5-2015

Three Essays on Panel Data Estimation

Alexander Houser
Western Michigan University, yalexhouser@yahoo.com

THREE ESSAYS ON PANEL DATA ESTIMATION

by

Alexander Houser

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Economics
Western Michigan University
May 2015

Doctoral Committee:

Debasri Mukherjee, PhD., Chair
Eskander Alvi, PhD.
Joeseph McKean, PhD.

THREE ESSAYS ON PANEL DATA ESTIMATION

Alexander Houser, Ph.D.

Western Michigan University, 2015

This work discusses various aspects of panel data estimation. In chapter one, an algorithm for semiparametric random effects estimation is proposed. The performance of bootstrap-based confidence intervals for the proposed estimators are examined and found reasonable. The algorithm is also applied to a set of U.S. state level medical expenditure data to estimate the medical Engel curve. In the second chapter, the predictive performance of various parametric and semiparametric panel data estimators is compared on the same dataset of U.S. state level medical expenditures as well as out of sample forecast performance and bootstrap bias-corrected mean square errors of the competing estimators are evaluated. In general, the estimator discussed in the first chapter is found to perform well. In the third chapter a generalized method of moments estimator is investigated under various norms.

ACKNOWLEDGEMENTS

I owe a great debt to many people who have served as guides and mentors over the last three years of graduate study. I also could not have made it to where I am today without the aid and generosity of countless people during both my graduate and undergraduate years. To list all those who I owe my thanks would require a document almost as long as this dissertation, but a few shorter acknowledgements are in order.

First of all, I would like to thank my parents who have helped me through so many of the complications that arise in the process of starting independent life. Through countless phone calls, emails, and visits you have helped me through the myriad of problems that have arisen from my having lived in three cities and moved six times in the last five years. It is your constant support and encouragement that keeps me going through the many ups and downs of graduate study.

I would also like to thank my grandparents. My roots are here in Kalamazoo, and it was not until I came to Western Michigan University that I realized how much virtue family connectedness has, and how much I am giving up when I leave for Connecticut. I will miss both of you tremendously and the countless dinners and conversations we've had. I would also like to thank Mike McCarville for providing me a very nice place to live after my lease expired. Your hospitality has been instrumental in this last month while I complete my degree.

I would like to thank both of my previous employers, Dr. Jason Bodnar and Dr. Hongjie Wang, it was an absolute pleasure to work under both of you. Before I began working under Jason I had only the vaguest notion of what I wanted from my studies, when I returned from Terumo I returned with both a new direction in my own life and vastly enhanced skill set with which to pursue that direction. Hongjie remains an inspiration in his knowledge, capabilities, and intellectual fervor. I've never before met someone with such a broad mastery of so many areas of inquiry and I have also never met someone so incredibly kind to his employees, interns, and friends. I am also very grateful for your help in arranging the position which I will be taking up upon completion of this project.

I would like to thank the members of my committee. Each of you have had a surprisingly large impact on the course of my life, although you may not know it. Dr. Joe McKean, you were the first statistician I ever took classes under, and I must admit that you gave me the firmest challenge I've received since starting graduate study. You introduced me to programming in R, a programming language I've since fallen in love with, and you've expanded my horizons much beyond the purview of traditional econometrics. Dr. Eskander Alvi you are similarly an inspiration in your broad knowledge of economics and your passion for conveying that knowledge to new groups of economists. When I first came to visit this program before enrolling, I took the time to stop by your office to discuss some past difficulties I had encountered as a student. Your encouragement prompted me towards enrollment and towards what has ultimately turned out to be a successful graduate career.

Finally, I would like to thank my advisor, Dr. Debasri Mukherjee, for her countless hours of devotion to helping me through this project. You are an inspiration in your exceptional knowledge, attention to detail, and commitment to helping others. I can say after working with you over the last two years, that you are a model of the researcher I hope to become. I deeply appreciate the hours we've spend going over simulation results and drafts of this manuscript, and I know that this process has made me a much better economist than I otherwise would have been. I am very lucky to have such a wonderful advisor and scholar to guide me through the final stage of acquiring my degree and to have provided me with such an excellent set of research opportunities. I wish you the best in the coming years for all the help you've given through the construction of this dissertation and the enormous impact you've had on my life.

Alexander Houser

TABLE OF CONTENTS

Table of Contents—Continued

LIST OF TABLES

LIST OF FIGURES

Introduction

This dissertation consists of three essays on panel data. Chapter 1 focuses on modeling nonlinearities in the regression function along with random effects in errors. The primary contribution of the chapter is the estimation of a partially linear (semiparametric) regression model under kernel smoothing, allowing for time invariant random effects in the error. The chapter provides a simple algorithm for estimating parametric and nonparametric marginal effects (of the covariates on response) as well as the nonparametric conditional mean function. A bootstrap based empirical method for constructing confidence intervals for the nonparametric components is also studied and the coverage probabilities of the confidence intervals have been assessed. The proposed technique has been illustrated using state level medical data for the United States. Chapter 2 examines performance of various panel data estimators. Using the same medical data, predictive performance of a number of parametric and semiparametric panel data estimators are examined. Cross-validated out-of-sample forecasts as well as bootstrapped bias corrected mean-square-errors are used for the purpose of evaluation. Compared to its various counterparts, the semiparametric random effects estimator as described in the first chapter is found to work very well in terms of predictive performance. Lastly in Chapter 3 the application of a number of distinct norms to the system GMM estimator are discussed and some improvement in robustness to mis-measurement in a simulated dataset by using these norms is shown.

The literature on semiparametric random effects regression under kernel smoothing is limited. Li and Stengos (1996) as well as Wang et al. (2005) discuss the consistency and efficiency of the parametric marginal effects in a semiparametric regression setting without considering any random effects structure in the error. Also any discussion on the nonparametric components (under such a setting) is missing. However, Ruckstuhl et al. (2000) and Henderson et al. (2005) consider purely nonparametric regression with random effects and discuss the consistent estimation of nonparametric components, but such discussion has not been extended to a partially linear (semiparametric) setting. The first chapter in this dissertation combines knowledge from these two distinct groups of research and builds upon the work of Ruckstuhl et al. (2000) and Li and Stengos (1996) to discuss the estimation of nonparametric components in a random effects partially linear (semiparametric) regression setting. Simulation results presented in this chapter suggest that the estimators of nonparametric components work well in terms of integrated mean-square errors, and their performance improves with increasing sample sizes, as expected. The marginal effect (on response) of the covariate modeled parametrically also performs satisfactorily. Ruckstuhl et al. (2000) in their study of a nonparametric random effects model observe that although random effects estimators can have smaller asymptotic variance compared to pooled estimators (ignoring random effects) in a nonparametric setting (under local kernel smoothing), there may be a cost in terms of possibly higher asymptotic bias. We note the same problem in our semiparametric setting as well and therefore the search for a local random effects estimator (in a nonparametric or semiparametric setting) which will consistently produce lower asymptotic mean-square-error (compared to local pooled estimators) remains open.

Additionally neither Henderson et al. (2005) nor Ruckstuhl et al. (2000) explicitly state how to form appropriate confidence intervals for the estimators of the nonparametric components. This chapter discusses such a method, taking lead from the work of Hardle et al. (2004). Hardle et al. (2004) use a wild bootstrap of the type attributed to Mammen (1993) to produce confidence intervals in a partially linear setting and we replicate this design. The coverage probabilities for such interval estimates show satisfactory results for the estimate of the nonparametric conditional mean function.

The proposed estimators are illustrated by examining the medical Engel curve for the fifty states of the United States (modeling medical expenditures as a function of income and percentage of older people in the population).

The second chapter uses the same medical dataset as in the first chapter and considers eight different types of panel data estimators. The purpose is to compare/evaluate the predictive performance of these estimators, including the semiparametric random effects estimator as discussed in Chapter 1. The evaluation is done via two different criteria - (i) an out-of-sample (cross validation) forecast in which a group of estimators computed using one dataset are applied to make predictions about medical spending in a different dataset, and (ii) a bootstrapped bias corrected in-sample mean square error. The evaluation criteria are based on Efron et al. (1993).

The estimators compared include standard panel data estimators; pooled OLS as well as linear fixed and random effects estimators, semiparametric pooled estimates as in Li and Stengos (1996), and the SPRE estimator proposed in the first chapter. Additionally we include three other popular estimators, a shrinkage estimator discussed by Maddala et al. (2001) and two finite mixture models for panel data as discussed by Deb and Trivedi (2013). The designs of the mixture models and shrinkage estimator do not readily permit one to evaluate their bootstrapped bias-corrected mean square error (MSE), so they evaluated only by using the first criterion.

The shrinkage estimator works by calculating ordinary least squares (time series) estimates for each cluster in the panel and then adjusting, or shrinking, these estimates towards the pooled estimate for the entire sample (combining clusters with time series). This is accomplished through an iterative averaging process which proceeds until convergence. The motivation for including the shrinkage model comes from several previous studies in the literature which find it to have strong and sometimes superior forecasting properties relative to other standard linear panel data estimators.

The mixture models operate on the premise of automatically assigning each cluster in the sample under consideration to one of a finite number of groups. There are standard ways of doing such sorting via the expectation-maximization (EM) algorithm, but an innovation apparently unique to Deb and Trevedi (2013) lies in allowing the fixed effects of each mixture to vary by cluster by taking the within transformation before applying the EM algorithm.

Based on the aforesaid evaluation criteria the semiparametric random effects model works very well, followed by linear fixed or random effects model; and the pooled OLS or semiparametric estimators seem to perform quite unsatisfactorily. Based on the first criterion (out-of-sample forecast), the shrinkage estimator works the best, followed by the semiparametric random effects estimators.

The third chapter of this dissertation discusses the computation of the system GMM (Blumdell and Bond, 1995) estimator under various alternative norms. The traditional GMM estimator works by minimizing the norm of a number of moment conditions derived from the assumption that future changes in the error term of the specified model will be uncorrelated with past values of the dependent variable and that changes in the past value of the dependent variable will be uncorrelated with future realizations of the error term. Traditionally this set of sample moment conditions is minimized using a weighted least squares norm with weights computed from the estimated covariance of the moment conditions. As an alternative, this chapter uses a pseudo-norm, called for our purposes a weighted nonlinear norm, which operates by constructing a transformation of the ranks of the dependent variables. Other alternative norms are also discussed including a generalized M norm which falls within the class of the weighted least squares norms discussed above and a similar estimator to the weighted nonlinear norm which blends in some qualities of the absolute value norm.

References

[1] Blundell, Richard and Bond, Stephan "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models" *Journal of Econometrics* 87.2 (1998):115-147. Print.

[2] Beck, Thorsten, Levine, Ross, and Loayza, Norma. "Finance and the Sources of Growth" *Journal of Financial Economics* 58.2 (2000): 261-300. Print.

[3] Deb, Partha, and Trivedi, Pravin "Finite Mixture for Panels with Fixed Effects" *Journal of Econometric Methods* 2.1 (2013): 35-51. Print.

[4] Di Matteo, Livio "The Income Elasticity of Health care spending. A comparison of parametric and nonparametric approaches." *The European Journal of Health Economics* 4 (2003): 23-29. Print.

[5] Henderson, D. and Ullah, A. "A nonparametric random effects estimator" *Economics Letters* 88.3 (2005): 403-407. Print.

[6] Li, Qi, and Stengos, Thanasis "Semiparametric Estimation of Partially Linear Panel Data Models" *Journal of Econometrics* 71.2 (1996): 389-397. Print.

[7] Maddala, G.S., Li, Hongyi, and Srivastava, V.K. "A Comparative Study of Different Shrinkage Estimators for Panel Data Models" *Annals of Economics and Finance* 2 (2001): 1-30. Print.

[8] Mammen, Enno, "Bootstrap and Wild Bootstrap for High Dimensional Linear Models" *The Annals of Statistics* 21.1 (1993):225-285. Print.

[9] Ruckstuhl, A.F., Welsh, A.H. and Carroll, R.J. "Nonparametric Function Estimation of the Relationship Between Two Repeatedly Measured Variables." *Statistica Sinica*, 10 (2000): 51-71. Print.

[10] Efron, Bradley and Tibshirani, Robert. "An Introduction to the Bootstrap" *Monographs on Statistics and Applied Probability* 57. Chapman & Hal (1993). Print.

[11] Wang, Naisyin, Carrol, Raymond, and Lin, Xihong. "Efficient Semiparametric Marginal Estimation for Longitudinal/Clustered Data" *Journal of the American Statistical Association* 100 (2005): 147-157. Print.

Chapter I: Semiparametric Random Effects Estimators and Their Confidence Intervals: An Examination of the Medical Engel Curve for the US States

Introduction

This chapter considers a semiparametric random effects regression model under kernel smoothing. The theoretical literature on semiparametric regression with panel/longitudinal data is surprisingly limited. Notable papers in nonparametric regression with random effects include Ruckstuhl et al. (2000) and Henderson and Ullah (2005), whereas prominent papers in the semiparametric panel data literature include Li and Stengos (1996) and Wang et al. (2005). Li and Stengos (1996) consider a pooled partially linear (semiparametric) model (in the sense of pooling cross-section and time-series together) without considering any presence of random-effects in the estimation. Following Robinson (1988), they establish root-n consistency of the marginal effect (on the response variable) of the covariate modeled parametrically in the regression with panel data. Wang et al. (2005) further show the efficiency of such marginal effects in a semiparametric regression setting. Ruckstuhl et al. (2000) consider a nonparametric random effects regression and show the consistency of the variance components as well as the nonparametric conditional mean. Henderson and Ullah (2005) discuss the consistency of the nonparametric conditional mean function as well as the marginal effects of the covariates on the response variable. However, to the best of our knowledge, there appears to be no result in the literature discussing the estimation of nonparametric conditional mean or the marginal effects of the nonparametrically modeled covariates in a *semiparametric random effect* regression setting, especially when one would like to incorporate the underlying random effects error structure.

This chapter contributes to the literature in the following ways: (1) it extends the result of Ruckstuhl et al. (2000) and Henderson and Ullah (2005) to a *partially linear regression* setting and presents an algorithm for estimating the nonparametric conditional mean as well as the marginal effects of the nonparametrically modeled covariates under the semiparametric setting. This algorithm provides feasible GLS estimators of the above mentioned nonparametric components incorporating random effects. (2) The chapter also suggests using the wild bootstrap technique to obtain interval estimates of the aforesaid nonparametric components. Note that in Ruckstuhl et al. (2000) and in Henderson and Ullah (2005) the interval estimation of the nonparametric components under random effects were not discussed. The coverage probabilities of the bootstrapped confidence intervals suggested are also checked via simulation. (3) The chapter illustrates the feasible GLS semiparametric estimators using data on US state level healthcare expenditures. The application focuses on the effect of income on medical spending vis-a-vis the medical Engel curve for the US states in particular. (4) Finally, the chapter also investigates a crucial point made in Ruckstuhl et al. (2000) and makes some interesting observations. A standard two step GLS estimator of the nonparametric conditional mean can certainly have lower asymptotic variance compared to that of the corresponding pooled estimator (that ignores random effects) but it may not have lower asymptotic bias. This contradicts our intuition from linear error component models where the random effects GLS estimators are always considered superior to their OLS counterparts if the true model exhibits random effects. This problem may very well be carried forward to a partially linear regression setting. A possible way to address the problem is by correcting for the aforementioned bias via bootstrap. Hardle et al. (2004) suggests that the wild bootstrap technique can be useful for bias correction and constructing improved confidence intervals in a generalized semiparametric additive regression setting with i.i.d. errors. Taking lead from this, a wild bootstrapped bias-corrected estimator of the nonparametric components is provided in this chapter. Although the bias correction seems to work for the nonparametric marginal effect in a purely nonparametric regression setting it does not work so well in a semiparametric regression setting and this needs further investigation.

1 Literature Review

A typical partially linear panel estimator, as in Li and Stengos (1996), with longitudinal data can be written as:

$$Y_{i,t} = X_{i,t}\beta + m(Z_{i,t}) + \nu_{i,t} \tag{1}$$

where $Y_{i,t}$ is the dependent variable, $X_{i,t}$ is a vector of variables which enters the model linearly and $Z_{i,t}$ is a vector of variables entering nonparametrically, i is the index of individuals (or clusters) within the panel, and t is the index of time periods (with the number of time periods T being fixed and with the number of individuals, n, going to infinity). This model does not include any random effects. Following methods similar to Robinson (1988), Li and Stengos (1996) discuss the root-nT consistency of $\hat{\beta}$ in this setting. Wang et al. (2005) discuss the efficiency of $\hat{\beta}$ without or even with cluster correlation in the errors assuming the true variance-covariance matrix is known. None of the papers above discuss the random effects specification in a partially linear regression setting especially when the true variance-covariance relationship is not known.

Ruckstuhl et al. (2000), however, consider a nonparametric model of the type

$$Y_{i,t} = m(Z_{i,t}) + \nu_{i,t} + u_i \tag{2}$$

where u_i captures the individual random effects. In a setting analogous to the traditional large n and small T, they discuss how to provide root-n consistent estimators of the variance components σ_ν^2 and σ_u^2, as well as the associated variance-covariance matrix of the composite error $\nu_{i,t} + u_i$. The estimated variance-covariance matrix in their setting is based on the spectral decomposition technique. They further discuss consistent and feasible estimation of $m(\cdot)$ using these variances. Henderson and Ullah (2005) consider feasible GLS estimation of $m(\cdot)$ as well as $m'(\cdot)$ (using a local linear model) in the above setting and discuss the consistency of these estimators. Although the consistencies of $m(\cdot)$ and $m'(\cdot)$ have been shown in the presence of random effects, there is no rigorous discussion on interval estimation of these components, without which inference and hypothesis testing remains problematic. This chapter combines the model in equation (1) with Ruckstuhl's specification, as in (2), and considers a partially linear model with random effects as:

$$Y_{it} = m(Z_{it}) + X_{it}\beta + u_i + \nu_{it} \tag{3}$$

The main focus here is on the interval estimation of $m(\cdot)$ in this semiparametric setting, using the same variance components and spectral decomposition technique used in Ruckstuhl et al. (2000), Henderson and Ullah (2005). Hardle et al. (2004) discuss a wild bootstrap technique for obtaining confidence intervals for the estimates of nonparametric components in a partially linear additive regression model. Note that wild bootstrap technique has been widely used in such nonparametric regression settings, see also Li and Wang (1998) and Hardle and Mammen (1993). This chapter uses the same technique for obtaining confidence interval for the nonparametric components. Since the setting is slightly different than that in Hardle et al. (2004), due to the inclusion of the random effects in the errors, we assess the coverage probabilities of the confidence interval estimates following Efron and Tibshirani (1993). The next section describes the algorithm.

2 Algorithm

As mentioned previously the model considered in the chapter takes the form:

$$Y_{it} = m(Z_{it}) + X_{it}\beta + u_i + \nu_{it} \tag{4}$$

Let X, Z, Y, be vectors containing the stacked elements of $X_{i,t}, Z_{i,t}, Y_{i,t}$ for $i \in \{1, ..., n\}$ and $t \in \{1, ..., T\}$ and ν, and u contain the stacked elements of the errors u_i and $\nu_{i,t}$ such that u contains each distinct element of u_i repeated T times. Further let $\epsilon_{i,t} = u_i + \nu_{i,t}$ and similarly let $\epsilon = u + \nu$. Therefore the model can be expressed as

$$Y = m(Z) + X\beta + u + \nu \tag{5}$$

Taking expectations of (5) under the assumption $E(u|X, Z) = 0$ and $E(\nu|X, Z) = 0$ we get

$$E(Y|Z) = E(X|Z) + m(Z) \tag{6}$$

Thus (5)-(6) gives us

$$Y - E(Y|Z) = (X - E(X|Z))\beta + u + \nu \tag{7}$$

Estimations of the the expectations $E(X|Z)$ and $E(Y|Z)$ are formed by local constant regression, but may be derived via any number of methods as discussed in Li and Stengos (1996). From (7) we can obtain the semiparametric estimator of β as

$$\hat{\beta}_{SP} = [(X - E(X|Z))'(X - E(X|Z))]^{-1}[(X - E(X|Z))'(Y - E(Y|Z))] \tag{8}$$

substituting this into equation (6) we get $Y - X\hat{\beta}_{SP} = m(Z) + errors$. Following the standard literature we consider local linear estimation of $m(\cdot)$. See, Li and Stengos (1996).

Let us define $\delta(z) = [m(z), m'(z)]$. Here $m(\cdot)$ and $m'(\cdot)$ are the nonparametric conditional mean and first partial in the local linear specification around an arbitrary fixed point of interest z. Let Z^* be a nTx2 matrix with 1's in the first column and $Z_{it} - z$ in the second column. The first step local linear estimator of $\delta(z)$ be obtained as

$$\hat{\delta}(z) = (Z^{*\prime}K(z)Z^*)^{-1}Z^{*\prime}K(z)(y - \hat{\beta}_{SP}X) \tag{9}$$

Where $K(z)$ is a nT x nT diagonal matrix whose (i,i)th element is

$$K_{i,i}(z) = \frac{1}{h}k(\frac{z_i - z}{h})$$

for an arbitrary kernel function $k(\cdot)$ symmetric about zero where $\int_{-\infty}^{\infty} k(\cdot) = 1$ where h is the bandwidth with assumptions as in Li and Stengos (1996) and Henderson et al. (2005). We use $K(z)$ taken as a standard normal kernel. Next we form an estimate of the residuals $\hat{\epsilon}$ and from this obtain estimates of σ_u^2 and σ_v^2 using spectral decomposition as in Henderson and Ullah (2005).

Let $\bar{\hat{\epsilon}}_i$ denote the average across t of the residuals for the data-points within the ith cluster and $\bar{\bar{\hat{\epsilon}}}$ denote the grand mean of the residuals across both clusters and time-periods. Using this we can estimate the variance components as

$$\hat{\sigma}_\nu^2 = \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}(\hat{\epsilon}_{it} - \bar{\hat{\epsilon}}_i)^2 \tag{10}$$

and

$$\hat{\sigma}_u^2 = \frac{n\sum_{i=1}^{n}(\bar{\hat{\epsilon}}_i - \bar{\bar{\hat{\epsilon}}})^2 - \hat{\sigma}_\nu^2}{T} \tag{11}$$

Define the variance-covariance matrix of ϵ as

$$\Omega = \sigma_\nu^2 I_{nT} + \sigma_u^2 I_n \otimes \mathbf{1_T} \tag{12}$$

where I_n is an n x n identity matrix and and $\mathbf{1_T}$ is a T x T matrix of ones. Similarly denote

$$\hat{\Omega} = \hat{\sigma}_\nu^2 I_{nT} + \hat{\sigma}_u^2 I_n \otimes \mathbf{1_T}. \tag{13}$$

Now the two step GLS estimate of $\delta(\cdot)$ can be obtained. We can estimate $\delta(z)$ as

$$\tilde{\delta}(z) = [Z^{*\prime}\sqrt{(K(z)})\hat{\Omega}^{-1}\sqrt{(K(z)})Z^*]^{-1}(Z^{*\prime}\sqrt{(K(z))}\hat{\Omega}^{-1}\sqrt{(K(z))}(Y - X\hat{\beta}_{SP})) \tag{14}$$

with $\tilde{m}(\cdot)$ and $\tilde{m}'(\cdot)$ being the analogous estimates of $m(\cdot)$ and $m'(\cdot)$ respectively where $\sqrt{K(z)}$ is simply the square root applied element wise to $K(z)$. It is found also to be helpful to update $\hat{\beta}$ by running GLS with the estimated covariance matrix plugged in for Ω i.e.

$$\tilde{\beta} = [(X - E(X|Z))'\hat{\Omega}^{-1}(X - E(X|Z))]^{-1}(X - E(X|Z))'\hat{\Omega}^{-1}(Y - E(y|Z)) \tag{15}$$

although it should be noted that the asymptotic advantage to such an update has yet to be shown[1].

3 Bootstrapped Confidence Intervals

The existing literature does not give a clear direction on how to obtain interval estimates of either nonparametric or semiparametric random effect estimators. To circumvent this problem a wild bootstrap technique is used to obtain percentile estimates for the aforementioned

[1]The consistency of $\tilde{\delta}(z)$ and $\widetilde{\beta}$ appears in the the working paper paper, Houser and Mukherjee (2015).

estimators. Note that it is fairly common in the literature to obtain interval estimates of nonparametric or semiparametric components by using the wild bootstrap technique. Also provided is a description of how this wild bootstrap technique works. Because of the panel nature of the estimator proposed in this paper, the bootstrapped confidence intervals below are constructed via re-sampling each individual (each i) within the panel. One can create a set of counterfactual samples each of the form (Y_b^*, X, Z), where only Y_b^* varies between replications while X and Z maintain their observed values. To generate Y_b^* the method uses

$$Y_{i,t,b}^* = \tilde{m}(Z_{i,t}) + X_{i,t}\beta + r_{b,i}^*(Y_{i,t} - \tilde{m}(Z_{i,t}) - X_{i,t}\beta) \tag{16}$$

where $r_{b,i}^* = -\frac{\sqrt{5}-1}{2}$ with $p = \frac{\sqrt{5}+1}{2\sqrt{5}}$ and $r_{b,i}^* = \frac{\sqrt{5}+1}{2}$ with $p = 1 - \frac{\sqrt{5}+1}{2\sqrt{5}}$. Note that the term $r_{b,i}^*$ depends only on i not on t. See, Li and Wang (1998) for details.

On each random sample, a bootstrapped confidence interval estimate is formed using the bandwidth calculated by the least squares cross-validation estimate for pooled panel estimates[2]. To determine a $1-\alpha\%$ confidence interval, I order the estimates of $\tilde{m}(z)$ and take the $((1+\alpha)/2)$th and $((1-\alpha)/2)$th values of $\tilde{m}_b^*(z)$ as the lower and upper bounds of the confidence interval. Let $L_\alpha(z)$ and $U_\alpha(z)$ denote these upper and lower bounds for the confidence interval of width $1-\alpha$.

In order to demonstrate the improved performance of our estimator in small samples as well as the quality of the underlying confidence interval, I perform a Monte Carlo simulation specified as follows. The data are simulated using two data generating processes: a cubic function $m(z) = \frac{(z-5)^3}{20}$ and a quadratic function $m(z) = \frac{(z-5)^2}{5}$ with $(T, n) = (4, 50)$ and $(T, n) = (4, 100)$. The data are generated according to the equation $Y_{i,t} = X_{i,t}\beta + m(Z_{i,t}) + u_i + \nu_{i,t}$, with $X_{i,t}, Z_{i,t} \sim N(0, 7/\sqrt{12})$, $u_i \sim N(0, \sigma_u^2)$, $\nu_{i,t} \sim N(0, 1)$, and $\beta = 0.5$. All simulation results reported here are from 400 Monte Carlo replications.

[2]See the 'np' package made available by Jeff Racine for the computation of least square cross validated bandwidth: http://cran.r-project.org/web/packages/np/np.pdf

4 Some Assessments

4.1 Integrated Mean Squared Error

The primary statistic used to assess the accuracy of the estimator is the integrated mean squared error (IMSE).

$$IMSE = \int_{Z_1}^{Z_2} (\tilde{m}(z) - m(z))^2 \tag{17}$$

In each Monte Carlo iteration, our estimates of the nonparametric component $\hat{m}(z)$ (pooled) and $\tilde{m}(z)$ (random effects) are calculated and compared to the true generating function $m(z)$. Since $\hat{m}(z)$ and $\tilde{m}(z)$ do not have defined functional forms, the integral in equation (17) is estimated via Riemann summation following the trapezoid rule. In Table 1 we see a consistent decrease in the integrated mean-square error from using the semiparametric random effects estimate of our $\tilde{m}(z)$ over the semiparametric pooled estimate $\hat{m}(z)$. As one might expect, as the number of clusters in the panel is increased from 50 to 100 the IMSE becomes lower for both $\hat{m}(z)$ and $\tilde{m}(z)$. As we turn to the marginal effect estimates, we find that our results are amplified considerably. Here we also see a consistent improvement in moving from $\hat{m}('z)$ to $\tilde{m}'(z)$, but this is typically much larger in scope, especially when the variance of the individual error component σ_u^2 is large. The parametric marginal effect also works relatively well, for example see Table 9 and Table 10.

4.2 Coverage Probabilities of Confidence Interval

I check the coverage probabilities of the bootstrapped confidence intervals along the lines of Efron and Tibshirani (1993). In order to test the coverage probabilities of the bootstrapped confidence interval I generate a confidence interval for each set of simulated data and record whether said interval contains $m(z)$ at each of three points 3,5,7. These points crudely corresponding to the median, first and third quantile of the data (to be more specific this is the first and third quartile rounded to the nearest integer away from the median.) As above this is performed for both the quadratic and cubic specifications and for n=50 and n=100 individuals (clusters)[3]. As outlined in Table 3, the 95% confidence intervals on $\tilde{m}(z)$ contains the true value for m(z) just around 95% of the time in the simulations listed below, showing almost perfect coverage. Typically the capture rates are a little improved as we move from $n = 50$ to $n = 100$, and in simulations not reported here we have also found the capture rates are improved by increasing the number of bootstrap replications used to compute these confidence intervals.

However, as we can see in Table 4, the confidence intervals for the bootstrap marginal effect are consistently showing over coverage. The only advantage to using this confidence interval for the marginal effect is that it may still be used for hypothesis testing as it consistently exceeds 95% coverage, being overly conservative, rather than demonstrating under-coverage.

4.3 Bootstrap Bias Correction

As noted in section one Ruckstuhl et al. (2000) made an interesting observation; while the two-step GLS estimator applied in a nonparametric context may potentially have a lower asymptotic variance than a pooled version of the same estimator its asymptotic bias may, in fact, be greater than that of its corresponding pooled version. This suggests that a fruitful course of investigation might be to use a bias-correction algorithm of the type discussed in Efron and Tibshirani (1993). Indeed Hardle et al. (2005) use just such an algorithm in a semiparametric additive model with i.i.d. errors. How one might extend Hardle's approach to this setting is discussed below.

Following the literature we can estimate the bias intrinsic in the estimator as $\tilde{\gamma}(z) = \sum_{b=1}^{B} \tilde{m}_b^*(z) - \tilde{m}(z)$ where B is the number of bootstrap resamples. Subtracting off this estimated bias term gives a new bias corrected estimate

$$\hat{\tilde{m}}(z) = \tilde{m}(z) - \tilde{\gamma}(z) = 2\tilde{m}(z) - \sum_{b=1}^{B} \tilde{m}_b^*(z) \qquad (18)$$

We can construct a bias corrected estimate of the marginal effect as in an analogous fashion.

[3]In these simulations $\sigma_u^2 = 1/4$

Shifting the confidence interval by this estimated bias term such that $L^*_\alpha(z) = L_\alpha(z) - \tilde{\gamma}(z)$ and $U^*_\alpha(z) = U_\alpha(z) - \tilde{\gamma}(z)$ gives us a confidence interval for these bias corrected estimates.

The coverages for the interval estimates of the nonparametric function $m(z)$ (in both a purely nonparametric setting as in Henderson et al., 2005 and the semiparametric setting as ours) appear almost perfect for $\tilde{m}(z)$ in our simulations. See Table 3 for example. Since the coverages of the estimates of $m'(z)$ do not appear as strong in either setting, see for example Table 4, the confidence intervals appearing too broad, we try to check if it improves with any bias correction of the aforesaid type. Interestingly, with the bias correction, the coverage probabilities of the interval estimates of the marginal effects appear almost perfect in the purely nonparametric setting (see Table 6), although the same do not improve in the semiparametric setting. The bias corrected interval estimates of $m(z)$, however, look the same in terms of coverages in both settings (nonparametric and semiparametric), being near perfect.

5 Application

To demonstrate the SPRE estimator in practice, it is applied to a sample of state-level medical expenditure and simple demographic data from the period between 2004 and 2007. Our estimating equation is

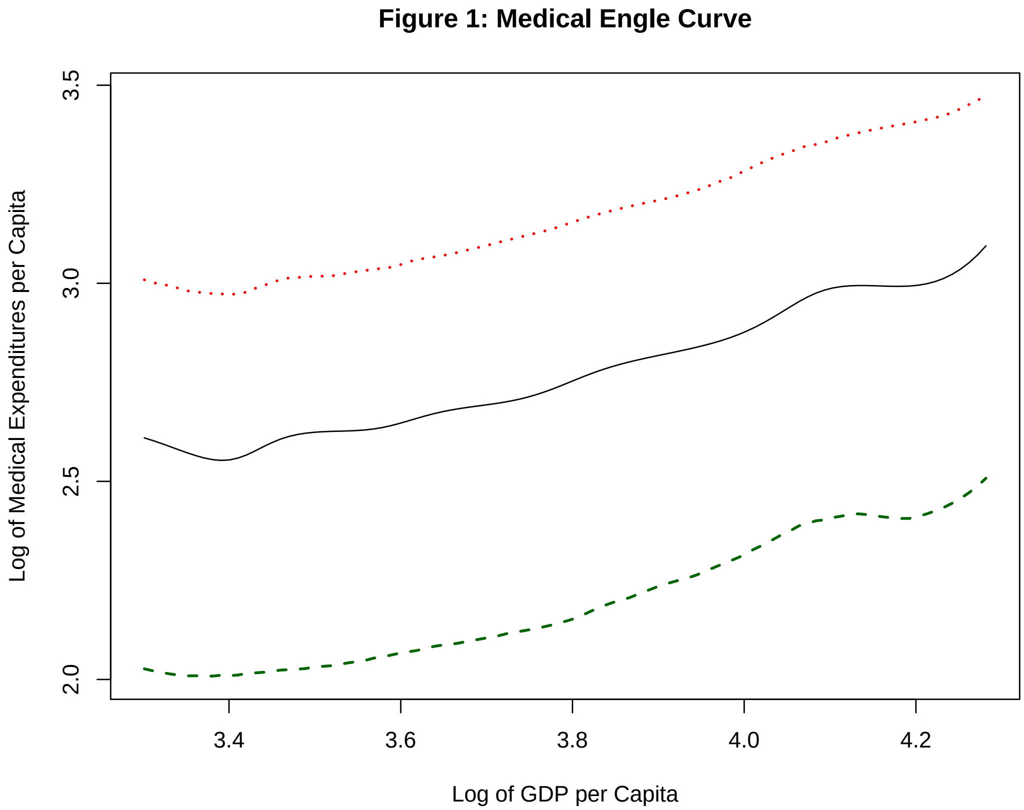
$$MedSpend_{i,t} = m(GDP_{i,t}) + \beta Above60_{i,t} + u_i + \nu_{i,t}$$

Where, $MedSpend_{i,t}$ is real medical spending per person in state i at year t, $GDP_{i,t}$ is the real state gross domestic product per capita, and $Above60_{i,t}$ is the percentage of the population above sixty years of age, and $\nu_{i,t}$ and u_i are i.i.d random variables defined as above with $\nu_{i,t}$ as the idiosyncratic error[4].

Attempting to address the impact that age and income have on healthcare spending has been very much on the national consciousness over the last ten to fifteen years. There have been many studies seeking to address the question of how the changing demographic profile of the United States is affecting our rate of healthcare expenditure. For purposes of levity, a

[4] Medical expense and GDP are in log terms.

Figure 1: Conditional mean (bandwidth: cross-validated) with confidence bands



comprehensive literature review is beyond the scope of this chapter, however readers seeking to acquaint themselves with further with the literature on this subject are directed to the excellent literature review by Martin et al. (2011). The studies also somewhat relevant to this chapter are those of Di Matteo (2003, 2005), and Okunde and Murthy (2002).

Di Matteo (2003) is interested in contrasting nonparametric estimates of the income elasticity of medical expenditures to ordinary least-squares estimates of the same. Using the LOWESS estimator on a panel of US state expenditures, Canadian province level expenditures, and OECD country level expenditures the author finds that typically lower levels of GDP per capita are associated with higher income elasticities, contrary to our findings in this chapter. They also find that income elasticities estimated using international data are typically larger than those estimated using inter-state or inter-province data. As a follow up study, Di Matteo (2005) performs the estimation again using linear parametric and LOESS estimation on US inter-state and Canadian inter-province data with a different set of covariates. The paper shows that that up to two thirds in the change in healthcare expenditures can also be attributed simply to time or technology. From this, Di Matteo (2005) concludes in favor of the so called "Newhouse Conjecture" namely that the recent increase in medical expenditures is partly due technological change. The Newhouse conjecture is difficult to confirm or deny empirically in our setting. Okunde and Murthy (2002), examining a time-series model of health sector research and development (R&D) expenditures, GDP per capita, and medical expenditure per capita find evidence consistent with the possibility of a cointegrating relationship between the three variables.

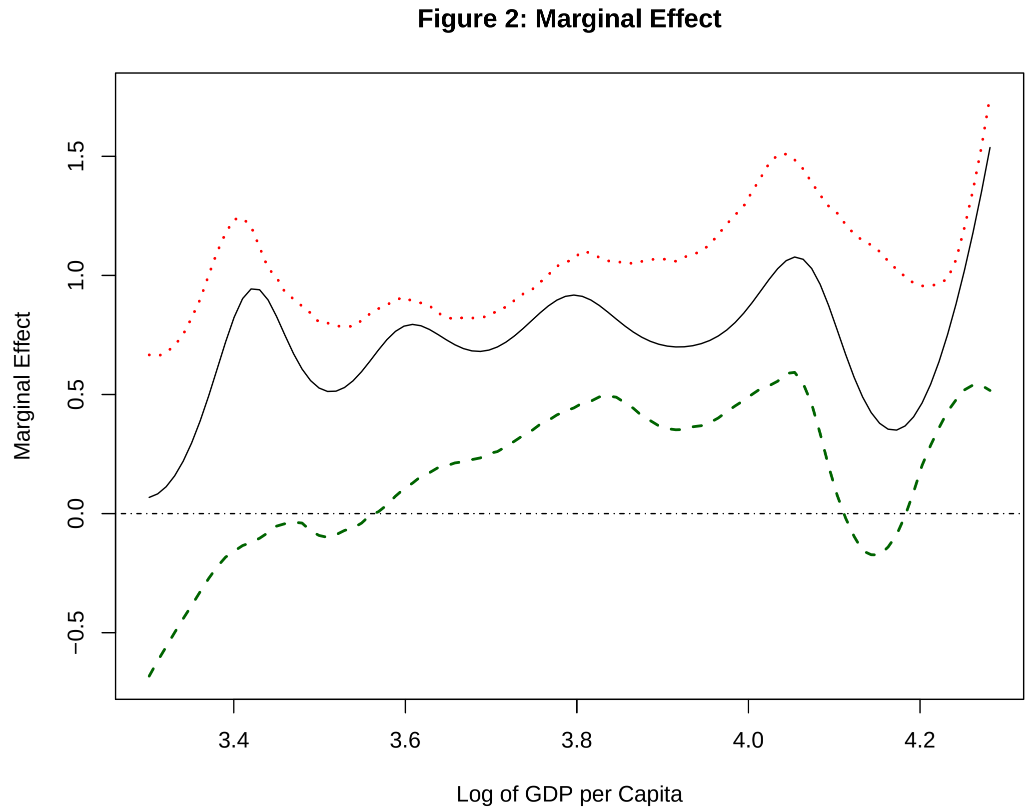
Note that the focus in this chapter is on state-wide health expenditure in a small time series context, an angle that has not explicitly been explored before. The emphasis is also on investigating the medical "Engel curve". I find that medical expenditures are typically rising with income, and our results further hint that this effect may be accelerating with GDP growth.

On the question of bandwidth selection both the least squares cross-validated bandwidth and Silverman's plug-in bandwidth are very similar for this data, about 0.064.[5]

From Figure 1 we can see that medical expenditures appear to bear a positive and significant relationship to GDP per capita, as expected, because medical expenditure is a non-inferior (normal or superior) good. However, interestingly medical expense as a good carries both 'necessity' (certain parts of the expenses can't be avoided) as well as 'luxury' (certain medical expenses/tests will be undertaken only if the individual achieves an income threshold) elements. Our marginal effects therefore seem to be positive but fairly flat perhaps due to the composite nature of the good being a necessity as well as a luxury item. See Figure 2 for illustration. However, the effects are almost always significant as the zero line remains outside the 95% confidence band of the estimators. Additionally we find the coefficient on the proportion of the population above sixty years of age to be positive and significant as expected with a value of 0.553 with an interval estimate between 0.247 and 0.758.

[5]See Li and Racine (2007) for details.

Figure 2: Marginal effect (bandwidth: cross-validated) with confidence bands



6 Conclusion

This paper extends the earlier work of Henderson and Ullah (2005) and Ruckstuhl et al. (2000) in several directions. First, by modifying these estimators to allow for a partially linear specification this paper improves the applicability of their methods to data with too many control variables. Furthermore, it is shown that the standard wild-bootstrap technique used in a nonparametric setting can be extended to our setting to create relatively well behaved confidence intervals for $m(z)$. In simulation studies we find that there is a consistent reduction in integrated mean-square error to be gained by using this semiparametric random effects estimator, as opposed to linear random effects estimators (or also the pooled semiparametric estimators), especially when the variance of the individual error term (σ_u^2) is large. The new estimators are used to examine the medical "Engel curve" for the US states, an application that has not been been explored previously.

References

[1] Di Matteo, Livio "The Income Elasticity of Health care spending. A comparison of parametric and nonparametric approaches." *The European Journal of Health Economics* 4 (2003): 23-29. Print.

[2] Di Matteo, Livio, "The macro determinants of health expenditure in the United States and Canada: assessing the impact of income, age distribution and time," *Health Policy* 71.1 (2005): 23-42. Print.

[3] Efron, Bradley and Tibshirani, Robert. "An Introduction to the Bootstrap" *Monographs on Statistics and Applied Probability* 57. Chapman & Hal (1993). Print.

[4] Henderson, D. and Ullah, A. "A nonparametric random effects estimator" *Economics Letters* 88.3 (2005): 403-407. Print.

[5] Hardle, W. and Mammen, E. "Comparing nonparametric versus parametric regression fits" *Annals of Statistics* 21.4 (1993): 1926-1947. Print.

[6] Hardle, Wolfgang, Huet, Sylvie, Mammen, Enno, and Sperlich, Stefan. "Bootstrap Inference in a Semiparametric Generalized Additive Models" *Econometric Theory* 20.2 (2004): 265-300. Print.

[7] Houser, A. and Mukherjee, D., "Semiparametric Random effects Estimators and their confidence intervals: Application to medical 'Engel curve' for the US data." *Working paper.*

[8] Li, Qi, and Stengos, Thanasis "Semiparametric Estimation of Partially Linear Panel Data Models" *Journal of Econometrics* 71.2 (1996): 389-397. Print.

[9] Li, Qi, and Wang, Suojin, "A simple consistent bootstrap test for a parametric regression function" *Journal of Econometrics* 87.1 (1998): 145-165. Print.

[10] Lin, Xihong, and Carroll, Raymond "Semiparametric Regression for Clustered Data Using Generalized Estimating Equations" *Journal of the American Statistical Association* 96 (2001): 1045-1056. Print.

[11] Martin, Jose, Gonzales, M. Puerto, and Garcia, Dolores "Review of the literature on the determinants of healthcare expenditure" *Applied Economics* 43.1 (2011): 19-46. Print.

[12] Mammen, Enno, "Bootstrap and Wild Bootstrap for High Dimensional Linear Models" *The Annals of Statistics* 21.1 (1993):225-285. Print.

[13] Okunade, A. A. and Murthy, V. N. "Technology as a major driver of health care costs: a cointegration analysis of the Newhouse conjecture" *Journal of Health Economics* 21 (2002): 147-159.

[14] Robinson, P.M. "Root-n Consistent Semiparametric Regression." *Econometrica* 54.4 (1988): 931-954. Print.

[15] Ruckstuhl, A.F., Welsh, A.H. and Carroll, R.J. "Nonparametric Function Estimation of the Relationship Between Two Repeatedly Measured Variables." *Statistica Sinica*, 10 (2000): 51-71. Print.

[16] Severini, T. A. and Staniswalis, J. G. . "Quasi-likelihood Estimation in Semiparametric Models" *Journal of the American Statistical Association* 89 (1994): 501-511.

[17] Li, QI, and Racine, Jeffrey, *Nonparametric Econometrics* Princeton University Press (2007). Print.

[18] Wang, Naisyin, Carrol, Raymond, and Lin, Xihong. "Efficient Semiparametric Marginal Estimation for Longitudinal/Clustered Data" *Journal of the American Statistical Association* 100 (2005): 147-157. Print.

7 Tables

Table 1: Semiparametric Integrated Mean Squared Error for $m(z)$ with Cross-validated Bandwidth

Key
Cubic: $m(z) = (z-5)^3/20$
Quadratic: $m(z) = (z-5)^2/5$

Set-up	SPRE	SP-Pooled	RE	(N, T)	σ_u^2/σ_ν^2
Cubic	0.156	0.158	1.054	(50, 4)	1/4
Quadratic	0.127	0.128	2.318	(50, 4)	1/4
Cubic	0.074	0.076	0.962	(100, 4)	1/4
Quadratic	0.062	0.064	2.291	(100, 4)	1/4
Cubic	0.582	0.609	1.025	(50, 4)	4
Quadratic	0.524	0.550	2.832	(50, 4)	4
Cubic	0.351	0.364	2.231	(100, 4)	4
Quadratic	0.311	0.323	4.166	(100, 4)	4
Cubic	1.741	1.856	1.261	(50, 4)	16
Quadratic	1.611	1.716	3.069	(50, 4)	16
Cubic	1.033	1.094	2.369	(100, 4)	16
Quadratic	0.938	0.999	4.260	(100, 4)	16

Table 2: Semiparametric Integrated Mean Squared Error for $m'(z)$ with Cross-validated Bandwidth

Key

Cubic: $m(z) = (z-5)^3/20$

Quadratic: $m(z) = (z-5)^2/5$

Set-up	SPRE	SP-Pooled	RE	(N, T)	σ_u^2/σ_ν^2
Cubic	0.721	0.750	1.054	(50, 4)	1/4
Quadratic	0.228	0.242	2.318	(50, 4)	1/4
Cubic	0.308	0.324	1.334	(100, 4)	1/4
Quadratic	0.100	0.108	1.336	(100, 4)	1/4
Cubic	1.031	1.124	1.386	(50, 4)	4
Quadratic	0.734	0.810	1.245	(50, 4)	4
Cubic	0.695	0.750	2.344	(100, 4)	4
Quadratic	0.442	0.485	1.460	(100, 4)	4
Cubic	1.936	2.246	1.388	(50, 4)	16
Quadratic	1.389	1.638	1.244	(50, 4)	16
Cubic	1.508	1.687	2.343	(100, 4)	16
Quadratic	1.003	1.142	1.460	(100, 4)	16

Table 3: Semiparametric Coverage Probabilities: Nonparametric Component $\tilde{m}(z)$ with Cross-Validated Bandwidth, $\sigma_u^2/\sigma_\nu^2 = 1/4$

Key

Cubic: $m(z) = (z-5)^3/20$

Quadratic: $m(z) = (z-5)^2/5$

Function	Coverage (left missed, right missed)	(N, T)	q^*
Cubic	0.9200 (0.0125, 0.0675)	(50,4)	3
Cubic	0.9550 (0.0250, 0.0200)	(50,4)	5
Cubic	0.9475 (0.0450, 0.0075)	(50,4)	7
Quadratic	0.9450 (0.0500, 0.0050)	(50,4)	3
Quadratic	0.9275 (0.0675, 0.0050)	(50,4)	5
Quadratic	0.9525 (0.0425, 0.0050)	(50,4)	7
Cubic	0.9375 (0.0200, 0.0425)	(100,4)	3
Cubic	0.9525 (0.0275, 0.0200)	(100,4)	5
Cubic	0.9425 (0.0525, 0.0050)	(100,4)	7
Quadratic	0.9325 (0.0450, 0.0225)	(100,4)	3
Quadratic	0.9450 (0.0300, 0.0250)	(100,4)	5
Quadratic	0.9475 (0.0325, 0.0200)	(100,4)	7

Table 4: Semiparametric Coverage Probabilities: Marginal Effect $\tilde{m}'(z)$ with Cross-Validated Bandwidth, $\sigma_u^2/\sigma_\nu^2 = 1/4$

Key

Cubic: $m(z) = (z-5)^3/20$

Quadratic: $m(z) = (z-5)^2/5$

Function	Coverage (left missed, right missed)	(N, T)	q^*
Cubic	0.9975 (0.0000, 0.0025)	(50,4)	3
Cubic	0.9750 (0.0250, 0.0000)	(50,4)	5
Cubic	0.9950 (0.0025, 0.0025)	(50,4)	7
Quadratic	0.9800 (0.0175, 0.0025)	(50,4)	3
Quadratic	0.9975 (0.0000, 0.0025)	(50,4)	5
Quadratic	0.9550 (0.0025, 0.0425)	(50,4)	7
Cubic	0.9850 (0.0100, 0.0050)	(100,4)	3
Cubic	0.9975 (0.0000, 0.0025)	(100,4)	5
Cubic	0.9975 (0.0000, 0.0025)	(100,4)	7
Quadratic	0.9825 (0.0125, 0.0050)	(100,4)	3
Quadratic	0.9975 (0.0000, 0.0025)	(100,4)	5
Quadratic	0.9950 (0.0000, 0.0050)	(100,4)	7

Table 5: Nonparametric Coverage Probabilities: Non-Bias Corrected Marginal Effect $\tilde{m}'(z)$ with Cross-Validated Bandwidth, $\sigma_u^2/\sigma_\nu^2 = 1/4$

Key

Cubic: $m(z) = (z-5)^3/20$

Quadratic: $m(z) = (z-5)^2/5$

Function	Coverage (left missed, right missed)	(N, T)	q^*
Cubic	0.991 (0.006, 0.003)	(50,4)	3
Cubic	0.989 (0.009, 0.002)	(50,4)	5
Cubic	0.994 (0.000, 0.006)	(50,4)	7
Quadratic	0.983 (0.014, 0.003)	(50,4)	3
Quadratic	0.997 (0.001, 0.002)	(50,4)	5
Quadratic	0.985 (0.000, 0.015)	(50,4)	7

Table 6: Nonparametric Coverage Probabilities: Bias-Corrected Marginal Effect $\hat{\hat{m}}'(z)$ with Cross-Validated Bandwidth, $\sigma_u^2/\sigma_\nu^2 = 1/4$

Key

Cubic: $m(z) = (z-5)^3/20$

Quadratic: $m(z) = (z-5)^2/5$

Function	Coverage (left missed, right missed)	(N, T)	q^*
Cubic	0.946 (0.022, 0.032)	(50,4)	3
Cubic	0.961 (0.024, 0.015)	(50,4)	5
Cubic	0.949 (0.013, 0.038)	(50,4)	7
Quadratic	0.949 (0.034, 0.017)	(50,4)	3
Quadratic	0.962 (0.018, 0.020)	(50,4)	5
Quadratic	0.944 (0.013, 0.043)	(50,4)	7

Table 7: Semiparametric Coverage Probabilities: Nonparametric Component $\tilde{m}(z)$ with Cross-Validated Bandwidth, $\sigma_u^2/\sigma_\nu^2 = 4$

Key
Cubic: $m(z) = (z-5)^3/20$
Quadratic: $m(z) = (z-5)^2/5$

Function	Coverage	(N, T)	q^*
Cubic	0.9575	(100,4)	3
Cubic	0.9475	(100,4)	5
Cubic	0.9625	(100,4)	7
Quadratic	0.9425	(100,4)	3
Quadratic	0.9450	(100,4)	5
Quadratic	0.9575	(100,4)	7

Table 8: Semiparametric Coverage Probabilities: Nonparametric Component $\tilde{m}(z)$ with Cross-Validated Bandwidth and Skewed Normal Errors, $\sigma_u^2/\sigma_\nu^2 = 1$, Skewness=1.5

Key
Cubic: $m(z) = (z-5)^3/20$
Quadratic: $m(z) = (z-5)^2/5$

Function	Coverage	(N, T)	q^*
Cubic	0.9575	(100,4)	3
Cubic	0.9375	(100,4)	5
Cubic	0.9525	(100,4)	7

Table 9: Semiparametric Coverage Probabilities: Linear Component $\tilde{\beta}$ with Cross-Validated Bandwidth, $\sigma_u^2/\sigma_\nu^2 = 1/4$

Key
Cubic: $m(z) = (z-5)^3/20$
Quadratic: $m(z) = (z-5)^2/5$

Function	Coverage (left missed, right missed)	(N, T)
Cubic	0.9450 (0.0250, 0.030)	(50,4)
Quadratic	0.9425 (0.0275, 0.030)	(50,4)

Table 10: Semiparametric Mean Square Error for β with Cross-Validated Bandwidth

Key

Cubic: $m(z) = (z-5)^3/20$

Quadratic: $m(z) = (z-5)^2/5$

Function	Sp-pooled	SPRE	(N, T)	σ_u^2/σ_ν^2
Cubic	0.0007	0.0007	(50,4)	1/4
Cubic	0.0004	0.0003	(100,4)	1/4
Cubic	0.0012	0.0007	(50,4)	1
Cubic	0.0006	0.0004	(100,4)	1
Quadratic	0.0007	0.0006	(50,4)	1/4
Quadratic	0.0004	0.0003	(100,4)	1/4
Quadratic	0.0012	0.0007	(50,4)	1
Quadratic	0.0006	0.0004	(100,4)	1

Chapter II: A Comparison of the Performance of Various Panel Data Estimators via Cross-Validation and Bootstrapping

Introduction

This chapter aims at comparing predictive performances of several competing models/estimation techniques and assessing the roles of the semiparametric random effects model (as discussed in the previous chapter) in this light. In order to perform such comparison standard out-of-sample prediction errors as well as bootstraped corrected prediction errors have been used for a sample of U.S. state-level longitudinal data on medical expenditure. The chapter also assesses predictive powers of "shrinkage models" and "gaussian mixture models" (to be discussed later).

Two classic papers comparing the performances of different panel data estimators are those of Baltagi and Griffen (1997) who estimate the country level demand for gasoline on a panel of 18 OECD countries between 1960 and 1990, and Baltagi, Griffin, and Xiong (2000) who estimate the state level demand for cigarettes on a panel of 46 american states between 1963 to 1992. The primary focus of both papers is comparing the performance of estimators of homogeneous models (such as the random effects estimator) against the performance of heterogenous estimators which come up with a separate set of coefficients for each state, such as ordinary least squares (OLS) run on each state individually. In both papers the comparison is done by breaking the sample into a set of training and test data, separated by time period, and calculating the mean-square prediction error for the trained estimator run on the test sample.

The general finding in Baltagi and Griffen (1997) is that homogeneous estimators typically outperform heterogeneous estimators with the shrinkage estimators (to be discussed below) performing somewhere in-between the two groups. In a follow up study of Baltagi, Griffen, and Xiong (2000), a similar pattern is shown with homogeneous estimators typically outperforming their heterogeneous peers in forecasting exercises. In both of these studies the underlying model is dynamic, that is to say including a lagged dependent variable in the righthand side of the regression equation, which typically recommends the use of further lagged values of the dependent variable as instruments in the model. As a result we see these instrumented estimators outperforming their uncorrected peers among the homogeneous estimators in these papers.

Another similar paper by Baltagi, Griffen, and Xiong (2002) on the same subject examines a forecasting model for natural gas prices. This paper compares estimators using a simple dynamic structure with lagged natural gas prices entering into the estimating equation. The finding in this paper is again that shrinkage type estimators and homogeneous estimators (when data are pooled across cross sections and time series, and a single marginal effect is estimated for each covariate) typically outperform heterogeneous estimators (when separate time series OLS regressions are run for each individual cluster) by a considerable margin. Again here because of the dynamic nature of the data, two stage least squares estimators outperform other methods generally as they are meant to deal with the bias introduced by the lagged dependent variable in the estimating equation.

In the area of economic geography, economists have applied similar forecasting studies to estimate the impact of numerous economic variables on cross-country population movements. The study of Brucker and Silverstovs (2006) on this subject concludes that fixed-effects style estimators appear to have the best out-of-sample performance among the various panel estimators discussed previously. Unique to this study is the inclusion of a hierarchical Bayes' estimator, which performs comparably to the fixed effects estimator in terms of generating accurate predictions.

The purpose of this chapter is not to compare homogenous versus heterogenous estimators (none of our estimators is heterogenous), but to compare the predictive performances of some standard models for panel data in particular all of which are mainly homogeneous. These models are (1) pooled least squares (OLS), (2) fixed effects (FE), (3) random effects (RE), (4) semiparametric pooled (SP) and (5) semiparametric random effects (SPRE). See chapter 1 for discussion on SPRE and Greene (2014) for discussions on all other models. Standard mean-square out-of-sample prediction error as well as bootstrap bias corrected mean-square prediction error, as outlined in Efron and Tibshirani (1993), have been employed as two main criteria for comparing predictive performances. Although there are several tests to choose between these models, for example, using the Hausman test (Hausman, 1978) to choose between fixed versus random effects models in linear setting, or using the test of Li et al. (2002) to choose between a parametric versus a semiparametric/partially linear model; such tests often do not work under various misspecifications/complications. For example, the Li et al. (2002) test will not be informative if we want to choose between parametric versus semiparametric models in the presence of random effects. Similarly the Hausman test, will not work if the model is nonlinear or partially linear. Therefore selecting a modeling strategy based on predictive performances can be an alternative, data-driven way to circumvent the problem. Note that the chapter also analyzes the out-of-sample forecast performances for "shrinkage models" (Maddala, 2001) as well as for "gaussian mixture models" (Deb and Trivedi, 2013).

The shrinkage estimators are a popular class of estimator premised on blending the strategy of independently estimating a separate set of coefficients for each cluster in the data with that of estimating a single set of coefficients across clusters. This is done through an iterative process which estimates coefficients for each cluster in the panel and then shrinks them towards the estimate for all variables in the panel. The Gaussian mixture models are a class of estimators popular outside of economics that have recently started to find favor within the community. They operate on the premise of estimating multiple regression equations by assigning distinct clusters within the data to each equation automatically. Two different variants of these estimators are studied, a pooled version of the estimator and an extension which allows for fixed effects.

In the section below we typically find that the semiparametric random effects estimator as discussed in chapter 1 performs consistently well and provides a demonstrable improvement over a conventional semiparametric pooled or parametric (FE or RE) estimators, for both the (a) cross-validated/out-of-sample exercise as well as (b) the bootstrapped-bias corrected prediction exercise. In (a) the pooled gaussian mixture model does not work well, although its fixed effects version does outperform the pooled variant. Lastly, the shrinkage estimator demonstrates the best performance in (a), although it exceeds the performance of the semiparametric random effects estimator only by a very small amount.

1 Data

The data used in this chapter stems from three primary sources. Data about population size and age composition is taken from the U.S. Census Bureau's intercensal estimates. Data on medical expenditures is provided by the Center for Medicare and Medicaid Services. Estimates of state level GDP are provided by the Bureau of Economic Analysis's Regional Economic Accounts.

According to the Bureau of Economic Analysis state level GDP's are traditionally estimated using the income method. The three major components of each state's income are captured by labor income, business taxes, and capital income. Labor and capital income is in turn estimated from the Bureau of Economic Analysis's state personal income accounts. Business taxes are computed from census estimates as well as federal and state tax agencies. Additional consideration is given to estimating government income from non-tax sources, such as licensing fees, which are done on a industry by industry basis using data provided by the regulatory agency responsible for that industry.

Estimates of medical expenditures, as mentioned above, are drawn from the Center for Medicare and Medicaid Services. They compile estimates for health spending from the medicare "state of provider" level estimates, adjusting for border crossing behavior on the part of regional residents.

The nature of the American system of insurance makes estimation of many health care expenses difficult, as the confidential nature of insurance payments to healthcare providers prevents a channel by channel analysis of hospital revenue. Gross revenue estimates are reported by all healthcare providers for tax purposes and these form the basis of healthcare expenditure data. Prescription drug and medical device expenditures encounter similar difficulties. Estimates of gross consumer expenditures on prescription drugs are provided by the Census Bureau every five years, with statistical interpolation used to estimate the intervening years. A similar method is used to estimate consumer expenditures on durable and non-durable medical products, with data provided from the national input/output tables compiled by the Bureau of Economic Analysis.

The intercensal estimates produced by the Census Bureau used here of state population and age statistics are estimated from the five year census population data with interpolated values being calculated from yearly statistics on births, deaths, and migration. Data on births and deaths are taken from the National Center for Health Statistics with location data matched up from the Federal-State Cooperative for Population Estimates. The final component of migration is determined from the social security NUMIDENT file for American citizens and with international migration statistics being estimated from the three-year American Community Survey (ACS).

Real values of income/GDP as well as medical expenditures are calculated by using the Bureau of Economic Analysis GDP deflator scaled so that 2009 is taken as the base year in order to control for the effect of inflation. To summarize, our data is taken from Bureau of Economic Analysis, Census Bureau, and Center for Medicare and Medicaid estimates.

2 Models Used for Comparison

The main goal of this chapter is to analyze the effect of per capita gdp/income ($GDP_{i,t}$) on per-capita health expenditures ($MED_{i,t}$) - the so called "medical Engel curve". The proportion of the population below 20 ($YOUNG_{i,t}$) and above 60 ($OLD_{i,t}$) are used as control covariates. Note that these two have been often used in the literature to measure the effect of the "dependency ratio" on medical expenses.

The estimators we seek to evaluate in this chapter are as follows:

(1) Ordinary least squares applied to pooled data (across states and time) or pooled OLS:

$$MED_{i,t} = \alpha + \beta_1 GDP_{i,t} + \beta_2 YOUNG_{i,t} + \beta_3 OLD_{i,t} + \epsilon_{i,t} \tag{1}$$

(2) The random effects model (RE):

$$MED_{i,t} = \alpha + \beta_1 GDP_{i,t} + \beta_2 YOUNG_{i,t} + \beta_3 OLD_{i,t} + \epsilon_{i,t} + \mu_i \tag{2}$$

Where $\epsilon_{i,t}$ is the idiosyncratic error and μ_i is the cluster specific error for the one way random effects error components model.

(3) The fixed effects model (FE):

$$MED_{i,t} = \alpha_i + \beta_1 GDP_{i,t} + \beta_2 YOUNG_{i,t} + \beta_3 OLD_{i,t} + \epsilon_{i,t} \tag{3}$$

(4) Semiparametric pooled

$$MED_{i,t} = m(GDP_{i,t}) + \beta_2 YOUNG_{i,t} + \beta_3 OLD_{i,t} + \epsilon_{i,t} \tag{4}$$

(5) Semiparametric random effects (similar to the one in chapter 1):

$$MED_{i,t} = m(GDP_{i,t}) + \beta_2 YOUNG_{i,t} + \beta_3 OLD_{i,t} + \epsilon_{i,t} + \mu_i \tag{5}$$

Additional models:

(6) Shrinkage estimators outlined in Maddala et al. (2001):

$$MED_{i,t} = \alpha_i + \beta_{1,i} GDP_{i,t} + \beta_{2,i} YOUNG_{i,t} + \beta_{3,i} OLD_{i,t} + \epsilon_{i,t} \tag{6}$$

here each of the slope coefficients as well as the intercept are estimated using OLS, and then "shrunk" towards the value of the estimator run on the pooled dataset using an iterative process. As standard notation let $\beta_i = [\alpha_i, \beta_{1,i}, \beta_{2,i}, \beta_{3,i}]$ be the vector of coefficients in the regression mentioned above and similarly let X_i be the matrix of covariates including a vector of ones for the intercept term. Let $\hat{\beta}_i^*$ be the estimated coefficients for the ith state in the panel, let $\hat{\beta}_W$ denote the pooled ordinary least squares estimates of the coefficients, and let $\hat{\beta}_i$ be the ordinary least squares estimate of β_i. Lastly let K be the number of covariates in the model. By iteratively estimating the following system of equations until convergence one can arrive at the shrinkage estimates of β_i.

$$\hat{\beta}_i^* = \left(\frac{1}{s_i^2} X_i' X_i + \hat{\Sigma}^{-1}\right)^{-1} \left(\frac{1}{s_i^2} X_i' X_i \hat{\beta}_i + \hat{\Sigma}^{-1} \hat{\beta}_W\right) \tag{7}$$

Where s_i^2 is iteratively updated as

$$s_i^2 = \frac{1}{T+2}(MED_i - X_i\hat{\beta}_i^*)'(MED_i - X_i\hat{\beta}_i^*) \tag{8}$$

and the estimated covariance between the pooled estimates of β and the iterative estimate, $\hat{\Sigma}$ is given as

$$\hat{\Sigma} = \frac{1}{n-K-1}\sum_{i=1}^{n}(\hat{\beta}_i^* - \hat{\beta}_W)(\hat{\beta}_i^* - \hat{\beta}_W)' \tag{9}$$

Each equation is estimated sequentially until the parameter estimates converge on consistent values. For initial values of β_i^* we use the OLS estimate of $\hat{\beta}$ for the ith state in the sample.

Finally, an additional estimator used in this chapter is the gaussian mixture model for panel linear regression first used by Gaffney and Smyth (1999) and implemented in the economics literature by Deb and Trivedi (2010). In this case one can assume the data can be partitioned into some finite number of distinct linear regression models of the form:

$$MED_{i,t,r} = \alpha_r + \beta_{1,r}GDP_{i,t,r} + \beta_{2,r}YOUNG_{i,t,r} + \beta_{3,r}OLD_{i,t,r} + \epsilon_{i,t,r} \tag{10}$$

where in this case the number of regression models, $r \in \{1, 2, 3\}$, is chosen at three using the Schwartz-Bayes model selection criteria. As is typical for mixture models the coefficients are fit using the expectation-maximization (EM) algorithm. The panel gaussian mixture regression model, as per its name, operates on the assumption that the error term, $\epsilon_{i,t,r} \sim N(0, \sigma_r^2)$ is normally distributed with σ_r^2 only varying between the three underlying mixtures.

As a variation on the gaussian mixture model, one can also introduce a fixed effects version which generally outperforms the finite version by a small margin. In this situation, α_i is allowed to vary between each state in the panel, giving a specification of the form:

$$MED_{i,t,r} = \alpha_i + \beta_{1,r}GDP_{i,t,r} + \beta_{2,r}YOUNG_{i,t,r} + \beta_{3,r}OLD_{i,t,r} + \epsilon_{i,t,r} \tag{11}$$

This is estimated by running the expectation-maximization algorithm on the within transformed data, that is to say that the estimated fixed effects are removed for each variable in the ith cluster by subtracting off the mean of that variable within the ith cluster. The intercept coefficients (α_i) can be estimated as the mean residual from each state. In the EM algorithm this presents a problem as the assignment of states to each underlying regression model is not always done with certainty, e.g. one state might belong to the model r=1 with 40% probability and model r=2 with 60% probability, but typically convergence of the coefficient estimates occurs only where states have been assigned to specific models with certainty. In order to handle the situation where a particular state is not assigned to a particular model with near certainty, the intercept term for state i is calculated as the weighted average of residuals across models for that state, where the weights are the probabilities that state i belongs to that particular regression mixture.

3 Prediction Errors

One can compare predictions from each of these models via two distinct methods for estimating the mean-square prediction error (MSPE). These are standard one fold out of sample cross-validation and boot-strap bias-corrected in sample mean-square prediction error. Although the estimated MSPE varies wildly between methods, the ranking of the models, which is the primary variable of interest in this study, does not. On this basis, one may therefore reasonably make recommendations as to the relative forecasting ability of models within this sphere.

For the simple one-fold cross-validation exercise the dataset is broken into two parts, a training set consisting of observations from 1991 to 2005 and a test set consisting of the annual observations between 2005 and 2009. The various estimators outlined above are run on the training set to develop estimates of the mean-square prediction error. For simplicity let $\hat{y}_{m,i,t}$ be the predicted values of the response from the mth model tested above for state i at time t and let $y_{i,t}$ be the observed value of the response. Then the cross-validated mean-square prediction error is

$$CVMSPE_m = \frac{\sum_{i=1}^{50} \sum_{t=2005}^{2009} (y_{i,t} - \hat{y}_{m,i,t})^2}{nT} \tag{12}$$

where n is the number of states (50) and T is the number of time periods within the test sample (5).

We use bootstrap bias corrected mean-squared-error forecasts to serve as an alternate evaluation strategy. See Efron and Tibshirani (1993). This is a very different strategy for estimating prediction error because it produces an estimate of prediction error in the n direction rather than the T direction. The strategy is as follows: Keeping T fixed, $b = 1, .., B$ bootstrap samples have been generated by resampling each cluster.

Let y and $\hat{y}_m$ be the true observed values of the left hand side variable and their predicted values from the mth model. Let y_b and $\hat{y}_{m,b}$ be the b-th bootstrapped sample of the left hand side variable and the prediction using the bootstrapped estimator as well as bootstrapped sample. Finally let $\tilde{y}_{m,b}$ be the prediction by using bootstrapped estimator and observed data. The bootstrap corrected prediction error (as in Efron and Tibshirani, 1993) is defined as:

$$bMSPE_m = \frac{(y-\hat{y})^2}{nT} + \frac{1}{nTB}\sum_{b=1}^{B}\left[(y-\tilde{y}_{m,b})^2 - (y_b - \hat{y}_{m,b})^2\right] \tag{13}$$

Where the quantity within brackets is the optimism/bias correction for the in-sample prediction error. For further on this as well as other cross-validation methods please see Efron and Tibshirani (1993).

4 Results

The overall results reveal a few interesting trends about the data. Each estimation method shows substantially different performance, with the shrinkage estimator leading the pack on the out of sample cross-validation problem and with the semi-parametric random effects estimator trailing closely behind.

Table 11: Out of Sample Prediction Results

Model	$CVMSPE_m$	Rank	Ratio to Benchmark (Pooled OLS)
shrinkage	0.402	1	0.325
semiparametric random effects	0.427	2	0.345
random effects	0.455	3	0.368
fixed effects	0.464	4	0.375
gaussian mixture model w/fixed effects	0.473	5	0.382
gaussian mixture model	0.515	6	0.416
semiparametric pooled	1.143	7	0.924
ordinary least squares	1.236	8	1

We do not compute the “Efron type” (as in the) bMSPE for estimators from shrinkage or gaussian mixture models. In case of shrinkage, the first stage OLS estimators for each state are computed using time series data, and then are Shrunk towards the OLS estimators from the pooled sample. The aforesaid bootstrapping is not applicable for a purely time series regression (as is used in the first stage). Similarly, for gaussian mixture, since different states may follow different distributions, it may not be wise to apply simple "Efron type" bootstrap correction (which rely on independence and identical distribution as the number of clusters increases).

Table 12: Bootstrap Bias Corrected MSE Results

Model	$bMSPE_m$	Rank	Ratio to Benchmark (Pooled OLS)
semiparametric random effects	0.153	1	0.281
fixed effects	0.154	2	0.281
random effects	0.158	3	0.288
semiparametric pooled	0.239	4	0.436
ordinary least squares	0.547	5	1

5 Conclusion

In line with Baltagi et al. (1997, 2000, 2002) this chapter shows that the shrinkage estimator outperforms the fixed and random effects estimators for medical expenditure. Mixture models, which break the data apart into a limited number of underlying regression mixtures, were shown to have weaker performance than simple fixed and random effects regression. We also see that the semiparametric random effects estimator discussed in the prior chapter performs well in this setting outperforming all of the linear regression models (FE, RE, OLS) and coming in only slightly below the shrinkage estimator in terms of $CVMSPE$.

In the bootstrapped study, we find that the semiparametric random effects model performs well, just slightly exceeding the fixed effects model in terms of $bMSPE$. It demonstrates a considerable improvement over both the semiparametric pooled model as well as pooled parametric least squares neither of which perform as well. Both the random effects estimates and the fixed effects estimates perform comparably for this dataset, which again suggests that neither offers a marked advantage over the other in terms of estimation. Our preliminary estimates indicate that there seems be little cost and perhaps a small advantage to performing estimation with the semiparametric random effects estimator as opposed to a linear model for this dataset.

This chapter takes lead from Baltagi et al. (1997, 2000, 2002) and compares predictive performances of various models. However, unlike these papers, instead of focusing on homogeneous versus heterogeneous models, or models with or without endogeneity, comparing among various alternative models (parametric or semiparametric) used widely for economic problems involving panel data. Predictive performance is also investigated from a different angle by computing bootstrapped bias corrected prediction errors. In addition, this chapter examines some new models/estimators such as shrinkage estimators or gaussian mixture models; although a more rigorous exercise for checking their predictive performances from various angles may form a future research agenda.

References

[1] Baltagi, Badi, Bresson, Georges, and Pirotte, Alain. "Comparison of Forecast Performance for Homogeneous, Heterogeneous and Shrinkage Estimators: Some Empirical Evidence from US Electricity and Natural-Gas Consumption" *Economics Letters* 76.3 (2002): 375-382. Print.

[2] Baltagi, Badi, and Griffen, James. "Pooled Estimators vs. Their Heterogeneous Counterparts in the Context of Dynamic Demand for Gasoline" *Journal of Econometrics* 77 (1997): 303-327. Print.

[3] Baltagi, Badi, Griffin, James, and Xiong, Weiwen "To Pool or Not to Pool: Homogenous Versus Heterogeneous Estimators Applied to Cigarette Demand" *The Review of Economics and Statistics* 82.1 (2000): 117-126. Print 2005.

[4] Brucker, Herbert, and Silverstovs, Boriss "On the Estimation and Forecasting of International Migration: How Relevant is Heterogeneity Across Countries?" *Empirical Economics* 31.3 (2006): 735-754.

[5] Deb, Partha, and Trivedi, Pravin "Finite Mixture for Panels with Fixed Effects" *Journal of Econometric Methods* 2.1 (2013): 35-51. Print

[6] Li, Qi, Huang, Cliff, Li, Dong, and Fu, Tsu-Tan "Semiparametric Smooth Coefficient Models" *Journal of Business and Economic Statistics* 20.3 (2002): 412-422. Print.

[7] Downey, George K. and Aman, Gerard P. "Gross Domestic Product by State Estimation Methodology" *BEA Papers*. Online.
(http://www.bea.gov/regional/pdf/gsp/GDPState.pdf)

[8] Efron, Bradley and Tibshirani, Robert. "An Introduction to the Bootstrap" *Monographs on Statistics and Applied Probability* 57.Ch17. Chapman & Hal (1993). Print.

[9] Gaffney, Scott and Smyth, Padhraic "Trajectory clustering with mixtures of regression models" *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (1999): 63-72. Print.

[10] Greene, William H. *Econometric Analysis*, Pearson Higher Ed (2014). Upper Saddle River, New Jersey. Print.

[11] Hausman J.A., "Specification Tests in Econometrics" *Econometrica* 46.6 (1978): 1251-1271. Print.

[12] Maddala, G.S., Li, Hongyi, and Srivastava, V.K. "A Comparative Study of Different Shrinkage Estimators for Panel Data Models" *Annals of Economics and Finance* 2 (2001): 1-30. Print.

[13] Swamy, P., and Ahora, S. "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models" *Econometrica* 40.2 (1971): 63-71. Print.

[14] Rapach, David and Wohar, Mark. "Testing the Monetary Model of Exchange rate Determination: a Closer Look at Panels" *Journal of International Money and Finance* 23.6 (2004): 867-895. Print

Chapter III: An Investigation of System GMM Estimation Under a Weighted Non-linear Norm

Introduction

The generalized method of moments estimation framework provides a powerful and broad conceptual device for imposing economic theory into econometric estimation. One such popular estimator, which will be the focus of this paper, is the system generalized method of moments estimator due to Arellano and Bover (1995). As will often be the case in a panel setting, we will want to include lagged values of the dependent variable as a control. A panel data model which includes such an autoregressive lag term is called a *dynamic panel*, for example

$$y_{i,t} = \alpha y_{i,t-1} + \boldsymbol{x}_{i,t}\boldsymbol{\beta} + \epsilon_{i,t} + \mu_i \tag{1}$$

Where $\epsilon_{i,t}$ varies across time and and μ_i does not. Let $\epsilon_{i,t}$ be an i.i.d random variables with expectation zero and defined variance. Let $y_{i,t}$ be our observed left-hand side variable in cluster i at time t, and let $\boldsymbol{x}_{i,t}$ be an observed vector of covariates of interest in cluster i at time t where $i \in \{1, ..., n\}$ and $t \in \{1, ..., T\}$. Lastly let α and β be constant parameters across both clusters and time. In such circumstances, taking the usual within transformation for fixed effects results in the common endogeneity problem that the transformed regressors will no longer be orthogonal to the transformed error term. As is well known, this invalidates OLS estimation, so instead we seek another method which both removes the fixed effect, μ_i and allows our regressors of interest to be corrected with the error term. As long as our idiosyncratic error term $\epsilon_{i,t}$ is uncorrelated with past values of the dependent variable, such a method exists. If we are willing to assume that prior values of $y_{i,t}$ are uncorrelated with the innovation, that is to say $E(y_{t-s,i}\epsilon_{i,t}) = 0$ for $s > 1$, then we also know $E(y_{i,t-s}\Delta\epsilon_{i,t}) = 0$. The difference GMM model operates by finding the values of α and β which minimize the sum of the sample analog of these moment conditions.

For convenience define $\boldsymbol{y}_i = [y_{i,1}, ..., y_{i,T}]'$ and $\boldsymbol{x}_i = [\boldsymbol{x}_{i,1}, ..., \boldsymbol{x}_{i,T}]'$ as the observations of $y_{i,t}$ and $\boldsymbol{x}_{i,t}$ across cluster i at time t. Let $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ be an arbitrary estimates of the parameters α and $\boldsymbol{\beta}$. Further let $\hat{\epsilon}_{i,t} = y_{i,t} - \hat{\alpha} y_{i,t-1} - \boldsymbol{x_{i,t}} \hat{\boldsymbol{\beta}}$ be the residual for the ith cluster at time period t. Let the first differenced residual be

$$\Delta\hat{\epsilon}_{i,t} = [y_{i,t} - \hat{\alpha} y_{i,t-1} - \boldsymbol{x_{i,t}}\hat{\boldsymbol{\beta}}] - [y_{i,t-1} - \hat{\alpha} y_{i,t-2} - \boldsymbol{x_{i,t-1}}\hat{\boldsymbol{\beta}}]$$

Define our vector of sample moment conditions for the ith cluster in terms of a vector valued function of our parameters and data as

$$g_i(\boldsymbol{y}_i, \boldsymbol{x}_i, \hat{\alpha}, \hat{\boldsymbol{\beta}}) = \begin{bmatrix} y_{i,1}\Delta\hat{\epsilon}_{i,3} \\ \vdots \\ y_{i,1}\Delta\hat{\epsilon}_{i,T} \\ y_{i,2}\Delta\hat{\epsilon}_{i,4} \\ \vdots \\ y_{i,2}\Delta\hat{\epsilon}_{i,T} \\ \vdots \\ \vdots \\ y_{i,T-2}\Delta\hat{\epsilon}_{i,T} \end{bmatrix} \tag{2}$$

lastly let $\mathbf{x} = [\mathbf{x_1}, ..., \mathbf{x_n}]'$ and $\mathbf{y} = [\mathbf{y_1}, ..., \mathbf{y_n}]'$ be the stacked values of all variables $x_{i,t}$ $y_{i,t}$. We can define our sample average of these moment conditions as

$$g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n} g_i(\boldsymbol{y}_i, \boldsymbol{x}_i, \hat{\alpha}, \hat{\boldsymbol{\beta}}) \tag{3}$$

The difference GMM estimator works by minimizing a norm of the above moment conditions, estimating α and β as

$$(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = argmin_{\hat{\alpha},\hat{\boldsymbol{\beta}}} ||g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})||_W \tag{4}$$

Where $|| \cdot ||_W$ is a potentially weighted least square (L2) norm of the form

$$||g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})||_W = g'(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}}) W g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}}) \tag{5}$$

and W is an arbitrary positive definite matrix called the weighting matrix.

This strategy has two drawbacks. Monte-Carlo studies, most famously that of Blundell and Bond (1998) have found these estimators to have poor finite sample properties. Furthermore, first-differencing the estimating equation rules out including time-invariant regressors, as they are eliminated along with the fixed effect term. This is not a drawback if one is only interested in including such regressors as controls, however, if one is interested in studying their coefficients it presents a problem.

As a potential remedy to both problems, Blumdell and Bond (1995) include an additional set of moment restrictions following Arellano and Bover (1995), requiring $E(\Delta y_{i,t-s}(\epsilon_{i,t} + \mu_i)) = 0$ for $S > 1$. When this assumption is reasonable it significantly improves quality of fit, but it is not always reasonable. Since $\epsilon_{i,t} + \mu_i$ contains the fixed effect μ_i, we are assuming the growth in the dependent variable is uncorrelated with the initial value of y_i. This will only be true when the model has arrived at a steady state of growth. If the model is non-stationary ($\alpha = 1$), it will never hold.

In terms of the framework above, building these added set of moment conditions into the model takes the form of lengthening the vector valued function of sample moment conditions for each cluster, $g_i(\boldsymbol{y}_i, \boldsymbol{x}_i, \hat{\alpha}, \hat{\boldsymbol{\beta}})$, such that, letting $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$.

$$g_i(\boldsymbol{y}_i, \boldsymbol{x}_i, \hat{\alpha}, \hat{\boldsymbol{\beta}}) = \begin{bmatrix} y_{i,1}\Delta\hat{\epsilon}_{i,3} \\ \vdots \\ y_{i,1}\Delta\hat{\epsilon}_{i,T} \\ y_{i,2}\Delta\hat{\epsilon}_{i,4} \\ \vdots \\ y_{i,2}\Delta\hat{\epsilon}_{i,T} \\ \vdots \\ \vdots \\ y_{i,T-2}\Delta\hat{\epsilon}_{i,T} \\ \Delta y_{i,2}\hat{\epsilon}_{i,3} \\ \vdots \\ \Delta y_{i,T-2}\hat{\epsilon}_{i,T} \end{bmatrix} \tag{6}$$

with the new set of average sample moment conditions across clusters

$$g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^{n} g_i(\boldsymbol{y}_i, \boldsymbol{x}_i, \hat{\alpha}, \hat{\boldsymbol{\beta}}) \tag{7}$$

defined as before.

Such models have, however, encountered considerable criticism in recent years for their poor finite sample properties (Newey and Windmeijer, 2009), and a handful of corrections emerged. Perhaps of equal concern to the practical econometrician, however, is robustness to misspecification/misreporting as well as improperly specified models. Unfortunately, as shown in simulations below, even when a handful of data points are misreported the system GMM estimator will go significantly awry from its targeted parameters.

The contribution of this chapter lies in using a weighted nonlinear norm to perform the same minimization. The result is an estimator which is not strictly speaking a GMM estimator, but still performs comparably in a computational study discussed below. Let $g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n} g_i(\boldsymbol{x}_i, y_i, \hat{\alpha}, \hat{\boldsymbol{\beta}})$ be defined as in equations (7) and (3), we define this norm as being

$$||g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})||_{WN} = a(R[g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})]) \cdot g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})) \tag{8}$$

Here $R[\cdot]$ is the rank function which takes a vector as an argument and returns vector stating the order of the components in the previous list and $a(\cdot)$ can be one of a number of potential score functions. From here on out in the chapter we will use

$$a(\nu) = \sqrt{12}\left(\frac{\nu}{(T-1)(T-2)+T+1} - \frac{1}{2}\right)$$

called the Wilcoxon scoring function, where the vector of ranks, ν, is of length $(T-1)(T-2)+T$.

Again we will estimate[6] our coefficients as:

$$(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = argmin_{\hat{\alpha},\hat{\boldsymbol{\beta}}} ||g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})||_{WN} \tag{9}$$

Although this estimation strategy can be used generally, as in many applications relying on robust GMM estimation, our interest is in applying this estimator where there are many moment conditions, and thus where one overly large aberrant moment condition is likely to cause issues. A common estimator which suffers from this problem, and consequently where work on improving the robustness of GMM has focused, is in the system GMM estimator due to Arellano and Bover.

1 Literature Review

This paper borrows on two enormous literatures, the first of which is generalized method of moments estimation and dynamic panel data estimation in particular, whereas the second pertains to the theory of partially-rank based estimation attributable mainly to the works of Thomas Hettmansberger, Joeseph McKean, and others (Tempstra and McKean, 2005). Within the second literature there have been several estimators which work on the premise of minimizing, instead of the squared deviations of the error term, the error terms scaled by their ranks. That is, for a linear model of the form $y_i = x_i\beta + \epsilon_i$ where $x = [x_1, .., x_n]'$ and $y = [y_1, ..., y_n]'$ we seek to minimize

$$||y - x\beta||_R = [a(R(y - x\beta))]\,(y - x\beta) \tag{10}$$

called an R norm. Here $R : \mathbb{R} \to \{a_i \in \mathbb{N} | a_i < N\}$ is a function which returns the order of the ith residual relative to the others (i.e. 1,2, 3 up to the number of moment conditions), while $a(\cdot)$ can be chosen as any non-decreasing function referred to as a score function. Such rank based norms can be shown to have smaller information loss relative to other robust

[6] In order to ensure this norm, which is in fact a pseudo-norm, forces $g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})$ towards a zero vector, it may be necessary and advisable to shift the values of y such that for a new set of y values, $y^*_{i,t} = y_{i,t} - y_{1,t}$ and to run using this set of $y^*_{i,t}$ values in place of $y_{i,t}$ in the above. In simulation results run before this has not proven necessary.

norms, such as the absolute value norm (Hettmansberger & McKean 2010, Ch.3).

Since the development of the dynamic panel GMM estimators, there have been several attempts to build more robust versions of these estimators. Unfortunately, almost all of these attempts remain unpublished. Lucas, van Dijk, and Kloek (1997) present a working paper which modifies the Arellano and Bond difference GMM matrix by reweighting each individual moment condition according to the observed error term by an outer-weighting matrix in addition to the standard inner weighting used in generalized method of moments estimators (Lucas et al., 1996). Frank Kleinbergen has written a recent working paper on correcting for the weak instrument problem in panel data estimation using resampling methods (Kleinbergen, 2011).

Within the published literature, the most successful attempt at building a robust GMM formula was done by Ronchetti and Trojani (2001) who opt for the comparatively simple method of placing a cap on the maximum allowable deviation of a moment condition from its mean. Beyond a certain amount the deviation of a moment condition from zero will have no influence on this robust GMM estimator. Although this method has seen some minor use in the financial literature, there have been no attempts to apply it to dynamic panel GMM estimators such as the method above, this is one potential area of future research. Most of the applications of the Ronchetti and Trojani algorithm have been financial in nature, as an example see Dell'Aquia et a. (2004).

Outside of the dynamic panel data and GMM robustness literatures, a very heterogenous literature on generalized method of moments and its uses has developed. One recent graduate level economics text, Hayeshi (2000), uses GMM as an organizing principle. Several more texts on generalized method of moments and dynamic panel data estimators are also available. As mentioned above, the most recent papers on GMM estimation has focused on the situation where there are many weak instruments[7], see Francis and Windmeijer (2009), and much recent research regarding GMM robustness has focused creating dynamic panel data estimates which will not be overly-biased even when the moment conditions are weakly correlated with the true parameter estimates, as an example see Kleinbergen and

[7] Instruments, in this sense, are transformations of our data which are uncorrelated with some variables within the data, for historical reasons GMM estimates are still closely associated with IV estimates, all IV estimates are types of GMM, the converse however is not true.

Mavroeidis (2012).

2 Simulated Data Description

In order to compare the weighted-nonlinear generalized method of moments estimator with the conventional GMM estimator, Monte-Carlo simulations under a variety of specifications have been examined. First, results are presented from a simulation where there is no intrinsic corruption in the sample, that is to say all data points are generated via the linear process above

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \epsilon_{i,t} \tag{11}$$

In all simulations presented the length and width of our simulated panel are set to T=10 and n=50, providing 500 data-points a large number of moment conditions. It should be noted the conventional GMM estimates presented below are one step estimates as, with the large number of moment conditions, there are insufficient degrees of freedom to estimate the traditional asymptotically efficient weighting matrix (Hayeshi, 2000). The error in the equation above can be decomposed into a large fixed effect and a smaller time-varying error such that $\nu_{i,t} = \epsilon_{i,t} + \mu_i$ where $\epsilon_{i,t} \sim N(0, 0.05)$ is a white-noise process, and $\mu_i \sim uniform(0, 1)$ is the fixed effect.

Additionally, as the system GMM estimator is commonly used in situations where investigators are concerned about endogeneity this is also taken into account. To capture this the mode is rerunl so that $E(x_{i,t}\epsilon_{i,t}) \neq 0$. In this specification, $x_{i,t}$ is originally generated from a uniform distribution on the interval $[0, 1]$ and then adjusted by 10% of ϵ, thus $E(x_{i,t}\nu_{i,t}) = E(x_{i,t}\epsilon_{i,t}) = 0.005$.

In order to test for an improvement in robustness measurement errors are simulated in two other scenarios: in the first situation 10 random value of y are chosen and whose observed values are randomly increased or decreased by an additional error term $\eta_{ij} \sim N(0, 1)$, The same procedure is repeated with 10 random values of x offset by the same factor instead. We anticipate that the weighted-nonlinear estimator will be generally outperformed by the conventional L2 estimator in the uncorrupted, outlier-free, sample but will be more robust to outliers in the samples containing these outlier data points.

These two sets of randomly altered data points are meant to capture outliers which exist outside of the scope of the model, for example typographic errors in data entry and reporting as well as miscalculations in computed secondary statistics. The effect of real outliers within the data generating process, such as natural or financial disasters, firm closings, and other unexpected economic shocks are also of interest when performing robust regression. These two types of outliers must be distinguished in this case because of the autoregressive term α which links past realizations of $y_{t,i}$ to future realizations $y_{t+s,i}$. For example, if our left-hand-side variable is gross domestic product per capita and receives a large negative shock such as the aforementioned natural disaster this will spill over into subsequent periods via the autoregressive term. This "real" outlier provides useful information for accurately estimating alpha, but potentially distorts the accuracy of beta.

To perform simulation using real outliers, a contaminated error distribution is used of the form

$$\epsilon_{i,t} = \nu_{i,t} + I_{i,t}\tau_{i,t} \tag{12}$$

where $\tau_{i,t} \sim N(0,1)$ and $\nu_{i,t}$ represents the uncontaminated error term as before. $I_{i,t}$ is a bernoulli random variable which takes the value one with $p = 0.05$ and zero with $p = 0.95$. In the tables below, this will be referred to as the contaminated error generating process, and is taken to resemble a testing procedure for outlier robustness used in Kloke and McKean (2014).

Table 13: Summary Statistics Across Simulations
($y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \nu_{i,t}$)

Simulation	Mean $y_{i,t}$	S.D. $y_{i,t}$	Max $y_{i,t}$	Min $y_{i,t}$	Mean $x_{i,t}$	S.D. $x_{i,t}$
No Added Error	0.25	0.21	1.38	-1.27	0.49	0.28
Error in $y_{i,t}$	0.25	0.24	1.59	-1.43	0.49	0.29
Error in $x_{i,t}$	0.26	0.28	3.41	-1.10	0.49	0.31
Contaminated error process	0.25	0.36	3.09	- 1.40	0.50	0.28

To measure the accuracy of our parameter estimates, the primary statistic of interest will the ratio of the root mean squared errors (rMSE's) in each parameter α and β in equation 4, between the two estimators.

$$V_\alpha = \sqrt{\frac{(\hat{\alpha}_{L2} - \alpha)^2}{(\hat{\alpha}_{WN} - \alpha)^2}} \qquad V_\beta = \sqrt{\frac{(\hat{\beta}_{L2} - \beta)^2}{(\hat{\beta}_{WN} - \beta)^2}} \tag{13}$$

where α and β are the true parameters in the model, $\hat{\alpha}_{L2}$ and $\hat{\beta}_{L2}$ are the estimates produced by the conventional norm, and $\hat{\alpha}_{WN}$ and $\hat{\beta}_{WN}$ are the estimates produced by the weighted-nonlinear norm. Values of $V_\alpha > 1$ indicate that the weighted-nonlinear norm outperforms the L2 norm, and values of $V_\alpha < 1$ imply the converse.

Finally we consider estimation under two additional norms. The first of these is a type of generalized-M norm as outlined in Kloke and McKean (2014) and the second of these is the weighted-nonlinear-norm as outlined above, except with a sign-score function instead of the Wilcoxon score function. Let let $g_m(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})$ be the mth element of $g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})$. In this case the GM norm is a weighted L2 norm of the sort typical to GMM estimation of the form $||g(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})||_{GM} = \sum_{i=1}^{N} w_i g_m^2(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})$, where $w_i = min(1, c/\nu_i)$ and

$$\nu_i = \frac{(\bar{g}_m(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}}) - g_m(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}}))^2}{v\hat{a}r(g_m(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}}))}$$

where c in this case is set, as is traditional, to the 95th percentile of a $\chi^2(1)$. In order to ensure robustness, here $\bar{g}_m(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}})$ and $v\hat{a}r(g_m(\boldsymbol{x}, \boldsymbol{y}, \hat{\alpha}, \hat{\boldsymbol{\beta}}))$ are estimated using the Minimum Covariance Determinant as calculated using the MCD package in R based on the algorithm of Rousseeuw, and van Driessen (1999) We also consider a sign norm of the form $a(\cdot) = \left(\frac{|\cdot|}{(T-1)(T-2)+T+1} - \frac{1}{2}\right)$. Although these alternative norms are not the focus of the study their root-mean-square errors are reported in Table 18 below.

3 Empirical Results

The empirical results of this analysis are mixed, but do confirm that the weighted-nonlinear-norm is more robust to outliers in the data. Both estimators can be shown to behave poorly when faced with even very small levels of corruption for the autocorrelation parameter α. As anticipated, when faced with measurement error in the data, the weighted nonlinear estimator does continue to provide sound estimates for β, where as the conventional estimator suffers more severely. The most immediate implication is that conventional system GMM estimators cannot be trusted to provide accurate estimates even when a small percentage of data points suffer from severe measurement error.

In the situation where there is no introduced measurement error in the outliers in x, unsurprisingly the conventional (L2) system GMM estimator, that is to say a weighted norm with the matrix of weights W set equal to the identity matrix, outperforms the weighted-nonlinear (WN) estimator described above, although the efficiency loss in this case is quite small.

Table 14: Estimator Comparisons for Simulation with no Introduced Measurement Error
$(y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \nu_{i,t})$
$corr(x_{i,t}\nu_{i,t}) = 0$

Parameter	$Mean(\hat{\theta}_{L2})$	$Mean(\hat{\theta}_{WN})$	$Mean[(\hat{\theta}_{L2} - \theta)^2]$	$Mean[(\hat{\theta}_{WN} - \theta)^2]$	$\frac{rMSE(\hat{\theta}_{L2})}{rMSE(\hat{\theta}_{L2})}$
$\alpha = .50$	0.493	0.492	0.011	0.012	0.927
$\beta = .25$	0.255	0.256	0.017	0.018	0.990

$corr(x_{i,t}\nu_{i,t}) = 0.10$

Parameter	$Mean(\hat{\theta}_{L2})$	$Mean(\hat{\theta}_{WN})$	$Mean[(\hat{\theta}_{L2} - \theta)^2]$	$Mean[(\hat{\theta}_{WN} - \theta)^2]$	$\frac{rMSE(\hat{\theta}_{L2})}{rMSE(\hat{\theta}_{L2})}$
$\alpha = .50$	0.493	0.492	0.011	0.012	0.926
$\beta = .25$	0.256	0.256	0.018	0.018	0.991

Both estimators perform quite well with very modest mean squared errors. Undisturbed by outliers, the L2 GMM estimator surpasses the weighted-nonlinear generalized method of moments estimator. The WN estimator operates at approximately 92% efficiency compared to the L2 estimator for alpha, and 99% efficiency of the L2 estimator for β, in line with expectations.

Introducing ten random severe mis-measurements among first the y and then the x variables significantly harms the efficiency of both estimators. Unfortunately the weighted-nonlinear based estimator usually does not offer sufficient protection in this case to offset the damage done. Given that only 2% of the y variables in our sample are misrepresented, this is sufficient to introduce a great deal of mistrust into the system GMM estimator

Both estimates of the autoregressive parameter α completely break down and in repeated simulations, not shown here, tend to converge towards zero. With regards to β, which measures the simultaneous effect of x on y, the weighted-nonlinear estimator finally starts

Table 15: Estimator Comparisons for Simulation with Introduced Measurement Error in $y_{i,t}$

$$(y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \nu_{i,t})$$
$$corr(x_{i,t}\nu_{i,t}) = 0$$

Parameter	$Mean(\hat{\theta}_{L2})$	$Mean(\hat{\theta}_{WN})$	$Mean[(\hat{\theta}_{L2} - \theta)^2]$	$Mean[(\hat{\theta}_{WN} - \theta)^2]$	$\frac{rMSE(\hat{\theta}_{L2})}{rMSE(\hat{\theta}_{WN})}$
$\alpha = .50$	0.018	0.013	0.504	0.506	0.996
$\beta = .25$	0.354	0.251	0.298	0.210	1.421

$$corr(x_{i,t}\nu_{i,t}) = 0.10$$

Parameter	$Mean(\hat{\theta}_{L2})$	$Mean(\hat{\theta}_{WN})$	$Mean[(\hat{\theta}_{L2} - \theta)^2]$	$Mean[(\hat{\theta}_{WN} - \theta)^2]$	$\frac{rMSE(\hat{\theta}_{L2})}{rMSE(\hat{\theta}_{WN})}$
$\alpha = .50$	-0.076	-0.045	0.371	0.334	1.11
$\beta = .25$	0.466	0.288	0.137	0.111	1.21

to perform, only being slightly biased by the outliers while again the conventional GMM Estimator breaks down completely. Since in most dynamic panel data models, β and not α is the parameter of interest, this is heartening.

In Table 16 as in Table 15, ten randomly selected variables are "corrupted" by adding large measurement noise to their true value.

Table 16: Estimator Comparisons for Simulation with Introduced Measurement Error in $x_{i,t}$

$$(y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \nu_{i,t})$$
$$corr(x_{i,t}\nu_{i,t}) = 0$$

Parameter	$Mean(\hat{\theta}_{L2})$	$Mean(\hat{\theta}_{WN})$	$Mean[(\hat{\theta}_{L2} - \theta)^2]$	$Mean[(\hat{\theta}_{WN} - \theta)^2]$	$\frac{rMSE(\hat{\theta}_{L2})}{rMSE(\hat{\theta}_{WN})}$
$\alpha = .50$	0.467	0.467	0.042	0.043	0.989
$\beta = .25$	0.368	0.185	0.086	0.081	1.062

$$corr(x_{i,t}\nu_{i,t}) = 0.10$$

Parameter	$Mean(\hat{\theta}_{L2})$	$Mean(\hat{\theta}_{WN})$	$Mean[(\hat{\theta}_{L2} - \theta)^2]$	$Mean[(\hat{\theta}_{WN} - \theta)^2]$	$\frac{rMSE(\hat{\theta}_{L2})}{rMSE(\hat{\theta}_{WN})}$
$\alpha = .50$	0.466	0.467	0.043	0.044	0.989
$\beta = .25$	0.178	0.185	0.087	0.082	1.061

As we can see here, both autoregressive parameter estimates are reduced downward, and remain very close to each other in specification. Now that the instruments used to estimate the autocorrelation coefficient are slightly weaker, this measurement continues to converge downward. When it comes to the estimate of β both estimators perform equivalently, having roughly the same sample mean across simulations and a very similar variance. The weighted-nonlinear norm clearly offers us no special protection over the conventional norm to outliers in our non-autoregressive variable x, nor does it seem to be at a particular disadvantage in this case.

Table 17: Estimator Comparisons for Simulation with contamination introduced in $\epsilon_{i,t}$
$(y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \nu_{i,t})$
$corr(x_{i,t}\nu_{i,t}) = 0$

Parameter	$Mean(\hat{\theta}_{L2})$	$Mean(\hat{\theta}_{WN})$	$Mean[(\hat{\theta}_{L2} - \theta)^2]$	$Mean[(\hat{\theta}_{WN} - \theta)^2]$	$\frac{rMSE(\hat{\theta}_{L2})}{rMSE(\hat{\theta}_{WN})}$
$\alpha = .50$	0.022	0.088	0.504	0.460	1.09
$\beta = .25$	0.368	0.280	0.289	0.226	1.282

$corr(x_{i,t}\nu_{i,t}) = 0.10$

Parameter	$Mean(\hat{\theta}_{L2})$	$Mean(\hat{\theta}_{WN})$	$Mean[(\hat{\theta}_{L2} - \theta)^2]$	$Mean[(\hat{\theta}_{WN} - \theta)^2]$	$\frac{rMSE(\hat{\theta}_{L2})}{rMSE(\hat{\theta}_{WN})}$
$\alpha = .50$	0.07	0.12	0.46	0.43	1.08
$\beta = .25$	0.50	0.37	0.37	0.26	1.45

In the contaminated model without endogeneity in the regressors, reported in Table 17, we see similar results as above with the weighted-nonlinear estimator outperforming the conventional GMM estimator in terms of accuracy, especially for the typical parameter of interest, beta. However, both estimators are very inaccurate with both the contaminated error distribution and with x being correlated with the error term. Thus in situations where concerns about endogeneity are present as well as large shocks to the error term in the model, neither estimator can be recommended.

Finally, we include the sign and GM norm estimators discussed in the section above. Let $\hat{\theta}_S$ be the estimated values of α and β from the sign-norm estimator and $\hat{\theta}_{GM}$ be the estimated values of α and β from the GM-norm estimates.

The sign norm, the weighted-nonlinear norm, and GM estimators perform similarly in each situation. The GM estimator appears to perform slightly better than the others across simulations in the presence of errors in the y space, although this effect is not large relative to the increased variance of all three estimators in this circumstance. Each estimator also performs comparably in the x-space, including the L2 norm estimator, showing that all three do not offer significant improvement in protection against distortions of the data in the x variable.

Table 18: Comparison of Alternative Norms

No Simulated Error

	$rMSE(\hat{\theta}_{L2})$	$rMSE(\hat{\theta}_{WN})$	$rMSE(\hat{\theta}_{GM})$	$rMSE(\hat{\theta}_{S})$
α	0.049	0.053	0.0490	0.036
β	0.013	0.011	0.012	0.012

Error in $y_{i,t}$

	$rMSE(\hat{\theta}_{L2})$	$rMSE(\hat{\theta}_{WN})$	$rMSE(\hat{\theta}_{GM})$	$rMSE(\hat{\theta}_{S})$
α	0.679	0.668	0.618	0.665
β	0.233	0.175	0.168	0.171

Error in $x_{i,t}$

	$rMSE(\hat{\theta}_{L2})$	$rMSE(\hat{\theta}_{WN})$	$rMSE(\hat{\theta}_{GM})$	$rMSE(\hat{\theta}_{S})$
α	0.171	0.187	0.205	0.195
β	0.072	0.069	0.072	0.070

4 Conclusion

Although in models without any data measurement errors the conventional L2 norm outperforms the weighted nonlinear norm by a very small margin, in models with measurement error a trade-off occurs where the weighted nonlinear norm becomes relatively more accurate for $\boldsymbol{\beta}$ and seems to hold to similar accuracy for α. This may seem like an even trade, but as accurate estimation of α is almost immediately rendered impossible by even a small number of outliers, it seems more immediately useful to concentrate on β, which is almost always the parameter of interest in dynamic panel data model estimation. We are less interested in studying the degree of conditional autocorrelation in y than we are in studying the impact that a change in our, loosely termed, exogenous variable x has on y. As long as the focus of estimation remains on β and not α, which we view as likely, the weighted nonlinear GMM estimator would seem preferable. As charted in the literature review, the lack of robustness of system GMM estimates has been a long known and studied problem. However, more general robust GMM frameworks such as those developed by Ronchetti and Trojani (2001) may prove applicable to the problem at hand, certainly their performance should be examined against the weighted-nonlinear GMM and conventional GMM estimates as part of future work.

The fact that the generalized-M estimator discussed above belongs to the class of weighted least-squares norms suggests that it should obey many of the properties of standardized generalized method of moments estimators. These include allowing construction of asymptotically valid confidence intervals via the usual methods, analytic tractability, and the applicability of Hansen's J test to the data generated by the estimator. Future work may also focus on extension and testing of the estimator described in Kloke et al. (2009) in such a dynamic panel setting.

Further potential extensions of the alternative system GMM estimators proposed here will work on moving away from using numerical minimization methods such as the NLM($\cdot$) function in R to compute estimates of $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$. The penchant of non-analytic minimizers to misestimate especially in the presence of erratic data is a concern, and this may be the source of some of the poor performance in the simulation, especially for estimates of the lagged dependent variable α. As an alternate approach, future consideration may be given to reworking this estimator so it might be expressed as a nonlinear estimation problem. In this situation standardized r-code is available for computing the estimates as discussed in Kloke and McKean (2014).

References

[1] Arellano, Manue, and Bover, Olympia "Another look at the instrumental variable estimation of error-components models" *Journal of Econometrics* 68.1 (1995): 29-51. Print.

[2] Blundell, Richard and Bond, Stephan "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models" *Journal of Econometrics* 87.2 (1998):115-147. Print.

[3] Dell'Aquila, Rosario, Ronchetti, Elvezio, and Trojani, Fabio. "Robust GMM analysis of models for the short Rate Process" *Journal of Empirical Finance* 10.3 (2003): 373-397. Print.

[4] Hansen, Lars Peter "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50.4 (1982):1029-1054. Print.

[5] Hayeshi, Fumio. *Econometrics* Princeton: Princeton University Press. (2000). Print.

[6] Hettmansperger, and Mckean, Joeseph. *Nonparametric Statistical Methods.* London: Chapman & Hall, (2010). Print.

[7] Kleinbergen, Frank "Improved accuracy of weak instrument robust GMM statistics through bootstrap and Edgeworth approximations" *Working Paper.* (2011). Unpublished.

[8] Kleibergen, Frank, and Mavroeidis, Sophocles. “Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve” *Journal of Business and Economic Statistics.* 27.3 (2009): 293-311. Print.

[9] Kloke, John and McKean, Joeseph. “Ch.7, Regression II” *Nonparametric Statistical Methods Using R* (2014), Chapman-Hall: Boca Raton, FL.

[10] Kloke, John, McKean, Joeseph, and Rashid, Mushfiqur. “Rank-Based Estimation and Associated Inferences for Linear Models With Cluster Correlated Errors” *Journal of the American Statistical Association* 104.485 (384-390). Print.

[11] Lucas, Andre, Van Dijk, Ronald, and Kloek, Teun. “Outlier robust GMM estimation of leverage determinants in Linear Dynamic Panel Data Models” *Working Paper.* (1997): Unpublished.

[12] Newey, Whitney and Windmeijer, Frank. ”GMM with Many Weak Moment Conditions.” *Econometrica* 77.3 (2009): 687-719. Print.

[13] Ronchetti, Elvezio, and Trojani, Fabio. “Robust Inference with GMM Estimators” *Journal of Econometrics* 101.1 (2001): 37-69. Print.

[14] Rousseeuw, P. J. and van Driessen, K. “A fast algorithm for the minimum covariance determinant estimator.” *Technometrics* 41 (1999): 212-223.

[15] Terpstra, Jeff, and Mckean, Joseph. “Rank-Based Analyses of Linear Models Using R.” *Journal of Statistical Software* 14.7 (2005): 1-26. Online.

Conclusion and Discussion of Future Work

Chapter 1 presents an algorithm for estimating nonparametric components in a semiparametric (partially linear) regression setting with individual/cluster random effects. While one group of papers in the existing literature have discussed semiparametric estimation with no random effects, a second group of works have discussed purely nonparametric regression with random effects. Chapter 1 contributes by combining these two, exploring semiparametric regression with random effects. It also contributes by proposing a bootstrap based technique for obtaining interval estimates of the nonparametric conditional mean function and the proposed interval estimates have almost accurate coverage probabilities. The proposed estimators seem to work well in terms of their integrated mean square error (IMSE) as well. An illustration has been provided by estimating the state level medical 'Engel curve' (medical expenditure as a function of income) for the United States. The interval estimates of the marginal effects of nonparametric components, however, seem to have over-coverage in various Monte-Carlo simulations considered. Potential remedies for this still need to be sought, perhaps in the form of alternative bootstrapping strategies.

Several interesting future research ideas stem from Chapter 1. One of the problems that has been identified in the nonparametric literature (Ruckstuhl et al., 2000) with random effects is that although random effects estimators seem to enjoy lower asymptotic variance compared to the pooled least squares estimators, they may suffer from higher asymptotic bias. This seems to contradict our conventional wisdom from the linear regression setting where random effects estimators are almost always found to be superior to the OLS estimators (which ignore random effects structure in the variance) in terms of the MSEs, as long as the true variance structure exhibits random effects. The aforementioned bias in the nonparametric components, as indicated in the literature (in the context of purely nonparametric regression), seems to be present in the context of semiparametric regression as well. Therefore, The search for a local nonparametric estimator preserving the underlying random effects structure (in the variance components) which will consistently produce a lower asymptotic MSE and a reasonably narrow confidence interval (compared to its "pooled" counterpart that does not preserve the underlying random effects structure in the clusters) remains open. Following Hardle et al. (2004), an attempt has been made in Chapter 1 to correct for the aforesaid bias in the nonparametric components. *Notably, coverage probabilities of the the interval estimates of nonparametric marginal effects improve (lowering the over coverage problem) with the bias correction.* More rigorous future research, therefore, needs to be undertaken in this important direction.

While performing simulations in the first chapter it was also noticed that for a particular fixed point z the accuracy of the estimate of the nonparametric conditional mean at that point, $m(z)$, could often be improved in simulations by replacing the estimate of $m(z)$ with the mean of the estimates from each bootstrapped sample. Whether this is a replicatable phenomenon or merely an idiosyncrasy of the data generating process remains an open question but some further investigation is perhaps warranted.

As a tie-in with the third chapter of this dissertation, it would be useful and fairly straight-foreword to extend the current simulation studies with a contaminated error distribution to gauge some sense of the robustness of the semiparametric estimators. At present almost all data generation is done using gaussian errors. Two particular extensions can be undertaken, namely simulation under distributions exhibiting positive skewness, such as as skewed normal distribution, and distributions drawn from a contaminated generating process where a handful of generated error terms are offset through the addition of a separate contaminating error term with an alternative distribution as in Ronchetti and Trojani (2001).

In the second chapter a variety of estimators were compared (including the standard pooled ordinary least squares, random effects, and fixed effects estimators) to the semiparametric random effects (SPRE) estimator as discussed in Chapter 1. In panel data settings, the existing literature has compared between estimators from pooled data (pooled across clusters and time) versus individual cluster wise estimators and the former seem to have much better predictive performance than the latter in terms of in-sample prediction error. This chapter contributes by extending this idea in two important directions. It compares predictive performance of various alternative pooled estimators that have not been explored before. The chapter also scrutinizes predictive performance of the alternative estimators from two entirely different and more robust angles - (i) out-of-sample predictive performance, as well as (ii) bootstrapped bias corrected predictive performance of the estimators. In addition a shrinkage type estimator and two different gaussian mixture models were examined. Of these, the shrinkage estimator was found to perform best in the out-of-sample cross validation exercise but with the SPRE estimator showing a similar level of accuracy. The bootstrap-based criterion assessing performance among the first five estimators discussed also ranked the semiparametric estimator best among those under consideration by a small margin.

Such comparison exercises always leave open the possibility of comparing additional classes of estimators. For brevity only eight distinct estimators were studied, but on the basis of this analysis other potentially interesting estimators could be reviewed. The shrinkage estimators performed very well in the simulation study, however there are many distinct variants of the shrinkage estimator as discussed by Maddala et al. (2001) and comparing the performance of these variants for forecasting medical spending may be informative. Similarly it may also be informative to compare other nonparametric and semiparametric models such as a purely nonparametric model, or a semiparametric model with more than one covariate modeled nonparametrically.

Note that this study only uses a small number of covariates. While this has the advantage of keeping the analysis simple and readily understood it would be worthwhile to experiment with the inclusion of other sets of demographic covariates, as in Di Matteo (2003), which might allow for a broader set of comparisons. One can examine how well the semiparametric random effects estimators work in the presence of many covariates. More specifically one might investigate how many covariates can be modeled nonparametrically for any given sample size exploring the "curse of dimensionality" in detail from the point of view of predictive performance.

In the third chapter estimation of a system GMM type model is discussed under various alternative norms and compared to the standard weighted least squares norm. The alternative weighted-nonlinear norm was found to offer some improvements in robustness to outliers among the dependent variable in the model and the error term, but not outliers in the independent variable. This final chapter is of a more investigatory nature and as such there are many different possibilities and alternative specifications remaining to be tried with regards to the problem it considers.

Given the exploratory nature of the third chapter, there are many remaining potential improvements. The most immediate next step involves moving away from numerically driven estimates of the parameter on the autoregressive term (α) and independent variable (β) to either an analytic or more sound iterative process for calculating the estimates of these parameters. Several possibilities present themselves, at present the most practical option may be to re-express the current minimization problem as a nonlinear estimator, (as each set of moment conditions is ultimately a transformation of the current data-generating process), at which point standard minimization algorithms are available which may serve to provide for more stable estimates.

Another robust norm which was discussed briefly in the prior chapter and may be improved upon would be the generalized-M norm. At present, it is a weighted norm of the type commonly used in GMM estimators and as such should be amenable to the usual methods of generating asymptotic confidence intervals as well as allowing for a Hansen test of the validity of the moment restrictions. However, at present, the numerical minimization algorithm used to compute the first stage estimates the weights are shifted at each iteration of the algorithm and therefore the applicability of these methods is unclear. If instead the weights were fixed, for example by a first stage estimate of the model using something like least trimmed squares, the derivation of a set of asymptotic confidence intervals may be available.

In conclusion, the thesis provides guidance on how to perform random effects estimation in a semiparametric regression setting with kernel smoothing and demonstrates a method for creating confidence intervals for the estimators. It further compares the predictive performance of the semiparametric random effects estimator to alternative panel data estimators and finds it to perform favorably. Lastly, an investigation of the behavior of the system generalized method of moments estimator under various alternative norms conducted, and some preliminary results on how to improve robustness of this estimator have been discussed.

References

[1] Blundell, Richard and Bond, Stephan "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models" *Journal of Econometrics* 87.2 (1998):115-147. Print.

[2] Di Matteo, Livio "The Income Elasticity of Health care spending. A comparison of parametric and nonparametric approaches." *The European Journal of Health Economics* 4 (2003): 23-29. Print.

[3] Hardle, Wolfgang, Huet, Sylvie, Mammen, Enno, and Sperlich, Stefan. "Bootstrap Inference in a Semiparametric Generalized Additive Models" *Econometric Theory* 20.2 (2004): 265-300. Print.

[4] Maddala, G.S., Li, Hongyi, and Srivastava, V.K. "A Comparative Study of Different Shrinkage Estimators for Panel Data Models" *Annals of Economics and Finance* 2 (2001): 1-30. Print.

[5] Ronchetti, Elvezio, and Trojani, Fabio. "Robust Inference with GMM Estimators" *Journal of Econometrics* 101.1 (2001): 37-69. Print.

[6] Ruckstuhl, A.F., Welsh, A.H. and Carroll, R.J. "Nonparametric Function Estimation of the Relationship Between Two Repeatedly Measured Variables." *Statistica Sinica*, 10 (2000): 51-71. Print.