left[-d_t^L, d_t^U\right]$.

It should also be noted that the lower bounds on the deviation cannot be negative ($d_t^L \geq 0$) and that the upper bounds must also not be negative ($d_t^U \geq 0$).

After these bounds have all been estimated, they are used as the support bounds for the GME procedure (Section 3.4).

3.7 DDE1 estimator

Both the DDE2 and the DDE1 methods involve the use of the interval-valued non-linear programming method of Wu (2008) to solve a LAD estimator. When the equations for this "bounded set regression" are used to give informative bounds before the use of the GME equation, we get the DDE2 estimator. However when these equations are integrated into the GME estimator itself, we get the one-step model, called DDE1 throughout this thesis.

The objective function of the DDE1 estimator is:

$$\begin{aligned} \min J = & \sum_{k=1}^{K}\sum_{c=1}^{C} P_{k,c}\ln(P_{k,c}) + \sum_{t=1}^{T}\sum_{d=1}^{D} W_{t,d}\ln(W_{t,d}) \\ & + \sum_{t=1}^{T} f_t^L \ln(f_t^L) + \sum_{t=1}^{T} f_t^U \ln(f_t^U) + \\ & \sum_{k=1}^{K} \tilde{\beta}_k^L \ln(\tilde{\beta}_k^L) + \sum_{k=1}^{K} \tilde{\beta}_k^U \ln(\tilde{\beta}_k^U) \end{aligned} \quad . \qquad (3.38)$$

Here, the probabilities $P_{k,c}$ and $W_{t,d}$ are decision variables computed through nonlinear optimisation, f_t^L is the lower bound and f_t^U is the upper bound of the deviation between the estimated regression line and the actual value of the dependent variable (i.e. f_t^L and f_t^U are the lower and upper bounds of the error term respectively), and $\tilde{\beta}_k^L$ and $\tilde{\beta}_k^U$ are the absolute value approximations of the lower (β_t^L) and upper (β_t^U) bounds of the interval-valued coefficients estimated in both DDE estimators.

The DDE1 objective function is made up of three parts. The first is the maximum entropy equation. This has been multiplied by -1; hence, it poses a minimisation, rather than a maximisation, problem. The second part minimises the entropy of the deviations of the error supports, and the third minimises the entropy of the absolute values of the bounds defined for the coefficients in the bounded set regression. Absolute values are used for the coefficients to ensure the natural logarithms in the entropy criterion are defined. The objective function in equation (3.38) is minimised, subject to the constraints set out in equations (3.39) – (3.51) below.

The constraint in equation (3.39) is an analogue of the data constraint specified in the standard GME model:

$$y_t = \sum_{k=1}^{K} X_{k,t} \sum_{c=1}^{C} P_{k,c} z_{k,c} + \sum_{d=1}^{D} W_{t,d} \quad \forall t \,. \tag{3.39}$$

The following data constraint is designed to estimate the lower support bounds for the coefficients and deviations in a data driven manner:

$$y_t = \sum_{k=1}^{K} \beta_k^L X_{k,t} - f_t^L \ \forall t . \tag{3.40}$$

The following data constraint is designed to estimate the upper support bounds for the coefficients and deviations in a data driven manner:

$$y_t = \sum_{k=1}^{K} \beta_k^U X_{k,t} + f_t^U \ \forall t . \tag{3.41}$$

Equation (3.42) sets the variable $z_{k,1}$ equal to the lower bound, while equation (3.43) sets the variable $z_{k,C}$ equal to the upper bound. Equation (3.44) generates the support values inside the interval $\left[z_{k,1}, z_{k,C}\right]$ if $C > 2$. Also note that the coefficient supports here are dependent on the coefficient bounds estimated earlier in equations (3.40) and (3.41).

$$z_{k,1} = \beta_k^L , \tag{3.42}$$

$$z_{k,C} = \beta_k^U , \tag{3.43}$$

$$z_{k,c} = z_{k,1} + \frac{(c-1)\left(\beta_k^U - \beta_k^L\right)}{C-1} \ \forall c > 1, c < C . \tag{3.44}$$

The error supports are generated through equations (3.45) – (3.47). Note that the error supports are dependent on the maximum deviation calculated previously in equations (3.40) and (3.41).

$$v_{t,1} = -f_t^L , \tag{3.45}$$

$$v_{t,D} = f_t^U \,. \tag{3.46}$$

The below equation generates values for the supports in the interval $\left[v_{t,1}, v_{t,D}\right]$ if $D > 2$:

$$v_{t,d} = v_{t,1} + \frac{(d-1)\left(f_t^L + f_t^U\right)}{D-1} \;\forall d > 1, d < D \,. \tag{3.47}$$

The two equations below, (3.48) and (3.49), are used to compute the absolute values of their respective β coefficients. These will both be greater than zero by definition, and are calculated to ensure positive numbers enter the entropy criterion in the objective function:

$$\tilde{\beta}_k^L = \sqrt{\left(\beta_k^L\right)^2 + \omega^2} \;\forall k \,, \tag{3.48}$$

$$\tilde{\beta}_k^U = \sqrt{\left(\beta_k^U\right)^2 + \omega^2} \;\forall k \,. \tag{3.49}$$

The below equations, (3.50) and (3.51), state that the sum of all of the probabilities used in the estimation of the coefficient are equal to 1, as are those in the disturbance term. This is necessary as a set of probabilities must sum to 1.

$$\sum_{c=1}^{C} P_{k,c} = 1 \;\forall k \,, \tag{3.50}$$

$$\sum_{d=1}^{D} W_{t,d} = 1 \;\forall t \,. \tag{3.51}$$

3.8 Estimator Comparison

Estimators are compared using two metrics. They are compared according to their prediction risk, and how precise their estimates are empirically (Mittelhammer et al., 2000). In the below equations, those coefficient estimates created over the course of the Monte Carlo simulations are designated $b_{k,s}$.

Empirical prediction risk is calculated using the mean squared error term (MSE) and is the ability of the estimator to reliably estimate a dependent variable based on the observed data. This is computed through:

$$MSE = \frac{\sum_{s=1}^{S}\left(\frac{\sum_{t=1}^{N}\left(\hat{e}_{t,s}\right)^2}{N-K}\right)}{S}, \tag{3.52}$$

where $\hat{e}_{t,s}$ is the estimated error for observation *t* in simulation *s*. A higher value of MSE tells us that a regression was, in general, less able to produce a good prediction of the dependent variable, when compared to a regression which had a lower value of MSE.

The level of empirical precision shown by an estimator is the ability of that estimator to reliably estimate true coefficients. This is measured using the mean squared error loss (MSEL):

$$MSEL = \frac{\sum_{k=1}^{K}\sum_{s=1}^{S}\left(b_{k,s}-\beta_k\right)^2}{K \cdot S}. \tag{3.53}$$

A higher MSEL tells us that a regression was less able to estimate the true parameters in this experiment, compared to a regression with a lower MSEL.

3.9 Response surface analysis

Response surface analysis has been shown to be a reasonable way of summarising the results of a set of Monte Carlo experiments (Campbell & Hill, 2005; Davidson & MacKinnon, 1993). This involves two regressions in this study. One regression between MSE and the characteristics of each individual Monte Carlo simulation run for each estimator, and one regression between MSEL and the characteristics of each individual Monte Carlo simulation run for each estimator.

In this study, we estimate the following response surface regression for the MSE computed for each estimator. This aims to identify how MSE changes under each set of circumstances:

$$MSE = \alpha_1 + \alpha_2 CN + \alpha_3 TR + \alpha_4 CHR + \alpha_5 MGE + \alpha_6 TE + \alpha_7 CHE , \qquad (3.54)$$

where α are coefficients, CN is the condition number, TR is equal to 1 if the $t_{(3)}$ data distribution was used and 0 if not, CHR is equal to 1 if the $\chi^2_{(5)}$ data distribution was used and 0 if not, MGE is equal to 1 if mixed Gaussian errors were used and 0 if not, TE is equal to 1 if $t_{(3)}$ distributed errors were used and 0 if not, and CHE is equal to 1 if $\chi^2_{(5)}$ were used and 0 if not. This regression was run for each estimator to enable us to see the effect that each of these variables has on the mean squared error (MSE) for each method.

Also, in this study, we estimate the following response surface regression for the MSEL computed for each estimator. This aims to identify how MSEL changes under each set of circumstances:

$$MSEL = \beta_1 + \beta_2 CN + \beta_3 TR + \beta_4 CHR + \beta_5 MGE + \beta_6 TE + \beta_7 CHE\,, \qquad (3.55)$$

where β are coefficients.

3.10 Summary

This Chapter has outlined the methodology used to assess the different estimators tested here, across varying degrees of multicollinearity. It described the numerical simulation process used, the structure of the estimators tested, and the metrics used for comparing the estimators, as well as giving an outline of the response surface analysis procedure to be used.

Chapter 4

Results and Discussion

4.1 Comparison of mean squared error for each estimator

Empirical prediction risk is calculated using the mean squared error term (MSE). It is the ability of the estimator to reliably estimate a dependent variable based on the observed data. A higher value of MSE tells us that a regression was in general less able to produce a good prediction of the dependent variable, when compared to a regression that had a lower value of MSE.

10 different condition numbers are tested to ensure robust results, however for the sake of parsimony, as well as to add focus to the results, only the five condition numbers that provide significant insight are discussed here. The five condition numbers discussed are 1, 40, 60, 100, and 500. To see the results for condition numbers 10, 20, 80, 250, and 1000 and the associated discussion, see Appendix 2.

4.1.1 Mean Squared Error for all models when condition number is 1

Table 4.1 presents MSE for all estimators when the condition number is 1.When looking at the estimators in this situation, i.e. no multicollinearity, we find that all of the estimators act relatively similarly. They are all at their highest points when the errors are drawn from the mixed Gaussian distribution, regardless of the data distribution that was drawn upon. This demonstrates the difficulty of estimating relationships when error distributions are contaminated with data from other processes.

Table 4.1: Mean squared error for estimators with a condition number of 1.

Mean squared error	Condition number: 1									
	Model									
Dist.[1]	OLS	Ridge-1	Ridge-2	RLS	GMEN	GMEW	DDE2	DDE1	Leuven-1	Leuven-2
e1d1	1.01	1.13	1.01	1.01	1.8	1.01	1.17	1.15	1.27	1.03
e1d2	1.01	1.15	1.01	1.01	1.83	1.02	1.17	1.16	1.3	1.03
e1d3	1	1.12	1	1	1.78	1.01	1.13	1.13	1.27	1.02
e2d1	1.8	1.93	1.8	1.8	2.65	1.8	2.01	2.02	2.06	1.81
e2d2	1.8	1.94	1.8	1.8	2.67	1.8	2.11	2.07	2.09	1.81
e2d3	1.8	1.93	1.8	1.8	2.62	1.81	2.04	2.07	2.07	1.81
e3d1	1.06	1.17	1.06	1.06	1.9	1.1	1.4	1.57	1.32	1.07
e3d2	1.04	1.18	1.04	1.04	1.88	1.06	1.47	1.62	1.33	1.06
e3d3	1.02	1.14	1.02	1.02	1.8	1.03	1.32	1.57	1.28	1.03
e4d1	1.07	1.19	1.07	1.07	1.85	1.07	1.26	1.59	1.32	1.08
e4d2	1.07	1.2	1.07	1.07	1.87	1.07	1.33	1.76	1.35	1.08
e4d3	1.07	1.19	1.07	1.07	1.84	1.08	1.29	1.68	1.33	1.09

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The GMEN estimator produces the highest values of MSE for all error and data distributions, with values ranging between 1.8 and 2.67 (Table 4.1). This could be because it has been shown previously that GME with relatively narrow support bounds will perform poorly when it comes to MSE levels, if the centre of the support bounds are not the true β values (Mittelhammer et al., 2013). This follows from the structure of the entropy formalism that most strongly pulls towards the most-conservative distribution (the uniform distribution). The mean of this distribution is the midpoint of the range defined by the supports; thus, if the centre of the bounds are not the true coefficient values, then the estimates will not be accurate.

The GMEN estimator is generally followed by either the DDE1, or the Leuven-1 estimator, with the DDE1 estimator not handling the $t_{(3)}$ or $\chi^2_{(5)}$ error distributions as well as the Leuven-1 estimator (Table 4.1). These estimators are followed by the DDE2 estimator.

OLS, ridge-1, RLS, GMEW, and Leuven-2 all produce similar results, however with a condition number of 1, the OLS, ridge-2, and RLS estimators all produce exactly the same MSE, regardless of the error or data distributions used (Table 4.1). This is to be expected in a case of no multicollinearity, with OLS and RLS using essentially the same equations and there not being enough multicollinearity to degrade the data and reduce the effectiveness of OLS. Also, ridge-2 is equivalent as the penalty associated with multicollinearity is not active, given the low condition number.

These results tell us that in cases with no multicollinearity, OLS, ridge-2, and RLS are the three estimators that would be expected to most accurately predict the dependent variable based on the observed data, whereas the other estimators are expected to be relatively less able to make this accurate prediction. This is in line with expectations, given that OLS, RLS, and ridge estimators contain a sole or strong focus upon minimising MSE in their objective functions, particularly relative to the entropy-based estimators. Nevertheless, the levels of MSE identified are very low for all instances, with no estimator clearly unsuitable in the presence of no multicollinearity.

4.1.2 Mean Squared Error for all models when condition number is 40

Table 4.2 presents MSE for all estimators when the condition number is 40. At this level, we should expect to see some level of degradation due to multicollinearity. As the multicollinearity level increases, it becomes more obvious that the MSE for OLS is going to be lower than the MSE of the alternative estimators. This is to be expected as the objective function of the OLS estimator aims primarily to minimise the empirical prediction risk.

Table 4.2: Mean squared error for estimators with a condition number of 40.

Mean squared error	**Condition number: 40**									
	Model									
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	1.01	1.06	1.01	1.01	1.41	1.04	1.1	1.13	1.27	2.17
e1d2	1	1.06	1.01	1.01	1.37	1.03	1.13	1.16	1.28	1.05
e1d3	1.01	1.07	1.01	1.02	1.42	1.03	1.11	1.18	1.28	1.06
e2d1	1.8	1.86	1.81	1.81	2.24	1.84	2	2.08	2.07	1.86
e2d2	1.78	1.84	1.8	1.8	2.18	1.82	2.05	2.13	2.07	1.84
e2d3	1.79	1.86	1.8	1.81	2.25	1.83	1.98	2.06	2.07	3.02
e3d1	1.06	1.11	1.06	1.06	1.55	1.16	1.28	1.59	1.32	2.23
e3d2	1.06	1.11	1.07	1.07	1.56	1.21	1.42	1.72	1.33	1.11
e3d3	1.07	1.13	1.07	1.07	1.96	1.3	1.34	1.72	1.33	1.11
e4d1	1.06	1.12	1.07	1.07	1.46	1.09	1.22	1.71	1.32	1.11
e4d2	1.07	1.12	1.08	1.08	1.42	1.1	1.27	1.81	1.34	1.12
e4d3	1.08	1.13	1.08	1.08	1.46	1.1	1.27	1.78	1.33	2.12

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

This is also the first time that any readily discernible difference has occurred between the ridge-2 and RLS estimators, as the multicollinearity reduces the effectiveness of the former (Table 4.2). However, in general, the estimators behave in much the same way as at the lower condition number.

At a condition number of 40, we continue to see DDE1 differing from the other estimators, in that while its highest values of MSE are for mixed Gaussian errors, it maintains MSE values nearly this high over the $t_{(3)}$ and $\chi^2_{(5)}$ distributed errors also (Table 4.2). Potentially, this is because the thicker tails of the $t_{(3)}$ error distribution and higher multicollinearity may place some of the error values outside of the calculated range during bound estimation.

4.1.3 Mean Squared Error for all models when condition number is 60

Table 4.3 presents MSE for all estimators when the condition number is 60; this signifies the cut-off denoting significant multicollinearity (Belsley, 1991) where we may expect to see degradation in our inferences. One of the most noticeable changes here is that the Leuven-2 estimator has continued to become more variable and move less in step with the other estimators. This is likely because of the entropy term being extended to the error specification in the Leuven-2 estimator, compared with the Leuven-1 model. However, this also causes this estimator to become more variable than other estimators.

Table 4.3: Mean squared error for estimators with a condition number of 60.

Mean squared error	Condition number: 60									
	Model									
Dist.[1]	OLS	Ridge-1	Ridge-2	RLS	GMEN	GMEW	DDE2	DDE1	Leuven-1	Leuven-2
e1d1	1.01	1.06	1.02	1.02	1.39	1.04	1.13	1.17	1.26	1.05
e1d2	1.01	1.06	1.02	1.02	1.35	1.04	1.14	1.18	1.28	1.06
e1d3	1.01	1.05	1.02	1.02	1.39	1.03	1.11	1.16	1.27	2.16
e2d1	1.81	1.86	1.82	1.83	2.23	1.85	1.98	2.13	2.08	1.87
e2d2	1.82	1.87	1.83	1.84	2.2	1.85	2.05	2.12	2.11	1.87
e2d3	1.8	1.85	1.82	1.82	2.22	1.84	1.97	2.09	2.08	2.87
e3d1	1.05	1.09	1.05	1.05	1.45	1.09	1.24	1.62	1.29	1.09
e3d2	1.06	1.1	1.07	1.07	1.59	1.29	1.41	1.75	1.33	2.23
e3d3	1.05	1.1	1.06	1.06	1.48	1.14	1.25	1.61	1.31	2.12
e4d1	1.06	1.11	1.07	1.07	1.43	1.09	1.22	1.76	1.31	1.11
e4d2	1.06	1.11	1.07	1.07	1.4	1.09	1.26	1.84	1.32	1.11
e4d3	1.06	1.1	1.06	1.07	1.43	1.09	1.21	1.7	1.32	1.1

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The OLS estimator has noticeably lower MSE values than the other estimators, as expected due to its objective function (Table 4.3). The DDE1 estimator still follows the OLS, ridge-1, ridge-2, RLS, GMEN, GMEW, DDE2 and Leuven-1 estimators in having a peak over the mixed Gaussian error distributions, thus illustrating the difficulty that corrupted sampling distributions add to estimation.

With this condition number, the GMEN estimator often provides relatively high MSE results, which is unsurprising, as illustrated in Section 4.1.4. However, GMEN does not always provide the largest MSE, with both the Leuven-2 and the DDE1 estimator producing the largest MSE value at times (Table 4.3). This infers that these estimators are generally the least accurate for predicting the explanatory

variable based upon the data at the condition numbers simulated thus far, regardless of the data and error distributions used.

4.1.4 Mean squared error for all models when condition number is 100

Table 4.4 presents MSE for all estimators when the condition number is 100, simulating severe multicollinearity. The results, however, tend to follow the same general pattern as with the lower condition numbers. The GMEN estimator again generally produces the highest MSE values, closely followed and sometimes surpassed by DDE1.

Table 4.4: Mean squared error for estimators with a condition number of 100.

Mean squared error	**Condition number: 100**									
	Model									
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	1.01	1.05	1.02	1.02	1.41	1.04	1.11	1.16	1.27	1.05
e1d2	1.01	1.05	1.03	1.03	1.4	1.04	1.12	1.17	1.29	1.06
e1d3	1	1.04	1.01	1.02	1.41	1.03	1.1	1.16	1.27	1.05
e2d1	1.79	1.84	1.81	1.82	2.24	1.83	1.93	2.07	2.07	1.85
e2d2	1.8	1.84	1.83	1.83	2.24	1.84	2	2.17	2.09	1.86
e2d3	1.8	1.85	1.82	1.82	2.28	1.84	1.99	2.12	2.09	1.85
e3d1	1.06	1.1	1.07	1.07	1.57	1.45	1.3	1.64	1.32	1.1
e3d2	1.06	1.1	1.07	1.08	1.66	1.3	1.39	1.74	1.34	1.11
e3d3	1.05	1.09	1.07	1.07	1.75	1.48	1.34	1.59	1.33	1.1
e4d1	1.07	1.11	1.08	1.08	1.46	1.09	1.21	1.68	1.32	1.11
e4d2	1.06	1.1	1.07	1.08	1.46	1.09	1.24	1.74	1.34	1.11
e4d3	1.07	1.11	1.08	1.09	1.49	1.1	1.25	1.83	1.34	2.05

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N(0,1)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N(0,1)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

It is unsurprising to find GMEN giving high MSE values, particularly when compared to GMEW, as a decrease in GME bounds leads to an increase in prediction losses, based on arguments presented in Golan et al. (1996).

In the case of DDE1, these relatively high values are also unsurprising, as the DDE1 estimator is not designed for minimising MSE as multicollinearity increases. The variability in the DDE1 estimator is likely due to the multicollinearity in the data causing a wider range of values arising from the bounded set regression employed within the framework. The MSE of the DDE1 estimator is nearly as high for the values utilising a $t_{(3)}$ or $\chi^2_{(5)}$ error distribution as for the mixed Gaussian errors (Table 4.4), which follows the pattern that has been emerging with the lower condition numbers.

With the condition number of 100, the estimator with the lowest MSE is again OLS, as would be expected due to its objective function, followed by ridge-2, RLS, and ridge-1 (Table 4.4). Ridge-2 seems to be a more effective estimator in the MSE sense than the MSEL sense (Section 4.2) in instances of high multicollinearity. This is likely due to the use of the *f-stat* in its ridge parameter, as it involves the residual sum of squares (Section 3.3) that is equivalent to MSE. .

It should however be noted that differences between estimators are relatively small, not showing any of the estimators to be completely unsuitable for making predictions of the dependent variable in the presence of multicollinearity, at least at the condition numbers simulated thus far.

4.1.5 Mean squared error for all models when condition number is 500

Table 4.5 presents MSE for all estimators when the condition number is 500; here we are dealing with extreme multicollinearity. In this situation, as shown in Table 4.5, the estimator that produces the lowest values of MSE is again, as expected, the OLS estimator. This estimator has values lower than 1.1, except for in the situations using the mixed Gaussian error distribution.

Table 4.5: Mean squared error for estimators with a condition number of 500.

Mean squared error	**Condition number: 500**									
	Model									
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	1	1.03	1.03	1.03	1.42	1.03	1.08	1.16	1.27	1.05
e1d2	1	1.02	1.02	1.02	1.37	1.03	1.12	1.17	1.28	1.05
e1d3	1.01	1.03	1.03	1.03	1.4	1.03	1.11	1.15	1.27	1.05
e2d1	1.8	1.83	1.84	1.83	2.27	1.84	1.95	2.08	2.07	1.86
e2d2	1.79	1.83	1.83	1.83	2.23	1.84	2	2.12	2.09	1.85
e2d3	1.78	1.82	1.83	1.82	2.24	1.83	1.93	2.1	2.07	1.85
e3d1	1.05	1.08	1.08	1.08	1.60	1.33	1.29	1.76	1.32	1.1
e3d2	1.06	1.08	1.08	1.08	1.56	1.97	1.4	1.83	1.34	1.11
e3d3	1.06	1.08	1.08	1.08	1.89	1.42	1.39	1.79	1.32	1.1
e4d1	1.07	1.09	1.09	1.09	1.47	1.1	1.21	1.77	1.33	1.11
e4d2	1.07	1.09	1.09	1.09	1.42	1.09	1.24	1.81	1.33	1.11
e4d3	1.07	1.1	1.1	1.09	1.46	1.10	1.24	1.77	1.33	1.12

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N(0,1)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N(0,1)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

This result is followed by ridge-1, ridge-2, and RLS, respectively, which seem to give values that are practically identical to each other in this instance. The next lowest values of MSE are given by the GMEW estimator, except for peaks across

the $t_{(3)}$ error distributions. This, combined with fluctuations in the GMEN estimator, together reflect the decrease in the usefulness of GME for non-normal error distributions without adjusting for the error bounds accordingly (Golan et al., 1996).

This is followed by the Leuven-2 estimator, then DDE2, and Leuven-1, and then DDE1. Indeed, DDE1 has continued its pattern of generally being less effective than the other estimators in the presence of any of the non-normal error distributions, except for GMEN, which produces higher MSE values than DDE1 in the presence of mixed Gaussian errors (Table 4.5).

4.1.6 Mean squared error: Comparison of means for condition numbers: 1, 40, 60, 100, 500

To make the results more general, average values of MSE for each estimator for each condition number presented have been computed in Table 4.6. This allows us to see, on average, at a given condition number, which estimator is going to be the most accurate at predicting the dependent variable based upon the data.

Table 4.6: Mean squared error average values for each estimator and condition numbers: 1, 40, 60, 100, 500.

	Model									
Cond. No.	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
1	1.23	1.35	1.23	1.23	2.04	1.24	1.47	1.62	1.5	1.24
40	1.23	1.29	1.24	1.24	1.69	1.3	1.43	1.67	1.5	1.65
60	1.23	1.28	1.24	1.25	1.63	1.29	1.42	1.68	1.5	1.63
100	1.23	1.27	1.25	1.25	1.7	1.34	1.41	1.67	1.51	1.36
500	1.23	1.26	1.26	1.26	1.69	1.38	1.41	1.71	1.5	1.28

This makes the results more useful, as while the specific results are helpful if you know the distributions of the data that you are working with, in many circumstances this is not the case. As such, these averages can advise an investigator as to which estimator would be the most accurate to use, depending upon the levels of multicollinearity observed in the data. Table 4.6 includes the average MSE values only of those condition numbers discussed in Chapter 4. For a full table including the average MSE values of all condition numbers see Appendix 2 (Section A2.3.1).

As can be seen from Table 4.6, when looking at the average values of MSE, in general the highest values of MSE are produced by either the GMEN estimator, or the DDE1 estimator, making these the least effective at accurately predicting the value of the dependent variable. This is likely due to GMEN using relatively narrow support bounds that perform poorly when it comes to MSE levels if the support bounds are not perfectly centred upon the true β values (Mittelhammer et al., 2013).

It is also interesting to note that in the presence of no multicollinearity (condition number of 1), the GMEN estimator is slightly less accurate at making these predictions than in cases of multicollinearity (Table 4.6). When multicollinearity is present, there is no significant change in MSE for this estimator. This aligns with the results of Golan et al. (1996) who found the GME estimator to be nearly invariant to the degree of ill-conditioning considered, though these authors tested only a small range of multicollinearity. The MSE values of the DDE1 estimator are also relatively high, likely due to the sensitivity of the bounded set regression to the effects of multicollinearity.

The OLS estimator provides the lowest levels of MSE; either lower than, or equal to, the MSE of all the other estimators, regardless of multicollinearity level (Table 4.6). This is unsurprising, as OLS is designed to minimise the values of MSE. This underlies why multicollinear relationships can still produce accurate fits. It is the precision of parameter identification that is primarily affected by multicollinearity, not estimation (Mittelhammer et al., 2000).

The ridge-2 and RLS estimators provide very similar MSE results, with ridge-2 giving slightly lower values on occasion (Table 4.6). The MSE for both of these estimators suffers only slightly from an increase in multicollinearity. In the case of ridge-2, this is likely because the multicollinearity corrupts the data upon which the ridge-2 parameter is based. However, due to the general construction of the model (the ridge parameter being specifically designed to reduce the effects of multicollinearity), the detrimental effect is minimal.

The ridge-1 estimator is less able to predict the dependent variable in situations with low multicollinearity than those with high multicollinearity. This is evidenced by its values of MSE decreasing from 1.35 to 1.25 as the condition number increases (Table 4.6). This is likely due to the fact that the addition of the ridge parameter to the regression to combat multicollinearity slightly offsets results in the cases where multicollinearity is not present.

The DDE2 estimator continues its trend of being roughly in the middle of the estimators looked at here, performing as well as many of the estimators studied. DDE2 has a similar trend to the ridge-1 estimator, with a small MSE decrease as condition number increases (Table 4.6); this aligns itself with the results in Doole (2013).

The Leuven-2 estimator seems to be relatively unpredictable when it comes to prediction as multicollinearity increases. Instead of either a constant increase or decrease in the average values of MSE as the condition number increases, the Leuven-2 estimator instead produces fluctuations (Table 4.6). This estimator produces relatively high values of MSE under some condition numbers, and relatively low MSE values under others (this result can be better seen in Section A2.3.1). This makes it difficult to assess the appropriateness of the Leuven-2 estimator.

It is, however, important to note that overall the differences between estimators here are minor, to the point that the relative standings of the estimators may vary widely across different experimental designs. However, it is reasonable to predict that OLS will remain the most accurate at prediction, with GME-based procedures, including DDE, providing less precise estimation.

4.2 Comparison of mean squared error loss for each estimator

Mean squared error loss (MSEL) is used as a test of empirical precision risk; this is the capacity of the estimator to reliably estimate the true value of the coefficients. Any estimator that produces a higher MSEL statistic is shown to be less able, relative to an estimator with a lower MSEL statistic for that situation, to reliably estimate the true parameters in the numerical experiment.

10 different condition numbers are tested to ensure robust results; however, for the sake of parsimony, as well as to add focus to the results, only the five condition numbers that significantly add to the results will be discussed here. The

five condition numbers to be discussed are 1, 40, 60, 100, and 500. To see the results for condition numbers 10, 20, 80, 250, and 1,000, please see Appendix 2.

4.2.1 Mean squared error loss for all models when condition number is 1

Table 4.7 presents MSEL for all estimators when the condition number is 1, i.e. no multicollinearity. DDE1 generally produces the highest value of MSEL in this instance. This shows it to be the least able to reliably estimate the true coefficients when no multicollinearity is present. This low level of effectiveness in cases with no multicollinearity may be caused through the process of regularising the estimator to attempt to deal with severe multicollinearity, introducing bias.

Table 4.7: Mean squared error loss for estimators with a condition number of 1.

Mean squared error loss				**Condition number: 1**						
					Model					
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	0.02	0.02	0.02	0.02	0.03	0.02	0.05	0.05	0.02	0.02
e1d2	0.03	0.03	0.03	0.03	0.05	0.02	0.06	0.06	0.03	0.02
e1d3	0.02	0.03	0.02	0.02	0.04	0.02	0.05	0.05	0.03	0.02
e2d1	0.04	0.03	0.04	0.04	0.04	0.04	0.08	0.08	0.03	0.04
e2d2	0.04	0.05	0.04	0.04	0.06	0.04	0.13	0.11	0.04	0.04
e2d3	0.04	0.04	0.04	0.04	0.04	0.04	0.09	0.09	0.04	0.04
e3 d1	0.02	0.02	0.02	0.02	0.05	0.04	0.12	0.15	0.02	0.02
e3 d2	0.03	0.03	0.03	0.03	0.07	0.04	0.16	0.19	0.04	0.03
e3 d3	0.03	0.03	0.03	0.03	0.05	0.04	0.1	0.17	0.03	0.03
e4 d1	0.02	0.02	0.02	0.02	0.03	0.02	0.05	0.09	0.03	0.02
e4 d2	0.03	0.04	0.03	0.03	0.05	0.03	0.08	0.13	0.04	0.03
e4 d3	0.02	0.03	0.02	0.02	0.04	0.02	0.06	0.09	0.03	0.02

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The other estimator that produces a MSEL that is slightly higher than the others is the DDE2 estimator. The GMEN estimator is also marginally higher than the remaining alternative estimators regardless of error and data distribution (Table 4.7), making these three estimators the least effective at accurately predicting the true coefficients in a situation with no multicollinearity.

The remaining estimators are all fairly similar, with Leuven-2 producing values slightly lower than those of OLS, ridge-1, ridge-2, RLS, GMEW, and Leuven-1. This gives it the lowest MSEL values and thus makes it the most effective at accurately predicting the true coefficients (Table 4.7). However, the levels of MSEL identified are very low for all instances, with no estimator clearly unsuitable in the presence of no multicollinearity.

4.2.2 Mean squared error loss for all models when condition number is 40

Table 4.8 presents MSEL for all estimators when the condition number is 40, signifying multicollinearity that is starting to become problematic. Here, we see that the DDE2 estimator has become much worse at accurately predicting the coefficients. This is unsurprising and is due to the fact that in the bounded set regression stage of the estimator, which defines the support bounds to be used, there is no accounting made for multicollinearity in the data. This leads to a situation where, as soon as multicollinearity is present, the method is less effective.

Table 4.8: Mean squared error loss for estimators with a condition number of 40.

Mean squared error loss			Condition number: 40							
					Model					
Dist.[1]	OLS	Ridge-1	Ridge-2	RLS	GMEN	GMEW	DDE2	DDE1	Leuven-1	Leuven-2
e1d1	10.62	0.67	6.07	3.77	0.57	0.77	23.82	1	0.61	0.86
e1d2	11.49	0.74	5.28	3.71	0.58	0.81	31.41	0.97	0.6	0.68
e1d3	9.33	0.98	4.51	3.35	0.57	0.88	24.14	1.04	0.62	0.67
e2d1	13.05	0.91	5.59	4.37	0.58	0.97	36.37	1.47	0.62	0.71
e2d2	20.79	1.32	6.06	5.46	0.59	1.2	59.51	1.41	0.62	0.73
e2d3	16.93	1.15	7.03	4.8	0.61	1.14	40.35	1.35	0.64	0.92
e3d1	8.51	0.64	4.44	3.24	0.62	0.89	46.77	1.28	0.61	0.85
e3d2	8.61	0.7	3.98	3.04	0.63	1.24	44.69	1.37	0.6	0.68
e3d3	6.56	0.66	3.52	3.05	0.7	0.89	46.41	1.32	0.62	0.67
e4d1	9.07	0.66	4.75	3.56	0.56	0.77	20.58	1.08	0.6	0.67
e4d2	10.14	0.7	4.22	3.52	0.57	0.72	29.99	1.04	0.6	0.67
e4d3	8.38	0.7	4.34	3.53	0.58	0.79	26.05	1.24	0.61	0.86

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

This is followed by the OLS estimator, there having been significant degradation in the ability of the OLS estimator to accurately predict the coefficients as multicollinearity has increased (Table 4.8). This is unsurprising and follows the result from Paris (2001) in showing the MSEL of OLS increasing with an increase in condition number.

The GMEN estimator generally provides the lowest values of MSEL (closely followed by Leuven-1), making it the most accurate of the estimators tested in this scenario. The high accuracy of GMEN is in line with previous results (Ciavolino & Al-Nasser, 2009; Golan et al., 1996). However, it is also shown that the GMEW estimator produces a higher MSEL value than the GMEN estimator (Table 4.8), highlighting that the parameter estimates are indeed very sensitive to

the choice in support bounds (Akdeniz., 2011; Golan et al., 1996; Leon., 1999; Paris & Caputo, 2001; Paris & Howitt, 1998).

4.2.3 Mean squared error loss for all models when condition number is 60

Table 4.9 presents MSEL for all estimators when the condition number is 60, which denotes that multicollinearity is now affecting estimation. In this case, the DDE2 estimator again produces the highest level of MSEL due to the aforementioned reasons. This shows it to be, in this situation, the least accurate in predicting the true regression coefficients.

Table 4.9: Mean squared error loss for estimators with a condition number of 60.

Mean squared error loss			**Condition number: 60**							
					Model					
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	15.85	0.78	6.51	5.02	0.6	0.78	47.34	1.02	0.64	0.72
e1d2	22.91	0.9	6.41	5.32	0.6	0.85	64.09	1.08	0.62	0.73
e1d3	16.19	0.71	5.67	5.05	0.59	0.75	39.07	0.85	0.63	0.91
e2d1	32.06	1.12	6.77	5.61	0.59	0.96	76.43	1.15	0.63	0.75
e2d2	36.74	1.21	5.32	5.32	0.58	0.88	117.49	1.03	0.61	0.74
e2d3	36.94	1.52	6.42	5.14	0.57	0.81	91.02	1.09	0.61	0.87
e3d1	14.5	0.79	5.21	4.48	0.62	0.77	67.35	1.12	0.65	0.73
e3d2	28.01	1.77	6.21	4.72	0.74	0.96	91.23	1.22	0.62	0.92
e3d3	20.5	0.78	5.82	5.24	0.61	0.79	92.23	1.2	0.62	0.89
e4d1	18.3	0.76	6.45	4.18	0.59	0.71	57.98	0.96	0.63	0.71
e4d2	21.09	0.72	4.91	4.55	0.59	0.7	66.63	0.93	0.61	0.71
e4d3	22.73	0.83	8.11	4.29	0.6	0.79	48.46	0.97	0.63	0.72

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The OLS regression unsurprisingly produces the second highest values of MSEL; this is followed by the ridge-2 estimator and then the RLS estimator (Table 4.9). Due to the restriction of its search for potential coefficients to the closed interval between -5 and 5, RLS has considerably less empirical precision risk than OLS for condition numbers involving multicollinearity, showing that these bounds are effective in reducing MSEL relative to OLS.

The DDE1 estimator generally produces the next highest MSEL values, except for 3 points (mixed Gaussian errors and either $t_{(3)}$ or $\chi^2_{(5)}$ data, and $t_{(3)}$ data and errors), where the ridge-1 estimator provides slightly higher MSEL values (Table 4.9).

The lowest MSEL values are in general given by the GMEN estimator. This is unsurprising, as both GME estimators force estimates to be consistent with the sample and prior information. This causes the GME estimators to be relatively invariant to increases in condition number. This is combined with the fact that as condition number increases, narrow bounds become more important for restricting the variability evident in estimated coefficients (Golan et al., 1996).

GMEN is again followed closely by Leuven-1. The effectiveness of Leuven-1 is unsurprising and aligns itself with results in Paris (2001, 2004) which show that the MSEL for the Leuven-1 model only changes slightly with an increase in condition number. They also show that the Leuven-1 estimator is considerably more stable in the presence of multicollinearity than either the Leuven-2 or OLS estimator.

4.2.4 Mean squared error loss for all models when condition number is 100

Table 4.10 presents MSEL for all estimators when the condition number is 100 (i.e. extreme multicollinearity). We see a significant increase in the MSEL for DDE2 (which produces the highest MSEL value) and OLS (which provides the second highest MSEL value). The MSEL for DDE2 increases as the condition number rises, due to multicollinearity in the data that its support bounds are estimated from. It is unsurprising to see this result for OLS, as it is well known that parameter estimation in OLS is generally highly sensitive to multicollinearity (Farrar & Glauber, 1967).

Table 4.10: Mean squared error loss for estimators with a condition number of 100.

Mean squared error loss				**Condition number: 100**						
					Model					
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	55.35	1.24	7.34	5.51	0.55	0.76	126.43	0.9	0.58	0.66
e1d2	58.19	1.19	6.14	5.76	0.55	0.77	179.85	0.9	0.56	0.67
e1d3	56.68	0.83	7.72	5.13	0.54	0.68	176.33	0.85	0.57	0.65
e2d1	92.77	1.92	6.42	6.26	0.55	0.78	191.33	1.04	0.59	0.69
e2d2	118.31	3.36	7.05	6.03	0.56	0.81	281	1.11	0.58	0.69
e2d3	72.26	1.77	5.27	5.99	0.56	0.77	205.62	1.06	0.6	0.73
e3d1	57.44	1.37	7.12	5.64	0.67	1.07	281.86	1.25	0.57	0.66
e3d2	58.98	1.16	6.34	5.82	0.75	1.18	396.09	1.32	0.57	0.67
e3d3	53.57	0.85	6.83	4.58	0.63	0.76	309.34	1.1	0.57	0.65
e4d1	38.75	0.82	5.29	5.48	0.55	0.71	129.50	0.9	0.58	0.67
e4d2	54.74	1.02	5.9	5.51	0.56	0.74	170.07	1.03	0.57	0.68
e4d3	48.35	0.84	6.26	4.53	0.54	0.66	175.61	0.96	0.57	0.82

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The GMEN and Leuven-1 models produce the lowest values of MSEL (Table 4.10), making them the best at predicting coefficients in the presence of this level of multicollinearity. The similarity between the two estimators approximates the results of Paris (2001), who found that GME with narrow support intervals (as is the case with GMEN) produced a stability similar to that of the Leuven-1 model. Compared to this, a GME model with wider supports (such as GMEW) is considerably less stable.

GMEW producing higher MSEL values than GMEN is also expected, as the sensitivity of GME to the choice of support bounds is well established, and wider coefficient support bounds giving larger MSEL values aligns itself with these expectations (Akdeniz et al., 2011; Golan et al., 1996; Leon et al., 1999; Paris & Caputo, 2001; Paris & Howitt, 1998). For many of the error and data distributions, the next lowest values of MSEL are given by the Leuven-2 estimator, followed by GMEW and then either ridge-1 or DDE1 (Table 4.10).

It is also interesting to note that both GMEN and GMEW generally produce their lowest MSEL values in the presence of a normal error distribution, with their highest MSEL values occurring when a $t_{(3)}$ error distribution is used (Table 4.10). This is because while the n-sigma rule guarantees a majority of the error observations to be within the chosen error bounds, the error bounds do not contain the true error values for all observations, and this is particularly obvious under the thicker tailed $t_{(3)}$ error distribution (Campbell & Hill, 2005).

4.2.5 Mean squared error loss for all models when condition number is 500

Table 4.11 presents MSEL for all estimators when the condition number is 500. We again see a major increase in the levels of MSEL for the DDE2 and OLS estimator (Table 4.11), as the bounded set regression used in DDE 2 continues to fail to provide well-conditioned bound estimates under severe multicollinearity. OLS continues to show the expected increase in MSEL values at high condition numbers (Paris, 2001).

Table 4.11: Mean squared error loss for estimators with a condition number of 500.

Mean squared error loss				**Condition number: 500**						
				Model						
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	1,325.22	13.71	2	6.82	0.57	0.74	3,028.34	1.01	0.6	0.7
e1d2	1,421.03	11.13	1.95	7.13	0.58	0.78	4,027.37	0.97	0.59	0.71
e1d3	1,182.46	6.88	1.88	6.08	0.57	0.69	3,008.03	0.9	0.6	0.68
e2d1	2,031.26	25.35	1.85	6.71	0.58	0.79	7,255.56	0.97	0.63	0.73
e2d2	2,405.66	73.85	1.68	6.24	0.6	0.82	10,155.43	1	0.62	0.75
e2d3	2,561.08	36.33	1.90	6.29	0.56	0.78	4,764.70	0.99	0.6	0.7
e3d1	1,327.76	21.23	2.19	6.48	0.69	0.96	5,710.58	1.23	0.6	0.69
e3d2	1,364.60	13.64	1.97	6.6	0.84	2.7	8,238.05	1.25	0.59	0.7
e3d3	1,115.61	7.08	2.08	6.36	0.65	0.94	7,188.63	1.06	0.6	0.67
e4d1	1,244.18	5.8	2.23	6.34	0.58	0.73	3,034.36	0.91	0.61	0.7
e4d2	1,553.33	7.02	1.79	6.49	0.58	0.74	5,623.90	1.03	0.59	0.7
e4d3	1,179.40	16.07	1.87	6.95	0.57	0.74	3,265.09	1	0.59	0.68

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

These are followed by the ridge-1 model (Table 4.11), which at this level of multicollinearity has also begun suffering severe problems, compared to its previous results. This may be because as multicollinearity increases, more of its effects—other than the common one associated with increased coefficient values—begin to become evident, thus decreasing the effectiveness of the ridge parameter that aims to penalise large coefficient values (Cule & De Iorio, 2013).

This is followed by RLS, which has a relatively high level of MSEL in the presence of a condition number of 500, compared to a majority of the estimators (although it is considerably lower than DDE2, OLS and generally ridge-1) (Table 4.11). This is likely because of the coarse way that this estimator combats multicollinearity, given that it does so by solely restricting the magnitude of the coefficients to be estimated.

The lowest values of MSEL are again produced by the Leuven-1 and GMEN estimators. The Leuven-1 model may be so accurate, even at this extreme level of multicollinearity, because the parameter probabilities in general do not approach a uniform distribution as is the case in the GME class of estimators. This enables the choice of probabilities allowing more accurate coefficient estimates, without needing to tend towards uniformity. This allows more flexibility in the Leuven-1 estimator (Paris, 2001).

4.2.6 Mean squared error loss: Comparison of means for condition numbers: 1, 40, 60, 100, 500

To make the results more general, the average values of MSEL for each estimator, for each condition number discussed above, are computed in Table 4.12. This allows us to see, on average, for a given condition number, which estimator is going to be the most accurate at predicting the estimated coefficients.

Table 4.12: Mean squared error loss average values for each estimator, and condition numbers: 1, 40, 60, 100, 500.

	Model									
Cond. No.	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
1	0.03	0.03	0.03	0.03	0.05	0.03	0.09	0.1	0.03	0.03
40	11.12	0.82	4.98	3.78	0.6	0.92	35.84	1.21	0.61	0.75
60	23.82	0.99	6.15	4.91	0.61	0.81	71.61	1.05	0.62	0.78
100	63.78	1.36	6.47	5.52	0.58	0.81	218.59	1.04	0.58	0.69
500	1,559.3	19.84	1.95	6.54	0.61	0.95	5,441.67	1.03	0.6	0.7

This makes the results more useful, as while the more specific results are more helpful if you know the distributions of the data that you are working with, in many circumstances this is not the case. As such, these averages can advise someone as to which estimator is expected to be the most accurate, depending upon the levels of multicollinearity in the data. Table 4.12 includes the average MSEL values only of those condition numbers discussed in Chapter 4. For a full table including the average MSEL values of all condition numbers tested in this thesis, see Appendix 2 (Section A2.3.2).

With a condition number of 1, or no multicollinearity, OLS is as accurate at predicting the correct coefficients as ridge-1, ridge-2, RLS, GMEW, Leuven-1 and Leuven-2, and more accurate than the other estimators looked at here, with these estimators all having an average MSEL of 0.03 (Table 4.12). This aligns itself with expectations as we know that OLS is very accurate in the absence of multicollinearity (Stock & Watson, 2007).

These estimators are followed by GMEN, DDE2 and then DDE1 (Table 4.12). However, as the condition number increases, the average MSEL for the DDE2 and OLS estimators quickly surpass those of the other estimators investigated here. Thus, from a condition number of 20 upwards (Section A2.3.2), average MSEL for the DDE2 model is higher than any other estimator, reaching a peak MSEL average of 21,331.82 (Section A2.3.2) and a peak of 5,441.67 in the condition numbers examined in this chapter (Table 4.12). This makes DDE2 the least accurate at predicting the correct coefficients and is due to the presence of multicollinearity, which is not accounted for in the estimation procedure used to endogenously generate the support bounds for DDE2.

DDE2 is followed by the OLS estimator, which gives a higher average MSEL value than any estimator other than DDE2 from a condition number of 40 upwards (Table 4.12) This is unsurprising as it is well known that the MSEL of OLS will increase in the presence of multicollinearity (Paris, 2001), making OLS a less suitable estimator for accurately ascertaining the true coefficients based on the data, when multicollinearity increases.

The estimators that appear to deal with multicollinearity the best in terms of estimation precision, are the GMEN and Leuven-1 estimators, with these being

very similar to each other (Table 4.12). This is logical as the Leuven-1 estimator was specifically designed for resolving multicollinearity, and these results align with those of Paris (2001, 2004). The GMEN estimator performs very well due to its narrow bounds (as described in Section 4.2.3); these enable it to have far more information when predicting coefficient values than the other estimators (except, of course, RLS). These are followed in general by the Leuven-2 estimator, which supports the results of Paris (2004) who found that the Leuven-2 model provided MSEL values close to, but not as low as those of the Leuven-1 model and the narrow-bounded GME model in cases of strong multicollinearity.

The ridge-1 estimator closely follows this in terms of average MSEL values up until a condition number of 80, at which point the differences between ridge-1 and Leuven-2 become large. The ridge-1 model handles the increases in multicollinearity relatively well initially, with it slowly increasing until a condition number of 100 is reached (Table 4.12). After this point, it sharply increases to a maximum average value of 243.63 with a condition number of 1000 (Section A2.3.2). This may be because, as multicollinearity increases, more of its effects—other than the common one associated with increased coefficient values—begin to become evident, thus decreasing the effectiveness of the ridge parameter that aims to penalise large coefficient values (Cule & De Iorio, 2013).

The GMEW estimator follows the results of ridge-1 until a condition number of 40 is reached. After that, it produces lower values of MSEL than the ridge-1 model. The GMEW model follows our expectations in that it is relatively invariant to the change in condition number (Golan et al., 1996). However unsurprisingly, it is not as good as GMEN, producing higher MSEL values

(Akdeniz et al., 2011; Golan et al., 1996; Leon et al., 1999; Paris & Caputo, 2001; Paris & Howitt, 1998).

GMEW is generally followed by the DDE1 estimator, which performs well in comparison to OLS, ridge-1, ridge-2, RLS, and DDE2, although not as well as GMEN, GMEW, Leuven-1 and Leuven-2. In the presence of no multicollinearity (condition number 1), the DDE1 estimator produces an average MSEL value of 0.10, which then increases as the condition number increases to 20 (Section A2.3.2). From this point onwards, the average MSEL values of DDE1 stay relatively constant as multicollinearity increases (Table 4.12), never reaching higher than 1.36 (Section A2.3.2).

The RLS estimator also begins with a low MSEL of 0.03 in the presence of no multicollinearity (condition number of 1) (Table 4.12); it then increases as the condition number increases, reaching a maximum average MSEL of 6.76 when the condition number is 1000 (Section A2.3.2). RLS performs very well in comparison to OLS, which of course has the same objective function. This is because RLS has the same advantage as the GMEN estimator; namely having a set range in which to look for its coefficient values, thus making it considerably more effective than OLS, provided of course that the bounds are reasonable. Indeed, in reality, these bounds are typically unavailable and RLS will be less successful in these instances. Even with these bounds RLS does not perform well compared to many of the other estimators, likely because of the coarse way that this estimator combats multicollinearity, given that it does so by solely restricting the magnitude of the coefficients to be estimated.

One of the more interesting results here is presented by the ridge-2 model, and can also be seen in Table 4.12. Ridge-2 initially begins at the same level of many estimators in the presence of no multicollinearity (MSEL of 0.03). It then increases as multicollinearity increases, however at a decreasing rate, before reaching a maximum average MSEL of 6.47 with a condition number of 100 (Table 4.12), and then decreasing again, essentially moving in a curve. This signifies how the results of a ridge regression can be difficult to predict, as the accuracy with which it estimates coefficients varies, depending on the interaction between the data and the specific technique used to establish the ridge parameter.

4.3 Response Surface Results

4.3.1 Response Surface Results for mean squared error

The first response surface analysis that is conducted focuses on MSE. When looking at these response surface results, one of the main things we are looking for are estimators that allow us to deal with multicollinearity. This ability would become apparent here in cases where the coefficient on the condition number is classed as insignificant. This would tell us that the condition number of the data (and so the multicollinearity level) has no significant effect on the MSE values for the estimator, and so no significant effect on the ability of that estimator to accurately predict the dependent variable based on the data.

An estimator is evidenced as being more robust when more of the response surface coefficients for it are insignificant. If the coefficient for any given data or error distribution is insignificant, then it means that that distribution being present will have no significant effect on the MSE values found by that estimator. This

makes an estimator with more insignificant response surface coefficients more robust in situations where error and data distributions are unknown. It is, however, important to note that just because an estimator is shown to be robust to changes in error or data distributions through this analysis, does not mean it is the most accurate estimator, as it may still provide higher MSE and/or MSEL values than a comparatively less robust estimator.

Table 4.13 presents the summary statistics for the MSE response surface analysis, to enable an overall picture of the MSE values calculated for each estimator. As can be seen, the overall average MSE value is the lowest for OLS (1.23), and is the highest for GMEN (1.72). This is unsurprising and supports the results found in Section 4.1.

Table 4.13: Summary Statistics for the MSE (Dependent Variable) used in the Response Surface Analysis for all Estimators.

Summary Statistics for the MSE										
	Model									
	OLS	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
Mean	1.23	1.29	1.24	1.25	1.72	1.35	1.43	1.67	1.5	1.49
StD	0.33	0.34	0.33	0.33	0.36	0.38	0.34	0.35	0.34	0.56
Var.	0.11	0.11	0.11	0.11	0.13	0.14	0.12	0.12	0.11	0.32
Min	1	1.02	1	1	1.32	1.01	1.08	1.13	1.24	1.02
Max	1.82	1.94	1.84	1.84	2.68	2.38	2.11	2.17	2.13	3.31

The standard deviations and thus the variances of the MSE values for all of the estimators are relatively similar, except for the Leuven-2 estimator (Table 4.13). The Leuven-2 estimator provides a significantly higher value, which is unsurprising and supports Section 4.1 in showing how variable and inconsistent

the Leuven-2 estimator is in comparison to the other estimators, as discussed in Section 4.1.3.

We can also see that GMEN provided the highest minimum MSE value, as well as the highest maximum (Table 4.13). This is unsurprising as it generally provided the highest overall MSE values in the Monte Carlo simulations. OLS provides the lowest minimum and maximum values (Table 4.13), which is again unsurprising as it provided the lowest MSE values throughout Section 4.1.

From the response surface analysis we see that the condition number, and so the level of multicollinearity, is not significant at any reasonable level for MSE for the OLS, GMEN, GMEW, Leuven-1, and the Leuven-2 estimators (Table 4.14). This means that there is no statistical evidence to support the fact that the condition number (and so level of multicollinearity) causes any change to the values of MSE for OLS, GMEN, GMEW, Leuven-1, or Leuven-2 models. This tells us that the predictive capacity of these estimators is not degraded by the level of multicollinearity, and is aligned with the MSE results previously discussed (Section 4.1).

The condition number is significant at the 5% level for the ridge-2, RLS, DDE2 and DDE1 models, with positive coefficients for ridge-2, RLS and DDE1, and a negative coefficient for DDE2 (Table 4.14). It is expected that DDE2 would perform much better in an MSE sense than an MSEL sense, as the bounded set regression used in the first step is based upon the Least Absolute Deviations (LAD) estimator. As such it focuses on minimising deviations (Bloomfield & Steiger, 1980), making its focus more prediction than precision based.

The increase in MSE when the condition number rises makes sense for RLS. As RLS employs user-defined bounds to prevent ill-conditioning, it provides little guidance overall for estimation. This effect is more pronounced the less information the practitioner has while defining the bounds. It also makes sense for the DDE1 model as DDE1 was created in an attempt to minimise MSEL values in instances of severe multicollinearity, not to shrink MSE values.

The condition number also has a significant effect (at the 10% level) on the ridge-1 model; this effect is negative, implying that as the condition number increases, the MSE for the ridge-1 model decreases (Table 4.14). This is because of the addition of the ridge parameter to the regression to combat multicollinearity. In instances with no multicollinearity this addition can slightly offset the results.

The coefficients for drawing data from a $t_{(3)}$ distribution are insignificant for OLS, ridge-1, ridge-2, RLS, GMEN and GMEW (Table 4.14). This makes sense for the GME estimators as we place bounds on the unknown parameters, but not the regressors (Campbell & Hill, 2005). They are positive and significant at the 5% level for DDE2, DDE1 and Leuven-1 models, as well as negative and significant at the 5% level for the Leuven-2 estimator. This shows that the MSE will increase for the DDE2, DDE1, and Leuven-1 models when the regressors are drawn from a $t_{(3)}$ distribution, compared to the other estimators that are immune to this problem. This is seen with the Leuven-2 estimator which, in fact, experiences a decrease in MSE when the regressors are drawn from a $t_{(3)}$ distribution.

Table 4.14: Response surface results for MSE.

Response Surface results for MSE										
	Model									
	OLS	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
Constant	1.01	1.06	1.01	1.01	1.42	0.99	1.1	1.13	1.27	1.23
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Cond. No.	-3.58E-06	-3.18E-05	2.11E-05	1.8E-05	-2.58E-05	6.3E-05	-3.75E-05	5.29E-05	-7.01E-06	1.1E-04
	(0.32)	(0.09)	(0)	(0)	(0.75)	(0.41)	(0)	(0.04)	(0.54)	(0.49)
$t_{(3)}$ Regressors	6.06E-04	7.93E-04	1.7E-03	9.8E-04	6.29E-03	5.04E-02	5.57E-02	5.88E-02	1.32E-02	-0.17
	(0.72)	(0.66)	(0.22)	(0.46)	(0.68)	(0.19)	(0)	(0)	(0)	(0.05)
$\chi^2_{(5)}$ Regressors	-1.62E-03	-1.92E-04	-6.97E-04	-1.23E-03	3.05E-02	3.57E-02	8.61E-03	8.58E-03	1.61E-03	0.19
	(0.34)	(0.92)	(0.62)	(0.36)	(0.05)	(0.35)	(0.19)	(0.33)	(0.41)	(0.04)
Mixed Gaussian Errors	0.79	0.81	0.8	0.8	0.87	0.82	0.88	0.94	0.81	0.79
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
$t_{(3)}$ Errors	4.66E-02	5.23E-02	4.94E-02	4.93E-02	0.27	0.42	0.23	0.5	4.68E-02	5.25E-02
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0.68)
$\chi^2_{(5)}$ Errors	6.19E-02	7E-02	6.32E-02	6.31E-02	5.95E-02	6.08E-02	0.13	0.6	6.03E-02	6.73E-02
	(0)	(0)	(0)	(0)	(0.06)	(0.28)	(0)	(0)	(0)	(0.59)
R^2	1	1	1	1	0.93	0.75	0.99	0.98	1	0.43

When looking at the coefficients for data drawn from the chi-square distribution, we find that they are insignificant for nearly all of the estimators examined here (Table 4.14). The GMEN and the Leuven-2 estimators are the only estimators for which the coefficients are significant, in this case at the 5% level, and both are positive. In the GMEN case, this is because the skewedness of the chi-square distribution causes a situation which approximates the bounds not being centred on the true values, which is known to make GME less effective (Mittelhammer et al., 2013). This result shows that the MSE of GMEN and Leuven-2 will increase when the regressors are drawn from a chi-square distribution, compared with the other estimators (OLS, ridge-1, ridge-2, RLS, GMEW, DDE2, DDE1, and Leuven-1) that are immune to this problem. OLS is immune to any changes in condition number or data distribution when it comes to its MSE levels, as it is designed entirely with the purpose of minimising the levels of MSE based on data.

The coefficients for having drawn errors from a mixed Gaussian distribution are positive and significant at the 5% level for all of the estimators (OLS, ridge-1, ridge-2, RLS, GMEN, GMEW, DDE2, DDE1, Leuven-1 and Leuven-2). This signifies the difficulty that corrupted sampling distributions add to estimation.

While the coefficients for the $t_{(3)}$ error distributions are all positive, they are insignificant for the Leuven-2 estimator, while being positive and significant at the 5% level for the rest of the estimators (OLS, ridge-1, ridge-2, RLS, GMEN, GMEW, DDE2, DDE1, and Leuven-1) (Table 4.14). This signifies that the Leuven-2 estimator handles errors defined over a distribution with thicker tails better than the other estimators. This is potentially because in the Leuven-2 estimator the entropy term is extended to the error term, as well the data term,

thus enabling it to better handle the thick tails of the $t_{(3)}$ distribution and the chi-square distribution discussed below.

When looking at the chi-square error distribution, we find a similar result. While all of the coefficients are positive, they are significant at the 5% level for OLS, ridge-1, ridge-2, RLS, DDE2, DDE1, and Leuven-1 estimators, significant at the 10% level for the GMEN estimator, and insignificant for the GMEW and Leuven-2 estimators (Table 4.14). The wide bounds on the GMEW estimator allow it to better handle the asymmetry of the chi-square distribution than GMEN, as does the Leuven-2 estimator's inclusion of entropy in the error terms. Thus, this study highlights that overall the GMEW and Leuven-2 estimators are the most robust methods overall, when it comes to prediction in the presence of different distributions for the regressor and error terms.

The R^2 is high (above 0.9) for all of the estimators, except for the GMEW estimator and the Leuven-2 estimator (Table 4.14). This highlights the capacity of the response surface to explain the output of the estimators as a function of the investigative runs. In this case, it signifies that the predictions provided by the estimators other than GMEW and Leuven-2 can be strongly related to condition number and distributions for the regressor and error terms. This has practical implications, in that less of the variance in the GMEW and Leuven-2 predictions can be traced back to variations in these data or error distributions, or changes in condition number. This means that there are other factors affecting these models, making accurate prediction with these models difficult without knowing which factors affect the variance.

4.3.2 Response Surface Results for mean squared error loss

The second set of response surface analysis conducted is on the MSEL values of the Monte Carlo experiments, where we are looking for the same thing as in Section 4.3.1. Insignificant coefficients for condition number show an estimator is invariant to changes in condition number, and the more insignificant coefficients an estimator has for different error and data distributions, the more robust the estimator is overall.

Table 4.15 presents the summary statistics for the MSEL response surface analysis, to enable an overall picture of the MSEL values calculated for each estimator. The Leuven-1 estimator has the lowest overall mean value of MSEL, closely followed by GMEN. This is unsurprising as these two estimators produce the lowest MSEL values throughout Section 4.2. The highest MSEL value is produced by DDE2, followed by OLS (Table 4.15), showing the inability of these estimators to accurately predict true coefficients in the presence of multicollinearity, as discussed in Section 4.2.

Table 4.15: Summary Statistics for the MSEL (Dependent Variable) used in the Response Surface Analysis for all Estimators.

Summary Statistics for the MSEL										
	Model									
	OLS	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
Mean	835.92	27.61	3.37	4.19	0.56	0.84	2849.2	1.02	0.55	0.65
StD	1,931.10	171.73	2.49	2.47	0.2	0.53	7013.3	0.37	0.18	0.25
Var.	3,729,300	29,491	6.18	6.12	0.04	0.28	49,186,000	0.14	0.03	0.06
Min	0.02	0.02	0.02	0.02	0.03	0.02	0.05	0.05	0.02	0.02
Max	10,025	1,858.5	11.11	7.34	0.94	2.96	41,124	1.88	0.71	1.15

DDE2 also produces the highest standard deviation in its MSEL values, again followed by OLS (Table 4.15). This shows the extreme variability in these estimators as the condition number increases. These are followed by ridge-1, which is considerably lower than either DDE2 or OLS, while still being high, reflecting the results found in Section 4.2.6. The estimators which produce the lowest standard deviation (and so variance) in MSEL values are unsurprisingly the Leuven-1 estimator, followed by the GMEN estimator (Table 4.15), as these two estimators provided consistent (and low) MSEL values throughout Section 4.2.

The lowest minimum MSEL value (0.02) is shared by OLS, ridge-1, ridge-2, RLS, GMEW, Leuven-1 and Leuven-2 (Table 4.15), whereas the largest minimum MSEL value (0.05) is shared by the DDE2 and DDE1 models. DDE2 produces the highest maximum MSEL value (41,124) followed by OLS (10,025) and ridge-1 (1,858.5), while the lowest maximums are produced by Leuven-1 (0.71) and GMEN (0.94) (Table 4.15).

From the response surface results for the MSEL of the Monte Carlo experiments, we see that the coefficients on the condition number are insignificant for GMEN, GMEW, DDE1, Leuven-1, and Leuven-2 estimators, thus showing that these estimators (in the MSEL sense) are relatively immune to changes in levels of multicollinearity (Table 4.16). This is expected and supports previous results that indicate the usefulness for these estimators in the presence of multicollinearity (Golan et al., 1996; Paris, 2001, 2004).

Table 4.16: Response surface results for MSEL.

Response Surface results for MSEL										
	Model									
	OLS	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
Constant	-617.04	-55.94	4.19	4.99	0.53	0.54	-2,735.1	0.75	0.57	0.59
	(0)	(0.13)	(0)	(0.03)	(0)	(0	(0)	(0)	(0)	(0)
Cond. No.	5.85	0.22	-3.28E-03	-2.26E-03	-8.09E-05	3.29E-04	19.72	1.05E-04	-1.38E-04	1.4E-04
	(0)	(0)	(0.06)	(0.02)	(0.47)	(0.16)	(0)	(0.64)	(0.12)	(0.38)
$t_{(3)}$ Regressors	154.73	4.95	-0.28	8.39E-02	1.29E-02	0.16	1131.7	5.2E-02	-1.52E-02	-2.22E-02
	(0.08)	(0.89)	(0.08)	(0.17)	(0.13)	(0.02)	(0.02)	(0)	(0.02)	(0.12)
$\chi^2_{(5)}$ Regressors	160.67	42.75	-0.25	-0.2	-1.14E-02	1.88E-03	-138.25	-9.1E-03	-1.61E-02	2.24E-02
	(0.07)	(0.24)	(0.12)	(0)	(0.19)	(0.98)	(0.76)	(0.62)	(0.01)	(0.13)
Mixed Gaussian Errors	370.81	14.38	0.32	0.22	-4.14E-03	0.15	1,825.8	0.23	-6.12E-03	2.57E-02
	(0.03)	(0.72)	(0.34)	(0.1)	(0.82)	(0.2)	(0.03)	(0)	(0.65)	(0.4)
$t_{(3)}$ Errors	213	68.1	-0.34	-0.61	8.75E-02	0.53	2,224.1	0.34	-3.64E-02	-7.06E-03
	(0.27)	(0.1)	(0.39)	(0)	(0)	(0)	(0.02)	(0)	(0.02)	(0.84)
$\chi^2_{(5)}$ Errors	-20.86	7.93	-0.57	-0.92	-4.6E-02	2.16E-02	675.24	7.41E-02	-5.82E-02	-2.33E-02
	(0.9)	(0.85)	(0.12)	(0)	(0.03)	(0.86)	(0.44)	(0.08)	(0)	(0.49)
R^2	0.92	0.19	0.78	0.93	0.85	0.47	0.85	0.86	0.87	0.8

The respective coefficients are negative and significant at the 5% level for the RLS estimator, and negative and significant at the 10% level for the ridge-2 estimator. The RLS result perhaps demonstrates the risk of achieving expected results when simple methods are used to combat multicollinearity. RLS employs user-defined bounds to prevent ill-conditioning, providing little guidance overall to the estimation to the best value for each coefficient. This will be made worse when the practitioner has little available data to inform the bounds of the closed intervals used to set these restrictions. Interestingly, as can be seen in Table 4.12, though the estimated coefficient is negative for the rigde-2 estimator, MSEL varies greatly over the simulated set of condition numbers.

The coefficients are positive and significant at the 5% level for OLS, ridge-1, and DDE2 (Table 4.16), so we can expect the level of MSEL for these estimators to increase as condition number increases. This decreases their ability to accurately predict the correct coefficients based on the data as multicollinearity increases. This increase in MSEL for OLS is due to the well-established failure of OLS to adequately deal with multicollinearity (Paris, 2001). The DDE2 MSEL increase is due to the fact that the first step in the DDE2 model (bounded set regression) is not able to provide well-conditioned estimates when severe multicollinearity is present, due to the fact that its design does not take potential multicollinearity into account.

The coefficients for drawing regressors from a $t_{(3)}$ distribution are insignificant for ridge-1, RLS, GMEN and Leuven-2 estimators, showing that the MSEL of these estimators is unaffected by the presence of a $t_{(3)}$ data distribution. They are significant at the 5% level for GMEW, DDE1, DDE2 and Leuven-1, with

GMEW, DDE1 and DDE2 being positive, and Leuven-1 being negative (Table 4.16). It is likely that the wider tails of this distribution, and thus the prevalence of values further from the mean, misguides parameter identification in the GMEW estimator. This effect may be negated in GMEN due to its narrow bounds. The coefficients are negative and significant at the 10% level for ridge-2, and positive at the 10% significance level for the OLS estimator.

When looking at the coefficients for drawing regressors from a chi-square distribution, we find them to be insignificant for a majority of the models, including the ridge-1, ridge-2, GMEN, GMEW, DDE2, DDE1, and Leuven-2 estimators (Table 4.16). The coefficient is negative and significant at the 5% level for RLS and Leuven-1 estimators, and positive and significant at the 10% level for OLS (Table 4.16). The latter shows that the ability of OLS to accurately predict the true coefficients is negatively impacted by the regressors being drawn from a chi-square distribution, while the RLS and Leuven-1 estimators are more accurate in this situation.

The coefficients for the use of mixed Gaussian errors are insignificant for ridge-1, ridge-2, GMEN, GMEW, Leuven-1, and Leuven-2 estimators (Table 4.16). They are positive and significant at the 5% level for OLS, DDE2 and DDE1, and positive and significant at the 10% level for RLS. This shows us that the MSEL values for OLS, DDE2, RLS, and DDE1 would all be expected to increase when the errors are drawn from a mixed Gaussian distribution, making them less able to accurately predict the true coefficients. The other estimators are immune to this problem. This demonstrates the difficulties caused when corrupt sampling distributions are used, with OLS, DDE2, RLS, and DDE1 being unable to

accurately handle this case. However, this result is considerably different than when we look at the Response Surfaces for MSE (Section 4.3.1), where the mixed Gaussian coefficient is significant at the 5% level for all estimators. This difference implies that corrupted sampling is more of an issue when making predictions, than when it comes to precision risk.

When looking at errors drawn from a $t_{(3)}$ error distribution, the coefficient is positive and significant at the 5% level for GMEN, GMEW, DDE2 and DDE1 (Table 4.16). Potentially, this is because the thicker tails of the $t_{(3)}$ error distribution place some of the error values outside of the specified range (GMEN, GMEW) or the range calculated during bound estimation (DDE2, DDE1), thus causing these estimators to be less able to predict true coefficients.

The coefficient for errors bring drawn from a $t_{(3)}$ distribution is negative and significant at the 5% level for RLS and Leuven-1 estimators, as well as being positive and significant at the 10% level for the ridge-1 estimator (Table 4.16). MSEL decreasing in the presence of a $t_{(3)}$ error distribution for RLS is expected and has been found before (Golan et al., 1996). OLS, ridge-2, and Leuven-2 estimators are immune to any impact in this case. The Leuven-2 estimator is immune to any impact here because the entropy term is extended to its error as well its data term, thus enabling it to better handle the thick tails of the $t_{(3)}$ distribution.

Most of the chi-square error distribution coefficients are insignificant (OLS, ridge-1, ridge-2, GMEW, DDE2, and Leuven-2). However, the coefficients for RLS, GMEN, and the Leuven-1 estimator are negative and significant at the 5%

level, while the coefficient for DDE1 is positive and significant at the 10% level. This shows that RLS, GMEN and Leuven-1 estimators are all more able to predict the true coefficients when the errors are drawn from a chi-square distribution, whereas DDE1 is less able.

Overall, these results show that the Leuven-2 estimator is the most robust of the estimators studied when looking at MSEL values (the ability of the estimator to accurately predict true coefficients). This estimator is not significantly affected by the choice of error or data distributions used (of those tested) or by the level of multicollinearity. This allows the Leuven-2 estimator to be confidently applied in situations where the actual distributions of errors or data are unknown or when multicollinearity is involved, at least based on the results of this assessment.

The R^2 is very high (above 0.9) for OLS and RLS, and moderately high (above 0.7) for ridge-2, GMEN, DDE2, DDE1, Leuven-1, and Leuven-2. However, the R^2 is much lower for ridge-1 and GMEW, with R^2 values of 0.19 and 0.47, respectively (Table 4.16). This highlights that the ability of the estimators, except ridge-1 and GMEW, to accurately predict true coefficients is strongly related to condition number and the distributions of the regressors and error terms. It therefore infers that if these factors are accounted for in a regression, we would still expect large amounts of variability from ridge-1 and GMEW in the presence of multicollinearity.

Chapter 5

Conclusion

5.1 Motivation and Method

Stochastic linear inverse problems are prevalent in Economics. A common form of ill-conditioning in these problems is multicollinearity. Multicollinearity causes issues interpreting regression results due to interdependency between parameters (Alin, 2010; Murray, 2006). It also makes estimates and model parameters less reliable through creating large standard errors and inflated variances (Farrar & Glauber, 1967; Greenberg & Parks, 1997; Kiers & Smilde, 2007; Lauridsen & Mur, 2006; Meloun et al., 2002; Stewart, 1987; Thomas, 1993). There are many forms of regression that attempt to deal with the multicollinearity problem, and this thesis has aimed to compare and contrast the performance of the main estimators used within Economics to deal with multicollinearity, as well as some new estimators that have received relatively little testing compared to established techniques.

The estimators which have been assessed here are: Ordinary Least Squares (OLS) (Campbell & Hill, 2005), ridge regression with the ridge parameter computed in two different ways (ridge-1, ridge-2) (Hoerl & Kennard, 1970; Mittelhammer et al., 2000), Restricted Least Squares (RLS) (Akdeniz et al., 2011), Generalised Maximum Entropy (GME) (Golan et al., 1996) with both narrow (GMEN) and wide support bounds (GMEW), a Two-Step Data Driven Entropy estimator (DDE2) (Doole, 2013), a One-Step Data Driven Entropy estimator (DDE1), a Leuven-1 regression (Paris, 2001, 2004), and a Leuven-2 regression (Paris, 2004).

Through the use of non-trivial, multi-factorial Monte Carlo experiments, these estimators have been tested across 10 different levels of multicollinearity, using data and errors drawn from different distributions (normal, $t_{(3)}$, $\chi^2_{(5)}$, and in the

case of errors, mixed Gaussian) to test robustness. The effectiveness of the alternative estimators in the presence of multicollinearity was tested through the comparison of their empirical prediction risk assessed through the mean squared error (MSE) statistic. This showed their ability to reliably estimate a dependent variable, based on the observed data. They were also tested according to their empirical precision risk, which was assessed through the mean squared error loss (MSEL) statistic. This indicates the ability of an estimator to predict the true parameters in an experiment, based upon the data. Response surface analysis (Campbell & Hill, 2005; Davidson & MacKinnon, 1993) was used to summarise the MSE and MSEL outputs obtained from the Monte Carlo experiments.

5.2 Key findings

The MSE results showed that the most accurate estimator for making predictions based on the data was the OLS estimator. This was the case regardless of the level of multicollinearity present or the combination of data and error distributions which were used. This was generally unsurprising, as OLS is structured specifically to minimise MSE in its objective function. This also ties in with theory that outlines that it is the precision of parameter identification that is affected by multicollinearity, not estimation (Mittelhammer et al., 2000). The highest values of MSE were generally given by either GMEN or DDE1. This illustrates the problems associated with estimators that attempt to combat multicollinearity through broad reformulation of the regression problem to one that does not involve the MSE criterion specifically. However, all of the MSE values produced in these experiments were relatively low, and the differences

between estimators were hence relatively small. Accordingly, it was not established that any estimators were completely unsuitable for making predictions of the dependent variable in the presence of multicollinearity.

The response surface results for MSE revealed that the predictive ability of OLS, GMEN, GMEW, DDE1, Leuven-1, and Leuven-2 was invariant to the presence of multicollinearity. As such, a researcher who is only interested in prediction and uses one of these aforementioned models, need not be too concerned about multicollinearity, at least based on the results of this assessment. These results also showed that the MSE of many of the estimators (namely OLS, ridge-1, ridge-2, RLS, and GMEW) were invariant to the data distribution which was used, establishing these estimators as useful for prediction when the data-generating distribution is unknown.

While the Leuven-2 estimator appears to be the estimator which is the most robust to changes in the error distribution, it also is the most variable with respect to this factor, as demonstrated in the standard deviation for the response surface results for MSE. This, combined with its low R^2, indicates that factors other than those analysed here account for variation in its MSE results. The implication of this is that while Leuven-2 is robust to the changes here, its variance is caused by factors not studied here. As such, without knowledge as to the factors which are causing the variance, and their relevance to the data set being studied, the use of the Leuven-2 estimator would not be advisable.

The MSEL results were far more variable than the MSE results in this study, illustrating that multicollinearity has a larger effect on parameter identification than prediction, as expected (Mittelhammer et al., 2000). When no

multicollinearity was present, all of the MSEL values were low, indicating that all of these estimators were appropriate for predicting the true coefficients when there was no multicollinearity in the data. When multicollinearity was present, two estimators stood out as being very ineffective at predicting the true parameters based on the data, these being DDE2, and OLS. This result shows the damaging effect that multicollinearity has on bounded set regression, which is used to identify parameter supports within DDE2. These results also reinforce the well-known inability of OLS to accurately predict parameters in the presence of multicollinearity. The extreme MSEL values of OLS and DDE2 at high levels of multicollinearity showed these two estimators to be entirely inappropriate for estimation in its presence.

Two estimators stood out in their ability to accurately predict the true coefficients in the presence of multicollinearity, and thus consistently produced the lowest MSEL values. These two estimators were the GMEN estimator, and the Leuven-1 estimator. Neither of these estimators was consistently better than the other, with fluctuation between which of the two produced the lowest MSEL values. It should, however, be noted that the GMEN estimator produced a marginally higher average MSEL value (0.05) than a majority of the estimators (0.03) in the presence of no multicollinearity. Moreover, this result relies heavily on very informed sets of bounds being set for the entropy estimator, which Leuven-1 did not require.

The response surface results for the MSEL regressions indicated that the GMEN, GMEW, DDE1, Leuven-1, and Leuven-2 estimators are all invariant to the condition number of the data being analysed, showing that a researcher interested

in determination of the true coefficients within a regression can reliably use these methods for estimation. Moreover, all of the estimators studied here were invariant to at least one type of data or error distribution. However, the Leuven-2 estimator was shown to be the most robust, with it being invariant to different specifications of the data and error generating distributions studied here, as well as the condition number. This allows the Leuven-2 estimator to be confidently applied in situations where the actual distribution of errors or data are unknown, or in the presence of multicollinearity, at least based on the results of this experiment. The Leuven-2 estimator is far more suitable for determining the true coefficients than for making predictions. The flexibility that this model has due to the presence of the entropy term in its error as well as data terms causes its predictions to be more variable. However this also allows the Leuven-2 estimator the flexibility to better handle changes in error distributions while determining true coefficients, leading to the robustness discussed above.

Based on the research conducted here, the Leuven-1 estimator (Paris, 2001, 2004) was found to overall be the best estimator to use in the presence of multicollinearity. The Leuven-1 performed reasonably, although not the best in the MSE results. However, it is still appropriate to be used as none of the MSE results were high enough to suggest unsuitability for prediction. In the MSEL case, the Leuven-1 estimator performed on par with the GMEN estimator, despite not having to use exogenous information in its formulation, thus making it more applicable to situations where this exogenous information is not available, and making it more comparable across studies.

5.3 Limitations and further research

As in any research, there are limitations and these provide insight into valuable areas for further research.

The RLS model is designed to be able to handle multicollinearity; however, this research showed RLS to be insufficient in most cases, though better than OLS. However, only one set of bounds for RLS were tested here, and as these bounds are determined exogenously, it would be interesting to see the effect of different sets of bounds on the effectiveness of the RLS estimator as a method for dealing with multicollinearity.

A second limitation is that in this research only the coefficient supports for the GME estimators were varied. It is known that the GME estimator is sensitive to changes in error supports as well (Paris & Caputo, 2001). As such, varying of these supports from the three-sigma rule may give a better insight into the potential variability of the GME estimator to differing levels of multicollinearity. Varying the error support may enable more of the true error values to be contained within the supports, potentially improving the robustness of the GME estimator.

A third limitation of the research is that a single sample size (n=50) was used for all of the Monte Carlo simulations because the main focus of this research was on different multicollinearity levels. Thus, extending this analysis to consider multiple sample sizes would be of interest.

A fourth limitation is that only 100 Monte Carlo simulations were run. It would be interesting to perform this experiment with a greater number of runs, though the impact is likely small.

A fifth limitation is that only a fixed number of coefficients have been tested. As part of the numerical experiments conducted here, the estimators were required to predict the true values of five coefficients, four of which were the same as those defined by Golan et al. (1996), Campbell & Hill (2005), and Doole (2013). As such, conducting experiments requiring the prediction of more or less coefficients may affect the abilities of various estimators. This would be an interesting area of research, as the relationship between parameter number and the relative value of alternative estimators is unknown.

A sixth limitation is that this research tests the abilities of two new estimators: DDE2 and DDE1. Due to their recency, it is possible that they are not yet specified in a manner that will cause them to perform to the best of their abilities. It would be interesting to perform more work, particularly with DDE1, in this area.

5.4 Final word

Overall, these results highlight that the Leuven-1 estimator is the most appropriate if a practitioner wishes to achieve high prediction accuracy and precision. In contrast, the GMEN model requires exogenous information and, as such, is much more problematic to accurately apply in different contexts. Nevertheless, it is critical that more attention is paid to the theoretical basis of the Leuven-1 estimator, as relating the definition of estimated coefficients back to the theory of light seems arbitrary. In fact if this relation were suitable it may be expected for Leuven-2 to be a more effective estimator than Leuven-1, instead of being more variable as seen here. This makes it possible that the Leuven-1 estimator has

simply performed well by chance and because its error specification is more flexible than in the Leuven-2 formulation. As such, a definitive answer regarding the best estimator to apply in the presence of multicollinearity ideally requires further testing, utilising diverse real and synthetic data sets. Thus, this is an important area for further research.

Literature Cited

Ahn, S. C., Perez, M. F., & Gadarowski, C. (2013). Two-pass Estimation of Risk Premiums with Multicollinear and Near-invariant Betas. Journal of Empirical Finance, 20, 1–17.

Agudo, L. F., & Gimeno, L. A. V. (2005). Effects of Multicollinearity on the Definition of Mutual Funds' Strategic Style: the Spanish case. Applied Economics Letters, 12, 553–556.

Akdeniz, F., Çabuk, A., & Güler, H. (2011). Generalized Maximum Entropy Estimators: Applications to the Portland Cement Dataset. The Open Statistics & Probability Journal, 3, 13–20.

Alauddin, M., & Nghiem, H. S. (2010). Do Instructional Attributes Pose Multicollinearity Problems? An Empirical Exploration. Economic Analysis & Policy, 40, 351–361.

Alin, A. (2010). Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics, 2, 370–374.

Beaver, S. E. (1974). The Effects of Causal Feedback on Ordinary Least-Squares Estimators. Sociological Methods & Research, 3, 189–208.

Belsley, D. A. (1991). Conditioning Diagnostics Collinearity and Weak Data in Regression. New York: Wiley-Interscience.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression Diagnostics. New York: John Wiley & Sons.

Bloomfield, P., & Steiger, W. (1980). Least Absolute Deviations Curve-Fitting. SIAM Journal on Scientific and Statistical Computing, 1, 290–301.

Bowles, S., & Levin, H. M. (1967). More on Multicollinearity and the Effectiveness of Schools. The Journal of Human Resources, 2, 393–400.

Buck, A. J., & Hakim, S. (1981). Appropriate Roles for Statistical Decision Theory and Hypothesis Testing in Model Selection. Regional Science and Urban Economics, 11, 135–147.

Burt, O. R. (1987). The Fallacy of Differencing to Reduce Multicollinearity. American Journal of Agricultural Economics, 69, 697–700.

Butler, M. R., & McNertney, E. M. (1991). Estimating Educational Production Functions: The problem of Multicollinearity. The Social Science Journal, 28, 489–499.

Campbell, R. C., & Hill, R. C. (2005). A Monte Carlo Study of the Effect of Design Characteristics on the Inequality Restricted Maximum Entropy Estimator. Review of Applied Economics, 1, 1–30.

Caputo, M. R., & Paris, Q. (2008). Comparative Statics of the Generalized Maximum Entropy Estimator of the General Linear Model. European Journal of Operational Research, 185, 195–203.

Cassidy, H. J. (1981). Using Econometrics: A Beginner's Guide. Reston VA: Reston Publishing Co.

Chennamaneni, P., Echambadi, R., Hess, J., & Syam, N. (2008). How Do You Properly Diagnose Harmful Collinearity in Moderated Regressions?

Ciavolino, E., & Al-Nasser, A. D. (2009). Comparing Generalised Maximum Entropy and Partial Least Squares Methods for Structural Equation Models. Journal of Nonparametric Statistics, 21, 1017–1036.

Cule, E., & De Iorio, M. (2013). Ridge Regression in Prediction Problems: Automatic Choice of the Ridge Parameter. Genetic Epidemiology, 37, 704–14.

Davidson, R., & MacKinnon, J. G. (1993). Estimation and Inference in Econometrics. New York: Oxford University Press.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. Journal of Marketing Research, 38, 269–277.

Doole, G. J. (2013). Robust Estimation of Support Bounds for the Generalised Maximum Entropy Estimator.

Ettarid, M. (1996). Variance Decomposition and its Application in Photogrammetry. International Archives of Photogrammetry and Remote Sensing, 31, 181–185.

Fabrycy, M. Z. (1973). Economic Theory and Multicollinearity. The American Economist, 17, 81–84.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. The Review of Economics and Statistics, 49, 92–107.

Feldstein, M. S. (1973). Multicollinearity and the Mean Square Error of Alternative Estimators. Econometrica, 41, 337–346.

Fourgeaud, C., Gourieroux, C., & Pradel, J. (1984). Some Theoretical Results for Generalized Ridge Regression Estimators. Journal of Econometrics, 25, 191–203.

Friendly, M., & Kwan, E. (2012). Where's Waldo? Visualizing Collinearity Diagnostics. The American Statistician, 63, 56–65.

Golan, A. (2006). Information and Entropy Econometrics - A Review and Synthesis. Foundations and Trends in Econometrics, 2, 1–145.

Golan, A., Judge, G., & Miller, D. (1996). Maximum Entropy Econometrics: Robust Estimation with Limited Data. New York: Wiley.

Golan, A., Judge, G., & Perloff, J. (1997). Estimation and Inference with Censored and Ordered Multinomial Response Data. Journal of Econometrics, 79, 23–51.

Golan, A., & Perloff, J. (2002). Comparison of Maximum Entropy and Higher-Order Entropy Estimators. Journal of Econometrics, 107, 195–211.

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. American Society for Quality, 21, 215–223.

Green, C. J., & Kiernan, E. (1989). Multicollinearity and Measurement Error in Econometric Financial Modelling. The Manchester School, 57, 357–369.

Greenberg, E., & Parks, R. P. (1997). A Predictive Approach to Model Selection and Multicollinearity. Journal of Applied Econometrics, 12, 67–75.

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing. Marketing Science, 23, 519–529.

Haitovsky, Y. (1969). Multicollinearity in Regression Analysis: Comment. The Review of Economics and Statistics, 51, 486–489.

Heckelei, T. (2002). Calibration and Estimation of Programming Models for Agricultural supply analysis. University of Bonn, Germany.

Hill, C., & Judge, G. (1987). Improved Prediction in the Presence of Multicollinearity. Journal of Econometrics, 35, 83–100.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12, 55–67.

Howitt, R. E., & Msangi, S. (2002). Reconstructing Disaggregate Production Functions. In Xth Congress of the European Association of Agricultural Economists.

Howitt, R. E., & Msangi, S. (2006). Estimating Disaggregate Production Functions: An Application to Northern Mexico. In Annual Meeting of the American Agricultural Economics Association.

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. Physical Review, 106, 620–630.

Johnston, J. (1984). Econometric Methods (3rd ed.). Singapore: McGraw-Hill.

Kalatzis, A. E. G., Bassetto, C. F., & Azzoni, C. R. (2011). Multicollinearity and Financial Constraint in Investment Decisions: a Bayesian Generalized Ridge Regression. Journal of Applied Statistics, 38, 287–299.

Kennedy, P. E. (1983). On an Inappropriate Means of Reducing Multicollinearity. Regional Science and Urban Economics, 13, 579–581.

Kiers, H. L., & Smilde, A. K. (2007). A Comparison of Various Methods for Multivariate Regression with Highly Collinear Variables. Statistical Methods and Applications, 16, 193–228.

Kreider, B., & Pepper, J. (2008). Inferring Disability Status from Corrupt Data. Journal of Applied Econometrics, 23, 329–349.

Kumar, K. T. (1975). Multicollinearity in Regression Analysis. The Review of Economics and Statistics, 57, 365–366.

Lauridsen, J., & Mur, J. (2006). Multicollinearity in Cross-Sectional Regressions. Journal of Geographical Systems, 8, 317–333.

Leamer, E. E. (1973). Multicollinearity: A Bayesian Interpretation. The Review of Economics and Statistics, 55, 371–380.

Leon, Y., Peeters, L., Quinqu, M., & Surry, Y. (1999). The Use of Maximum Entropy to Estimate Input-Output Coefficients from Regional Farm Accounting Data. Journal of Agricultural Economics, 50, 425–439.

Li, Y. F., Xie, M., & Goh, T. N. (2010). Adaptive Ridge Regression System for Software Cost Estimating on Multi-collinear Datasets. Journal of Systems and Software, 83, 2332–2343.

Liao, D., & Valliant, R. (2012). Condition Indexes and Variance Decompositions for Diagnosing Collinearity in Linear Model Analysis of Survey Data. Survey Methodology, 38, 189–202.

Maccini, L. J. (1981). Adjustment Lags, Economically Rational Expectations and Price Behaviour. The Review of Economics and Statistics, 63, 213–222.

Macedo, P., Scotto, M., & Silva, E. (2010). A General Class of Estimators for the Linear Regression Model Affected by Collinearity and Outliers. Communications in Statistics - Simulation and Computation, 39, 981–993.

Mahajan, V., Jain, A. K., & Bergier, M. (1977). Parameter Estimation in Marketing Models in the Presence of Multicollinearity: An Application of Ridge Regression. Journal of Marketing Research, XIV, 586–592.

Mason, R., & Brown, W. G. (1975). Multicollinearity Problems in Sociological and Ridge Regression. Social Science Research, 4, 135–149.

Meloun, M., Militký, J., Hill, M., & Brereton, R. G. (2002). Crucial Problems in Regression Modelling and their Solutions. The Analyst, 127, 433–450.

Mishra, S. K. (2004). Multicollinearity and Maximum Entropy Leuven Estimator. Economics Bulletin, 3, 1–11.

Mittelhammer, R. C., Judge, G. G., & Miller, D. J. (2000). Econometric Foundations. Cambridge: Cambridge University Press.

Mittelhammer, R., Cardell, N., & Marsh, T. (2013). The Data-Constrained Generalized Maximum Entropy Estimator of the GLM: Asymptotic Theory and Inference. Entropy, 15, 1756–1775.

Mittelhammer, R. C., Young, D. L., Damrongsak, T., & Donnelly, J. T. (1980). Mitigating the Effects of Multicollinearity Using Exact and Stochastic Restrictions: The Case of an Aggregate Agricultural Production Function in Thailand. American Journal of Agricultural Economics, 62, 199–210.

Murray, M. P. (2006). Econometrics A Modern Introduction. Boston: Addison-Wesley.

O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity, 41, 673–690.

Paris, Q. (2001). Multicollinearity and Maximum Entropy Estimators. Economics Bulletin, 3, 1–9.

Paris, Q. (2004). Maximum Entropy Leuven Estimators and Multicollinearity. Statistica, LXIV, 3–22.

Paris, Q., & Caputo, M. R. (2001). Sensitivity of the GME Estimates to Support Bounds. Department of Agricultural and Resource Economics, University of California.

Paris, Q., & Howitt, R. E. (1998). An Analysis of Ill-Posed Production Problems Using Maximum Entropy. Agricultural and Applied Economics Association, 80, 124–138.

Paul, R. K. (2008). Multicollinearity: Causes, Effects and Remedies. Indian Agricultural Statistics Research Institute, 1-14.

Rockwell, R. C. (1975). Assessment of Multicollinearity: The Haitovsky Test of the Determinant. Sociological Methods & Research, 3, 308–320.

Salinas, T. S., & Hillmer, S. C. (1987). Multicollinearity Problems in Modeling Time Series with Trading-Day Variation. Journal of Business & Economic Statistics, 5, 431–436.

Schroeder, M. A., Lander, J., & Levine-Silverman, S. (1990). Diagnosing and Dealing with Multicollinearity. Western Journal of Nursing Research, 12, 175–187.

Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27, 623–656.

Silvey, S. D. (1969). Multicollinearity and Imprecise Estimation. Journal of the Royal Statistical Society, 31, 539–552.

Stewart, G. W. (1987). Collinearity and Least Squares Regression. Statistical Science, 2, 68–84.

Stock, J. H., & Watson, M. W. (2007). Introduction to Econometrics. (D. Clinton, Ed., 2nd ed.). International Edition: Pearson Education.

Thomas, R. L. (1993). Introductory Econometrics: Theory and Applications (2nd ed.). New York: Longman Publishing.

Wan, A. T. K. (2002). On Generalized Ridge Regression Estimators under Collinearity and Balanced Loss. Applied Mathematics and Computation, 129, 455–467.

Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and Correlation among Local Regression Coefficients in Geographically Weighted Regression. Journal of Geographical Systems, 7, 161–187.

Wu, H. C. (2008). On Interval-Valued Nonlinear Programming Problems. Journal of Mathematical Analysis and Applications, 338, 299–316.

Wu, X. (2009). A Weighted Generalized Maximum Entropy Estimator with a Data-driven Weight. Entropy, 11, 917–930.

Yamagata, T., & Orme, C. D. (2005). On Testing Sample Selection Bias Under the Multicollinearity Problem. Econometric Reviews, 24, 467–481.

Appendix 1: GAMS code

The purpose of this appendix is to present the GAMS code used for the analysis in this thesis. The following GAMS code is split into two parts, each part preceded by its file name. The first file is used for importing and correctly structuring the data from Microsoft Excel to enable its use by GAMS. The second file is used for setting up the estimators and running the simulations. The italicised sections in this appendix are explanations of the purpose of the GAMS code which follows (in standard font).

A1.1 Importing of the data from Microsoft Excel.

File name: genST50

The following piece of code creates sets which can be used to store data, enabling reference to a set of similar data as opposed to a single data point.

```
sets
nr          number of regression types/1*9/
sc          number of supports for coefficients/1*5/
sd          number of supports for error terms/1*5/
ns          number of simulations/1*100/
nm          number of multicollinearity scenarios/1*10/
nb          number of data generation scenarios/1*3/
ne          number of error generation scenarios/1*4/
nv          number of variables/1*6/
np(nv)      number of dependent variables/1*1/
nk(nv)      number of explanatory variables/2*6/
```

nd number of data points/1*50/

lu sum of lower and upper bounds/1*2/;

In the created sets nr=1 is the Ordinary Least Squares regression, nr=2 is Ridge Regression 1, nr=3 is Ridge Regression 2, nr=4 is a regression using Restricted Least Squares, nr=5 is a standard Generalised Maximum Entropy regression, nr=6 is the Two-Step Data Driven Entropy estimator, nr=7 is the One-Step Data Driven Entropy estimator, nr= 8 is the Leuven-1 estimator, and nr=9 is the Leuven-2 estimator.

The following creates a matrix called "data" with the indices "ns, nm, nb, ne, nv and nd", which are explained above, and creates a data matrix of this information of the appropriate size. It then uses the GDXXRW utility to read the data in Excel format from the specified file and uses it to fill a data matrix within a GDX file.

parameter data(ns,nm,nb,ne,nv,nd) data matrix for a given sample size

$Call GDXXRW.exe "C:\Users\Luke\Dropbox\Masters\Appendix 1/ST50.xlsx" par=data rng=Sheet1!a1:bc72001 Rdim=5 Cdim=1

The following opens the previously created GDX file and loads the data matrix from there into the matrix object "data" within GAMS.

$GDXIN ST50.gdx

$LOAD data

$GDXIN

A1.2 DDE ST50 Programming for Monte Carlo simulations

File Name: DDE_ST50

Overall, the following section of code performs Monte Carlo simulations (using the data previously imported from Excel) for an Ordinary Least Squares regression, Two Ridge Regression models which calculate the ridge parameter differently, Restricted Least Squares, Generalised Maximum Entropy, Two-Step Data Driven Entropy, One-Step Data Driven Entropy, Leuven-1, and Leuven-2 estimators.

The following suppresses the reference map (list of all the named items i.e. sets, parameters etc as well as identifying them by type etc) and the symbol list (list of

all named items along with any text to explain them). This is important to reduce the size of the output file, so that the Monte Carlo simulation does not make it too large to read.

$offsymxref offsymlist

The first two commands define which solver to select in a given scenario, in this case for linear programming to use the conopt solver and for nonlinear programming to also use the conopt solver. Solprint=off stops the listing of solutions after solving, and solslack=1 allows the inclusion of "slacks" in output (i.e. the "slack" for the equation AX≤b is b-AX). The slack option is used, as this is easier to interpret than the default method used in GAMS.

options lp=conopt, nlp=conopt, solprint=off, solslack=1, limrow=0, limcol=0;

Creates an alias for the ns, nd, and nk sets to allow the use of a set more than once in a single statement without causing a circularity problem.

alias (ns,ss);

alias (nd,dd);

alias (nk,ak);

Names all parameters before regressions are performed.

parameters

bt(nk) true values of coefficients

States which variables are the dependent or explanatory variables for the regressions.

y(ns,nm,nb,ne,nd) dependent variable

x(ns,nm,nb,ne,nk,nd) explanatory variables

Constructs an individual data matrix for a given Monte Carlo simulation.

xs(nk,nd data matrix within a given simulation

Sets the dependent variable for a given Monte Carlo simulation.

ys(nd) dependent variable within a given simulation

Creates the supports for coefficients and error terms for use in the entropy models.

z(nk,sc) supports for coefficients in entropy model

v(nd,sd) supports for error terms in entropy model

Creates a parameter for the error support bound.

q error support bound

Defines a parameter that sets the number of standard deviations of the dependent variable to be used in the generation of the error support bounds.

nes number of standard deviations of dependent variable used to generate error supports

Creates a parameter for the coefficient support to be used in the Maximum Entropy regression.

csu coefficient support for entropy regression

Creates a parameter for the bound for the explanatory variables.

bb bound for explanatory variables;

Sets the true values for the various β's (coefficients) at -4, 2, 1, -3, and 2.

bt("2")=-4;bt("3")=2;bt("4")=1;bt("5")=-3; bt("6")=2;

Sets the explanatory variable bound at 5. This is used in the Restricted Least Squares regression, and states for the purpose of that regression that the coefficients are all between -5 and 5. This follows the procedure used by Golan et al. (1996).

bb=5;

Sets the number of standard deviations of the dependent variable used in error support generation at 3. This is due to the use of the 3σ rule in error support generation (see Sections 2.5 and 3.4).

nes=3;

Sets the initial coefficient support used in the entropy regressions at 5. This is based upon the practitioner broadly knowing the correct estimates, and is a standard means of deciding these supports based on Golan et al. (1996). After the first run this was changed to 20 to test the difference between wide and narrow bounds.

csu=5;

Assigns values from the matrix "data" to each set of the dependent and explanatory variables.

y(ns,nm,nb,ne,nd)=data(ns,nm,nb,ne,"1",nd);

x(ns,nm,nb,ne,nk,nd)=data(ns,nm,nb,ne,nk,nd);

The next stage of coding defines each of the regression models examined in this study.

The first set of text sets up the Ordinary Least Squares regression.

The following creates three variables: the total sum of squares, coefficient estimates, and error estimates in the Ordinary Least Squares regression. These variables are all set as free variables, this allows them to be either positive or negative.

free variables

tols total sum of squares for ols

bols(nk) estimated coefficients for ols

eols(nd) error estimates for ols;

The following section creates the equations used in the Ordinary Least Squares regression.

equations

POLS minimise TOLS

COLS(nd) data constraint;

The following section defines the equations used in the Ordinary Least Squares estimation. For more information on any of the equations mentioned in this appendix, see the Methods section (Section 3.1-3.8).

POLS.. tols=e=sum(nd,power(eols(nd),2));

COLS(nd).. ys(nd)=e=sum(nk,bols(nk)*xs(nk,nd))+eols(nd);

The following line tells GAMS that the Ordinary Least Squares model requires two equations, these being POLS and COLS as defined above.

model MOLS/pols,cols/;

The following section of code describes the Ridge Regression models.

Creates a set of parameters for use in the Ridge Regression model. These are required since the Ridge Regression requires computations not present in standard Ordinary Least Squares.

parameters

kk number of explanatory variables

my mean of dependent variable

sy standard deviation of dependent variable

vy variance of dependent variable

rd ridge parameter in first ridge regression

rd2 ridge parameter in second ridge regression

ess explained sum of squares

rss residual sum of squares

fstat f-statistic;

Creates three new variables. These are: total sum of squares for the Ridge Regression, the estimated coefficients for the Ridge Regression, and the error estimates for the Ridge Regression. These variables are once again all "free" variables, as described previously.

free variables

trid total sum of squares for ridge regression

brid(nk) estimated coefficients for ridge regression

erid(nd) error estimates for ridge regression;

Creates two equations for use in the Ridge Regression estimation.

equations

PRID minimise objective function in first ridge regression

PRID2 minimise objective function in second ridge regression

CRID(nd) data constraint;

Defines the equations used in both ridge regression models.

PRID.. trid=e=sum(nd,power(erid(nd),2))+rd*sum(nk,power(brid(nk),2));

PRID2.. trid=e=sum(nd,power(erid(nd),2))+rd2*sum(nk,power(brid(nk),2));

CRID(nd).. ys(nd)=e=sum(nk,brid(nk)*xs(nk,nd))+erid(nd);

Tells GAMS that the Ridge Regression 1 model requires the two previously defined equations; prid and crid.

model MRID1/prid,crid/;

Tells GAMS that the Ridge Regression 2 model requires the two previously defined equations; prid2 and crid.

model MRID2/prid2,crid/;

This section describes the Generalised Maximum Entropy model.

This creates one new free variable: Total entropy.

free variable

ten total entropy;

The two variables created below are defined as positive variables. This means that the values of these variables must never be less than 0. This is intuitive since probabilities can never be negative.

positive variables

p(nk,sc) probabilities for parameter coefficients in entropy model

w(nd,sd) probabilities for error terms in entropy model

The following creates the four equations required for the running of the Generalised Maximum Entropy model.

equations

sel objective function to maximise entropy

cel(nd) data constraints in entropy model

pel(nk) normalisation constraint for parameter coefficients

wel(nd) normalisation constraint for error coefficients;

The following defines the equations required for the running of the Generalised Maximum Entropy model.

sel.. ten =e= sum{nk,sum[sc,p(nk,sc)*log(p(nk,sc)+1.e-8)]}+sum{nd,sum[sd,w(nd,sd)*log(w(nd,sd)+1.e-8)]};

cel(nd).. ys(nd) =e= sum{nk,xs(nk,nd)*sum[sc,p(nk,sc)*z(nk,sc)]}+sum{sd,w(nd,sd)*v(nd,sd)};

pel(nk).. sum{sc,p(nk,sc)} =e= 1;

wel(nd).. sum{sd,w(nd,sd)} =e= 1;

The following tells GAMS that all four equations sel, cel, pel and wel are required to run the Generalised Maximum Entropy model.

model MENT/sel,cel,pel,wel/;

The following section defines the set up for the Two-Step Data Driven Entropy estimator.

The following creates three new free variables: the estimated coefficients of the lower regression line, the estimated coefficients of the upper regression line, and the total sum of deviations for both the lower and upper regression lines.

free variables

bl(nk) estimated coefficients for the lower regression line

bu(nk) estimated coefficients for the upper regression line

tsl total sum of deviations for the lower and upper regression lines;

The following creates two new variables which are again defined positive variables, restricting the values of these variables to those which are non-negative.

positive variables

dl(nd) negative deviation for the lower regression line (negativity established in equation definition)

du(nd) positive deviation for the upper regression line;

Creates the three equations required for BSR estimation in the Two-Step Data Driven Entropy Estimator.

equations

sbd minimise absolute deviations

lbl(nd) compute regression line below the data

hbl(nd) compute regression line above the data;

Defines the equations required for the Data Driven Entropy Estimator.

sbd.. tsl=e=sum(nd,dl(nd)+du(nd));

lbl(nd).. ys(nd) =e= sum(nk,xs(nk,nd)*bl(nk))-dl(nd);

hbl(nd).. ys(nd) =e= sum(nk,xs(nk,nd)*bu(nk))+du(nd);

Tells GAMS that the three equations; sbd, lbl, and hbl; are required for running the BSR estimation in the Two-Step Data Driven Entropy Estimator.

model MBSR/sbd,lbl,hbl/;

The following section defines the information and equations required for the running of the One-Step Data Driven Entropy Estimator.

Creates four new free variables.

free variables

zz(nk,sc) variable representing support for coefficients

vv(nd,sd) variable representing support for errors

bld(nk) lower bound of coefficients in one-step model

bud(nk) upper bound of coefficients in one-step model;

Creates five new variables which are defined as positive and so are restricted to non-negative values.

positive variables

abl(nk) absolute value of lower bound coefficients

abu(nk) absolute value of upper bound coefficients

dnd(nd) lower bound deviation in one-step model

dpd(nd) upper bound deviation in one-step model

dxd(nd) absolute value of largest deviation;

The following stage creates the fourteen equations which are required to run the One-Step Data Driven Entropy estimator.

equations

obd objective function to minimise total entropy formulation

cd1(nd) data constraint

cd2(nd) lower deviation constraint

cd3(nd) upper deviation constraint

cd4(nk,sc) define first support for each coefficient

cd5(nk,sc) define interior supports for each coefficient

cd6(nk,sc) define last support for each coefficient

cd7(nd,sd) define first support for each error

cd8(nd,sd) define interior supports for each error

cd9(nd,sd) define last support for each error

cd10(nk) define absolute value of lower bound coefficient

cd11(nk) define absolute value of upper bound coefficient

cd12(nk) normalise probabilities for each coefficient

cd13(nd) normalise probabilities for each error term;

Defines the values of the previously created equations to allow the running of the One-Step Data Driven Entropy Estimator.

obd.. ten =e= sum{nk,sum[sc,p(nk,sc)*log(p(nk,sc)+1.e-8)]}+sum{nd,sum[sd,w(nd,sd)*log(w(nd,sd)+1.e-8)]}+sum{nd,dnd(nd)*log(dnd(nd)+1.e-8)}+sum{nd,dpd(nd)*log(dpd(nd)+1.e-8)}+sum{nk,abl(nk)*log(abl(nk)+1.e-8)}+sum{nk,abu(nk)*log(abu(nk)+1.e-8)};

```
cd1(nd)..        ys(nd) =e=
sum{nk,xs(nk,nd)*sum[sc,p(nk,sc)*zz(nk,sc)]}+sum{sd,w(nd,sd)*vv(nd,sd)};

cd2(nd)..        ys(nd) =e= sum{nk,xs(nk,nd)*bld(nk)}-dnd(nd);

cd3(nd)..        ys(nd) =e= sum{nk,xs(nk,nd)*bud(nk)}+dpd(nd);

cd4(nk,"1")..   zz(nk,"1") =e= bld(nk);

cd5(nk,sc)$(ord(sc) gt 1 and ord(sc) lt card(sc))..    zz(nk,sc) =e=
zz(nk,"1")+(ord(sc)-1)*(bud(nk)-bld(nk))/(card(sc)-1);

cd6(nk,sc)$(ord(sc) eq card(sc))..     zz(nk,sc) =e= bud(nk);

cd7(nd,"1")..   vv(nd,"1") =e= -dnd(nd);

cd8(nd,sd)$(ord(sd) gt 1 and ord(sd) lt card(sd))..    vv(nd,sd) =e=
vv(nd,"1")+(ord(sd)-1)*(dnd(nd)+dpd(nd))/(card(sd)-1);

cd9(nd,sd)$(ord(sd) eq card(sd))..     vv(nd,sd) =e= dpd(nd);

cd10(nk)..       abl(nk)=e=sqrt(power(bld(nk),2)+power(0.0001,2));

cd11(nk)..       abu(nk)=e=sqrt(power(bud(nk),2)+power(0.0001,2));

cd12(nk)..       sum{sc,P(nk,sc)} =e= 1;

cd13(nd)..       sum{sd,W(nd,sd)} =e= 1;
```

Tells GAMS that the One-Step Data Driven Entropy estimator requires the use of the fourteen equations defined above.

```
model DDE1/obd,cd1,cd2,cd3,cd4,cd5,cd6,cd7,cd8,cd9,cd10,cd11,cd12,cd13/;
```

The following section defines the set up for the Leuven 1&2 estimators.

The next section sets up both the Leuven 1 and the Leuven 2 estimators.

Firstly, three free variables (unrestricted variables that can take on any values) are created.

```
free variable

tenl          total entropy

cl(nk)        cofficient matrix for leuven
```

er(nd) error;

Four positive variables are then created, positive variables that must at all times remain non-negative.

positive variables

pl(nk) probabilities for coefficients

wl(nd) probabilities for errors

Lp leuven variable for coefficient

Lw leuven variable for each error;

The following creates the seven equations which are required to allow the running of Leuven 1 and Leuven 2 estimators.

equations

sel1 objective function to minimise in Leuven 1 model

sel2 objective function to minimise in Leuven 2 model

cell(nd) data constraints in entropy model

del define leuven variable for coefficient

dep(nk) define probabilities for each coefficient

rel define leuven variable for residual

rep(nd) define probabilities for each residual;

The following defines the equations created above to allow the running of the Leuven 1 and Leuven 2 estimators.

sel1.. ten =e= sum{nk,p(nk)*log(p(nk)+1.e-4)}+{Lp*log(Lp+1.e-4)}+sum(nd,power(er(nd),2));

sel2.. tenl =e= sum{nk,pl(nk)*log(pl(nk)+1.e-8)}+{Lp*log(Lp+1.e-8)}+sum{nd,wl(nd)*log(wl(nd)+1.e-8)}+(Lw*log(Lw+1.e-8));

cell(nd).. ys(nd) =e= sum{nk,xs(nk,nd)*cl(nk)}+er(nd);

del.. Lp=e=sum(nk,power(cl(nk),2));

dep(nk).. pl(nk)=e=power(cl(nk),2)/Lp;

rel.. Lw=e=sum(nd,power(er(nd),2));

rep(nd).. wl(nd)=e=power(er(nd),2)/Lw;

The following tells GAMS that running the Leuven 1 estimator requires the use of the sel1, cel, del, and dep equations.

model MLE1/sel1,cell,del,dep/;

The following tells GAMS that the Leuven 2 estimator requires the use of the sel2, cel, del, dep, rel, and rep equations.

model MLE2/sel2,cell,del,dep,rel,rep/;

Sets the lower bound for the Leuven variable for coefficients to 1.e-5. This is necessary to prevent taking the log of a zero quantity.

Lp.lo=1.e-5;

Sets the lower bound for the Leuven variable for errors to 1.e-5. This is necessary to prevent taking the log of a zero quantity.

Lw.lo=1.e-5;

The following creates the structures to receive the output.

Creates the parameters to store the data computed in the Monte Carlo simulations.

parameters

Creates parameter eb that stores the estimated coefficients for each regression.

eb(nr,nk,ns,nm,nb,ne) estimated coefficients in each regression type

Creates parameter ee that stores the estimated errors for each regression type.

ee(nr,nd,ns,nm,nb,ne) estimated errors in each regression type

Creates parameter ey that stores the estimated dependent variables for each regression type.

ey(nr,nd,ns,nm,nb,ne) estimated dependent variable in each regression type

Creates a parameter that stores the feasibility of each regression type.

ef(nr,ns,nm,nb,ne) feasibility of each regression

Creates a parameter that stores the estimated bias for each regression type.

es(nr,nm,nb,ne) estimated bias in each regression type

Creates a parameter that stores the estimated squared bias for each regression type.

ses(nr,nm,nb,ne) estimated squared bias in each regression type

Creates a parameter that stores the estimated variance for each regression type.

ev(nr,nm,nb,ne) estimated variance in each regression type

Creates a parameter that stores the estimated mean squared error loss for each regression type.

emsel(nr,nm,nb,ne) estimated mean squared error loss in each regression type

Creates a parameter that stores the estimated mean squared error for each regression type.

emse(nr,nm,nb,ne) estimated mean squared error in each regression type

Creates a parameter that stores the feasibility of the first step in the Two-Step Data Driven Entropy estimator.

efBSR(ns) feasibility of bounded set regression

Creates a parameter that stores the estimated upper and lower bounds for the coefficients which were identified through use of the Two-Step Data Driven Entropy estimator.

ebbsr(lu,nk,ns,nm,nb,ne) estimated lower and upper bounds identified using BSR

Creates a parameter that stores the upper and lower bounds for the deviation terms (deviations between the observed data and estimated values) identified through the use of the Two-Step Data Driven Entropy estimator.

edbsr(lu,nd,ns,nm,nb,ne) estimated lower and upper bounds for deviations identified using BSR;

The next stage of coding involves the actual running of the models. This code creates a loop function, which runs all models where ns (number of simulations) is less than 101, nm (number of multicollinearity scenarios) is less than 11,

number of data generating scenarios is less than 4, and the error generating scenario is less than 5.

loop((ns,nm,nb,ne)$(ord(ns) lt 101 and ord(nm) lt 11 and ord(nb) lt 4 and ord(ne) lt 5),

The following sets the dependent variable parameter for each simulation equal to the dependent variables generated for that simulation.

ys(nd)=y(ns,nm,nb,ne,nd);

The following sets the data matrix for a given simulation to equal the data values generated for that simulation.

xs(nk,nd)=x(ns,nm,nb,ne,nk,nd);

The following sets the value for the mean of the dependent variables.

my=sum(nd,ys(nd))/card(nd);

The following sets the value for the standard deviation of the dependent variables.

sy=sqrt(sum(nd,sqr(ys(nd)-my))/(card(nd)-1));

The following solves the OLS Regression.

Sets lower and upper bounds for estimated Ordinary Least Squares coefficients at –infinity and infinity respectively.

bols.lo(nk)=-inf;

bols.up(nk)=+inf;

Tells GAMS to solve the Ordinary Least Squares model (mols) by minimising the total sum of squared residuals, using nonlinear programming.

solve mols minimising TOLS using nlp;

Stores the estimated coefficients.

eb("1",nk,ns,nm,nb,ne)=bols.l(nk);

Stores the estimated errors.

ee("1",nd,ns,nm,nb,ne)=eols.l(nd);

Stores the estimated dependent variable.

ey("1",nd,ns,nm,nb,ne)=sum(nk,xs(nk,nd)*bols.l(nk));

Stores the feasibility of the model, so it can be seen later whether the model solved effectively or otherwise.

ef("1",ns,nm,nb,ne)=mols.modelstat;

The following solves Ridge Regression 1.

The following code allows for the solving of the Ridge Regression model.

Firstly the Ridge parameter is computed.

Defines the number of explanatory variables as 5.

kk=5;

Sets the formula for working out the variance of the dependent variable.

vy=power(sy,2);

Calculates the ridge parameter.

rd=kk*vy/sum(nk,bols.l(nk)*bols.l(nk));

Solves the Ridge Regression 1 model by minimising the total sum of squares using nonlinear programming.

solve mrid1 minimising TRID using nlp;

Stores the estimated coefficients.

eb("2",nk,ns,nm,nb,ne)=brid.l(nk);

Stores the estimated errors.

ee("2",nd,ns,nm,nb,ne)=erid.l(nd);

Stores the estimated dependent variable.

ey("2",nd,ns,nm,nb,ne)=sum(nk,xs(nk,nd)*brid.l(nk));

Stores the feasibility of the model so it can be seen later whether the model solved effectively or otherwise.

ef("2",ns,nm,nb,ne)=mrid1.modelstat;

The following solves Ridge regression 2.

The first steps in this formulation involve an alternative method of computing the ridge parameter used, relative to the Ridge Regression 1 model.

Computes explained sum of squares for the Ridge Regression 2 formulation, based on the OLS estimation.

```
ess=sum(nd,power(ey("1",nd,ns,nm,nb,ne)-my,2));
```

Computes residual sum of squares for the Ridge Regression 2 formulation, based on the OLS estimation.

```
rss=sum(nd,power(eols.l(nd),2));
```

Sets the formula for working out the F statistic (based upon the explained sum of squares and residual sum of squares calculated above).

```
Fstat=(ess/4)/(rss/(card(nd)-5));
```

Displays the Fstat parameter for viewing.

```
display Fstat;
```

Sets the value of the new ridge parameter based upon the above calculations.

```
rd2=1/Fstat;

solve mrid2 minimising TRID using nlp;
```

Stores the estimated coefficients.

```
eb("3",nk,ns,nm,nb,ne)=brid.l(nk);
```

Stores the estimated errors.

```
ee("3",nd,ns,nm,nb,ne)=erid.l(nd);
```

Stores the estimated dependent variable.

```
ey("3",nd,ns,nm,nb,ne)=sum(nk,xs(nk,nd)*brid.l(nk));
```

Stores the feasibility of the model, so it can be seen later whether the model solved effectively or otherwise.

```
ef("3",ns,nm,nb,ne)=mrid2.modelstat;
```

The following solves the Restricted Least Squares regression.

Sets the lower and upper bounds for the estimated coefficients to –bb and bb respectively. This is currently set as 5 and -5; however, by using the variable bb, it enables it to be easily changed.

```
bols.lo(nk)=-bb;

bols.up(nk)=+bb;
```

Solves the Restricted Least Squares model thorough minimising the sum of squares using nonlinear programming.

```
solve mols minimising TOLS using nlp;
```

Stores the estimated coefficients.

```
eb("4",nk,ns,nm,nb,ne)=bols.l(nk);
```

Stores the estimated errors.

```
ee("4",nd,ns,nm,nb,ne)=eols.l(nd);
```

Stores the estimated dependent variable.

```
ey("4",nd,ns,nm,nb,ne)=sum(nk,xs(nk,nd)*bols.l(nk));
```

Stores the feasibility of the model, so it can be seen later whether the model solved effectively or otherwise.

```
ef("4",ns,nm,nb,ne)=mols.modelstat;
```

The following solves the Generalised Maximum Entropy regression.

The following code is used to solve the Generalised Maximum Entropy model.

First, it uses a loop to generate the support bounds for the estimated coefficients.

```
loop(nk,

z(nk,"1")=-csu;

z(nk,sc)$(ord(sc) eq card(sc))=csu;

loop(sc$(ord(sc) gt 1 or ord(sc) lt card(sc)),
```

```
z(nk,sc)=z(nk,"1")+(ord(sc)-1)*(csu+csu)/(card(sc)-1);

));
```

It next uses the 3σ rule to compute the error supports for the model.

```
q=nes*sy;

loop(nd,

v(nd,"1")=-q;

v(nd,sd)$(ord(nd) eq card(nd))=q;

loop(sd$(ord(sd) gt 1 or ord(sd) lt card(sd)),

v(nd,sd)=-q+(ord(sd)-1)*(q+q)/(card(sd)-1);

););
```

It then solves the maximum entropy model by maximising TEN (which is total entropy), again using nonlinear programming.

```
solve ment minimising TEN using nlp;
```

Stores the estimated coefficients.

```
eb("5",nk,ns,nm,nb,ne)=sum{sc,p.l(nk,sc)*z(nk,sc)};
```

Stores the estimated errors.

```
ee("5",nd,ns,nm,nb,ne)=sum{sd,w.l(nd,sd)*v(nd,sd)};
```

Stores the estimated dependent variable.

```
ey("5",nd,ns,nm,nb,ne)=sum(nk,xs(nk,nd)*sum{sc,p.l(nk,sc)*z(nk,sc)});
```

Stores the feasibility of the model so it can be seen later whether the model solved effectively or otherwise.

```
ef("5",ns,nm,nb,ne)=ment.modelstat;
```

The following solves the Two-Step Data Driven Entropy regression.

This solves the Two-Step Data Driven Entropy Estimator.

Sets the supports for coefficients and errors to 0, given that these are non-zero following their use in the maximum entropy model above.

z(nk,sc)=0;

v(nd,sd)=0;

Solves the first step of the 2 step Data Driven Entropy estimator by minimising the total sum of deviations using linear programming.

solve mbsr minimising TSL using lp;

Stores the feasibility of the model, so it can be seen later whether the model solved effectively or otherwise.

efBSR(ns)=mbsr.modelstat;

Stores the estimated coefficients for the lower and upper regression lines.

ebbsr("1",nk,ns,nm,nb,ne)=BL.L(nk);

ebbsr("2",nk,ns,nm,nb,ne)=BU.L(nk);

Sets the estimated lower bounds for the deviations to the negative maximum of the previously identified levels of the negative and positive deviations for the lower and upper regression line.

edbsr("1",nd,ns,nm,nb,ne)=-max(DL.L(nd),DU.L(nd));

Sets the estimated upper bounds for the deviations to the positive maximum of the previously identified levels of the negative and positive deviations for the lower and upper regression line.

edbsr("2",nd,ns,nm,nb,ne)=max(DL.L(nd),DU.L(nd));

Uses a loop to generate the support bounds for the estimated coefficients.

loop(nk,

z(nk,"1")=ebbsr("1",nk,ns,nm,nb,ne);

z(nk,sc)$(ord(sc) eq card(sc))=ebbsr("2",nk,ns,nm,nb,ne);

loop(sc$(ord(sc) gt 1 or ord(sc) lt card(sc)),

z(nk,sc)=z(nk,"1")+(ord(sc)-1)*(ebbsr("2",nk,ns,nm,nb,ne)-
ebbsr("1",nk,ns,nm,nb,ne))/(card(sc)-1);

));

Uses a loop to generate the support bounds for the errors.

loop(nd,

v(nd,"1")=edbsr("1",nd,ns,nm,nb,ne);

v(nd,sd)$(ord(nd) eq card(nd))=edbsr("2",nd,ns,nm,nb,ne);

loop(sd$(ord(sd) gt 1 or ord(sd) lt card(sd)),

v(nd,sd)=edbsr("1",nd,ns,nm,nb,ne)+(ord(sd)-
1)*(edbsr("2",nd,ns,nm,nb,ne)+edbsr("2",nd,ns,nm,nb,ne))/(card(sd)-1);

););

Uses the previous information to solve the Maximum Entropy estimator (part two of the Two-Step Data Driven Entropy estimator), by minimising total entropy using nonlinear programming.

solve ment minimising ten using nlp;

Stores the estimated coefficients.

eb("6",nk,ns,nm,nb,ne)=sum{sc,p.l(nk,sc)*z(nk,sc)};

Stores the estimated errors.

ee("6",nd,ns,nm,nb,ne)=sum{sd,w.l(nd,sd)*v(nd,sd)};

Stores the estimated dependent variable.

ey("6",nd,ns,nm,nb,ne)=sum(nk,xs(nk,nd)*sum{sc,p.l(nk,sc)*z(nk,sc)});

Stores the feasibility of the model, so it can be seen later whether the model solved effectively or otherwise.

ef("6",ns,nm,nb,ne)=ment.modelstat;

The following solves the One-Step Data Driven Entropy regression.

Solves the One-Step Data Driven Entropy estimator.

Sets the value of the starting level for the absolute value of the lower bound coefficient, the absolute value of the upper bound coefficient, the lower bound deviation, and the upper bound deviation.

abl.l(nk)=bols.l(nk)-1;

abu.l(nk)=bols.l(nk)+1;

dnd.l(nd)=eols.l(nd);

dpd.l(nd)=eols.l(nd);

Solve the DDE1 estimator by minimising total entropy using nonlinear programming.

solve DDE1 minimising ten using nlp;

Stores the estimated coefficients.

eb("7",nk,ns,nm,nb,ne)=sum{sc,p.l(nk,sc)*zz.l(nk,sc)};

Stores the estimated errors.

ee("7",nd,ns,nm,nb,ne)=sum{sd,w.l(nd,sd)*vv.l(nd,sd)};

Stores the estimated dependent variable.

ey("7",nd,ns,nm,nb,ne)=sum(nk,xs(nk,nd)*sum{sc,p.l(nk,sc)*zz.l(nk,sc)});

Stores the feasibility of the model so it can be seen later whether the model solved effectively or otherwise.

ef("7",ns,nm,nb,ne)=dde1.modelstat;

The following solves the Leuven 1 model.

Tells GAMS to solve the mle1 (Leuven 1) equation by minimising the total entropy and using nonlinear programming.

solve mle1 minimising tenl using nlp;

Stores the estimated coefficients.

eb("8",nk,ns,nm,nb,ne)=cl.l(nk);

Stores the estimated errors.

ee("8",nd,ns,nm,nb,ne)=er.l(nd);

Stores the estimated dependent variable.

ey("8",nd,ns,nm,nb,ne)=sum(nk,xs(nk,nd)*cl.l(nk));

Stores the feasibility of the model so it can be seen later whether the model solved effectively or otherwise.

ef("8",ns,nm,nb,ne)=mle1.modelstat;

The following solves the Leuven 2 model.

Tells GAMS to solve the mle2 (Leuven 2) equation by minimising the total entropy and using nonlinear programming

solve mle2 minimising tenl using nlp;

Stores the estimated coefficients.

eb("9",nk,ns,nm,nb,ne)=cl.l(nk);

Stores the estimated errors.

ee("9",nd,ns,nm,nb,ne)=er.l(nd);

Stores the estimated dependent variable.

ey("9",nd,ns,nm,nb,ne)=sum(nk,xs(nk,nd)*cl.l(nk));

Stores the feasibility of the model so it can be seen later whether the model solved effectively or otherwise.

ef("9",ns,nm,nb,ne)=mle2.modelstat;

The following bracket ends the loop.

);

The final section of coding for this program works to process and appropriately display the results of the above calculations.

Fills the estimated bias for each regression type with the actual estimated biases.

es(nr,nm,nb,ne)=sum(nk,(sum(ns,eb(nr,nk,ns,nm,nb,ne))/card(ns))-
bt(nk))/card(nk);

Fills the squared estimated bias from each regression type with the actual squared estimated biases.

ses(nr,nm,nb,ne)=power(es(nr,nm,nb,ne),2);

Fills the estimated variance in each regression type with the actual estimated variances.

ev(nr,nm,nb,ne)=sum(nk,sum(ns,power(eb(nr,nk,ns,nm,nb,ne)-
(sum(ss,eb(nr,nk,ss,nm,nb,ne))/card(ns)),2))/card(ns))/card(nk);

Fills the estimated mean squared error loss in each regression type with the actual estimated mean squared error losses.

emsel(nr,nm,nb,ne)=ses(nr,nm,nb,ne)+ev(nr,nm,nb,ne);

Fills the estimated mean squared error in each regression type with the actual mean squared errors.

emse(nr,nm,nb,ne)=sum(ns,((sum(nd,power(ee(nr,nd,ns,nm,nb,ne),2))/(card(nd)-
card(nk)))))/card(ns);

States that at the end of all of the regression runs, GAMS will display the feasibility, estimated bias for each regression type, the estimated variance for each regression type, the estimated mean squared error loss for each regression type, and the estimated mean squared error for each regression type.

display ef,es,ev,emsel,emse;

Appendix 2: Extended Results

Appendix 2 presents the MSE and the MSEL results for five of the condition numbers studied here. These condition numbers are: 10, 20, 80, 250, and 1000, and are not included in Section 4 for the sake of parsimony. It also presents a table showing the average MSE values for each estimator, for all 10 condition numbers studied, and a table showing the same for MSEL values.

A2.1 Comparison of mean squared error for each estimator

A2.1.1 Mean squared error for all models when condition number is 10

Table A2.1 presents MSE for all estimators when the condition number is 10, thus giving slight multicollinearity. In this case, we see broadly similar results to those in Table 4.1. As can be seen in Table A2.1, all of the estimators generally peak during the mixed Gaussian error distribution, except for the GMEW estimator. The GMEW estimator peaks here, as well as experiencing a second peak when dealing with $t_{(3)}$ data, or $\chi^2_{(5)}$ data and a $t_{(3)}$ error distribution. This secondary peak of the GMEW estimator in the presence of a $t_{(3)}$ error distribution is found in many of the multicollinearity scenarios studied here, and reflects the inability of broad support ranges to deal with non-normal distributions.

Table A2.1: Mean squared error for estimators with a condition number of 10.

Mean squared error	Condition number: 10									
	Model									
Dist.[1]	OLS	Ridge-1	Ridge-2	RLS	GMEN	GMEW	DDE2	DDE1	Leuven-1	Leuven-2
e1d1	1.01	1.09	1.01	1.01	1.4	1.01	1.12	1.13	1.3	1.05
e1d2	1	1.09	1	1	1.43	1.01	1.12	1.15	1.3	1.05
e1d3	1	1.09	1	1	1.42	1.01	1.11	1.14	1.31	1.05
e2d1	1.8	1.9	1.8	1.8	2.27	1.81	2.03	2.05	2.13	1.85
e2d2	1.8	1.91	1.8	1.8	2.3	1.82	2.06	2.1	2.13	1.86
e2d3	1.81	1.9	1.81	1.81	2.27	1.82	2.01	2.06	2.13	2.97
e3d1	1.05	1.13	1.05	1.05	1.64	1.15	1.34	1.58	1.34	1.1
e3d2	1.06	1.15	1.06	1.06	1.71	2.07	1.43	1.56	1.36	1.11
e3d3	1.06	1.14	1.06	1.06	1.68	1.91	1.36	1.46	1.36	1.11
e4d1	1.07	1.16	1.07	1.07	1.5	1.09	1.27	1.65	1.39	1.13
e4d2	1.07	1.16	1.07	1.07	1.52	1.08	1.3	1.73	1.38	1.12
e4d3	1.07	1.15	1.07	1.07	1.49	1.08	1.27	1.68	1.37	1.12

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

OLS, ridge-2, and RLS again achieve the lowest MSE, and are again equal with each other, regardless of error and data distributions (Table A2.1). This indicates that multicollinearity has not yet increased to a point that degrades RLS and the ridge-2 regression. With the condition number of 10, the Leuven-2 estimator begins to show some variability, with a peak of 2.97 when mixed Gaussian errors and Chi squared data generation are used (Table A2.1). This gives it the highest single MSE value with this condition number. In general, the GMEN estimator produces the highest levels of MSE, followed by Leuven-1, DDE1, and DDE2.

A2.1.2 Mean squared error for all models when condition number is 20

Table A2.2 presents MSE for all estimators when the condition number is 20. Here, there is some multicollinearity, but at this stage it would not yet be expected to cause any significant degradation in the results. We see many of the same trends as previously, with all of the estimators other than Leuven-2 peaking in the region where mixed Gaussian error generation is used. However, the Leuven-2 estimator seems less consistent than previously, with its MSE fluctuating more than these other estimators. This is due to the extension of the entropy specification to the error term causing inaccuracies in the predictions of the dependant variable under multicollinearity.

Table A2.2: Mean squared error for estimators with a condition number of 20.

Mean squared error	**Condition number: 20**									
	Model									
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	1	1.07	1	1	1.36	1.02	1.11	1.15	1.26	1.05
e1d2	1	1.06	1	1	1.32	1.01	1.14	1.15	1.26	1.05
e1d3	1.01	1.08	1.01	1.01	1.37	1.03	1.12	1.15	1.27	1.06
e2d1	1.81	1.89	1.82	1.82	2.22	1.84	1.99	2.09	2.09	1.87
e2d2	1.82	1.89	1.83	1.83	2.19	1.85	2.07	2.17	2.1	1.88
e2d3	1.81	1.88	1.81	1.81	2.21	1.83	1.98	2.05	2.08	1.86
e3d1	1.05	1.11	1.05	1.06	1.57	1.23	1.28	1.6	1.31	1.1
e3d2	1.03	1.1	1.04	1.04	1.58	1.16	1.39	1.71	1.31	1.09
e3d3	1.07	1.14	1.07	1.07	1.74	1.82	1.36	1.47	1.33	2.21
e4d1	1.07	1.14	1.08	1.08	1.45	1.09	1.24	1.73	1.34	2.35
e4d2	1.07	1.13	1.07	1.07	1.41	1.09	1.3	1.84	1.34	1.12
e4d3	1.07	1.13	1.07	1.07	1.43	1.09	1.24	1.69	1.33	2.21

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

GMEN again generally produces the highest values of MSE (Table A2.2). After this, the DDE1 model generally has the next highest MSE, followed by Leuven-1 and DDE2 (Table A2.2). The ridge-1, ridge-2 and RLS estimators are once again the same, however this is the first time we see OLS having MSE values which are lower than those of all other estimators, as some of the OLS MSE values are (very marginally) lower than those created by ridge-2 and RLS.

A2.1.3 Mean squared error for all models when condition number is 80

Table A2.3 presents MSE for all estimators when the condition number is 80, which represents relatively severe multicollinearity. The Leuven-2 estimator, while having previously often given high and inconsistent values of MSE, seems to be a more effective estimator in the MSE sense with the condition number of 80, mainly giving lower MSE values than GMEN, DDE1, Leuven-1 and DDE2.

Table A2.3: Mean squared error for estimators with a condition number of 80.

Mean squared error	Condition number: 80									
	Model									
Dist.[1]	OLS	Ridge-1	Ridge-2	RLS	GMEN	GMEW	DDE2	DDE1	Leuven-1	Leuven-2
e1d1	1.01	1.05	1.02	1.02	1.37	1.04	1.1	1.16	1.26	1.06
e1d2	1.02	1.06	1.03	1.03	1.36	1.04	1.14	1.18	1.27	1.06
e1d3	1.01	1.04	1.02	1.02	1.34	1.03	1.1	1.14	1.25	1.05
e2d1	1.79	1.85	1.82	1.83	2.22	1.85	1.97	2.1	2.07	3.31
e2d2	1.79	1.85	1.82	1.83	2.2	1.85	2.03	2.12	2.08	1.87
e2d3	1.79	1.84	1.81	1.81	2.2	1.83	1.98	2.1	2.07	1.85
e3d1	1.06	1.09	1.06	1.06	1.51	1.18	1.28	1.75	1.29	1.1
e3d2	1.04	1.08	1.05	1.05	1.48	1.16	1.34	1.84	1.29	1.08
e3d3	1.05	1.08	1.06	1.06	1.74	2.1	1.32	1.66	1.29	1.09
e4d1	1.06	1.1	1.07	1.08	1.43	1.09	1.21	1.75	1.31	1.11
e4d2	1.07	1.1	1.08	1.08	1.41	1.1	1.26	1.84	1.32	1.11
e4d3	1.08	1.12	1.09	1.09	1.42	1.1	1.21	1.76	1.32	1.12

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

In general, the other estimators follow the pattern which has been evident so far, with very low MSE when using a normal error distribution and peaks when using the mixed Gaussian error distribution. Lower values again are seen when errors are produced using either the $t_{(3)}$ or $\chi^2_{(5)}$ distribution (Table A2.3).

As the condition number increases, the capacity of the MSE for the DDE1 estimator to drop significantly for the $t_{(3)}$ and $\chi^2_{(5)}$ error distribution lessens, with its values for the $t_{(3)}$ and $\chi^2_{(5)}$ error distributions being only marginally lower than those with the mixed Gaussian error distribution (Table A2.3). The GMEN estimator again commonly produces high levels of MSE, while OLS acts as expected again in producing the lowest MSE values.

A2.1.4 Mean squared error for all models when condition number is 250

Table A2.4 presents MSE for all estimators when the condition number is 250, which represents a severe multicollinearity situation. The GMEN and GMEW both exhibit a pattern of a second peak during the $t_{(3)}$ error distribution as the large multicollinearity level enables the thick tails of the distribution to place errors outside of expectations and reduce the effectiveness of these maximum entropy estimators.

Table A2.4: Mean squared error for estimators with a condition number of 250.

Mean squared error	**Condition number: 250**									
	Model									
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	1.01	1.04	1.03	1.02	1.35	1.03	1.09	1.14	1.24	1.05
e1d2	1.01	1.03	1.03	1.03	1.33	1.03	1.12	1.17	1.26	1.05
e1d3	1.01	1.04	1.03	1.03	1.36	1.04	1.1	1.17	1.25	3.28
e2d1	1.8	1.84	1.84	1.84	2.19	1.85	1.96	2.12	2.05	2.98
e2d2	1.8	1.84	1.84	1.84	2.17	1.85	1.99	2.12	2.06	1.86
e2d3	1.79	1.83	1.84	1.83	2.2	1.84	1.96	2.09	2.05	1.86
e3d1	1.06	1.08	1.07	1.07	1.78	2.38	1.32	1.61	1.29	2.08
e3d2	1.08	1.1	1.1	1.09	1.95	2.3	1.38	1.75	1.32	1.12
e3d3	1.06	1.09	1.08	1.08	1.86	1.52	1.32	1.74	1.3	1.1
e4d1	1.07	1.1	1.09	1.09	1.43	1.1	1.23	1.82	1.31	1.12
e4d2	1.08	1.1	1.1	1.1	1.41	1.11	1.25	1.88	1.33	1.12
e4d3	1.08	1.1	1.09	1.09	1.42	1.1	1.26	1.84	1.31	2.19

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The DDE1 model continues to handle the non-normal error distributions noticeably less well than the normal error distributions (Table A2.4). The DDE2 method is shown to have a lower MSE than the DDE1 method for all data and error distributions with this condition number. OLS once again has the lowest MSE, followed by RLS.

A2.1.5 Mean squared error for all models when condition number is 1000

Table A2.5 presents MSE for all estimators when the condition number is 1000. With a condition number of 1000, the multicollinearity represented in this round of the experiment is very extreme, and it would be expected that we would see significant degradation in results. At this condition number, as seen in Table A2.5, the lowest values of MSE are again, as expected, produced by the OLS estimator due to its objective function, followed by the ridge-1, ridge-2 and RLS estimators, which produce results very similar to one another.

Table A2.5: Mean squared error for estimators with a condition number of 1000.

Mean squared error	Condition number: 1000									
	Model									
Dist.[1]	OLS	Ridge-1	Ridge-2	RLS	GMEN	GMEW	DDE2	DDE1	Leuven-1	Leuven-2
e1d1	1.01	1.03	1.03	1.03	1.42	1.04	1.09	1.14	1.27	2.12
e1d2	1	1.03	1.03	1.03	1.42	1.04	1.1	1.17	1.28	1.06
e1d3	1	1.03	1.03	1.03	1.4	1.03	1.09	1.16	1.27	2.06
e2d1	1.81	1.84	1.84	1.84	2.25	1.85	1.98	2.12	2.08	1.86
e2d2	1.81	1.84	1.84	1.84	2.25	1.85	2.02	2.14	2.1	1.86
e2d3	1.8	1.84	1.84	1.84	2.25	1.85	1.97	2.14	2.08	1.86
e3d1	1.06	1.08	1.08	1.08	1.7	1.32	1.31	1.66	1.32	1.11
e3d2	1.06	1.08	1.09	1.09	1.91	1.86	1.45	1.72	1.34	1.11
e3d3	1.02	1.04	1.06	1.06	1.55	1.17	1.21	1.73	1.3	2.08
e4d1	1.06	1.08	1.09	1.09	1.49	1.12	1.2	1.79	1.32	1.11
e4d2	1.06	1.08	1.09	1.09	1.52	1.13	1.23	1.88	1.33	1.11
e4d3	1.06	1.08	1.08	1.08	1.46	1.1	1.2	1.73	1.31	2.09

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The DDE2 estimator consistently provides lower MSE values than the DDE1 estimator which is unsurprising in that the DDE1 estimator is designed to reduce MSEL not MSE in the case of multicollinearity. However, the Leuven-2 estimator is very inconsistent in terms of condition number, fluctuating from levels below those of the DDE2 estimator to values considerably higher than those of the DDE1 estimator (Table A2.5). This makes it difficult to judge where to place the Leuven-2 estimator in terms of the accuracy of prediction with extreme multicollinearity. The GMEW estimator follows the same pattern as at previous condition numbers, as does the GMEN estimator which consistently provides higher MSE values than all the estimators other than Leuven-2, and DDE1.

It should however be noted that differences between estimators are relatively small, not showing any of the estimators to be completely unsuitable for making predictions of the dependent variable in the presence of multicollinearity.

A2.2 Comparison of mean squared error loss for each estimator

A2.2.1Mean squared error loss for all models when condition number is 10

Table A2.6 presents MSEL for all estimators when the condition number is 10, which represents slight multicollinearity, however not enough that we would expect to see any significant degradation in estimation ability. With a condition number of 10, we see that the DDE2 estimator produces the highest values of MSEL (Table A2.6), making it the least able to accurately predict the true coefficients. After DDE2, the next highest MSEL is generally produced by the DDE1 estimator. OLS generally produces the next highest levels of MSEL as multicollinearity begins to degrade its accuracy, followed generally by ridge-2, Leuven-1, and RLS. These all peak at the point where errors are produced from a mixed Gaussian distribution and data from a $t_{(3)}$ distribution.

Table A2.6: Mean squared error loss for estimators with a condition number of 10.

Mean squared error loss			Condition number: 10							
					Model					
Dist.[1]	OLS	Ridge-1	Ridge-2	RLS	GMEN	GMEW	DDE2	DDE1	Leuven-1	Leuven-2
e1d1	0.39	0.36	0.38	0.39	0.41	0.25	1.22	0.51	0.49	0.35
e1d2	0.68	0.43	0.65	0.6	0.45	0.36	1.4	0.64	0.49	0.41
e1d3	0.58	0.38	0.56	0.57	0.42	0.32	1.34	0.67	0.5	0.37
e2d1	0.9	0.44	0.85	0.84	0.45	0.5	2.85	1.28	0.51	0.44
e2d2	1.35	0.46	1.2	1.11	0.46	0.55	3.53	1.33	0.5	0.44
e2d3	0.97	0.45	0.89	0.89	0.44	0.51	2.64	1.16	0.5	0.64
e3d1	0.45	0.39	0.37	0.38	0.56	0.99	2.26	1.05	0.5	0.39
e3d2	0.52	0.42	0.48	0.5	0.59	0.7	3	1.48	0.5	0.41
e3d3	0.47	0.38	0.44	0.45	0.63	0.51	3.14	1.32	0.51	0.38
e4d1	0.64	0.37	0.61	0.62	0.42	0.38	1.76	0.86	0.48	0.36
e4d2	0.69	0.42	0.63	0.63	0.44	0.44	2.05	1.05	0.48	0.4
e4d3	0.7	0.41	0.65	0.65	0.42	0.44	1.76	0.9	0.49	0.38

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The GMEN and GMEW models also fit in this area, being lower than RLS, but generally higher than ridge-1 and Leuven-2, which in general produce the lowest level of MSEL for this condition number (Table A2.6). Their values are very similar to one another, showing that for this condition number ridge-1 and Leuven-2 are generally the most able to accurately predict the correct coefficients.

A2.2.2 Mean squared error loss for all models when condition number is 20

Table A2.7 presents MSEL for all estimators when the condition number is 20. Here we see a dramatic change in the MSEL of the DDE2 estimator, with it now ranging from 4.51 up to 17.01 when $t_{(3)}$ error and data generation are used, thus

again making it the least accurate at predicting the true coefficients. The Leuven-1 model produces relatively constant values of MSEL, making it in some ways more useful than some of the other estimators which are more variable, such as the ridge-1, DDE2, GMEN, and GMEW approaches.

Table A2.7: Mean squared error loss for estimators with a condition number of 20.

Mean squared error loss			**Condition number: 20**							
					Model					
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	2.05	0.61	1.68	1.43	0.58	0.56	4.51	0.87	0.64	0.62
e1d2	2.74	0.65	2.09	1.76	0.59	0.66	8.74	1.04	0.63	0.65
e1d3	2.03	0.58	1.68	1.44	0.59	0.54	5.33	0.96	0.64	0.63
e2d1	3.9	0.69	2.81	2.24	0.62	0.83	6.52	1.46	0.66	0.68
e2d2	4.56	0.72	2.98	2.3	0.61	0.82	11.29	1.34	0.64	0.67
e2d3	3.39	0.76	2.4	2.3	0.61	0.86	8.22	1.33	0.66	0.68
e3d1	3.19	1.05	1.49	1.39	0.75	2.07	10.8	1.75	0.65	0.69
e3d2	4.18	1.24	1.78	1.65	0.84	2.05	17.01	1.88	0.65	0.72
e3d3	2.21	0.61	1.55	1.33	0.75	0.99	16.18	1.78	0.64	0.82
e4d1	2.26	0.69	1.62	1.28	0.58	0.63	7.28	1.27	0.64	0.81
e4d2	3.09	0.73	2.11	1.71	0.59	0.76	9.43	1.29	0.64	0.66
e4d3	2.57	0.65	2.1	1.83	0.6	0.7	7.28	1.44	0.64	0.85

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N(0,1)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N(0,1)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The next lowest MSEL value is, in general, given by the DDE1 estimator (Table A2.7). OLS produces the second highest level of MSEL with this condition number, making it the second worst at predicting the correct coefficients. Due to fluctuations in the results of these models, there is no estimator which consistently gives the lowest MSEL values at this condition number.

A2.2.3 Mean squared error loss for all models when condition number is 80

Table A2.8 presents MSEL for all estimators when the condition number is 80. The DDE2 model again provides the highest levels of MSEL for all data and error distributions, as can be seen in Table A2.8. This is followed by the OLS estimator which peaks over the range where mixed Gaussian errors are used, similarly to DDE2.

Table A2.8: Mean squared error loss for estimators with a condition number of 80.

Mean squared error loss			**Condition number: 80**							
					Model					
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	27.87	0.98	6.2	5.1	0.64	0.85	63.78	0.99	0.67	0.75
e1d2	31.8	1.39	5.55	5.13	0.64	0.95	98.67	0.97	0.65	0.76
e1d3	28.81	0.81	6.22	5.1	0.63	0.78	74.49	0.95	0.66	0.74
e2d1	78.26	2.91	11.11	5.33	0.66	1.12	166.44	1.21	0.69	1.01
e2d2	84.72	4.3	7.76	6.43	0.66	1.29	215.87	1.37	0.68	0.83
e2d3	63.91	1.77	7.32	6.02	0.65	0.98	164.39	1.28	0.67	0.8
e3d1	20.53	0.83	5.1	4.41	0.71	1.43	106.5	1.38	0.67	0.75
e3d2	33.46	1.38	4.32	4.44	0.75	1.53	173.75	1.5	0.67	0.79
e3d3	42.17	1.88	5.28	4.8	0.78	2.32	165.79	1.4	0.65	0.74
e4d1	31.91	2.93	6.5	4.84	0.63	0.82	65.21	1.09	0.67	0.75
e4d2	42.07	2.23	7	5.32	0.63	0.8	89.96	1.13	0.65	0.75
e4d3	29.23	0.91	3.72	3.87	0.63	0.86	84.49	1	0.65	0.74

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

These are followed by the ridge-2 and RLS estimators, with the ridge-2 estimator producing higher MSEL values. The lowest values of MSEL are again produced at some points by the GMEN estimator, and at some by the Leuven-1 estimator,

followed by the Leuven-2 and GMEW estimators. At this level of multicollinearity, the results of the ridge-2 estimator are considerably more variable than previously (Table A2.8).

A2.2.4Mean squared error loss for all models when condition number is 250

Table A2.9 presents MSEL for all estimators when the condition number is 250. We again see a marked increase in the MSEL values of DDE2 and OLS due to this extremely severe multicollinearity, with DDE2 peaking at an MSEL value of 2284.51 and OLS peaking at 611.06 (Table A2.9).

Table A2.9: Mean squared error loss for estimators with a condition number of 250.

Mean squared error loss			**Condition number: 250**							
					Model					
Dist.[1]	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
e1d1	253.11	2.33	3.57	6.57	0.68	0.86	641.37	0.98	0.7	0.8
e1d2	398.69	7.91	2.74	7.17	0.68	0.9	1,026.89	0.99	0.68	0.81
e1d3	355.93	4.77	4.2	6.66	0.68	0.85	718.08	0.97	0.7	1.15
e2d1	576.56	8.61	3.44	7.11	0.68	0.87	1,087.08	1.01	0.71	0.99
e2d2	725.51	12.64	2.9	7.34	0.69	0.94	1,711.47	1.06	0.7	0.83
e2d3	611.06	15.4	2.93	7.04	0.68	0.96	1,510.13	1.14	0.71	0.83
e3d1	265.64	3.62	4.44	6.71	0.94	1.85	1,550.02	1.38	0.7	0.97
e3d2	387.85	6.04	3.52	7.07	0.89	2.53	2,284.51	1.45	0.68	0.8
e3d3	319.72	3.24	5.5	7	0.94	2.96	1,486.48	1.22	0.69	0.8
e4d1	343.21	3.35	3.95	6.86	0.69	0.9	844.17	1.11	0.71	0.81
e4d2	407.06	7.65	2.85	6.69	0.69	0.89	1,317.53	1.24	0.69	0.81
e4d3	249.84	1.63	3.07	6.84	0.67	0.88	916.69	1.02	0.69	0.97

[1]Distributions presented show error number then data number (e.g. e1d1). Error 1 (e1) = $N\left(0,1\right)$ distribution, error 2 (e2) = mixed Gaussian distribution, error 3 (e3) = $t_{(3)}$ distribution, and error 4 (e4) = $\chi^2_{(5)}$ distribution. Data 1(d1) = $N\left(0,1\right)$ distribution, data 2 (d2) = $t_{(3)}$ distribution, and data 3 (d3) = $\chi^2_{(5)}$ distribution**.**

The Leuven-1 and GMEN models are again the estimators with the lowest values of MSEL (Table A2.9), making them the most accurate at predicting the coefficients with this level of multicollinearity. The Leuven-1 model is potentially so accurate, even at this extreme of multicollinearity, because the parameter probabilities in general do not approach a uniform distribution, thus allowing more flexibility in the estimator (Paris, 2001). The GMEN model is still benefiting from its search area for the true coefficients being constrained to such a small interval. These estimators are followed again by the Leuven-2 estimator in most cases, then generally the GMEW estimator.

A2.2.5 Mean squared error loss for all models when condition number is 1000

Table A2.10 presents MSEL for all estimators when the condition number is 1000. This is the most extreme case of multicollinearity that has been analysed here, and we see a continuation of some of the patterns noticed earlier. In relation to DDE2 and OLS, both have MSEL values that have increased greatly in comparison to the values associated with a condition number of 500. The DDE2 values now range between 11,557.17 to 41,123.61, and OLS results range from 4,238.94 to 10,024.87 (Table A2.10).

Table A2.10: Mean squared error loss for estimators with a condition number of 1000.

Mean squared error loss			Condition number: 1000							
					Model					
Dist.[1]	OLS	Ridge-1	Ridge-2	RLS	GME (-5,5)	GME (-20,20)	DDE2	DDE1	Leuven 1	Leuven 2
e1d1	4,238.94	17.1	1.38	6.68	0.57	0.7	11,557.17	0.95	0.6	0.84
e1d2	6,005.22	47.01	2.13	7.12	0.56	0.72	15,800.81	0.91	0.58	0.67
e1d3	5,346.94	24.62	1.45	7.19	0.57	0.74	12,178.93	0.93	0.6	0.86
e2d1	6,334.43	84.39	1.2	7.12	0.57	0.79	25,736.25	0.98	0.61	0.72
e2d2	7,257.65	156.08	1.16	6.66	0.57	0.81	33,920.93	1.05	0.59	0.71
e2d3	7,227.61	154.77	1.37	6.09	0.58	0.79	21,471.09	1.11	0.62	0.71
e3d1	5,412.25	147.81	1.45	6.39	0.76	0.8	26,503.21	1.19	0.61	0.68
e3d2	6,744.46	128.96	1.7	6.89	0.79	2.82	41,123.61	1.26	0.59	0.67
e3d3	10,024.87	1,858.52	1.32	7.21	0.63	0.83	21,549.54	1.19	0.6	0.85
e4d1	5,688.18	107.74	1.68	6.63	0.58	0.81	13,659.84	1.01	0.61	0.69
e4d2	6,083.23	166.4	1.69	6.65	0.59	1.02	18,551.08	1.11	0.6	0.69
e4d3	4,596.82	30.15	1.36	6.5	0.57	0.77	13,929.36	0.87	0.61	0.86

[1]Distributions presented show error number then data number (e1d1). Error 1 = $N\left(0,1\right)$ distribution, error 2 = mixed Gaussian distribution, error 3 = $t_{(3)}$ distribution and error 4 = $\chi^2_{(5)}$ distribution. Data 1 = $N\left(0,1\right)$ distribution, data 2 = $t_{(3)}$ distribution and data 3 = $\chi^2_{(5)}$ distribution**.**

The ridge-1 estimator also again provides the next highest levels of MSEL, with most of its values ranging between 17.10 and 166.40 (Table A2.10) as at such a high level of multicollinearity, effects other than large condition numbers have come into play. The lowest values of MSEL, and thus the estimators best able to accurately predict coefficients, are the GMEN and Leuven-1 estimators (with GMEN generally producing the lower MSEL values except for during the $t_{(3)}$ error distribution). However, as the GMEN model is likely only providing such a low MSEL value due to its exogenous support bounds, the Leuven-1 model is generally more reliable.

A2.3 Average tables for each condition number and estimator

A2.3.1 Average MSE values for each estimator over all 10 studied condition numbers

Table A2.11: Mean squared error average values for each condition number and estimator.

	Model									
Cond. No.	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
1	1.23	1.35	1.23	1.23	2.04	1.24	1.47	1.62	1.5	1.24
10	1.23	1.32	1.23	1.23	1.72	1.41	1.45	1.61	1.54	1.38
20	1.23	1.3	1.24	1.24	1.65	1.34	1.44	1.65	1.5	1.57
40	1.23	1.29	1.24	1.24	1.69	1.3	1.43	1.67	1.5	1.65
60	1.23	1.28	1.24	1.25	1.63	1.29	1.42	1.68	1.5	1.63
80	1.23	1.27	1.24	1.25	1.64	1.36	1.41	1.7	1.49	1.4
100	1.23	1.27	1.25	1.25	1.7	1.34	1.41	1.67	1.51	1.36
250	1.23	1.27	1.26	1.26	1.7	1.51	1.41	1.7	1.48	1.73
500	1.23	1.26	1.26	1.26	1.69	1.38	1.41	1.71	1.5	1.28
1000	1.23	1.25	1.26	1.26	1.72	1.36	1.4	1.7	1.5	1.62

A2.3.2 Average MSEL values for each estimator over all 10 studied condition numbers

Table A2.12: Mean squared error loss average values for each condition number and estimator.

	Model									
Cond. No.	**OLS**	**Ridge-1**	**Ridge-2**	**RLS**	**GMEN**	**GMEW**	**DDE2**	**DDE1**	**Leuven-1**	**Leuven-2**
1	0.03	0.03	0.03	0.03	0.05	0.03	0.09	0.1	0.03	0.03
10	0.69	0.41	0.64	0.64	0.47	0.5	2.25	1.02	0.49	0.41
20	3.01	0.75	2.02	1.72	0.64	0.96	9.38	1.37	0.64	0.71
40	11.12	0.82	4.98	3.78	0.6	0.92	35.84	1.21	0.61	0.75
60	23.82	0.99	6.15	4.91	0.61	0.81	71.61	1.05	0.62	0.78
80	42.89	1.86	6.34	5.07	0.67	1.14	122.44	1.19	0.67	0.78
100	63.78	1.36	6.47	5.52	0.58	0.81	218.59	1.04	0.58	0.69
250	407.85	6.43	3.59	6.92	0.74	1.28	1,257.87	1.13	0.7	0.88
500	1,559.3	19.84	1.95	6.54	0.61	0.95	5,441.67	1.03	0.6	0.7
1000	6,246.72	243.63	1.49	6.76	0.61	0.97	21,331.82	1.04	0.6	0.75