Inference in Linear Panel Data Models with Serial Correlation and an Essay on the Impact of 401(k) Participation on the Wealth Distribution

by

Christian Bailey Hansen

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2004

Author ....
Department of Economics
May 15, 2004

Certified by ....
Whitney Newey
Professor of Economics
Thesis Supervisor

Certified by ....
Victor Chernozhukov
Assistant Professor of Economics
Thesis Supervisor

Accepted by ....
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Inference in Linear Panel Data Models with Serial Correlation and an Essay on the Impact of 401(k) Participation on the Wealth Distribution

by

Christian Bailey Hansen

Submitted to the Department of Economics
on May 15, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis considers inference issues in serially correlated multilevel and panel data and presents a separate essay that examines the impact of 401(k) participation on wealth.

The first chapter examines generalized least squares (GLS) estimation in data with a grouped structure where the groups may be autocorrelated. The analysis presents computationally convenient methods for obtaining GLS estimates in large multilevel data sets and discusses estimation of covariance parameters for use in GLS when the shock follows an AR(p) process. Standard estimates of the AR coefficients will typically be biased due to the inclusion of group level fixed effects, so a simple bias correction for the AR coefficients is offered which will be valid in the presence of fixed effects and group specific time trends. The chapter concludes with a simulation study that illustrates the usefulness of the derived methods.

The second chapter further explores inference in serially correlated panel data by considering the asymptotic properties of a robust covariance matrix estimator which is advocated for use in panel data. The estimator has good properties when the cross-section dimension, n, grows large with the time dimension, T, fixed. However, many panel data sets are characterized by a non-negligible time dimension. Chapter 2 extends the usual analysis to cases where $T \to \infty$, showing that t and F tests based on the robust covariance matrix estimator display their usual limiting behavior as long as $n \to \infty$ with T. When $T \to \infty$ with n fixed, the results show that t and F statistics can be used for inference despite the fact that the robust covariance matrix estimator is not consistent but converges to a limiting random variable. The properties of tests based upon the robust covariance matrix estimator are examined in a short simulation study.

The final chapter uses instrumental variables quantile regression to examine the effects of participating in a 401(k) on wealth. The results show the effects of 401(k) participation on net financial assets are positive and significant over the entire range of the asset distribution and that the increase in the lower tail appears to translate completely into an increase in wealth. However, there is evidence of substitution between net financial assets and other forms of wealth in the upper tail of the distribution. The results demonstrate that estimates of treatment effects which focus on a single feature of the outcome distribution may fail to capture the full impact of the treatment and that examining additional features may enhance our understanding of the economic relationships involved.

Thesis Supervisor: Whitney Newey
Title: Professor of Economics

Thesis Supervisor: Victor Chernozhukov
Title: Assistant Professor of Economics

Acknowledgments

I am extremely grateful to a number of people who have helped make completing this dissertation possible. I am greatly indebted to Victor Chernozhukov who has been an excellent advisor and has also allowed me to write papers with him from my first year as a graduate student. I would definitely not be where I am without Victor's support, encouragement, and patience. Whitney Newey has also been a great advisor, and I have benefited tremendously from his wisdom both as an economist and as an individual.

Many other faculty members both at MIT and elsewhere have also been important in developing this thesis. It was through conversations with Josh Angrist that I became interested in the topics discussed in the first two chapters of this thesis, and Josh provided numerous useful comments during their development. Interest in the topic of 401(k)'s came through Jim Poterba's public finance course, and Jim has been very generous with his time in providing comments on my dissertation and in providing helpful career advice. I will always be grateful to Jim McDonald who allowed me to be his research assistant while I was an undergraduate at BYU and thereby introduced me to research in economics and econometrics. I have also benefitted greatly through conversations and interaction with Daron Acemoglu, David Autor, Esther Duflo, Jerry Hausman, and Guido Kuersteiner at MIT and Gary Solon at the University of Michigan.

I was supported by an MIT fellowship during my first year as a graduate student and by a National Science Foundation Graduate Research Fellowship during my final three years as a graduate student. I am grateful to both MIT and the NSF for the financial support.

The best thing about graduate school has been the interaction with my fellow students, and I am grateful to all of them. I am especially grateful to Angus Armstrong, Bill Kerr, Youngjin Hwang, and David Lyle who helped me get through the first two years of classes. David Lyle has been an especially close friend, and I am grateful to him for his help, encouragement, faith, and prayers. I have benefitted enormously from interaction with Byron Lutz who helped develop the question examined in Chapter 2 of this thesis and provided invaluable comments on all of the essays in this thesis. I am also thankful to Melissa Boyle, Daniel Chen, Ivan Val-Fernandez, Joanna Lahey, Ashley Lester, Stavros Panageas, Josh Rauh, and Andrew Sweeting.

I am especially grateful to my family for the love and support they have given me during the last four years. Both my parents and my in-laws have provided much needed financial support and even more needed emotional support throughout my graduate school experience. The support and friendship of my brothers and sister have also been invaluable.

Of course, my biggest debt of gratitude is to my wonderful wife, Nicole, and to my two sons, Conner and Jude. They have provided a constant source of comfort, joy, and stability through the ups and downs of my graduate school experience. Nicole has had to deal with homework, exams, generals, the struggle to develop a dissertation topic, and the long hours of work every bit as much as I have and has provided me with love and encouragement throughout. I will certainly never be able to repay her or the boys for all they have sacrificed for me during the last four years, and I am eternally grateful to them for all they have done.

Contents

Introduction

In this thesis, I explore inference issues in multilevel and panel data and present a separate essay that examines the impact of 401(k) participation on wealth and financial assets. The first two chapters study different inference methods that will be valid in multilevel and panel data with serially correlated disturbances. The first chapter discusses efficient estimation and inference when the disturbances follow an AR(p) process, and the second chapter considers the behavior of a robust covariance matrix estimator for panel data which allows arbitrary intertemporal correlation within individuals and heterogeneity across individuals. The final chapter of the thesis presents an empirical analysis of the impact of 401(k) participation on wealth and financial assets based on instrumental variable quantile regression methods.

The first two chapters of this thesis consider inference in serially correlated multilevel and panel data. The first chapter explicitly focuses on inference for a time series of multilevel data, that is, several years of data with variation at both the individual and a more aggregate level. This data structure is quite common in empirical economics; for example, in differences-in-differences and policy analysis studies, the outcome often varies at the individual level while the covariate of interest is a policy that affects all individuals within a group. One potential complication to inference in this type of data is that the presence of an unobserved group specific shock may result in severe biases to conventional OLS inference, and most applied researchers routinely make use of corrections for this "clustering" problem. However, it is far less common to account for group level autocorrelation which may arise if the unobserved group level shock is serially correlated, and ignoring this source of correlation may also result in highly misleading inference even after adjusting for "clustering". In addition, OLS is not the Gauss Markov estimator under these conditions, and tests based on OLS estimates may have significantly less power than more efficient alternatives.

In Chapter 1, I offer one potential solution to this problem by providing accurate and powerful inference methods for data with a grouped structure where the groups may be autocorrelated. In particular, I consider GLS estimation in this setting and present computationally convenient methods for obtaining the GLS estimates. One complication which arises is that standard estimates of the components of the variance matrix necessary for obtaining the GLS estimates will typically be biased and inconsistent due to the inclusion of group level fixed effects. Focusing on the case where the group level shock follows an AR(p) process, I offer a bias correction for the AR coefficients which will be valid in the presence of fixed effects and group-specific time trends. I develop asymptotic properties of the FGLS estimator and the bias-corrected AR coefficients in asymptotics where the number of groups goes to infinity with the time dimension fixed and in asymptotics where the number of groups and the number of time periods go to infinity jointly. The final section of the chapter illustrates the usefulness of GLS and the derived bias-correction for the parameters of the autoregressive process through a simulation study which uses data from the Current Population Survey Merged Outgoing Rotation Group files. I find that the GLS estimates have relatively accurate size and much higher power than OLS with standard errors robust to the presence of a serially-correlated unobserved group level shock.

Chapter 2 examines an alternate OLS-based approach for performing inference in serially correlated panel models. The chief drawback of the GLS-methods explored in Chapter 1 is that their implementation requires the imposition of a parametric model on the disturbance process. An alternative strategy is to use OLS to estimate the model and to then base inference on estimates of the covariance matrix of the OLS parameters which are robust to the presence of serial correlation and heteroskedasticity.

I consider the asymptotic properties of a robust covariance matrix estimator which is commonly advocated for use in panel data. The estimator is a generalization of the conventional heteroskedasticity consistent covariance matrix estimator for panel data which allows for arbitrary correlation within each individual and heterogeneity across individuals. Under conventional panel asymptotics where the cross-section dimension, n, grows large with the time series dimension, T, fixed, this estimator has good properties while allowing an essentially unconstrained time series pattern of correlation because the short time dimension is ignorable. However, many panel data sets are characterized by a non-negligible time dimension. In these cases, it is not obvious that the robust covariance matrix estimator will even be consistent.

I extend the usual analysis performed under asymptotics where $n \to \infty$ with T fixed to cases where n and T go to infinity jointly, considering both non-mixing and mixing cases, and show that conventional t and F tests based on the robust covariance matrix estimator are consistent and have the usual limiting behavior in these cases. In addition, when $T \to \infty$ with n fixed and other regularity conditions are satisfied, I show that the usual t and F statistics can be used for inference despite the fact that the robust covariance matrix estimator is not consistent but converges in distribution to a limiting random variable. In addition to its use in performing inference about regression coefficients, the robust covariance matrix can be used to test the specification of simple parametric models of the error process, and I also outline the properties of such a test in asymptotics where n and T go to infinity jointly. The properties of the robust covariance matrix estimator and tests based upon it are examined in a short Monte Carlo study in the final section of the chapter. The simulation results suggest that inference based on the robust covariance matrix is quite accurate regardless of the relative size of n and T and that it is accurate even for small n if T is moderate and the dependence between observations decreases as the distance between them increases.

Chapter 3 shifts away from inference issues in panel data models and reports results from an empirical examination of the impact of 401(k) participation on various measures of wealth. The analysis makes use of instrumental variables quantile regression methods which allow estimation of the impact of participating in a 401(k) on multiple points of the outcome distribution. Knowing the treatment effect at many points in the outcome distribution provides a more full characterization of the treatment impact and gives additional insights into the economic relationships involved.

Specifically, the paper makes use of data from Wave 4 of the 1990 Survey of Income and Program Participation. The data include measures of financial assets and total wealth as well as data on 401(k) participation and eligibility for a 401(k), which is used to instrument for 401(k) participation. The instrumental variables quantile regression estimates indicate that there is considerable heterogeneity in the effect of 401(k) participation on net financial assets, with the treatment effect increasing monotonically as one moves from the lower to the upper tail of the asset distribution. The results are also uniformly positive and significant, suggesting that 401(k) participation positively impacts net financial assets across the entire distribution. The effect of participation on total wealth is positive and approximately constant for all quantiles. In addition, it is of the same magnitude as the effect of participation on net financial assets for low quantiles, but is substantially smaller than the effect of participation on the upper quantiles of net financial assets. These results suggest that the increase in net financial assets observed in the lower tail of the conditional assets distribution can be interpreted as an increase in wealth, while the increase in the upper tail of the distribution is mitigated by substitution with some other component of wealth. The effect of participation on net financial assets excluding 401(k) assets is uniformly insignificant, which suggests there is little substitution for 401(k) assets along this dimension of wealth. The heterogeneity of the treatment effect and substitution patterns are obscured by analyses which focus on a single point in the outcome distribution and illustrate the usefulness of quantile regression for examining economic outcomes.

Chapter 1

Generalized Least Squares Inference in Multilevel Models with Serial Correlation and Fixed Effects

1.1 Introduction

Many economic analyses are characterized by regressions involving both aggregate and individual level data, that is, multilevel data. This is especially prevalent in differences-in-differences (DD) estimation, and policy analysis more generally, where the dependent variable is often an individual level outcome and the covariate of interest is a policy which applies to all individuals within a group. For example, in a study of the impact of the minimum wage on employment, the dependent variable may be the employment in a firm in state s at time t, and the policy would be the minimum wage in state s at time t. While this sampling design does not pose any serious problems for estimation of the linear model, it may lead to serious problems for inference. In particular, the sampling design gives rise to potential sources of correlation between observations, termed here the "clustering problem" and the "policy autocorrelation problem", that would be ignored in computing conventional least squares standard errors. The clustering problem is caused by the presence of a common unobserved random shock at the group level that will lead to correlation between all observations within each group. The policy autocorrelation problem arises if the groups (not necessarily the individuals within the groups) are followed over time and the group level shocks are serially correlated, which will result in correlation between individuals from the same group at different time periods.[1] In general, ignoring these correlations will bias conventional least squares standard errors and lead to misleading inference. The purpose of this paper is to provide accurate, powerful, and easily computable inference methods for data which are potentially affected by both the clustering problem and the policy autocorrelation problem.

The clustering problem has long been recognized in the econometric literature on panel data, and has more recently been emphasized in economics in other contexts which involve multi-stage sampling or multilevel data. There are a number of methods for dealing with this problem which are available in most statistical packages. The most common approach is to estimate a linear model with OLS and then correct the standard errors for the intracluster correlation as in Moulton (1986), Arellano (1987), or Kezdi (2002). Feasible Generalized Least Squares (FGLS)[2] estimation may also be performed easily and will asymptotically result in a more efficient estimator and more powerful tests than OLS.

The policy autocorrelation problem has received considerably less attention from applied economic researchers. In a survey of DD papers published in six leading applied economics journals from 1990-2000,[3] Bertrand, Duflo, and Mullainathan (2003) found that only five of 65 articles with a potential serial correlation problem explicitly address it; and in simulations based on individual level Current Population Survey Merged Outgoing Rotation Group (CPS-MORG) data, they found a 44% rejection rate for a 5% level test using standard techniques which correct only for the intragroup correlation. To focus on the policy autocorrelation problem, they also performed a simulation based on data from the CPS-MORG aggregated to the state-year level. In the aggregate data, they found that using simple parametric models for the serial correlation did not correct the size distortion, but that tests based on the OLS estimator which flexibly account for serial correlation, such as the bootstrap or using a variance matrix robust to arbitrary correlation at the state instead

[1]For simplicity, I will typically refer to groups as states and time periods as years.

[2]Throughout, I refer to GLS as the infeasible estimator which assumes that the variance matrix of the disturbances is known and to FGLS as a feasible estimator which uses estimates of the elements of the variance matrix.

[3]The journals are *The American Economic Review*, *Industrial and Labor Relations Review*, *The Journal of Labor Economics*, *The Journal of Political Economy*, *The Journal of Public Economics*, and *The Quarterly Journal of Economics*.

of the state-year level, had approximately correct size. However, while these tests appear to have correct size, the results of Bertrand, Duflo, and Mullainathan (2003) also suggest that they have low power against relevant alternatives.

In this paper, I contribute to the existing literature by offering computationally attractive FGLS-based estimation and inference procedures that deliver accurate and powerful inference in settings that are subject to both the clustering problem and the policy autocorrelation problem. In particular, I explicitly consider FGLS estimation in a general model for grouped individual and aggregate level data which incorporates standard DD and panel models and could easily be extended to cases with additional levels of variation. I present a differencing strategy and an aggregation strategy building upon Amemiya (1978) which recover the FGLS estimates and are computationally convenient in large data sets. I then focus on estimation and inference under the assumption that the state-year shock follows a stationary AR(p) process. FGLS estimation and inference based on parametric time series models for the error process is complicated by the relatively short time dimension available in many policy analyses and the inclusion of state-level fixed effects. It is well-known that estimates of the parameters of time series models in panel data with fixed effects are biased when the time dimension is short due to the incidental parameters problem. Using a strategy due to Nickell (1981) and Solon (1984), I derive a bias correction for the coefficients of an AR(p) model. I then develop asymptotic inference results for the bias-corrected coefficients and the corresponding FGLS estimator which cover both conventional panel asymptotics where the number of states goes to infinity with the time dimension fixed and asymptotics where the number of states and time periods go to infinity jointly as in Phillips and Moon (1999), Hahn and Kuersteiner (2002), and Hahn and Newey (2002). The usefulness of the bias correction and the FGLS procedure are then demonstrated through a simulation study based on the CPS-MORG.

The results from the simulation study strongly support the use of the FGLS procedure with bias-corrected AR coefficients for performing inference in settings with combined individual and grouped data where the groups are potentially autocorrelated. As in Bertrand, Duflo, and Mullainathan (2003), I find that conventional OLS and, to a lesser extent, conventional FGLS suffer from severe size distortions in the presence of the policy autocorrelation and clustering problems. This size distortion is essentially removed by OLS with standard errors clustered by state and the bias-corrected FGLS procedure. However, the FGLS procedure clearly dominates OLS with standard errors robust to arbitraty correlation within states in terms of both power and confidence interval length. For example, in a simulation performed by resampling directly from the CPS-MORG, I find that conventional OLS has a rejection rate of 37% for a 5% level test. In contrast, OLS with standard errors clustered by state rejects 6.6% of the time, and FGLS based on bias-corrected AR(3) coefficients has a 6.4% rejection rate. At the same time, the power of the bias-corrected FGLS-based procedure versus the alternative that the treatment increases the dependent variable by 2% is 0.788 compared to 0.344 from OLS with clustered standard errors. Similarly, the length of the FGLS confidence interval is 0.028 compared to the OLS interval length of 0.050.

The remainder of this paper is organized as follows. In Section 2, I briefly review GLS estimation in settings involving both individual and aggregate data and present a computationally attractive procedure for obtaining the GLS estimates which will be valid as the group size grows large within each state-year cell. Section 3 presents a bias-correction for fixed effects estimates of the parameters of a pth order autoregressive model which will be used in FGLS estimation. Simulation results comparing the FGLS estimator to other estimators are presented in Section 4, and Section 5 concludes.

1.2 GLS Estimation and Policy Analysis

1.2.1 Overview of GLS with Correlated Error Components

Estimates in DD and policy analysis studies are often obtained using a linear model defined by

$$y_{ist} = w_{ist}'\beta_0 + C_{st} + u_{ist} \tag{1.1}$$

and

$$C_{st} = x_{st}'\beta_1 + z_{st}'\beta_2^s + v_{st}, \tag{1.2}$$

where $s = 1, ..., S$, $t = 1, ..., T$, $i = 1, ..., N_{st}$ for each s and t, C_{st} are state-year effects, w_{ist} are covariates that vary at the individual level, x_{st} are covariates that vary at the state-year level and have constant coefficients, z_{st} are covariates that vary at the state-year level and have state-specific coefficients[4], y_{ist} is the outcome of interest which varies at the individual level,

[4]For the theoretical development, z_{st} will be assumed to be nonstochastic and identical across states.

and v_{st} and u_{ist} are unobservable random variables which are uncorrelated with the observed explanatory variables and with each other and have zero means. Typical specifications of z_{st} include $z_{st} = 1$, the fixed effects model, and $z_{st} = [1, t]$, the fixed effects model with state-specific time trends. It is also standard in DD models to include time effects in x_{st}. In addition, it is often assumed that $E[v_{st}v_{s't'}] = 0$ for all $s \neq s'$. Conventionally, estimation and inference are performed on the model formed by combining (1.1) and (1.2) as

$$y_{ist} = w'_{ist}\beta_0 + x'_{st}\beta_1 + z'_{st}\beta_2^s + \epsilon_{ist} \tag{1.3}$$

where $\epsilon_{ist} = v_{st} + u_{ist}$. The clustering problem then results from the fact that $E[\epsilon_{ist}\epsilon_{jst}] = \sigma_v^2$ for all $i \neq j$, and the policy autocorrelation problem arises from $E[\epsilon_{ist}\epsilon_{js(t-k)}] = \gamma(k) \neq 0$ if v_{st} is serially correlated. In most studies, the model is estimated using OLS, and the estimated standard errors are adjusted to account for the presence of correlation between individuals within state-year cells. While this approach has a number of appealing features, it will yield incorrect standard error estimates and tests if there are other sources of correlation, such as correlation within states over time due to a correlated state-specific shock. In addition, if the errors are correlated, OLS is not the Gauss-Markov estimator, and more efficient estimates and more powerful tests may be obtained through GLS.

To facilitate discussion of the GLS estimator, note that equations (1.3) may be stacked and represented in matrix form as

$$Y = \Phi\theta + \epsilon, \tag{1.4}$$

where $\Phi = [W, X, Z]$, $Y = (Y_1', ..., Y_S')'$, $Y_s = (Y_{s1}', ..., Y_{sT}')'$, $Y_{st} = (y_{1st}, ..., y_{N_{st}st})'$, and W, X, Z, and ϵ are defined similarly. ϵ may be written as $DV + U$ for $V = (v_{11}, v_{12}, ..., v_{ST})'$, U defined as Y, and $D = [d_{11}\ d_{12}\ \cdots\ d_{ST}]$ with d_{st} a dummy variable indicating the observation belongs to state s at time t, so under the assumption that V and U are uncorrelated, $E[\epsilon\epsilon'] = \Sigma = D\Omega D' + \Lambda$ where $E[VV'] = \Omega$ and $E[UU'] = \Lambda$. Given the parameters of Ω and Λ, the best linear unbiased estimator of θ is the GLS estimator

$$\widehat{\theta}_{GLS} = (\Phi'\Sigma^{-1}\Phi)^{-1}\Phi'\Sigma^{-1}Y. \tag{1.5}$$

Given Ω and Λ ($\widehat{\Omega}$ and $\widehat{\Lambda}$), implementation of the GLS (FGLS) estimator may proceed in a straightforward fashion for moderately sized data sets by numerically obtaining Σ^{-1} ($\widehat{\Sigma}^{-1}$) and computing $\widehat{\theta}_{GLS}$ ($\widehat{\theta}_{FGLS}$) directly. However, for larger scale problems, such as the one considered in the simulation section, this procedure is computationally burdensome due to the size of Σ. Fortunately, there are also numerically convenient approaches available for compution of $\widehat{\theta}_{GLS}$. One approach recognizes that the structure of the problem implies that $\widehat{\theta}_{GLS}$ may be computed as a least squares regression on quasi-differenced data. This method will provide the GLS estimates of all parameters in θ and will generally reduce the computational burden from the more brute force implementation. The second approach, due to Amemiya (1978), uses the fact that equation (1.3) is equivalent to the model defined by (1.1) and (1.2),

$$y_{ist} = w'_{ist}\beta_0 + C_{st} + u_{ist}$$

and

$$C_{st} = x'_{st}\beta_1 + z'_{st}\beta_2^s + v_{st},$$

to reduce the dimension of the problem of finding the GLS estimates of β_1 and β_2^s from $\sum_s \sum_t N_{st}$ to ST. In addition, this approach provides intuition for asymptotic results regarding the parameters that vary at the state-year level and suggests a simple estimation method that will be asymptotically equivalent to GLS as $N_{st} \to \infty$.

1.2.2 GLS as OLS on Quasi-Differenced Data

One potentially tractable method which may be used to obtain the GLS estimates of θ is based on the fact that the GLS estimator of model (1.3) is equivalent to OLS regression on appropriately differenced data. In particular, straightforward regression algebra yields the following result.

Proposition 1 *Let $E[VV'] = \Omega$, $E[UU'] = \Lambda$, and $D = [d_{11}\ d_{12}\ \cdots\ d_{ST}]$ with d_{st} a dummy variable indicating the observation belongs to state s at time t. Define $\widetilde{V}_B = (D'\Lambda^{-1}D)^{-1} + \Omega$, $\widetilde{Y} = \Lambda^{-1/2}Y$, $\widetilde{\Phi} = \Lambda^{-1/2}\Phi$, and $\widetilde{D} = \Lambda^{-1/2}D$. Then the GLS estimator of model (1.3) is equivalent to OLS regression of $\widetilde{y}_{ist} - \tilde{\bar{y}}_{ist} + \tilde{\bar{y}}^B_{ist}$ on $\widetilde{\phi}_{ist} - \tilde{\bar{\phi}}_{ist} + \tilde{\bar{\phi}}^B_{ist}$, where $\widetilde{y}_{ist}$ is the ist element of $\widetilde{Y}$, $\tilde{\bar{y}}_{ist}$ is the ist element of $\widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}'\widetilde{Y}$, $\tilde{\bar{y}}^B_{ist}$ is the ist element of $\widetilde{D}(\widetilde{D}'\widetilde{D})^{-1/2}\widetilde{V}_B^{-1/2}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}'\widetilde{Y}$, and $\widetilde{\phi}_{ist}$, $\tilde{\bar{\phi}}_{ist}$, and $\tilde{\bar{\phi}}^B_{ist}$ are defined similarly.*

While this result suggests that a differencing strategy will be cumbersome for general Λ, there is substantial simplification if Λ is diagonal, as is conventionally assumed in policy analysis and DD models. In particular, letting $E[u_{ist}^2] = (\sigma_{ist}^u)^2$, the previous result implies that the GLS estimate of θ may be obtained through the linear regression of $\frac{1}{\sigma_{ist}^u}(y_{ist} - \bar{y}_{st} + \bar{y}_{st}^B)$ on $\frac{1}{\sigma_{ist}^u}(\phi_{ist} - \bar{\phi}_{st} + \bar{\phi}_{st}^B)$ where $\bar{y}_{st}$ is the variance weighted average of the individual observations in state s in year t, $\bar{y}_{st}^B$ is the st element of $(\widetilde{D}'\widetilde{D})^{-1/2}\widetilde{V}_B^{-1/2}\bar{Y}$, $\bar{Y}$ is the $ST \times 1$ vector of the within state-year variance weighted averages of the individual observations, and $\bar{\phi}_{st}$ and $\bar{\phi}_{st}^B$ are defined similarly. The chief difficulty in implementing this strategy is then computation of $(\widetilde{D}'\widetilde{D})^{-1/2}\widetilde{V}_B^{-1/2}$ which will typically be simple given the relatively small s and t dimensions. Further simplifications are also available. For example, in the simulations reported in Section 4 of this paper, it is assumed that U is homoskedastic and that V is uncorrelated across states, implying that Ω is block diagonal. This results in the variance weighted averages being replaced by simple averages and reduces the computation of $(\widetilde{D}'\widetilde{D})^{-1/2}\widetilde{V}_B^{-1/2}$ to operations on T-dimensional matrices, which impose little computational burden.

1.2.3 Modeling the State-Year Fixed Effects

Another convenient approach which will provide the GLS estimates of β_1 and β_2^s, the coefficients on the covariates that vary at the state-year level, is based on the decomposition of equation (1.3) into equations (1.1) and (1.2). Amemiya (1978) demonstrated that estimates of β_1 and β_2^s from the following two-step procedure are numerically identical to the GLS estimates obtained from estimating model (1.3):

1. Estimate equation (1.1), $y_{ist} = w_{ist}'\beta_0 + C_{st} + u_{ist}$ by GLS to obtain estimates of C_{st}, $\widehat{C}_{st}$.

2. Obtain estimates of β_1 and β_2^s by estimating the equation $\widehat{C}_{st} = x_{st}'\beta_1 + z_{st}'\beta_2^s + \nu_{st}$, where $\nu_{st} = v_{st} + (\widehat{C}_{st} - C_{st})$ by GLS.

This approach will typically be computationally easier than directly computing Σ^{-1}. If, as is conventionally assumed, Λ is diagonal, the first step may be computed by weighted least squares, and the second step only requires inversion of an $ST \times ST$ matrix.

In addition to providing a tractable method for obtaining GLS estimates of β_1, this approach also clearly illustrates that β_1 is not consistent as $N_{st} \to \infty$ with S and T fixed. In particular, we see that consistency and asymptotic normality of β_1 requires that $ST \to$

∞.[5] The use of Amemiya's (1978) approach and the inconsistency of estimates of β_1 in asymptotics with S and T fixed has recently been emphasized in work of Donald and Lang (2001) in the context of DD estimation when serial correlation is not present.

Finally, Amemiya's (1978) results also suggest a simple estimation strategy which will be equivalent to GLS as $N_{st} \to \infty$ for all s and t:

1'. Estimate equation (1.1), $y_{ist} = w'_{ist}\beta_0 + C_{st} + u_{ist}$ by OLS or GLS to obtain estimates of C_{st}, $\widehat{C}_{st}$.

2'. Obtain estimates of β_1 and β_2^s by estimating the equation $\widehat{C}_{st} = x'_{st}\beta_1 + z'_{st}\beta_2^s + v_{st}$, by GLS.

Note that 1' and 2' differ from 1 and 2 above in that 1' does not require the first step to be estimated by GLS and 2' ignores the fact that the dependent variable in the second step, $\widehat{C}_{st}$, was estimated. The equivalence of this approach to GLS for estimating β_1 and β_2^s as $N_{st} \to \infty$ follows from numeric equivalence of Amemiya's (1978) two-step approach to GLS and consistency of $\widehat{C}_{st}$ for C_{st} as $N_{st} \to \infty$.[6] This result also implies that estimation of $\widehat{C}_{st}$ may be ignored and GLS estimates of β_1 may be obtained through standard panel methods when N_{st} is sufficiently large. This is particularly useful since data used in many DD problems are characterized by rather large cell sizes, and the approach outlined above is easy to implement. Throughout most of the simulation section, I focus on this method of estimating β_1.

1.3 Bias Correction for pth Order Autoregressive Coefficients in Fixed Effects Models

In order to operationalize GLS estimation and inference in practice, parameters of the covariance matrices, Ω and Λ, must be estimated. Estimation of the parameters of Λ may

[5]This is also straightforward to demonstrate using the GLS estimator for all the parameters outlined in the previous section.

[6]If the first step is estimated using GLS, it may also be shown that the estimates of β_1 converge to the GLS estimates of β_1 as $N_{st} \to \infty$ for all s and t.

generally proceed in a straightforward fashion from equation (1.1) or may be bypassed completely in asymptotics where $N_{st} \to \infty$ by using the aggregation method discussed above, so I focus on estimation of Ω. While there are numerous approaches that one could consider for estimating the parameters of Ω,[7] I adopt the simple strategy of assuming that v_{st} follows a stationary pth order autoregressive process:

$$v_{st} = \sum_{j=1}^{p} \alpha_j v_{s(t-j)} + \eta_{st}.$$

I also focus on the case where the data have been aggregated to the state-year level using the approach of Amemiya (1978), though the method outlined below could be easily extended to treat other cases. In addition, I assume estimation of $\widehat{C}_{st}$ is ignored.[8] The results here may also be adapted easily to bias-correct AR coefficients in dynamic panel models without covariates. Throughout the remainder of the development it is assumed that v_{st} has zero mean and constant variance which do not depend on X or Z for all s and t.[9] Without loss of generality, the variance of v_{st} is set equal to 1.

Under the assumptions outlined in the preceding paragraph, an obvious approach to estimating the process of v_{st} would be to use the residuals from estimation of

$$\widehat{C}_{st} = x'_{st}\beta_1 + z'_{st}\beta_2^s + v_{st} \tag{1.6}$$

to estimate the $\alpha = (\alpha_1, ...\alpha_p)'$ using least squares. These estimates of α will be consistent as both S and T approach infinity. However, estimation of α is complicated by the presence of z_{st} in equation (1.6) and the fairly short time series dimension available in most applications, which may result in substantial bias in the estimates of α. Bertrand, Duflo, and Mullainathan (2003), in a survey of published differences-in-differences papers, find an average time series length of only 16.5 periods. They also find significant bias in autoregressive parameter estimates in their simulations.

To address this problem, I use the arguments of Nickell (1981) and Solon (1984) to derive the bias of α as $S \to \infty$. I then use this to form a bias-corrected estimator of α. A simple

[7]See, for example, Kiefer (1980), Macurdy (1982), Solon (1984), Lancaster (2002), and Hausman and Kuersteiner (2003).

[8]In the appendix, I give a modification of the formula presented below which accounts for estimation of $\widehat{C}_{st}$.

[9]These assumptions are formalized below in Assumptions 1 and 2.

one-step estimator based on this strategy removes the bias from the asymptotic distribution of the estimator of α as long as $\frac{S}{T^3} \to 0$. In addition, I show that an iterative procedure is consistent as $S \to \infty$ even with T fixed. The basic approach is similar to that of Hahn and Newey (2002) and Hahn and Kuersteiner (2002). In fact, for the AR(1) model, it is straightforward to show that the difference between the Hahn and Kuersteiner (2002) bias reduction and the one-step bias reduction derived here is $O\left(\frac{1}{T^2}\right)$.

1.3.1 The Bias Correction

The least squares estimator of α using the residuals from the estimation of equation (1.6), $\widetilde{v}_{st}$, is

$$\widehat{\alpha} = \left(\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \widetilde{v}_{st}^{-} \widetilde{v}_{st}^{-\prime}\right)^{-1} \left(\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \widetilde{v}_{st}^{-} \widetilde{v}_{st}\right) \tag{1.7}$$

where $\widetilde{v}_{st}^{-\prime} = (\widetilde{v}_{s(t-p)}, ..., \widetilde{v}_{s(t-1)})$. The nature of this estimator as a double sum over s and t makes it simple to analyze as $S \to \infty$. Let $E[v_{st} v_{s(t-k)}] = \gamma_k(\alpha)$, and let

$$\Gamma_p = \begin{bmatrix} \gamma_0(\alpha) & \gamma_1(\alpha) & \cdots & \gamma_{p-1}(\alpha) \\ \gamma_1(\alpha) & \gamma_0(\alpha) & & \gamma_{p-2}(\alpha) \\ \vdots & & \ddots & \vdots \\ \gamma_{p-1}(\alpha) & \gamma_{p-2}(\alpha) & \cdots & \gamma_0(\alpha) \end{bmatrix}.$$

Then, using calculations similar to those found in Nickell (1981) and Solon (1984) and assuming regularity condition collected in Assumption 1 below hold, one can show that as $S \to \infty$ with T fixed $\widehat{\alpha} \xrightarrow{p} \alpha_T(\alpha) = (\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}(A(\alpha) + \frac{1}{T-p}\Delta_A(\alpha))$. with $A(\alpha) = (\gamma_1(\alpha), ..., \gamma_p(\alpha))'$, $\Delta_\Gamma(\alpha)$ a $p \times p$ matrix with

$$\begin{aligned}[\Delta_\Gamma(\alpha)]_{[i,j]} = {} & \text{trace}\left(Z_s'\Gamma(\alpha)Z_s(Z_s'Z_s)^{-1}Z_{s,-i}'Z_{s,-j}(Z_s'Z_s)^{-1}\right) \\ & -\text{trace}\left(Z_s'\Gamma_{-i}(\alpha)Z_{s,-j}(Z_s'Z_s)^{-1}\right) - \text{trace}\left(Z_s'\Gamma_{-j}(\alpha)Z_{s,-i}(Z_s'Z_s)^{-1}\right),\end{aligned} \tag{1.8}$$

and $\Delta_A(\alpha)$ a $p \times 1$ vector with

$$\begin{aligned}[\Delta_A(\alpha)]_{[i,1]} = {} & \text{trace}\left(Z_s'\Gamma(\alpha)Z_s(Z_s'Z_s)^{-1}Z_{s,-i}'Z_{s,-0}(Z_s'Z_s)^{-1}\right) \\ & -\text{trace}\left(Z_s'\Gamma_{-i}(\alpha)Z_{s,-0}(Z_s'Z_s)^{-1}\right) - \text{trace}\left(Z_s'\Gamma_{-0}(\alpha)Z_{s,-i}(Z_s'Z_s)^{-1}\right),\end{aligned} \tag{1.9}$$

where $\Gamma(\alpha) = \mathrm{E}[V_s V_s']$, $\Gamma_{-k}(\alpha) = \mathrm{E}[V_s V_{s,-k}']$, $V_{s,-k} = (v_{s(p+1-k)}, v_{s(p+2-k)}, \ldots, v_{s(T-k)})'$ and $Z_{s,-k}$ is defined similarly. Thus, $\mathrm{Bias}(\widehat{\alpha}) = -\alpha + \alpha_T(\alpha)$. This suggests that the bias of $\widehat{\alpha}$ may be estimated as $-\widehat{\alpha} + \alpha_T(\widehat{\alpha})$ and that a bias corrected estimator of α may be constructed as

$$\widehat{\alpha}^{(1)} = \widehat{\alpha} - [-\widehat{\alpha} + \alpha_T(\widehat{\alpha})]. \tag{1.10}$$

In addition, $\widehat{\alpha} \xrightarrow{p} \alpha_T(\alpha)$ suggests that a consistent (in S alone) estimate of α may be obtained by inverting $\alpha_T(\alpha)$ to obtain $\widehat{\alpha}^{(\infty)} = \alpha_T^{-1}(\widehat{\alpha})$. This estimator can be calculated by iterating $\widehat{\alpha}^{(k+1)} = \widehat{\alpha} - [\alpha_T(\widehat{\alpha}^{(k)}) - \widehat{\alpha}^{(k)}]$ to convergence, since, denoting $\widehat{\alpha}^{(\infty)}$ as the point that the procedure converges to, $\widehat{\alpha}^{(\infty)} = \widehat{\alpha} - [\alpha_T(\widehat{\alpha}^{(\infty)}) - \widehat{\alpha}^{(\infty)}] \Rightarrow \alpha_T(\widehat{\alpha}^{(\infty)}) = \widehat{\alpha} \Rightarrow \widehat{\alpha}^{(\infty)} = \alpha_T^{-1}(\widehat{\alpha})$. Bhargava, Franzini, and Narendranathan (1982) suggest a similar iterative procedure based on the Durbin-Watson statistic to remove the bias of autoregressive parameter estimates in AR(1) models with fixed effects, though no formal asymptotic results are presented and the extension to models beyond the AR(1) is not clear.

Asymptotic properties of the estimators are collected in the next two sections, which will make use of the following notation. Let

$$C_{st} = x_{st}'\beta_1 + z_{st}'\beta_2^s + v_{st}, \tag{1.11}$$

or, in vector notation, $C_s = X_s\beta_1 + Z_s\beta_2^s + V_s$, where $C_s = [C_{s1}, \ldots, C_{sT}]'$ is $T \times 1$, $X_s = [x_{s1}, \ldots, x_{sT}]'$ is $T \times k_1$, $Z_s = [z_{s1}, \ldots, z_{sT}]'$ is $T \times k_2$, and $V_s = [v_{s1}, \ldots, v_{sT}]'$ is $T \times 1$. Also, let x_{sth} be the h^{th} element of x_{st} so that $x_{st}' = [x_{st1}, \ldots, x_{stk_1}]$, and define z_{sth} similarly. Define $\ddot{v}_{st} = v_{st} - z_{st}'(Z_s'Z_s)^{-1}Z_s'V_s$, $\ddot{x}_{st}' = x_{st}' - z_{st}'(Z_s'Z_s)^{-1}Z_s'X_s$, $\ddot{V}_s = [\ddot{v}_{s1}, \ldots, \ddot{v}_{sT}]'$, and $\ddot{X}_s = [\ddot{x}_{s1}, \ldots, \ddot{x}_{sT}]'$. Let v_{st}^- be a $p \times 1$ vector with $v_{st}^- = [v_{s(t-p)}, \ldots, v_{s(t-1)}]'$, and define $\ddot{v}_{st}^-$ similarly.

1.3.2 Asymptotics as $S \to \infty$ with T Fixed

To establish the asymptotic properties of the least squares estimate of α as $S \to \infty$ with T fixed, I impose the following conditions in addition to model (1.11).

Assumption 1 ($S \to \infty$, T fixed) *Suppose the data are generated by model (1.11) and*

N1. $v_{st} = v_{st}^{-\prime}\alpha + \eta_{st}$, *where* η_{st} *is strictly stationary in* t *for each* s, $\mathrm{E}[\eta_{st}^2] = \sigma_\eta^2$, $\mathrm{E}[\eta_{st}\eta_{s\tau}] = 0$ *for* $t \neq \tau$, *and the roots of* $1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p = 0$ *have modulus greater than 1.*

N2. $\{X_s, V_s, \eta_s\}$ *are iid across s.* $\{Z_s\}$ *are nonstochastic and identical across s.*

N3. *(i)* $\text{Rank}(\sum_{t=1}^{T} \text{E}[\ddot{x}_{st}\ddot{x}'_{st}]) = \text{Rank}(\text{E}[\ddot{X}'_s \ddot{X}_s]) = k_1$. *(ii)* $\text{Rank}(Z'_s Z_s) = k_2 \;\forall\; s$.

N4. $\text{E}[V_s|X_s] = 0$, $\text{E}[V_s V'_s|X_s] = \Gamma(\alpha)$.

N5. $\text{E}[\eta_{st}^4] = \mu_4 < \infty$ *and* $\text{E}[x_{sth}^4] \leq \Delta < \infty \;\forall\; s,\; t,\; h.$

Remark 1.3.1 The majority of the conditions imposed in Assumption 1 are standard for fixed effects panel models, with the key difference being the imposition of the AR(p) structure on the error term. Also, the existence of fourth moments and strict stationarity of η_{st} imply the existence of first through fourth absolute moments of v_{st} under the stationarity condition that the roots of $1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p = 0$ have modulus greater than 1. (See, for example, Hamilton (1994) Proposition 7.10.)

Under Assumption 1, Proposition 2 and Proposition 3 are obtained.

Proposition 2 *If Assumption 1 is satisfied,* $\widehat{\alpha} \xrightarrow{p} \alpha_T(\alpha)$, *where*

$$\alpha_T(\alpha) = \text{E}[\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}]^{-1}\text{E}[\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{v}_{st}] = (\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}(A(\alpha) + \frac{1}{T-p}\Delta_A(\alpha)).$$

Remark 1.3.2 This proposition simply formalizes the bias result of the previous section, verifying that $\widehat{\alpha}$ is inconsistent as $S \to \infty$ with T fixed.

Proposition 3 *Suppose* $\alpha_T(\alpha)$ *is continuously differentiable in* α *and that* $H = D\alpha_T(\alpha)$ *is invertible for all* α *such that N1 is satisfied, where* $D\alpha_T(\alpha)$ *is the derivative matrix of* $\alpha_T(\alpha)$ *in* α. *Define* $\widehat{\alpha}^{(\infty)} = \alpha_T^{-1}(\widehat{\alpha})$. *Then, if Assumption 1 is satisfied,* $\widehat{\alpha}^{(\infty)} - \alpha \xrightarrow{p} 0$ *and* $\sqrt{S}(\widehat{\alpha}^{(\infty)} - \alpha) \xrightarrow{d} \frac{1}{T-p}H^{-1}(\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}N(0, \Xi)$, *where*

$$\Xi = \text{E}[\sum_{t_1=p+1}^{T} \sum_{t_2=p+1}^{T} \ddot{v}_{st_1}^{-}\ddot{\mu}_{st_1}\ddot{\mu}_{st_2}\ddot{v}_{st_2}^{-\prime}]$$

and $\ddot{\mu}_{st} = \ddot{v}_{st} - \ddot{v}_{st}^{-\prime}\alpha_T(\alpha)$.

Remark 1.3.3 The condition that $\alpha_T(\alpha)$ is continuously differentiable in α and that $H = D\alpha_T(\alpha)$ is invertible for all α such that N1 is satisfied guarantees the existence of the inverse of $\alpha_T(\alpha)$ and seems reasonable in many settings. For example, it is satisfied when Z only contains fixed effects. However, this is an asymptotic result, and even if all the assumptions are satisfied, sampling variation in $\widehat{\alpha}$ may result in there not being a solution to $\alpha_T(\alpha) = \widehat{\alpha}$ in any given sample. The effects of this are illustrated in the simulation section.

Proposition 3 verifies that $\widehat{\alpha}^{(\infty)}$ is consistent and asymptotically normal as $S \to \infty$ even if T is fixed, demonstrating that the inconsistency due to the incidental parameters may be completely removed through the use of an iterative bias-correction. The result may also be of considerable interest in the dynamic panel context, where it provides a simple alternative to GMM methods, although it does rely on strong exogeneity assumptions.

1.3.3 Asymptotics as $S,\ T \to \infty$

To establish the asymptotic properties of the least squares estimate of α as $S,\ T \to \infty$, I impose the following conditions in addition to model (1.11).

Assumption 2 ($S,\ T \to \infty$) *Suppose the data are generated by model (1.11) and*

NT1. $v_{st} = \bar{v}_{st}^{-\prime}\alpha + \eta_{st}$, *where η_{st} is strictly stationary in t for each s,* $\mathrm{E}[\eta_{st}^2] = \sigma_\eta^2$, $\mathrm{E}[\eta_{st}\eta_{s\tau}] = 0$ *for $t \neq \tau$, and the roots of* $1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p = 0$ *have modulus greater than 1.*

NT2. $\{X_s, V_s, \eta_s\}$ *are iid across s.* $\{Z_s\}$ *are nonstochastic and identical across s.*

NT3. (i) $[X, Z]$, *where* $X = [X_1', \ldots, X_S']'$ *and* $Z = \mathrm{diag}(Z_1, ..., Z_S)$ *has full rank. (ii)* $Z_s'Z_s$ *is uniformly positive definite with minimum eigenvalue* $\lambda_s \geq \lambda > 0$ *for all s.*

NT4. $\mathrm{E}[V_s|X_s] = 0$, $\mathrm{E}[V_s V_s'|X_s] = \Gamma(\alpha)$.

NT5. $\{X_{st}, V_{st}, \eta_{st}\}$ *is α-mixing of size* $\frac{-3r}{r-4}$, $r > 4$, *and* $z_{ith}^2 \leq \Delta < \infty$, $\mathrm{E}|x_{ith}^2|^{r+\delta} \leq \Delta < \infty$, *and* $\mathrm{E}|\eta_{it}^2|^{r+\delta} \leq \Delta < \infty$ *for some* $\delta > 0$ *and all* i, t, h.

Remark 1.3.4 The majority of the conditions imposed in Assumption 2 are standard for fixed effects panel models, with the key differences being the imposition of the AR(p) structure and the mixing conditions. Note that existence of absolute moments of order $2(r+\delta)$ and strict stationarity of η_{st} imply the existence of absolute moments of order $2(r+\delta)$ for v_{st} under the stationarity condition that the roots of $1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p = 0$ have modulus greater than 1.

Remark 1.3.5 A simple modification of model (1.11) is necessary for the results to accommodate trends. In particular, by redefining the coefficient on the trend as $\beta_{2h}^{sT} = T\beta_{2h}^{s}$ and the trend in each time period t as $\frac{t}{T}$, the trend becomes a uniform variable and the conditions in NT5 apply. It is straightforward to verify that the estimates of β_1, α, and other components of β_2^s obtained with the transformed data are numerically identical to the original estimates of β_1, α, and β_2^s. In addition, standard results for the coefficient on the trend are obtained by considering $\widehat{\beta}_{2h}^{s} - \beta_{2h}^{s} = \frac{1}{T}(\widehat{\beta}_{2h}^{sT} - \beta_{2h}^{sT})$.

Under Assumption 2, the following result is obtained.

Proposition 4 *If Assumption 2 is satisfied, $\sqrt{ST}(\widehat{\alpha}-\alpha) \xrightarrow{d} N(\rho B(\alpha), \Gamma_p^{-1}\Xi\Gamma_p^{-1})$ if $\frac{S}{T} \to \rho \geq 0$, where $\Xi = \lim_{T\to\infty} \frac{1}{T-p}\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \mathrm{E}[v_{st_1}^{-}\eta_{st_1}\eta_{st_2}v_{st_2}^{-'}]$. In addition, if η_{st} are independent for all s and t, $\Xi = \sigma_\eta^2\Gamma_p$.*

Remark 1.3.6 Proposition 4 provides the limiting distribution of $\widehat{\alpha}$ when $\frac{S}{T} \to \rho \geq 0$, confirming that there is bias in the limiting distribution even when S and T grow at the same rate - $\frac{S}{T} \to \rho > 0$.

In order to consider the properties of the bias correction as S, $T \to \infty$ jointly, the following additional assumption will be used.

Assumption 3 *For $B(\alpha, T)$ defined in Lemma 1.9.5,*

NT6. $B(\alpha, T)$ is continuously differentiable in α with bounded derivative uniformly in T for all α satisfying the stationarity condition given in NT1.

Remark 1.3.7 This assumption imposes that the derivatives of $\alpha_T(\alpha)$ are well-behaved as $T \to \infty$. This seems to be a reasonable assumption and is satisfied, for example, in the model where Z includes only fixed effects.

Proposition 5 *If Assumption 2 is satisfied,* $\sqrt{ST}(\widehat{\alpha}^{(1)} - \alpha) \xrightarrow{d} N(0, \Gamma_p^{-1}\Xi\Gamma_p^{-1})$ *for* Ξ *defined in Proposition 4 if* $\frac{S}{T} \to \rho \geq 0$. *If Assumption 3 is also satisfied, then* $\sqrt{ST}(\widehat{\alpha}^{(1)} - \alpha) \xrightarrow{d} N(0, \Gamma_p^{-1}\Xi\Gamma_p^{-1})$ *if* $\frac{S}{T^3} \to 0$.

Remark 1.3.8 This result demonstrates that $\widehat{\alpha}^{(1)}$ removes the bias from the asymptotic distribution of $\widehat{\alpha}$ as long as S grows more slowly than T^3. $\frac{S}{T^3} \to 0$ may be a good approximation in many situations, such as the CPS data examined in the simulation section. It would also be straightforward to demonstrate that, for a finite number of iterations k, $\widehat{\alpha}^{(k)}$ removes the bias from the asymptotic distribution of $\widehat{\alpha}$ as long as S grows more slowly than T^{2k+1}.

Proposition 6 *If Assumptions 2 and 3 are satisfied and* $H = D\alpha_T(\alpha)$ *is invertible for all* α *satisfying the stationarity condition given in NT1 uniformly in* T, *where* $D\alpha_T(\alpha)$ *is the derivative matrix of* $\alpha_T(\alpha)$ *in* α, $\sqrt{ST}(\widehat{\alpha}^{(\infty)} - \alpha) \xrightarrow{d} N(0, \Gamma_p^{-1}\Xi\Gamma_p^{-1})$ *for* Ξ *defined in Proposition 4.*

Remark 1.3.9 This confirms the result obtained with T fixed, verifying that the iterative bias-correction removes the bias from the limiting distribution of $\widehat{\alpha}$ as S, $T \to \infty$ for any S, T sequence. The additional condition imposed in the proposition guarantees that the inverse of $\alpha_T(\alpha)$ exists for all T and seems to be a reasonable assumption. For example, it is satisfied when Z includes only fixed effects.

Propositions 4 and 5 mirror similar results from Hahn and Kuersteiner (2002) and Hahn and Newey (2002), demonstrating that bias remains in the limiting distribution of $\widehat{\alpha}$ even when T grows as fast as S, but that this bias is removed if the one-step bias correction is used as long as T grows fast enough relative to S. Hahn and Newey (2002) also suggest iterating the bias correction in nonlinear models and provide some simulation results, though without asymptotic theory.

It is important to note that Propositions 4, 5, and 6 ignore the estimation of time effects, which would further complicate the analysis. The time effects will be $\sqrt{S}$-, not $\sqrt{ST}$-, consistent, which will generally add an $O\left(\frac{1}{S}\right)$ bias to the estimator. This will not affect the stated results as long as $\frac{S}{T} \to \infty$. However, if S and T grow at the same rate, the inclusion of the time effects will result in bias in the limiting distribution of all the estimators, including the bias-corrected ones. In many applications, this should not be a large source of bias as S is typically larger than T. In addition, the neglected term is a sum over states not over time, so the additional bias will include only contemporaneous correlations. Overall, it seems that any bias coming from the time fixed effect is likely to be small, and the simulations in the next section also suggest that ignoring this source of bias provides a reasonable approximation.

1.3.4 Implications for FGLS Estimation

While the bias-correction results presented above may be interesting for a number of reasons, the chief motivation for their development in this paper is for use in FGLS estimation as outlined in Section 1.2.

To develop the properties of the GLS estimator of β_1, define $\widetilde{X}_s = \Gamma(\alpha)^{-1/2}X_s$ and $\widetilde{Z}_s = \Gamma(\alpha)^{-1/2}Z_s$ where $\Gamma(\alpha) = \mathrm{E}[V_sV_s']$. Then, under standard conditions and using conventional arguments, it follows that the GLS estimator of β_1, $\widehat{\beta}_1(\alpha)$, is consistent and that $\sqrt{S}(\widehat{\beta}_1(\alpha) - \beta_1) \xrightarrow{d} N(0, V(\alpha)^{-1})$ where $V(\alpha) = \mathrm{E}[\widetilde{X}_s'\widetilde{X}_s - \widetilde{X}_s'\widetilde{Z}_s(\widetilde{Z}_s'\widetilde{Z}_s)^{-1}\widetilde{Z}_s'\widetilde{X}_s]$ as $S \to \infty$ with T fixed and that $\widehat{\beta}_1(\alpha)$ is consistent and that $\sqrt{ST}(\widehat{\beta}_1(\alpha) - \beta_1) \xrightarrow{d} N(0, V(\alpha)^{-1})$ for $V(\alpha) = \lim_{T\to\infty} \mathrm{E}[\widetilde{X}_s'\widetilde{X}_s - \widetilde{X}_s'\widetilde{Z}_s(\widetilde{Z}_s'\widetilde{Z}_s)^{-1}\widetilde{Z}_s'\widetilde{X}_s]$ as $S,\ T \to \infty$. Furthermore, the GLS estimator is the Gauss-Markov estimator and hence is efficient among linear estimators.

Then letting $\widehat{\beta}_1(\widetilde{\alpha})$ denote the FGLS estimate of β_1 when the covariance matrix is constructed using $\widetilde{\alpha}$ and using standard arguments, it is straightforward to show that $\sqrt{S}(\widehat{\beta}_1(\widehat{\alpha}^{(\infty)}) - \widehat{\beta}_1(\alpha)) = o_p(1)$ in asymptotics where $S \to \infty$ and that $\sqrt{ST}(\widehat{\beta}_1(\widetilde{\alpha}) - \widehat{\beta}_1(\alpha)) = o_p(1)$ for any $\widetilde{\alpha} \xrightarrow{p} \alpha$ in asymptotics where $S,\ T \to \infty$. This result indicates that, as $S \to \infty$ with T fixed, FGLS based on the iteratively bias-corrected estimator of α will yield efficiency gains relative to the other estimators of α considered above, but that there is no efficiency gain from using a bias-corrected estimate of α in asymptotics with $S,\ T \to \infty$.

While these results suggest that there is little motivation to use bias-corrected estimates of α in performing FGLS estimation and inference in circumstances where $S,\ T \to \infty$ asymptotics provide a good approximation, there are at least two reasons that it may still be preferable. First, letting $\widehat{V}(\widetilde{\alpha}) = \frac{1}{ST}\sum_{s=1}^{S} X_s'\Gamma(\widetilde{\alpha})^{-1}X_s - X_s'\Gamma(\widetilde{\alpha})^{-1}Z_s(Z_s'\Gamma(\widetilde{\alpha})^{-1}Z_s)^{-1}Z_s'\Gamma(\widetilde{\alpha})^{-1}X_s$, it is straightforward to show that $\sqrt{ST}(\widehat{V}(\widetilde{\alpha}) - V(\alpha)) = \sum_{i=1}^{p} \frac{\partial(\widehat{V}(\widetilde{\widetilde{\alpha}}))}{\alpha_i} \sqrt{ST}(\widetilde{\alpha}_i - \alpha_i) + \sqrt{ST}(\widehat{V}(\alpha) - V(\alpha))$ where $\widetilde{\widetilde{\alpha}}$ is an intermediate value between $\widetilde{\alpha}$ and α. In other words, using $\widetilde{\alpha}$ to estimate the variance matrix of the FGLS estimator will result in bias in the estimate of the same order as the bias in $\widetilde{\alpha}$, implying that the use of the bias-corrected estimates will be preferable for performing inference on β_1. Second, it seems likely that there would be higher order improvements to the FGLS estimator of β_1 if a bias-corrected estimate of α is used. Exploration of this possibility is left to future research, though the hypothesis receives some support from the simulation results.

As a practical matter, it is important to note that, while unnecessary for estimation of α and β_1, knowledge of σ_η^2 is necessary for performing inference on β_1. A natural estimator can be constructed as $\widehat{\sigma}_\eta^2 = \frac{\widehat{\eta}'\widehat{\eta}}{S(T-k_2)-k_1}$ where $\widehat{\eta}$ are residuals obtained from regressing $\sigma_\eta \Gamma(\widetilde{\alpha})^{-1/2} Y$ on $\sigma_\eta \Gamma(\widetilde{\alpha})^{-1/2} \widetilde{X}$ and $\sigma_\eta \Gamma(\widetilde{\alpha})^{-1/2} \widetilde{Z}$. It is straightforward to demonstrate that this estimator will be consistent for σ_η^2 if $\widetilde{\alpha} = \widehat{\alpha}^{(\infty)}$ in asymptotics where $S \to \infty$ with T fixed. In asymptotics where S, $T \to \infty$ jointly, $\widehat{\sigma}_\eta^2$ will be consistent for any $\widetilde{\alpha} \xrightarrow{p} \alpha$, but the estimator will be biased to the same order as $\widetilde{\alpha}$.

Finally, it should be observed that the validity of the results presented here requires stringent assumptions about the exact nature of the error process. In practice, a researcher may be concerned that these assumptions are not satisfied. For example, one may suspect that there is temporal heteroskedasticity or that the AR process is not constant across states. In these cases, the FGLS estimates obtained assuming homoskedasticity and constant AR coefficients will still generally be consistent and asymptotically normal and may still offer efficiency gains over OLS, although use of a robust variance matrix will be necessary for correct inference.[10] This approach is examined in the simulation section.

1.4 Monte Carlo Evidence

In order to provide evidence on the performance of the proposed methods, I performed a Monte Carlo experiment based on data from the CPS-MORG. The data are for women in their fourth interview month for the years 1979 to 2001, and the sample is restricted to women

[10]The use of FGLS with a robust variance matrix has also been suggested by Wooldridge (2003) and Liang and Zeger (1986).

aged 25 to 50 who report positive weekly earnings.[11] With these restrictions imposed, the total sample size is 600,941 observations in 1173 state-year cells, which generates an average cell size of approximately 512 observations. The dependent variable, y_{ist} is defined as the log of the weekly wage, and covariates include a quartic in age, four education dummies, and state and time fixed effects. Iteratively bias-corrected AR(4) parameter estimates (standard errors) in the actual data are $\widehat{\alpha}_1 = 0.397$ (0.032), $\widehat{\alpha}_2 = 0.268$ (0.034), $\widehat{\alpha}_3 = 0.146$ (0.034), and $\widehat{\alpha}_4 = 0.058$ (0.032), where the last coefficient is insignificant at the 95% level.

I consider three different simulation designs. In Design 1, I draw from the actual data by resampling states and include a randomly generated treatment which varies at the state-year level, x_{st}, as a regressor which enters the model with $\beta_1 = 0$. In Design 2 and Design 3, I aggregate the data to state-year cells by estimating equation (1.1) and saving the estimated fixed effects. I then regress the estimated fixed effects on all regressors which are constant within state-year cells and treat these coefficient estimates as the "true" parameters. Then for each simulation iteration, I construct C_{st} from model (1.2) using the parameters estimated in the previous step and a randomly generated treatment, x_{st} which enters with coefficient $\beta_1 = 0$. In Design 2, I assume v_{st} follows an AR(1) with $\alpha_1 = 0.8$, and in Design 3, v_{st} follows an AR(2) with $\alpha_1 = 0.43$ and $\alpha_2 = 0.30$. In both cases, I construct the error term so that its variance is similar to the empirical variance in the sample, and all estimates are constructed ignoring estimation of C_{st}. In all cases, I generate the treatment by randomly selecting 26 states to be treated and then randomly selecting a start date for the treatment, which may be any but the first period. The treatment variable is a dummy variable which equals one in the treatment year and all years following, and I allow the treatment date to be different in each treated state. When simulating from the actual data with $T = 12$ ($T = 6$), I use the most recent 12 (6) years of data. In the simulated data, the time blocks are drawn randomly.

[11] These are the same sample selection criteria as used in Bertrand, Duflo, and Mullainathan (2002), though the Bertrand, Duflo, and Mullainathan (2002) study only had data from 1979 to 1999. In addition, the data are aggregated to state-year cells using different methods in their paper. However, the OLS and clustered results I report are similar to those in Bertrand, Duflo, and Mullainathan (2002).

1.4.1 Bias of AR(p) Parameter Estimates

Before turning to inference on the treatment effect, it is useful to consider the bias of uncorrected and bias-corrected estimates of the AR parameters. Tables 1 and 2 contain the bias and MSE (in parentheses) when the model is specified as (1.2) and the model is simulated using Design 2. The model in Table 1 contains only a fixed effect, while the model in Table 2 contains a fixed effect and state-specific time trend. In the columns corresponding to $\widehat{\alpha}^{(\infty)}$, the number in brackets represents the number of instances in which $\alpha_T^{-1}(\alpha)$ did not exist. In these cases, $\widehat{\alpha}^{(\infty)}$ was set equal to $\widehat{\alpha}^{(1)}$.

The results in Table 1 clearly demonstrate that the uncorrected estimates suffer from substantial bias, even when T is reasonably large. In addition, the results show that both the one-step and iterative bias-corrections eliminate a large portion of the bias in all cases considered, though a sizable bias remains in the one-step estimator for small T. The results also illustrate the consistency of the iterative bias-correction as $S \to \infty$ with T fixed, though bias goes away much more slowly in S than in T. The results which include state-specific trends, Table 2, follow essentially the same pattern, though the biases are larger in all cases with substantial biases remaining in even the iteratively bias-corrected estimator for small S and T. Overall, the results suggest that both of the derived bias-corrections are effective in removing a large component of the bias in the AR parameter estimates, though the iterative procedure dominates in terms of both bias and MSE.

1.4.2 Inference on the Treatment Effect

The results on the bias of estimates of α in the fixed effects model presented above are encouraging, and results for the use of the bias-correction in performing inference on the estimated treatment effect also suggest the bias-correction is useful in obtaining tests with accurate size and good power in clustered samples with autocorrelation.

Results for inference about the treatment parameter are contained in Tables 3 to 7. In each table, the first three columns use the full sample of 51 states and 23 years, while the middle three columns use 12 years of data and the last three use only 6 years of data. Rows labeled OLS contain test results from the OLS estimates without any adjustment to the standard errors, and rows labeled cluster use variance matrices that are robust to correlation within groups at the specified levels. E.g. a row labeled with "Cluster by State" uses a variance matrix which is robust to arbitrary correlation among all observations within a state. The use of robust variance matrices in the individual level data allowing for correlation at the level of the aggregate data (the state-year level in this case) is probably the most commonly used method for accounting for possible correlations arising due to the use of aggregate and individual level data. Bertrand, Duflo, and Mullainathan (2003) suggest using this correction, clustering at the state level instead of the state-year level, and find that this procedure yields tests with approximately correct size in their simulation study. The row labeled random effects reports results from the standard random effects estimator allowing for correlation at the state-year level, and rows labeled "FGLS-U" use the FGLS approach suggested in Kiefer (1980) which does not constrain the variance matrix over time within states but assumes the variance matrix is identical across states. The remaining rows contain test results based on FGLS where the state-year shock is assumed to follow the specified process; the "bc" subscript indicates the use of the iteratively bias-corrected AR parameter estimates in the FGLS estimation and inference. The rows designated "AR(p)-Cluster by state" estimate the model using FGLS based on an AR(p) process and then use a robust variance matrix clustered at the state level for inference, while the rows labeled "AR(p)" use the standard GLS formula to estimate the variance matrix. Within each table, I report results from conventional inference methods in Panel A and results which use the bias-correction procedure developed in this paper in Panel B.

Table 3 contains results regarding the variance of the estimated treatment parameter, $\widehat{\beta}_1$. The columns labeled $\widehat{\sigma}^2$ report the mean of the estimated variance of $\widehat{\beta}_1$, while the corresponding asymptotic variance and variance of the simulation estimates of β_1 are contained in the columns labeled σ_a^2 and σ_s^2, respectively. Simulation results are for data generated using Design 2 described above. For readability, all results are multiplied by 1000.

The results summarized in Table 3 provide strong evidence supporting the use of FGLS estimation with bias-corrected estimators of the AR-parameters. While the difference between the asymptotic variance and the mean of the estimated variance is small for all of the estimators considered with the exception of "FGLS-U" and the unadjusted OLS estimator, the variances estimated from FGLS with bias-corrected AR-parameters are always approximately unbiased for the asymptotic variance of the estimator. There is also some weak evidence that the bias in the estimated variance based on the uncorrected AR-parameter estimates remains even for moderate T, though the result is not striking. Unsurprisingly, it appears that the asymptotic approximation of the FGLS estimator performs substantially better when the FGLS is based on the bias-corrected AR parameters than when the uncorrected estimates are used, and relative to FGLS with conventional AR parameter estimates, there does appear to be a sizable efficiency gain to using the bias-corrected estimates.

The results for variance estimation presented in Table 3 suggest that the use of FGLS with bias-corrected AR parameter estimates may result in substantial improvements in inference over OLS-based methods or FGLS with uncorrected AR parameter estimates. Tables 4 to 7, which contain results for size and power of hypothesis tests about the treatment parameter as well as confidence interval lengths, provide further evidence on the potential gains to using FGLS with bias-corrected AR parameters. In all cases, size and power are for 5% level tests, and power is versus the alternative that $\beta_1 = 0.02$[12]. The reported interval length is the confidence interval length divided by two.

Tables 4 and 5 summarize the results for the simulation based on Design 1 outlined above. Table 4 reports results from estimation in the individual level data, while Table 5 contains the results from estimation in data aggregated using the aggregation method of Amemiya (1978) outlined in Section 1.2.3 and ignoring the first stage estimation of C_{st}.

The results in Table 4 clearly illustrate the potential pitfalls in using individual level data with aggregate level variables. As expected, the uncorrected OLS estimates have large size distortions for moderate T, though the size distortion is modest when $T = 6$. The rejection rates for a 5% level test are 0.594 with $T = 23$, 0.398 with $T = 12$, and 0.072 with $T = 6$. Mirroring results from Bertrand, Duflo, and Mullainathan (2003), I also find that, for $T = 12$ and $T = 23$, tests which allow for correlation within state-year cells but not over time suffer from severe size distortions, but that tests based on OLS with standard errors clustered at the state level remove much of the distortion, rejecting 7.8% of the time for a 5% level test in both cases. For $T = 23$, the tests based on parametric FGLS with bias-corrected coefficients also remove much of the size distortion, producing similar rejection rates to tests based on OLS with clustered standard errors. For $T = 12$, the FGLS estimates remain more distorted than the OLS-based test using robust standard errors, though the robust FGLS tests have similar size to the robust OLS-based test. As would be anticipated, all the FGLS-based tests,

[12]The dependent variable is the log of the weekly wage, so an impact of .02 represents an approximate 2% increase in weekly wages. This is the magnitude of the effect considered in Bertrand, Duflo, and Mullainathan (2003).

including those which use robust standard errors, do have substantially more power against an alternative of .02 than the test using OLS and clustering standard errors at the state level. In addition, the confidence intervals of OLS with standard errors clustered by state are substantially longer than the FGLS intervals. It is interesting that, with $T = 6$, serial correlation does not appear to play much of a role. In this case, none of the size distortions are large, and the random effects estimator has correct size and good power relative to the other tests.

The results summarized in Table 5 follow a similar pattern to those in Table 4, though in most cases the size distortions are smaller.[13] In general, tests based on OLS with clustered standard errors, tests based on bias-corrected FGLS, and tests based on FGLS with robust standard errors have similar sizes. However, the FGLS tests are more powerful against the alternative that $\beta_1 = .02$ and have shorter confidence intervals. In many cases, tests based on FGLS with bias-corrected AR parameters and robust standard errors are more size distorted than the corresponding tests without robust standard errors. This distortion seems likely to be due to the small sample bias of the robust standard errors discussed in Bell and McCaffrey (2002) and illustrated in Table 3. Also, as in Table 4, serial correlation does not seem to pose a serious problem to inference with $T = 6$. In this case, the unadjusted OLS has correct size as does the OLS test which uses clustered standard errors. Finally, it is interesting that FGLS estimation using a variance matrix which is unconstrained within states ("FGLS-U") does poorly in all cases. While this is unsurprising for moderate T, the poor performance with $T = 6$ suggests that even with a reasonably short time series dimension the added variability induced by estimating an unconstrained variance matrix poses a serious problem for inference. Also, comparing across Tables 4 and 5, it appears that the loss of efficiency due to aggregating is small and that tests performed in the aggregate data suffer from smaller size distortions, suggesting that performing inference in the aggregate data may be preferable to using the individual level data.

Tables 6 and 7 summarize the results from the simulation models based on Design 2 and Design 3. These data are simulated without taking into account estimation of C_{st} and so are representative of standard panel data. The results follow the same general pattern of those presented in Table 5, though the sizes are generally closer to the actual size of the test. In particular, the results show a substantial bias in the uncorrected OLS tests which

[13] Dickens (1990) presents some arguments for why this may be so in a different but related context.

is largely eliminated by clustering or the use of FGLS with bias-corrected AR coefficients. A comparison of the power and interval lengths of FGLS and OLS with clustered standard errors clearly demonstrates the large potential efficiency gain to using FGLS, and the results also indicate that the use of an unconstrained variance matrix is problematic even for small T.

Overall, the simulation results support the use of FGLS methods for performing inference in the type of models examined here. Tests based on bias-corrected FGLS do not appear to be substantially more size-distorted than the OLS tests with standard errors robust to arbitrary correlation within states but have much higher power and shorter confidence intervals in the majority of cases. This improved performance also appears to hold when estimation is performed with FGLS and robust standard errors are used, though in some cases this does result in a larger size distortion to the test. It would be interesting to see if performance in these cases could be further improved using the bias-reduction and degrees of freedom corrections outlined in Bell and McCaffrey (2002).

1.5 Conclusion

Many policy analyses rely on data which vary at both the individual and aggregate level. The grouped structure of the data gives rise to many potential sources of correlation between individual observations. In particular, the presence of group level shocks will result in correlation among all individuals within a group. In addition, if groups are followed over time, correlation between individuals in the same group at different times may arise due to serial correlation in the group level shock. While there are numerous solutions to the first source of correlation, relatively little attention has been paid to the potential problems which may be caused by the second. Bertrand, Duflo, and Mullainathan (2003) illustrate that serial correlation in the group level shock may cause conventional tests to be highly misleading, and offer several OLS-based strategies which yield tests with correct size, but have low power against relevant alternatives.

In this paper, I explore FGLS estimation in data with a grouped structure where the groups may be autocorrelated and present a simple method for obtaining the FGLS estimates which will be valid as the number of individual observations within each aggregate cell grows large. I then focus on the case where the group level shock follows an AR(p) process.

In this case, standard estimates of the AR coefficients will typically be biased due to the incidental parameters problem. I offer a simple bias correction for these coefficients which will be valid in the presence of fixed effects or other variables with coefficients that vary at the group level. The usefulness of FGLS and the derived bias-correction for the AR parameters is demonstrated through a simulation study based on data from the CPS-MORG. The simulation results show that the proposed bias-correction removes a substantial portion of the bias from the AR parameter estimates. The results also demonstrate that tests based on FGLS using bias-corrected AR parameter estimates have approximately correct size. In addition, the simulations confirm that the FGLS-based tests have much higher power and yield much shorter confidence intervals than their OLS-based counterparts.

The simulation results clearly illustrate the potential usefulness of FGLS procedures in data which include aggregate and individual level variables. Useful extensions would include considering more general time series models and allowing for cross-sectional as well as intertemporal correlation. It also seems that pursuing more flexible methods of estimating the covariance matrix for use in FGLS, as in Hausman and Kuersteiner (2003), may be worthwhile. In addition, it would be interesting to extend these results to estimation of nonlinear models when the data are serially correlated and have a grouped structure.

1.6 Appendix 1. Verification of Proposition 1

Proposition 1 follows from straightforward manipulation of the GLS objective function. Recall that $\Sigma = \Lambda + D\Omega D'$, and let $\widetilde{D} = \Lambda^{-1/2}D$. Then

$$\begin{aligned}
\Sigma^{-1} &= \Lambda^{-1} - \Lambda^{-1}D\left[\Omega^{-1} + D'\Lambda^{-1}D\right]^{-1}D'\Lambda^{-1} \\
&= \Lambda^{-1/2}[I - \widetilde{D}[\Omega^{-1} + \widetilde{D}'\widetilde{D}]^{-1}\widetilde{D}']\Lambda^{-1/2} \\
&= \Lambda^{-1/2}[I - \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}' + \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}[(\widetilde{D}'\widetilde{D})^{-1} + \Omega]^{-1}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}']\Lambda^{-1/2} \\
&= \Lambda^{-1/2}[M_{\widetilde{D}} + \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{V}_B^{-1}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}']\Lambda^{-1/2}
\end{aligned}$$

where $\widetilde{V}_B = (\widetilde{D}'\widetilde{D})^{-1} + \Omega$ and $M_{\widetilde{D}} = I - \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}'$.

Then the result follows immediately from $\widehat{\theta}_{GLS} = \arg\min_{\theta}(Y - \Phi\theta)'\Sigma^{-1}(Y - \Phi\theta)$ and

$$\begin{aligned}(Y - \Phi\theta)'\Sigma^{-1}(Y - \Phi\theta) &= (Y - \Phi\theta)'\Lambda^{-1/2}[M_{\widetilde{D}} + \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{V}_B^{-1}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}']\Lambda^{-1/2}(Y - \Phi\theta) \\ &= (\widetilde{Y} - \widetilde{\Phi}\theta)'(M_{\widetilde{D}} + \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1/2}\widetilde{V}_B^{-1/2}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}')'\times \\ &\quad (M_{\widetilde{D}} + \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1/2}\widetilde{V}_B^{-1/2}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}')(\widetilde{Y} - \widetilde{\Phi}\theta).\end{aligned}$$

1.7 Appendix 2. Notation

The following notation will be used throughout Appendix 1.8 and Appendix 1.9.

Suppose the panel has a cross-sectional dimension $s = 1, \ldots, S$ and a time-series dimension $t = 1, \ldots, T$ with $T > 2p$ where p is the order of the autoregressive process. Let

$$C_{st} = x'_{st}\beta_1 + z'_{st}\beta_2^s + v_{st}, \tag{1.12}$$

or, in vector notation, $C_s = X_s\beta_1 + Z_s\beta_2^s + V_s$, where $C_s = [C_{s1}, \ldots, C_{sT}]'$ is $T \times 1$, $X_s = [x_{s1}, \ldots, x_{sT}]'$ is $T \times k_1$, $Z_s = [z_{s1}, \ldots, z_{sT}]'$ is $T \times k_2$, and $V_s = [v_{s1}, \ldots, v_{sT}]'$ is $T \times 1$. Also, let x_{sth} be the h^{th} element of x_{st} so that $x'_{st} = [x_{st1}, \ldots, x_{stk_1}]$, and define z_{sth} similarly.

Let v_{st}^- be a $p \times 1$ vector with $v_{st}^- = [v_{s(t-p)}, \ldots, v_{s(t-1)}]'$.

Define $\ddot{v}_{st} = v_{st} - z'_{st}(Z'_sZ_s)^{-1}Z'_sV_s$, $\ddot{x}'_{st} = x'_{st} - z'_{st}(Z'_sZ_s)^{-1}Z'_sX_s$, $\ddot{V}_s = [\ddot{v}_{s1}, \ldots, \ddot{v}_{sT}]'$, and $\ddot{X}_s = [\ddot{x}_{s1}, \ldots, \ddot{x}_{sT}]'$.

Throughout, let $\|A\| = [\text{trace}(A'A)]^{1/2}$ be the Euclidean norm of a matrix A.

1.8 Appendix 3. Proof of Proposition 2 and Proposition 3

Proposition 2 is then verified by combining Lemmas 1.8.3 and 1.8.4 below, and Proposition 3 follows from Proposition 2 and Lemma 1.8.6.

All results presented below are for asymptotics where $S \to \infty$ with T fixed.

Proof of Proposition 2. Immediate from Lemma 1.8.3 and Lemma 1.8.4. ■

Proof of Proposition 3. That $\alpha_T(\alpha)$ is continuously differentiable in α and that $H = D\alpha_T(\alpha)$ is invertible for all α such that N1 is satisfied imply that $\alpha_T(\alpha)$ is invertible for all α such that N1 is satisfied by the Inverse Function Theorem. (See, e.g. Fitzpatrick (1996) Theorem 16.9.) $\widehat{\alpha}^{(\infty)} - \alpha \xrightarrow{p} 0$ then follows immediately from the definition of $\widehat{\alpha}^{(\infty)}$ and Proposition 2.

To verify the asymptotic normality, expand $\widehat{\alpha}^{(\infty)}$ about $\widehat{\alpha} = \alpha_T(\alpha)$. This gives

$$\widehat{\alpha}^{(\infty)} = \alpha_T^{-1}(\alpha_T(\alpha)) + H^{-1}|_{\widetilde{\alpha_T(\alpha)}}(\widehat{\alpha} - \alpha_T(\alpha)),$$

where $\widetilde{\alpha_T(\alpha)}$ is an intermediate value between $\widehat{\alpha}$ and $\alpha_T(\alpha)$. The conclusion then follows from continuity of H, Proposition 2, and Lemma 1.8.6. ∎

1.8.1 Lemmas

Lemma 1.8.1 *Let $\widehat{\beta}_1$ be the ordinary least squares estimate of β_1. Then if the conditions of Assumption 1 are satisfied, $\widehat{\beta}_1 - \beta_1 \xrightarrow{p} 0$ and $\sqrt{S}(\widehat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, M^{-1}\Omega M^{-1})$, where $M = \mathrm{E}[\ddot{X}_s'\ddot{X}_s]$ and $\Omega = \mathrm{E}[\ddot{X}_s'\Gamma_(\alpha)\ddot{X}_s]$.*

Proof. Using $\|AB\| \leq \|A\|\|B\|$ and $I - Z_s(Z_s'Z_s)^{-1}Z_s$ positive semi-definite, $\mathrm{E}\|\ddot{X}_s'\ddot{X}_s\| \leq \mathrm{E}\|\ddot{X}_s\|^2 = \mathrm{E}[\mathrm{trace}(X_s'X_s - X_s'Z_s(Z_s'Z_s)^{-1}Z_s'X_s)] \leq \mathrm{trace}(\mathrm{E}X_s'X_s) = \mathrm{trace}(\sum_{t=1}^T \mathrm{E}[x_{st}x_{st}']) < \infty$ by N5. Also, $\|\ddot{X}_s'\ddot{V}_s\| \leq (\mathrm{E}\|\ddot{X}_s'\|^2\mathrm{E}\|\ddot{V}_s\|^2)^{1/2} < \infty$ from the Cauchy-Schwarz inequality, N5, and the same argument as above. The Khinchin LLN then yields $\frac{1}{S}\sum_{s=1}^S \ddot{X}_s'\ddot{X}_s \xrightarrow{p} M$ and $\frac{1}{S}\sum_{s=1}^S \ddot{X}_s'\ddot{V}_s \xrightarrow{p} 0$, from which $\widehat{\beta}_1 - \beta_1 \xrightarrow{p} 0$ follows.

To show asymptotic normality of $\widehat{\beta}_1$ first note that $\ddot{X}_s'\ddot{V}_s$ is iid and has mean zero by N2 and N4. Also, $\mathrm{E}\|\ddot{X}_s'\ddot{V}_s\ddot{V}_s'\ddot{X}_s\| \leq (2\mathrm{E}\|X_s\|^4\mathrm{E}\|V_s\|^4)^{1/2} < \infty$ by N5, $\|AB\| \leq \|A\|\|B\|$, the Cauchy-Schwarz inequality, and

$$\begin{aligned}\mathrm{E}\|\ddot{X}_s\|^4 &= \mathrm{E}[(\mathrm{trace}(X_s'X_s))^2 - 2\mathrm{trace}(X_s'X_s)\mathrm{trace}(X_s'Z_s(Z_s'Z_s)^{-1}Z_s'X_s) \\ &\quad + (\mathrm{trace}(X_s'Z_s(Z_s'Z_s)^{-1}Z_s'X_s))^2] \\ &\leq \mathrm{E}[2(\mathrm{trace}(X_s'X_s))^2] = 2\mathrm{E}\|X_s\|^4,\end{aligned}$$

where the inequality follows from $X_s'X_s$, $X_s'Z_s(Z_s'Z_s)^{-1}Z_s'X_s$, and $I - Z_s(Z_s'Z_s)^{-1}Z_s$ positive semi-definite. It then follows from the Lindeberg-Levy CLT that $\frac{1}{\sqrt{S}}\sum_{s=1}^S \ddot{X}_s'\ddot{V}_s \xrightarrow{d}$

$N(0,\Omega)$ since $\mathrm{E}[\ddot{X}_s'\ddot{V}_s\ddot{V}_s'\ddot{X}_s] = \mathrm{E}[\ddot{X}_s'V_sV_s'\ddot{X}_s] = \mathrm{E}[\ddot{X}_s'\Gamma(\alpha)\ddot{X}_s]$, from which $\sqrt{S}(\widehat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, M^{-1}\Omega M^{-1})$ is obtained. ■

Lemma 1.8.2 *Define $\widetilde{v}_{st}$ to be the residual from least squares regression of (1.11); i.e. $\widetilde{v}_{st} = C_{st} - x_{st}'\widehat{\beta}_1 - z_{st}'\widehat{\beta}_2^s = v_{st} - x_{st}'(\widehat{\beta}_1 - \beta_1) - z_{st}'(\widehat{\beta}_2^s - \beta_2) = \ddot{v}_{st} - \ddot{x}_{st}'(\widehat{\beta}_1 - \beta_1)$, where $\widehat{\beta}_1$ and $\widehat{\beta}_2^s$ are least squares estimates of β_1 and β_2^s. Let $\widetilde{v}_{st}^-$ be a $p \times 1$ vector with $\widetilde{v}_{st}^- = [\widetilde{v}_{s(t-p)}, \ldots, \widetilde{v}_{s(t-1)}]'$, and let $\ddot{v}_{st}^-$ be a $p \times 1$ vector with $\ddot{v}_{st}^- = [\ddot{v}_{s(t-p)}, \ldots, \ddot{v}_{s(t-1)}]'$. Under the conditions of Assumption 1, $\frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \widetilde{v}_{st}^-\widetilde{v}_{st}^{-\prime} = \frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \ddot{v}_{st}^-\ddot{v}_{st}^{-\prime} + o_p(S^{-1/2})$, and $\frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \widetilde{v}_{st}^-\widetilde{v}_{st} = \frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \ddot{v}_{st}^-\ddot{v}_{st} + o_p(S^{-1/2})$.*

Proof.

$$\frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \widetilde{v}_{st}^-\widetilde{v}_{st} = \frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T (\ddot{v}_{st}^-\ddot{v}_{st} - \ddot{x}_{st}^-(\widehat{\beta}_1 - \beta_1)\ddot{v}_{st} - \ddot{v}_{st}^-\ddot{x}_{st}'(\widehat{\beta}_1 - \beta_1) + \ddot{x}_{st}^-(\widehat{\beta}_1 - \beta_1)(\widehat{\beta}_1 - \beta_1)'\ddot{x}_{st}),$$

where $\ddot{x}_{st}^-$ is a $p \times k_1$ matrix with $\ddot{x}_{st}^- = [\ddot{x}_{s(t-p)}, \ldots, \ddot{x}_{s(t-1)}]'$. Note

$$\mathrm{vec}(\frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \ddot{x}_{st}^-(\widehat{\beta}_1 - \beta_1)\ddot{v}_{st}) = \{\frac{1}{S}\sum_{s=1}^S [\sum_{t=p+1}^T (\ddot{v}_{st} \otimes \ddot{x}_{st}^-)]\}(\widehat{\beta}_1 - \beta_1) = o_p(1)O_p(S^{-1/2})$$

by Lemma 1.8.1, Assumption 1, and the Khinchin LLN. Similarly,

$$\mathrm{vec}(\frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \ddot{v}_{st}^-\ddot{x}_{st}'(\widehat{\beta}_1 - \beta_1)) = o_p(1)O_p(S^{-1/2}),$$

and

$$\mathrm{vec}(\frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \ddot{x}_{st}^-(\widehat{\beta}_1 - \beta_1)(\widehat{\beta}_1 - \beta_1)'\ddot{x}_{st}) = O_p(1)O_p(S^{-1/2})O_p(S^{-1/2}).$$

It then follows that $\frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \widetilde{v}_{st}^-\widetilde{v}_{st} = \frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \ddot{v}_{st}^-\ddot{v}_{st} + o_p(S^{-1/2})$ and, by a similar argument, $\frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \widetilde{v}_{st}^-\widetilde{v}_{st}^{-\prime} = \frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \ddot{v}_{st}^-\ddot{v}_{st}^{-\prime} + o_p(S^{-1/2})$. ■

Lemma 1.8.3 *Under the conditions of Assumption 1,*

$$\frac{1}{S}\sum_{s=1}^S \sum_{t=p+1}^T \ddot{v}_{st}^-\ddot{v}_{st}^{-\prime} \xrightarrow{p} \mathrm{E}[\sum_{t=p+1}^T \ddot{v}_{st}^-\ddot{v}_{st}^{-\prime}] = (T-p)(\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha)),$$

and

$$\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st} \xrightarrow{p} \mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}] = (T-p)(A(\alpha)+\frac{1}{T-p}\Delta_A(\alpha)),$$

where $\Gamma_p(\alpha)+\frac{1}{T-p}\Delta_\Gamma(\alpha)$ *is a* $p\times p$ *matrix with* i,j *element*

$$\begin{aligned}\mathrm{E}\left[\frac{1}{T-p}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}\right]_{[i,j]} &= \gamma_{|i-j|}(\alpha)-\frac{1}{T-p}\mathrm{trace}\left(Z_s'\Gamma_{-i}(\alpha)Z_{s,-j}(Z_s'Z_s)^{-1}\right)\\ &- \frac{1}{T-p}\mathrm{trace}\left(Z_s'\Gamma_{-j}(\alpha)Z_{s,-i}(Z_s'Z_s)^{-1}\right)\\ &+ \frac{1}{T-p}\mathrm{trace}\left(Z_s'\Gamma(\alpha)Z_s(Z_s'Z_s)^{-1}Z_{s,-i}'Z_{s,-j}(Z_s'Z_s)^{-1}\right),\end{aligned} \tag{1.13}$$

$A(\alpha)+\frac{1}{T-p}\Delta_A(\alpha)$ *is a* $p\times 1$ *vector with* i^{th} *element*

$$\begin{aligned}\mathrm{E}\left[\frac{1}{T-p}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}\right]_{[i,1]} &= \gamma_i(\alpha)-\frac{1}{T-p}\mathrm{trace}\left(Z_s'\Gamma_{-i}(\alpha)Z_{s,-0}(Z_s'Z_s)^{-1}\right)\\ &- \frac{1}{T-p}\mathrm{trace}\left(Z_s'\Gamma_{-0}(\alpha)Z_{s,-i}(Z_s'Z_s)^{-1}\right)\\ &+ \frac{1}{T-p}\mathrm{trace}\left(Z_s'\Gamma(\alpha)Z_s(Z_s'Z_s)^{-1}Z_{s,-i}'Z_{s,-0}(Z_s'Z_s)^{-1}\right),\end{aligned} \tag{1.14}$$

$\Gamma(\alpha)=\mathrm{E}[V_sV_s']$, $\Gamma_{-k}(\alpha)=\mathrm{E}[V_sV_{s,-k}']$, $\gamma_i(\alpha)=\mathrm{E}[v_{st}v_{s(t-i)}]$, $V_{s,-k}=(v_{s(p+1-k)},v_{s(p+2-k)},\ldots,v_{s(T-k)})'$ *and* $Z_{s,-k}$ *is defined similarly.*

Proof. The proof is given for $\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime} \xrightarrow{p} \mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}]=(T-p)(\Gamma_p(\alpha)+\frac{1}{T-p}\Delta_\Gamma(\alpha))$; $\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st} \xrightarrow{p} \mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}]=(T-p)(A(\alpha)+\frac{1}{T-p}\Delta_A(\alpha))$ follows by a similar argument.

$\frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}$ is a $p\times p$ matrix with $[i,j]$ element

$$\begin{aligned}\left[\frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}\right]_{[i,j]} &= \frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}\sum_{t=p+1}^{T}v_{s(t-i)}v_{s(t-j)}\\ &- \frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}\sum_{t=p+1}^{T}v_{s(t-i)}z_{s(t-j)}'(Z_s'Z_s)^{-1}Z_s'V_s\\ &- \frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}\sum_{t=p+1}^{T}v_{s(t-j)}z_{s(t-i)}'(Z_s'Z_s)^{-1}Z_s'V_s\end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}\sum_{t=p+1}^{T} z'_{s(t-i)}(Z'_sZ_s)^{-1}Z'_sV_sV'_sZ_s(Z'_sZ_s)^{-1}z_{s(t-j)} \\
&= \quad \frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}\sum_{t=p+1}^{T} v_{s(t-i)}v_{s(t-j)} \\
&- \quad \frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}V'_{s,-i}Z_{s,-j}(Z'_sZ_s)^{-1}Z'_sV_s \\
&- \quad \frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}V'_{s,-j}Z_{s,-i}(Z'_sZ_s)^{-1}Z'_sV_s \qquad (1.15)\\
&+ \quad \frac{1}{S}\sum_{s=1}^{S}\frac{1}{T-p}Z_{s,-i}(Z'_sZ_s)^{-1}Z'_sV_sV'_sZ'_s(Z'_sZ_s)^{-1}Z'_{s,-j} \\
&\xrightarrow{p} \quad \mathrm{E}[\frac{1}{T-p}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st}]_{[i,j]}
\end{aligned}
$$

by the Khinchin LLN and repeated application of the triangle and Cauchy-Schwarz inequalities since

(i) $[\frac{1}{T-p}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st}]_{[i,j]}$ is iid by N2,

(ii) $\mathrm{E}|v_{s(t-i)}v_{s(t-j)}| < \infty$ by N1,

(iii) $\mathrm{E}|v_{s(t-i)}z'_{s(t-j)}(Z'_sZ_s)^{-1}Z'_sV_s| \leq \left(\mathrm{E}|z'_{s(t-j)}(Z'_sZ_s)^{-1}Z'_sV_s|^2\mathrm{E}|v_{s(t-i)}|^2\right)^{1/2}$,

(iv) $\mathrm{E}|z'_{s(t-i)}(Z'_sZ_s)^{-1}Z'_sV_sV'_sZ_s(Z'_sZ_s)^{-1}z_{s(t-j)}|$
$\leq \left(\mathrm{E}|z'_{s(t-i)}(Z'_sZ_s)^{-1}Z'_sV_s|^2\mathrm{E}|z'_{s(t-j)}(Z'_sZ_s)^{-1}Z'_sV_s|^2\right)^{1/2}$, and

(v) by N1 and N5,

$$
\begin{aligned}
\mathrm{E}|z'_{s(t-j)}(Z'_sZ_s)^{-1}Z'_sV_s|^2 &= \mathrm{E}|z'_{s(t-j)}(Z'_sZ_s)^{-1}(Z'_sZ_s)(Z'_sZ_s)^{-1}Z'_sV_s|^2 \\
&\leq \|z'_{s(t-j)}(Z'_sZ_s)^{-1}Z'_s\|^2\mathrm{E}\|Z_s(Z'_sZ_s)^{-1}Z'_sV_s\|^2 \\
&= \mathrm{trace}(z'_{s(t-j)}(Z'_sZ_s)^{-1}z_{s(t-j)}) \times \\
&\qquad \mathrm{E}\left[\mathrm{trace}(V'_sZ_s(Z'_sZ_s)^{-1}Z'_sV_s)\right] \\
&\leq k_2\mathrm{E}[\mathrm{trace}(V'_sV_s)] = k_2\mathrm{E}[\sum_{t=p+1}^{T} v_{st}^2] < \infty,
\end{aligned}
$$

from which it follows that $\mathrm{E}|\left[\frac{1}{T-p}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st}\right]_{[i,j]}| < \infty$. ■

Lemma 1.8.4 *Let* $\widehat{\alpha} = (\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\widetilde{v}^{-}_{st}\widetilde{v}^{-\prime}_{st})^{-1}(\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\widetilde{v}^{-}_{st}\widetilde{v}_{st})$ *be the least squares estimate of* α *using the least squares residuals,* $\widetilde{v}_{st}$ *from estimating* β_1. *If Assumption 1 is satisfied,*

$$\widehat{\alpha} = (\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st})^{-1}(\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}_{st}) + o_p(S^{-1/2}).$$

Proof. By Lemma 1.8.2,

$$\widehat{\alpha} \;=\; (\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st} + M_1)^{-1}(\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}_{st} + M_2),$$

where $M_1 = o_p(S^{-1/2})$ and $M_2 = o_p(S^{-1/2})$. After some algebra, it then follows that

$$\begin{aligned}
\widehat{\alpha} &= (\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st})^{-1}(\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}_{st}) \\
&+ (\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st} + M_1)^{-1}M_2 \\
&- (\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st} + M_1)^{-1}M_1(\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st})^{-1}(\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}_{st}) \\
&= (\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}^{-\prime}_{st})^{-1}(\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{v}_{st}) + O_p(1)o_p(S^{-1/2}) + O_p(1)o_p(S^{-1/2})O_p(1)O_p(1)
\end{aligned}$$

where the last equality is by Lemma 1.8.3, which yields the conclusion. ■

Lemma 1.8.5 *Define* $\ddot{\mu}_{st} = \ddot{v}_{st} - \ddot{v}^{-\prime}_{st}\alpha_T(\alpha)$. *If Assumption 1 is satisfied,*

$$\frac{1}{\sqrt{S}}\sum_{s=1}^{S}(\sum_{t=p+1}^{T}\ddot{v}^{-}_{st}\ddot{\mu}_{st}) \xrightarrow{d} N(0,\Xi),$$

where $\Xi = \mathrm{E}[\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T}\ddot{v}^{-}_{st_1}\ddot{\mu}_{st_1}\ddot{\mu}_{st_2}\ddot{v}^{-\prime}_{st_2}]$.

Proof. $\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{\mu}_{st}$ are iid by N2. Also,

$$\begin{aligned}
\mathrm{E}[\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{\mu}_{st}] &= \mathrm{E}[\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}(\ddot{v}_{st} - \ddot{v}_{st}^{-\prime}\alpha_T(\alpha))] \\
&= \mathrm{E}[\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{v}_{st}] - \mathrm{E}\{\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}(\mathrm{E}[\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}])^{-1}\mathrm{E}[\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{v}_{st}]\} \\
&= 0,
\end{aligned}$$

where the second equality comes from $\alpha_T(\alpha) = \mathrm{E}[\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}]^{-1}\mathrm{E}[\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{v}_{st}]$ from Propostion 2. Also

$$\begin{aligned}
&\mathrm{E}[\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \ddot{v}_{s(t_1-i)}\ddot{\mu}_{st_1}\ddot{\mu}_{st_2}\ddot{v}_{s(t_2-j)}] \\
&\quad = \mathrm{E}[\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \ddot{v}_{s(t_1-i)}\ddot{v}_{st_1}\ddot{v}_{st_2}\ddot{v}_{s(t_2-j)}] \\
&\quad\quad - \mathrm{E}[\sum_{k=1}^{p}\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \ddot{v}_{s(t_1-i)}\ddot{v}_{s(t_1-k)}\ddot{v}_{st_2}\ddot{v}_{s(t_2-j)}\alpha_T(\alpha)_k] \\
&\quad\quad - \mathrm{E}[\sum_{k=1}^{p}\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \ddot{v}_{s(t_1-i)}\ddot{v}_{st_1}\ddot{v}_{s(t_2-k)}\ddot{v}_{s(t_2-j)}\alpha_T(\alpha)_k] \\
&\quad\quad + \mathrm{E}[\sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \ddot{v}_{s(t_1-i)}\ddot{v}_{s(t_1-k_1)}\ddot{v}_{s(t_2-k_2)}\ddot{v}_{s(t_2-j)}\alpha_T(\alpha)_{k_1}\alpha_T(\alpha)_{k_2}] \\
&\quad < \infty
\end{aligned}$$

since $\mathrm{E}|\ddot{v}_{st_1}\ddot{v}_{st_2}\ddot{v}_{st_3}\ddot{v}_{st_4}| \leq (\mathrm{E}|\ddot{v}_{st_1}|^4\mathrm{E}|\ddot{v}_{st_2}|^4\mathrm{E}|\ddot{v}_{st_3}|^4\mathrm{E}|\ddot{v}_{st_4}|^4)^{1/4} < \infty$ by repeated application of the Cauchy-Schwarz inequality and

$$\begin{aligned}
\mathrm{E}\|\ddot{v}_{st}\|^4 &= \mathrm{E}\|e_t'(V_s - Z_s(Z_s'Z_s)^{-1}Z_s'V_s)\|^4 \\
&\leq \mathrm{E}(\|e_t\|^4\|V_s\|^4\|I_T - Z_s(Z_s'Z_s)^{-1}Z_s'\|^4) \\
&= (T-k_2)^2\mathrm{E}\|V_s\|^4 < \infty
\end{aligned}$$

by N5 and T fixed, where e_t is the t^{th} unit vector. Then the conclusion follows by the Lindeberg-Levy CLT since

$$\mathrm{E}[\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \ddot{v}_{st_1}^{-}\ddot{\mu}_{st_1}\ddot{\mu}_{st_2}\ddot{v}_{st_2}^{-\prime}]_{[i,j]} = \mathrm{E}[\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \ddot{v}_{s(t_1-i)}\ddot{\mu}_{st_1}\ddot{\mu}_{st_2}\ddot{v}_{s(t_2-j)}] < \infty \ \forall\ i,j. \ \blacksquare$$

Lemma 1.8.6 *Suppose Assumption 1 holds, then*

$$\sqrt{S}(\widehat{\alpha} - \alpha_T(\alpha)) \xrightarrow{d} \frac{1}{T-p}(\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}N(0, \Xi).$$

Proof. For $\ddot{\mu}_{st}$ defined in Lemma 1.8.5,

$$\widehat{\alpha} - \alpha_T(\alpha) = (\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime})^{-1}(\frac{1}{S}\sum_{s=1}^{S}\sum_{t=p+1}^{T} \ddot{v}_{st}^{-}\ddot{\mu}_{st}) + o_p(S^{-1/2})$$

by Lemma 1.8.4. The conclusion is then immediate from Lemmas 1.8.3 and 1.8.5. ■

1.9 Appendix 5. Proof of Propositions 4, 5, and 6

The proofs of Propositions 4, 5, and 6 are collected below. All results presented below are for asymptotics where $S,\ T \to \infty$.

Proof of Proposition 4. $\sqrt{ST}(\widehat{\alpha} - \alpha) = \sqrt{ST}(\widehat{\alpha} - \alpha_T(\alpha) + \alpha_T(\alpha) - \alpha) = \sqrt{ST}(\widehat{\alpha} - \alpha_T(\alpha)) + \sqrt{\frac{S}{T}}B(\alpha, T)$ by Lemma 1.9.5, from which the conclusion follows by Lemmas 1.9.5 and 1.9.8. ■

Proof of Proposition 5.

$$\begin{aligned} \sqrt{ST}(\widehat{\alpha}^{(1)} - \alpha) &= \sqrt{ST}[\widehat{\alpha} - (-\widehat{\alpha} + \alpha_T(\widehat{\alpha})) - \alpha_T(\alpha) + \alpha_T(\alpha) - \alpha] \\ &= \sqrt{ST}(\widehat{\alpha} - \alpha_T(\alpha)) + \sqrt{\frac{S}{T}}(B(\alpha, T) - B(\widehat{\alpha}, T)). \end{aligned}$$

The first conclusion then follows from Lemmas 1.9.5, 1.9.8, 1.9.9, and the Continuous Mapping Theorem, and the second conclusion follows from Lemmas 1.9.5, 1.9.8, 1.9.9, and a Taylor expansion of $B(\widehat{\alpha}, T)$ about $\widehat{\alpha} = \alpha$. ■

Proof of Proposition 6. Recall $\widehat{\alpha}^{(\infty)} = \alpha_T^{-1}(\widehat{\alpha}) = \alpha_T^{-1}(\alpha_T(\alpha)) + H^{-1}|_{\widetilde{\alpha_T(\alpha)}}(\widehat{\alpha} - \alpha_T(\alpha))$. From Lemma 1.9.5, $\alpha_T(\alpha) = \alpha + \frac{1}{T-p}B(\alpha, T)$ which implies $H^{-1} \to I$ as $T \to \infty$ by NT6. The conclusion then follows from Lemma 1.9.8. ■

1.9.1 Lemmas

Lemma 1.9.1 *Let $\widehat{\beta}_1$ be the ordinary least squares estimate of β_1. Then if the conditions of Assumption 2 are satisfied, $\widehat{\beta}_1 - \beta_1 \xrightarrow{p} 0$ and $\sqrt{ST}(\widehat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, M^{-1}\Omega M^{-1})$, where*

$$\begin{aligned} M &= M_{XX} - M_{XZ}M_{ZZ}^{-1}M'_{XZ}, \\ \Omega &= \lim_{T\to\infty} \mathrm{E}[\frac{1}{T}X_s'\Gamma(\alpha)X_s] - M_{XZ}M_{ZZ}^{-1}(\lim_{T\to\infty}\mathrm{E}[\frac{1}{T}Z_s'\Gamma(\alpha)X_s]) \\ &\quad -(\lim_{T\to\infty}\mathrm{E}[\frac{1}{T}X_s'\Gamma(\alpha)Z_s])M_{ZZ}^{-1}M'_{XZ} + M_{XZ}M_{ZZ}^{-1}(\lim_{T\to\infty}[\frac{1}{T}Z_s'\Gamma(\alpha)Z_s])M_{ZZ}^{-1}M'_{XZ}, \end{aligned}$$

with $M_{XX} = \lim_{T\to\infty}\mathrm{E}[\frac{1}{T}X_s'X_s]$, $M_{XZ} = \lim_{T\to\infty}\mathrm{E}[\frac{1}{T}X_s'Z_s]$, and $M_{ZZ} = \lim_{T\to\infty}[\frac{1}{T}Z_s'Z_s]$.

Proof. Let $Q_s = I_T - Z_s(Z_s'Z_s)^{-1}Z_s'$. Then

$$\begin{aligned} \mathrm{E}\|\frac{1}{T}X_s'Q_sV_s\|^{1+\delta} &\le \mathrm{E}(\|\frac{1}{\sqrt{T}}Q_sX_s\|^{1+\delta}\|\frac{1}{\sqrt{T}}Q_sV_s\|^{1+\delta}) \\ &\le (\mathrm{E}\|\frac{1}{\sqrt{T}}X_s\|^{2+2\delta}\mathrm{E}\|\frac{1}{\sqrt{T}}V_s\|^{2+2\delta})^{1/2} \\ &\le k_1^{\frac{1+\delta}{2}}\Delta, \end{aligned}$$

where the first inequality is from $\|AB\| \le \|A\|\|B\|$, the second from the Cauchy-Schwarz inequality and $\|Q_sA\| \le \|A\|$, and the third from

$$\begin{aligned} \mathrm{E}\|\frac{1}{\sqrt{T}}X_s\|^{2+2\delta} &= \mathrm{E}[\frac{1}{T}\mathrm{trace}(X_s'X_s)]^{1+\delta} \\ &= \frac{1}{T^{1+\delta}}\mathrm{E}[\sum_{t=1}^{T}\sum_{h=1}^{k_1}x_{ith}^2]^{1+\delta} \\ &\le \frac{1}{T^{1+\delta}}[\sum_{t=1}^{T}\sum_{h=1}^{k_1}(\mathrm{E}|x_{ith}^2|^{1+\delta})^{\frac{1}{1+\delta}}]^{1+\delta} \\ &\le \frac{1}{T^{1+\delta}}(Tk_1\Delta^{\frac{1}{1+\delta}})^{1+\delta} = k_1^{1+\delta}\Delta, \end{aligned}$$

where the first inequality follows from Minkowski's inequality and the second from NT5, and $\mathrm{E}\|\frac{1}{\sqrt{T}}V_s\|^{2+2\delta} \le \Delta$ by a similar argument. It also follows that $\mathrm{E}\|\frac{1}{T}X_s'Q_sX_s\|^{1+\delta} \le k_1^{1+\delta}\Delta$ by the same reasoning. So $\mathrm{E}\|\frac{1}{T}X_s'Q_sX_s\|$ and $\mathrm{E}\|\frac{1}{T}X_s'Q_sV_s\|$ are uniformly integrable in T. (See, e.g. Billingsley (1995).) Also, under NT3-NT5,

$$\frac{1}{T}\sum_{t=1}^{T} x_{st}x_{st}' - \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}[x_{st}x_{st}'] \xrightarrow{p} 0, \quad \frac{1}{T}\sum_{t=1}^{T} x_{st}z_{st}' - \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}[x_{st}z_{st}'] \xrightarrow{p} 0$$
$$\frac{1}{T}\sum_{t=1}^{T} z_{st}z_{st}' \rightarrow \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} z_{st}z_{st}', \quad \frac{1}{T}\sum_{t=1}^{T} x_{st}v_{st}' \xrightarrow{p} 0, \text{ and } \frac{1}{T}\sum_{t=1}^{T} z_{st}v_{st}' \xrightarrow{p} 0$$

by a LLN, e.g. White (2001) Corollary 3.48. It then follows from Phillips and Moon (1999) Corollary 1 that $\frac{1}{S}\sum_{s=1}^{S}\frac{1}{T}X_s'Q_sX_s \xrightarrow{p} M$ and $\frac{1}{S}\sum_{s=1}^{S}\frac{1}{T}X_s'Q_sV_s \xrightarrow{p} 0$. Hence, $\widehat{\beta}_1 - \beta_1 \xrightarrow{p} 0$.

Let K denote a generic constant. To verify asymptotic normality, note that

$$\begin{aligned}\mathrm{E}\|\frac{1}{\sqrt{T}}X_s'V_s\|^{2+\frac{\delta}{2}} &\leq \frac{C}{T^{\frac{2+\frac{\delta}{2}}{2}}}\max\{\sum_{t=1}^{T}(\mathrm{E}|x_{st}v_{st}|^{2+\frac{\delta}{2}+\epsilon})^{\frac{2+\frac{\delta}{2}}{2+\frac{\delta}{2}+\epsilon}}, [\sum_{t=1}^{T}(\mathrm{E}|x_{st}v_{st}|^{2+\epsilon})^{\frac{2}{2+\epsilon}}]^{\frac{2+\frac{\delta}{2}}{2}}\} \\ &\leq CK < \infty \end{aligned} \quad (1.16)$$

by NT4-NT5 and Doukhan (1994) Theorem 2. Also,

$$\begin{aligned}\mathrm{E}\|(\frac{1}{T}X_s'Z_s)(\frac{1}{T}Z_s'Z_s)^{-1}(\frac{1}{\sqrt{T}}Z_s'V_s)\|^{2+\frac{\delta}{2}} &\leq (k_2M)^{2+\frac{\delta}{2}}(\mathrm{E}\|(\frac{1}{T}X_s'Z_s)\|^{4+\delta}\mathrm{E}\|(\frac{1}{\sqrt{T}}Z_s'V_s)\|^{4+\delta})^{1/2} \\ &\leq CK < \infty, \end{aligned} \quad (1.17)$$

where the second inequality follows from $\|AB\| \leq \|A\|\|B\|$, the Cauchy-Schwarz inequality, and NT3, and the third inequality is from NT4-NT5 and Doukhan (1994) Theorem 2. (1.16) and (1.17) imply $\mathrm{E}\|\frac{1}{\sqrt{T}}X_s'Q_sV_s\|^{2+\frac{\delta}{2}} < \infty$ since

$$\mathrm{E}\|\frac{1}{\sqrt{T}}X_s'Q_sV_s\|^{2+\frac{\delta}{2}} \leq [(\mathrm{E}\|\frac{1}{\sqrt{T}}X_s'V_s\|^{2+\frac{\delta}{2}})^{\frac{1}{2+\frac{\delta}{2}}} + (\mathrm{E}\|\frac{1}{\sqrt{T}}X_s'Z_s(Z_s'Z_s)^{-1}Z_s'V_s\|^{2+\frac{\delta}{2}})^{\frac{1}{2+\frac{\delta}{2}}}]^{2+\frac{\delta}{2}}$$

by Minkowski's inequality which yields $\mathrm{E}\|\frac{1}{\sqrt{T}}X_s'Q_sV_s\|^2$ uniformly integrable in T. (See, e.g. Billingsley (1995).) $\frac{1}{\sqrt{T}}X_s'Q_sV_s \xrightarrow{d} N(0,\Omega)$ from White (2001) Theorem 5.20, from which it follows that $\Omega_T = \frac{1}{T}\mathrm{E}[X_s'Q_s\Gamma(\alpha)Q_sX_s] \rightarrow \Omega$. (E.g. Billingsley (1995) Corollary 25.12.) Then Phillips and Moon (1999) Theorem 3 gives $\frac{1}{\sqrt{ST}}\sum_{s=1}^{S}\sum_{t=1}^{T}X_s'Q_sV_s \xrightarrow{d} N(0,\Omega)$ which implies $\sqrt{ST}(\widehat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, M^{-1}\Omega M^{-1})$. ■

Lemma 1.9.2 *Define $\widetilde{v}_{st}$ to be the residual from least squares regression of (1.11); i.e. $\widetilde{v}_{st} = C_{st} - x_{st}'\widehat{\beta}_1 - z_{st}'\widehat{\beta}_2^s = v_{st} - x_{st}'(\widehat{\beta}_1 - \beta_1) - z_{st}'(\widehat{\beta}_2^s - \beta_2) = \ddot{v}_{st} - \ddot{x}_{st}'(\widehat{\beta}_1 - \beta_1)$, where $\widehat{\beta}_1$ and $\widehat{\beta}_2^s$ are least* *squares estimates of* β_1 *and* β_2^s. *Let* $\widetilde{v}_{st}^-$ *be a* $p \times 1$ *vector with* $\widetilde{v}_{st}^- = [\widetilde{v}_{s(t-p)}, \ldots, \widetilde{v}_{s(t-1)}]'$, *and let* $\ddot{v}_{st}^-$ *be a* $p \times 1$ *vector with* $\ddot{v}_{st}^- = [\ddot{v}_{s(t-p)}, \ldots, \ddot{v}_{s(t-1)}]'$. *Under the conditions of Assumption 2,*

$$\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \widetilde{v}_{st}^- \widetilde{v}_{st}^{-\prime} = \frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st}^{-\prime} + o_p((ST)^{-1/2}),$$

and

$$\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \widetilde{v}_{st}^- \widetilde{v}_{st} = \frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st} + o_p((ST)^{-1/2}).$$

Proof. Using calculations similar to those found in the proof of Lemma 1.9.1, it can be shown that $\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{v}_{st} \ddot{x}_{s(t-j)h} = o_p(1)$ and $\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{x}_{st} \ddot{x}_{s(t-j)h} = O_p(1)$ for $j = -p, \ldots, p$ and $h = 1, \ldots, k_1$. The conclusion then follows by Lemma 1.9.1 and the same arguments used in proving Lemma 1.8.2. ■

Lemma 1.9.3 *Under the conditions of Assumption 2,*

(i) $\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st}^{-\prime} \xrightarrow{p} \Gamma_p(\alpha)$, *and* $\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st} \xrightarrow{p} A(\alpha)$.

(ii) $\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st}^{-\prime} = \frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} v_{st}^- v_{st}^{-\prime} + O_p(\frac{1}{T})$, *and* $\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st} = \frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} v_{st}^- v_{st} + O_p(\frac{1}{T})$.

Proof. The proof is given for

$$\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st}^{-\prime} \xrightarrow{p} \Gamma_p(\alpha);$$

$\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st} \xrightarrow{p} A(\alpha)$ follows by a similar argument.

$\frac{1}{T-p} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st}^{-\prime}$ is a $p \times p$ matrix with i, j element

$$\left[\frac{1}{T-p} \sum_{t=p+1}^{T} \ddot{v}_{st}^- \ddot{v}_{st}^{-\prime}\right]_{[i,j]} = \frac{1}{T-p} \sum_{t=p+1}^{T} v_{s(t-i)} v_{s(t-j)} - \frac{1}{T-p} \sum_{t=p+1}^{T} v_{s(t-i)} z_{s(t-j)}' (Z_s' Z_s)^{-1} Z_s' V_s$$

$$
\begin{aligned}
&- \frac{1}{T-p}\sum_{t=p+1}^{T} v_{s(t-j)} z'_{s(t-i)} (Z'_s Z_s)^{-1} Z'_s V_s \\
&+ \frac{1}{T-p}\sum_{t=p+1}^{T} z'_{s(t-i)} (Z'_s Z_s)^{-1} Z'_s V_s V'_s Z_s (Z'_s Z_s)^{-1} z_{s(t-j)} \\
&= \frac{1}{T-p}\sum_{t=p+1}^{T} v_{s(t-i)} v_{s(t-j)} \\
&- \frac{1}{T-p} V'_{s,-i} Z_{s,-j} (Z'_s Z_s)^{-1} Z'_s V_s \\
&- \frac{1}{T-p} V'_{s,-j} Z_{s,-i} (Z'_s Z_s)^{-1} Z'_s V_s \qquad (1.18) \\
&+ \frac{1}{T-p} Z_{s,-i} (Z'_s Z_s)^{-1} Z'_s V_s V'_s Z'_s (Z'_s Z_s)^{-1} Z'_{s,-j},
\end{aligned}
$$

where $V_{s,-k} = (v_{s(p+1-k)}, v_{s(p+2-k)}, \ldots, v_{s(T-k)})'$ and $Z_{s,-k}$ is defined similarly.

Consider

$$
\begin{aligned}
&\mathrm{E}\|V'_s Z_s (Z'_s Z_s)^{-1} Z'_{s,-i} Z_{s,-j} (Z'_s Z_s)^{-1} Z'_s V_s\|^{1+\frac{\delta}{2}} \\
&= \frac{T-p}{T}\mathrm{E}\|(\frac{1}{\sqrt{T}} V'_s Z_s)(\frac{1}{T} Z'_s Z_s)^{-1}(\frac{1}{T-p} Z'_{s,-i} Z_{s,-j})(\frac{1}{T} Z'_s Z_s)^{-1}(\frac{1}{\sqrt{T}} Z'_s V_s)\|^{1+\frac{\delta}{2}}. \qquad (1.19)
\end{aligned}
$$

Note that

$$
\|\frac{1}{T-p} Z'_{s,-i} Z_{s,-j}\| \;\le\; k_2 \Delta < \infty \qquad (1.20)
$$

by NT5. Also,

$$
\begin{aligned}
\mathrm{E}\|(\frac{1}{\sqrt{T}} Z'_s V_s)\|^{2+\delta} &= \frac{1}{T^{\frac{2+\delta}{2}}}\mathrm{E}\|(\sum_{t=1}^{T} z_{st} v_{st})\|^{2+\delta} \\
&\le \frac{1}{T^{\frac{2+\delta}{2}}} C \max\{\sum_{t=1}^{T}(\mathrm{E}|z_{st} v_{st}|^{2+\delta+\epsilon})^{\frac{2+\delta}{2+\delta+\epsilon}}, [\sum_{t=1}^{T}(\mathrm{E}|z_{st} v_{st}|^{2+\epsilon})^{\frac{2}{2+\epsilon}}]^{\frac{2+\delta}{2}}\} \\
&\le K < \infty \qquad (1.21)
\end{aligned}
$$

by Doukhan (1994) Theorem 2 and NT5. It then follows from the Holder's inequality, the Cauchy-Schwarz inequality, $\|AB\| \le \|A\|\|B\|$, (1.20), (1.21), and NT3-NT5 that (1.19) is bounded away from ∞, which gives $\|V'_s Z_s (Z'_s Z_s)^{-1} Z'_{s,-i} Z_{s,-j} (Z'_s Z_s)^{-1} Z'_s V_s\|$ uniformly integrable in T. Similarly, it can be shown that $\|V'_{s,-i} Z_{s,-j} (Z'_s Z_s)^{-1} Z'_s V_s\|$ and $\|\frac{1}{T-p}\sum_{t=p+1}^{T} v_{s(t-i)} v_{s(t-j)}\|$ are uniformly integrable in T.

Also, as $T \to \infty$,

1. $\frac{1}{T-p}\sum_{t=p+1}^{T} v_{s(t-i)}v_{s(t-j)} \xrightarrow{p} \gamma_{|i-j|}(\alpha)$,
2. $V'_{s,-i}Z_{s,-j}(Z'_sZ_s)^{-1}Z'_sV_s \xrightarrow{d} \Psi'_{ij}M_{ZZ}^{-1}\Psi$,
3. $V'_sZ_s(Z'_sZ_s)^{-1}Z'_{s,-i}Z_{s,-j}(Z'_sZ_s)^{-1}Z'_sV_s \xrightarrow{d} \Psi'M_{ZZ}^{-1}M_{ij}M_{ZZ}^{-1}\Psi$,

where $M_{ZZ} = \lim_{T\to\infty}\frac{1}{T}[Z'_sZ_s]$, $M_{ij} = \lim_{T\to\infty}\frac{1}{T}[Z'_{s,-i}Z_{s,-j}]$, $\Psi \sim N(0, \lim_{T\to\infty}\frac{1}{T}[Z'_s\Gamma(\alpha)Z_s]$, $\Psi_{ij} \sim N(0, \lim_{T\to\infty}\frac{1}{T-p}[Z'_{s,-j}\Gamma(\alpha)Z_{s,-j})$, $\Gamma_{-k} = \mathrm{E}[v_sv'_{s,-k}]$, and $\mathrm{E}[\Psi\Psi'_{ij}] = \lim_{T\to\infty}\frac{1}{T}[Z'_s\Gamma_{-i}(\alpha)Z_{s,-j}]$. (See White (2001) Corollary 3.48 and Theorem 5.20.) Combining 1., 2., and 3. and (1.18) gives

$$\frac{1}{T-p}\sum_{t=p+1}^{T} \ddot{v}_{s(t-i)}\ddot{v}_{s(t-j)} \xrightarrow{p} \gamma_{|i-j|}(\alpha).$$

It then follows from Phillips and Moon (1999) Corollary 1 that

$$\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T} \ddot{v}_{s(t-i)}\ddot{v}_{s(t-j)} \xrightarrow{p} \gamma_{|i-j|}(\alpha),$$

which yields the first conclusion. Also, using Phillips and Moon (1999) Corollary 1 and considering only 2., 3. and (1.18) yields the second conclusion. ■

Lemma 1.9.4 *Let* $\widehat{\alpha} = (\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\widetilde{v}_{st}^{-}\widetilde{v}_{st}^{-\prime})^{-1}(\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\widetilde{v}_{st}^{-}\widetilde{v}_{st})$ *be the least squares estimate of* α *using the least squares residuals,* $\widetilde{v}_{st}$ *from estimating* β_1. *If Assumption 2 is satisfied,*

$$\begin{aligned}\widehat{\alpha} &= (\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime})^{-1}(\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}) + o_p((ST)^{-1/2}) \\ &= (\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}v_{st}^{-}v_{st}^{-\prime})^{-1}(\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}v_{st}^{-}v_{st}) + O_p(T^{-1}) + o_p((ST)^{-1/2}).\end{aligned}$$

Proof. Follows immediately from Lemmas 1.9.2 and 1.9.3 using the same argument as in Lemma 1.8.4. ■

Lemma 1.9.5 *For* $\alpha_T(\alpha)$ *defined in Lemma 2,* $\alpha_T(\alpha) - \alpha = \frac{1}{T-p}B(\alpha,T)$, *where* $B(\alpha,T) \to B(\alpha)$ *as* $T \to \infty$, *and* $\alpha_T(\widehat{\alpha}) - \widehat{\alpha} = \frac{1}{T-p}B(\widehat{\alpha},T)$ *if the conditions of Assumption 2 are met.*

Proof. Recall $\alpha_T(\alpha) = (\Gamma_p(\alpha)+\frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}(A(\alpha)+\frac{1}{T-p}\Delta_A(\alpha))$ and that $\alpha = \Gamma_p(\alpha)^{-1}A(\alpha)$. Then $\alpha_T(\alpha) - \alpha = \frac{1}{T-p}(\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}(\Delta_A(\alpha) + \Delta_\Gamma(\alpha)\alpha) = \frac{1}{T-p}B(\alpha, T)$. That $B(\alpha, T) \to B(\alpha)$ as $T \to \infty$ follows from the proof of Lemma 1.9.3. $\alpha_T(\widehat{\alpha}) - \widehat{\alpha} = \frac{1}{T-p}B(\widehat{\alpha}, T)$ follows in the same fashion. ∎

Lemma 1.9.6 *Define $\ddot{\mu}_{st} = \ddot{v}_{st} - \ddot{v}_{st}^{-\prime}\alpha_T(\alpha)$. If Assumption 2 is satisfied,*

$$\frac{1}{\sqrt{S(T-p)}}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{\mu}_{st} = \frac{1}{\sqrt{S(T-p)}}\sum_{s=1}^{S}\sum_{t=p+1}^{T}v_{st}^{-}\eta_{st} + o_p(1).$$

Proof. $\frac{1}{\sqrt{S(T-p)}}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{\mu}_{st} = \frac{1}{\sqrt{S(T-p)}}\sum_{s=1}^{S}\sum_{t=p+1}^{T}(\ddot{v}_{st}^{-}\ddot{\mu}_{st} - \mathrm{E}[\ddot{v}_{st}^{-}\ddot{\mu}_{st}])$ since

$$\begin{aligned}
\mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{\mu}_{st}] &= \mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}(\ddot{v}_{st} - \ddot{v}_{st}^{-\prime}\alpha_T(\alpha))] \\
&= \mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}] - \mathrm{E}\{\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}(\mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}])^{-1}\mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}]\} \\
&= 0,
\end{aligned}$$

where the second equality comes from $\alpha_T(\alpha) = \mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime}]^{-1}\mathrm{E}[\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}]$ from Propostion 2. Also note that $\ddot{\mu}_{st} = \eta_{st} - (v_{st} - \ddot{v}_{st}) - v_{st}^{-\prime}(\alpha_T(\alpha) - \alpha) - (\ddot{v}_{st}^{-} - v_{st}^{-})'\alpha_T(\alpha)$ and $\ddot{v}_{st}^{-} = v_{st}^{-} - Z_{st}^{-}(Z_s'Z_s)^{-1}Z_s'V_s$ where $Z_{st}^{-} = [z_{s(t-1)}, \ldots, z_{s(t-p)}]'$. So,

$$\begin{aligned}
\ddot{v}_{st}^{-}\ddot{\mu}_{st} - \mathrm{E}[\ddot{v}_{st}^{-}\ddot{\mu}_{st}] &= (v_{st}^{-} - Z_{st}^{-}(Z_s'Z_s)^{-1}Z_s'V_s) \times \\
&\quad (\eta_{st} - z_{st}'(Z_s'Z_s)^{-1}Z_s'V_s - v_{st}^{-\prime}(\alpha_T(\alpha) - \alpha) + V_s'Z_s(Z_s'Z_s)^{-1}Z_{st}^{-\prime}\alpha_T(\alpha)) \\
&\quad -\mathrm{E}[(v_{st}^{-} - Z_{st}^{-}(Z_s'Z_s)^{-1}Z_s'V_s) \times \\
&\quad (\eta_{st} - z_{st}'(Z_s'Z_s)^{-1}Z_s'V_s - v_{st}^{-\prime}(\alpha_T(\alpha) - \alpha) + V_s'Z_s(Z_s'Z_s)^{-1}Z_{st}^{-\prime}\alpha_T(\alpha))] \\
&= v_{st}^{-}\eta_{st} - (v_{st}^{-}z_{st}'(Z_s'Z_s)^{-1}Z_s'V_s - \mathrm{E}[v_{st}^{-}z_{st}'(Z_s'Z_s)^{-1}Z_s'V_s]) \\
&\quad -(v_{st}^{-}v_{st}^{-\prime} - \mathrm{E}[v_{st}^{-}v_{st}^{-\prime}])(\alpha_T(\alpha) - \alpha) \\
&\quad +(v_{st}^{-}V_s'Z_s(Z_s'Z_s)^{-1}Z_{st}^{-\prime} - \mathrm{E}[v_{st}^{-}V_s'Z_s(Z_s'Z_s)^{-1}Z_{st}^{-\prime}])\alpha_T(\alpha) \\
&\quad -(Z_{st}^{-}(Z_s'Z_s)^{-1}Z_s'V_s\eta_{st} - \mathrm{E}[(Z_{st}^{-}(Z_s'Z_s)^{-1}Z_s'V_s\eta_{st}]) \\
&\quad +(Z_{st}^{-}(Z_s'Z_s)^{-1}Z_s'V_s z_{st}'(Z_s'Z_s)^{-1}Z_s'V_s \qquad (1.22) \\
&\quad -\mathrm{E}[Z_{st}^{-}(Z_s'Z_s)^{-1}Z_s'V_s z_{st}'(Z_s'Z_s)^{-1}Z_s'V_s])
\end{aligned}$$

$$+(Z^-_{st}(Z_s'Z_s)^{-1}Z_s'V_sv^{-'}_{st} - \mathrm{E}[Z^-_{st}(Z_s'Z_s)^{-1}Z_s'V_sv^{-'}_{st}])(\alpha_T(\alpha)-\alpha)$$
$$-(Z^-_{st}(Z_s'Z_s)^{-1}Z_s'V_sV_s'Z_s(Z_s'Z_s)^{-1}Z^{-'}_{st}$$
$$-\mathrm{E}[Z^-_{st}(Z_s'Z_s)^{-1}Z_s'V_sV_s'Z_s(Z_s'Z_s)^{-1}Z^{-'}_{st}])\alpha_T(\alpha)).$$

Now consider

$$Y_{S,T} = \frac{1}{\sqrt{S(T-p)}}\sum_{s=1}^{S}\sum_{t=p+1}^{T}(Z^-_{st}(Z_s'Z_s)^{-1}Z_s'V_sV_s'Z_s(Z_s'Z_s)^{-1}Z^{-'}_{st}$$
$$-\mathrm{E}[Z^-_{st}(Z_s'Z_s)^{-1}Z_s'V_sV_s'Z_s(Z_s'Z_s)^{-1}Z^{-'}_{st}]).$$

$Y_{S,T}$ is a $p\times p$ matrix with i,j element

$$[Y_{S,T}]_{[i,j]} = \frac{1}{\sqrt{S(T-p)}}\sum_{s=1}^{S}\sum_{t=p+1}^{T}(z'_{s(t-i)}(Z_s'Z_s)^{-1}Z_s'V_sV_s'Z_s(Z_s'Z_s)^{-1}z_{s(t-j)}$$
$$-\mathrm{E}[z_{s(t-i)}(Z_s'Z_s)^{-1}Z_s'V_sV_s'Z_s(Z_s'Z_s)^{-1}z_{s(t-j)}])$$
$$= \frac{1}{\sqrt{S(T-p)}}\sum_{s=1}^{S}(V_s'Z_s(Z_s'Z_s)^{-1}Z'_{s,-i}Z_{s,-j}(Z_s'Z_s)^{-1}Z_s'V_s$$
$$-\mathrm{E}[V_s'Z_s(Z_s'Z_s)^{-1}Z'_{s,-i}Z_{s,-j}(Z_s'Z_s)^{-1}Z_s'V_s]),$$

and

$$\mathrm{E}\|[Y_{S,T}]_{[i,j]}\|^2 = \frac{1}{T-p}\mathrm{E}\|V_s'Z_s(Z_s'Z_s)^{-1}Z'_{s,-i}Z_{s,-j}(Z_s'Z_s)^{-1}Z_s'V_s\|^2$$
$$-\frac{1}{T-p}\|\mathrm{E}[V_s'Z_s(Z_s'Z_s)^{-1}Z'_{s,-i}Z_{s,-j}(Z_s'Z_s)^{-1}Z_s'V_s]\|^2 \to 0,$$

where the equality is from independence assumption NT2, since $\mathrm{E}\|A\|^2 \geq \|E[A]\|^2$ and

1. $\mathrm{E}\|V_s'Z_s(Z_s'Z_s)^{-1}Z'_{s,-i}Z_{s,-j}(Z_s'Z_s)^{-1}Z_s'V_s\|^2$
   $\leq \frac{T-p}{T}(\mathrm{E}\|\frac{1}{\sqrt{T}}V_s'Z_s\|^4\|(\frac{1}{T}Z_s'Z_s)^{-1}(\frac{1}{T-p}Z'_{s,-i}Z_{s,-j})(\frac{1}{T}Z_s'Z_s)^{-1}\|^4\mathrm{E}\|\frac{1}{\sqrt{T}}Z_s'V_s\|^4)^{1/2}$
   by the Cauchy-Schwarz inequality.

2. $\|(\frac{1}{T}Z_s'Z_s)^{-1}(\frac{1}{T-p}Z'_{s,-i}Z_{s,-j})(\frac{1}{T}Z_s'Z_s)^{-1}\|^2 \leq (k_2M)^2(\|\frac{1}{T-p}Z'_{s,-i}Z_{s,-j}\|^2) \leq (k_2M)^2k_2\Delta$ by NT3 and NT5.

3. $\mathrm{E}\|\frac{1}{\sqrt{T}}Z_s'V_s\|^4 \leq T^{-2}C\max\{\sum_{t=1}^{T}(\mathrm{E}|z_{st}v_{st}|^{4+\epsilon})^{\frac{4}{4+\epsilon}}, [\sum_{t=1}^{T}(\mathrm{E}|z_{st}v_{st}|^{2+\epsilon})^{\frac{2}{2+\epsilon}}]^2\} \leq K$ by NT5 and Doukhan (1994) Theorem 2.

Hence, $[Y_{S,T}]_{[i,j]} \xrightarrow{p} 0$ by Chebychev's inequality, and it follows that $Y_{S,T} = o_p(1)$. That all other terms, except $\frac{1}{\sqrt{S(T-p)}} \sum_{s=1}^{S} \sum_{t=p+1}^{T} (v_{st}^{-} v_{st}^{-\prime} - \mathrm{E}[v_{st}^{-} v_{st}^{-\prime}])(\alpha_T(\alpha) - \alpha)$, are $o_p(1)$ follows similarly.

To show $\frac{1}{\sqrt{S(T-p)}} \sum_{s=1}^{S} \sum_{t=p+1}^{T} (v_{st}^{-} v_{st}^{-\prime} - \mathrm{E}[v_{st}^{-} v_{st}^{-\prime}])(\alpha_T(\alpha) - \alpha) = o_p(1)$, note that $(\alpha_T(\alpha) - \alpha) = O(\frac{1}{T})$ from Lemma 1.9.5. Also, $\mathrm{vec}\{\frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} (v_{st}^{-} v_{st}^{-\prime} - \mathrm{E}[v_{st}^{-} v_{st}^{-\prime}])\} \xrightarrow{d} \Psi \sim N(0, \Omega)$ where

$$\Omega = \lim_{T \to \infty} \frac{1}{T-p} \sum_{t_1=p+1}^{T} \sum_{t_2=p+1}^{T} [\mathrm{E}(v_{st_1}^{-} v_{st_2}^{-\prime} \otimes v_{st_1}^{-} v_{st_2}^{-\prime}) - \mathrm{E}(v_{st_1}^{-} \otimes v_{st_1}^{-}) \mathrm{E}(v_{st_2}^{-\prime} \otimes v_{st_2}^{-\prime})]$$

as $T \to \infty$ follows from a CLT (e.g. White (2001) Theorem 5.20) and the Cramer-Wold device, and $\| \frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \mathrm{vec}(v_{st}^{-} v_{st}^{-\prime} - \mathrm{E}[v_{st}^{-} v_{st}^{-\prime}]) \|^2 \xrightarrow{d} \|\Psi\|^2$ by the Continuous Mapping Theorem. In addition,

$$\begin{aligned}
&\mathrm{E}\| \frac{1}{\sqrt{S(T-p)}} \sum_{s=1}^{S} \sum_{t=p+1}^{T} (v_{st}^{-} v_{st}^{-\prime} - \mathrm{E}[v_{st}^{-} v_{st}^{-\prime}]) \|^2 \\
&= \mathrm{trace}\{\mathrm{E}[\frac{1}{T-p} \sum_{t_1=p+1}^{T} \sum_{t_2=p+1}^{T} ((v_{st_1}^{-} \otimes v_{st_1}^{-}) - \mathrm{E}(v_{st_1}^{-} \otimes v_{st_1}^{-}))((v_{st_2}^{-} \otimes v_{st_2}^{-}) - \mathrm{E}(v_{st_2}^{-} \otimes v_{st_2}^{-}))']\} \\
&= \frac{1}{T-p} \sum_{t_1=p+1}^{T} \sum_{t_2=p+1}^{T} [\mathrm{E}(v_{st_1}^{-} v_{st_2}^{-\prime} \otimes v_{st_1}^{-} v_{st_2}^{-\prime}) - \mathrm{E}(v_{st_1}^{-} \otimes v_{st_1}^{-}) \mathrm{E}(v_{st_2}^{-\prime} \otimes v_{st_2}^{-\prime})] \to \mathrm{trace}(\Omega) = \mathrm{E}\|\Psi\|^2
\end{aligned}$$

as $T \to \infty$. It follows that $\| \frac{1}{\sqrt{S(T-p)}} \sum_{s=1}^{S} \sum_{t=p+1}^{T} (v_{st}^{-} v_{st}^{-\prime} - \mathrm{E}[v_{st}^{-} v_{st}^{-\prime}]) \|^2$ is uniformly integrable in T. (For example, Billingsley (1995) 16.14.) Then, using Phillips and Moon (1999) Theorem 3,

$$\frac{1}{\sqrt{S(T-p)}} \sum_{s=1}^{S} \sum_{t=p+1}^{T} (v_{st}^{-} v_{st}^{-\prime} - \mathrm{E}[v_{st}^{-} v_{st}^{-\prime}]) \xrightarrow{d} N(0, \Omega),$$

which implies that

$$\frac{1}{\sqrt{S(T-p)}} \sum_{s=1}^{S} \sum_{t=p+1}^{T} (v_{st}^{-} v_{st}^{-\prime} - \mathrm{E}[v_{st}^{-} v_{st}^{-\prime}]) = O_p(1),$$

and the conclusion follows immediately. ■

Lemma 1.9.7 *If Assumption 2 is satisfied,* $\frac{1}{\sqrt{S(T-p)}}\sum_{s=1}^{S}\sum_{t=p+1}^{T} v_{st}^{-}\eta_{st} \xrightarrow{d} N(0,\Xi)$, *where*

$$\Xi = \lim_{T\to\infty}\frac{1}{T-p}\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \mathrm{E}[v_{st_1}^{-}\eta_{st_1}\eta_{st_2}v_{st_2}^{-\prime}].$$

In addition, if η_{st} *are independent for all* s *and* t, $\Xi = \sigma_\eta^2\Gamma$.

Proof. As $T\to\infty$, $\frac{1}{\sqrt{T-p}}\sum_{t=p+1}^{T} v_{st}^{-}\eta_{st} \xrightarrow{d} \Psi \sim N(0,\Xi)$ by a CLT (e.g. White (2001) Theorem 5.20) and the Cramer-Wold device, and $\|\frac{1}{\sqrt{T-p}}\sum_{t=p+1}^{T} v_{st}^{-}\eta_{st}\|^2 \xrightarrow{d} \|\Psi\|^2$ by the Continuous Mapping Theorem. Also,

$$\mathrm{E}\|\frac{1}{\sqrt{T-p}}\sum_{t=p+1}^{T} v_{st}^{-}\eta_{st}\|^2 = \mathrm{trace}(\frac{1}{T-p}\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \mathrm{E}v_{st_1}^{-}\eta_{st_1}\eta_{st_2}v_{st_2}^{-\prime}) \to \mathrm{trace}(\Xi) = \mathrm{E}\|\Psi\|^2,$$

which implies that $\mathrm{E}\|\frac{1}{\sqrt{T-p}}\sum_{t=p+1}^{T} v_{st}^{-}\eta_{st}\|^2$ is uniformly integrable in T. (For example, Billingsley (1995) 16.14.) Then, using Phillips and Moon (1999) Theorem 3,

$$\frac{1}{\sqrt{S(T-p)}}\sum_{s=1}^{S}\sum_{t=p+1}^{T} v_{st}^{-}\eta_{st} \xrightarrow{d} N(0,\Xi).$$

Also, if the η_{st} are independent for all s and t, $\mathrm{E}v_{st_1}^{-}\eta_{st_1}\eta_{st_2}v_{st_2}^{-\prime} = 0\ \forall\, t_1 \neq t_2$, and $\mathrm{E}v_{st}^{-}\eta_{st}^2 v_{st}^{-\prime} = \sigma_\eta^2\Gamma$. ■

Lemma 1.9.8 $\sqrt{ST}(\widehat{\alpha}-\alpha_T(\alpha)) \xrightarrow{d} N(0,\Gamma^{-1}\Xi\Gamma^{-1})$ *for* Ξ *defined in Lemma 1.9.7 if Assumption 2 is satisfied.*

Proof. From Lemma 1.9.4,

$$\begin{aligned}\widehat{\alpha}-\alpha_T(\alpha) &= (\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime})^{-1}(\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}) - \alpha_T(\alpha) + o_p((ST)^{-1/2}) \\ &= (\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{v}_{st}^{-\prime})^{-1}(\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T}\ddot{v}_{st}^{-}\ddot{\mu}_{st}) + o_p((ST)^{-1/2}).\end{aligned}$$

The conclusion then follows immediately from Lemmas 1.9.3, 1.9.6, and 1.9.7. ■

Lemma 1.9.9 *Under Assumption 2,* $\widehat{\alpha} - \alpha = \max\{O_p(T^{-1}), O_p((ST)^{-1/2})\}$.

Proof. Lemma 1.9.4 gives

$$\begin{aligned}\widehat{\alpha} &= \left(\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T} v_{st}^{-} v_{st}^{-\prime}\right)^{-1}\left(\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T} v_{st}^{-} v_{st}\right) + O_p(T^{-1}) + o_p((ST)^{-1/2}) \\ &= \left(\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T} v_{st}^{-} v_{st}^{-\prime}\right)^{-1}\left(\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T} v_{st}^{-} \eta_{st}\right) + \alpha \\ &\quad + O_p(T^{-1}) + o_p((ST)^{-1/2}).\end{aligned}$$

Then Lemma 1.9.8 yields

$$\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T} v_{st}^{-} \eta_{st} = O_p((ST)^{-1/2}),$$

and

$$\frac{1}{S(T-p)}\sum_{s=1}^{S}\sum_{t=p+1}^{T} v_{st}^{-} v_{st}^{-\prime} = O_p(1)$$

follows immediately from a LLN, e.g. White (2001) Corollary 3.48. ■

1.10 Appendix 6. Bias Correction when $\widehat{C}_{st}$ is Estimated

An additional complication will arise if estimation of C_{st} is considered. In this case, the error in equation (1.6) is $v_{st} + \widehat{C}_{st} - C_{st}$, not v_{st}, and implementation of the FGLS estimator will require estimation of $\mathrm{E}[(V_s + \widehat{C}_s - C_s)(V_s + \widehat{C}_s - C_s)'] = \sigma_v^2\Gamma(\alpha) + \mathrm{E}[(\widehat{C}_s - C_s)(\widehat{C}_s - C_s)']$. The presence of $\widehat{C}_{st} - C_{st}$ adds an additional $O\left(\frac{1}{N}\right)$ bias to the OLS estimates of α, where for simplicity I have assumed $N_{st} = N$ for all s and t. For moderate or large N, this bias will likely be a small concern, and ignoring it may be preferable.[14] However, if N is small, the researcher may wish to account for this bias. To this end, note that if the groups are distinct then the only correlation between $\widehat{C}_{st}$ across states comes from the fact that β_1 is

[14] For a discussion in a different but related context, see Dickens (1990).

estimated rather than known, and the covariance between $\widehat{C}_{st}$ and $\widehat{C}_{s't'}$ will be zero if a separate estimate of β_{1s} is calculated within each group. Let $\widehat{C}_s$ denote the vector of time effects for state s obtained in this manner, and define $V_1(\widehat{C})$ as a $p \times p$ matrix with $[i,j]$ element

$$\begin{aligned} V_{1[i,j]}(\widehat{C}) &= \frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \operatorname{Cov}(\widehat{C}_{s(t-i)}, \widehat{C}_{s(t-j)}) \\ &- \frac{1}{T-p} \operatorname{trace}\left(\left[Z_s' V_{s,-i}(\widehat{C}) Z_{s,-j} (Z_s' Z_s)^{-1}\right]\right) \\ &- \frac{1}{T-p} \operatorname{trace}\left(\left[Z_s' V_{s,-j}(\widehat{C}) Z_{s,-i} (Z_s' Z_s)^{-1}\right]\right) \\ &+ \frac{1}{T-p} \operatorname{trace}\left(\left[Z_s' V_s(\widehat{C}) Z_s (Z_s' Z_s)^{-1} Z_{s,-i}' Z_{s,-j} (Z_s' Z_s)^{-1}\right]\right) \end{aligned}$$

and $V_2(\widehat{C})$ as a $p \times 1$ vector with i^{th} element

$$\begin{aligned} V_{2[i,1]}(\widehat{C}) &= \frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \operatorname{Cov}(\widehat{C}_{st}, \widehat{C}_{s(t-i)}) \\ &- \frac{1}{T-p} \operatorname{trace}\left(\left[Z_s' V_{s,-i}(\widehat{C}) Z_{s,-0} (Z_s' Z_s)^{-1}\right]\right) \\ &- \frac{1}{T-p} \operatorname{trace}\left(\left[Z_s' V_{s,-0}(\widehat{C}) Z_{s,-i} (Z_s' Z_s)^{-1}\right]\right) \\ &+ \frac{1}{T-p} \operatorname{trace}\left(\left[Z_s' V_s(\widehat{C}) Z_s (Z_s' Z_s)^{-1} Z_{s,-i}' Z_{s,-0} (Z_s' Z_s)^{-1}\right]\right), \end{aligned}$$

where $\operatorname{Cov}(\widehat{C}_{st}, \widehat{C}_{s't'})$ estimates the covariance between $\widehat{C}_{st}$ and $\widehat{C}_{s't'}$, $V_s(\widehat{C})$ estimates $\mathrm{E}[(\widehat{C}_s - C_s)(\widehat{C}_s - C_s)']$, $V_{s,-k}(\widehat{C})$ estimates $\mathrm{E}[(\widehat{C}_s - C_s)(\widehat{C}_{s,-k} - C_{s,-k})']$, and $\widehat{C}_{s,-k} = (\widehat{C}_{s(p+1-k)}, \widehat{C}_{s(p+2-k)}, \ldots, \widehat{C}_{s(T-k)})'$. It is then straightforward to demonstrate that

$$\begin{aligned} \widetilde{\alpha} &= \left(\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \widetilde{v}_{st}^{-} \widetilde{v}_{st}^{-\prime} - V_1(\widehat{C})\right)^{-1} \left(\frac{1}{S(T-p)} \sum_{s=1}^{S} \sum_{t=p+1}^{T} \widetilde{v}_{st}^{-} \widetilde{v}_{st} - V_2(\widehat{C})\right) \\ &\xrightarrow{p} \alpha_T(\alpha) \end{aligned}$$

as $S \to \infty$ with T fixed as long as a law of large numbers applies and $\operatorname{Cov}(\widehat{C}_{st}, \widehat{C}_{s't'})$ is an unbiased estimate of the covariance between $\widehat{C}_{st}$ and $\widehat{C}_{s't'}$. The use of a consistent estimate instead of an unbiased estimate will remove the $O\left(\frac{1}{N}\right)$ bias.

σ_v^2 also needs to be estimated in this case. A natural choice to estimate σ_v^2 is

$$\begin{aligned}\widehat{\sigma}_v^2 &= \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T}\widetilde{v}_{st}'\widetilde{v}_{st} - \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T}\mathrm{E}(\widehat{C}_{st}-C_{st})^2 + \frac{1}{T}\mathrm{trace}(Z_s'V_s(\widehat{C})Z_s(Z_s'Z_s)^{-1} \\ &\xrightarrow{p} \sigma_v^2 - \frac{\sigma_v^2}{T}\mathrm{trace}(Z_s'\Gamma(\alpha)Z_s(Z_s'Z_s)^{-1})\end{aligned}$$

from which a consistent estimator may easily be recovered.

Bibliography

AMEMIYA, T. (1978): "A Note on a Random Coefficient Model," *International Economic Review*, 19(3), 793–796.

ARELLANO, M. (1987): "Computing Robust Standard Errors for Within-Groups Estimators," *Oxford Bulletin of Economics and Statistics*, 49(4), 431–434.

BELL, R. M., AND D. F. MCCAFFREY (2002): "Computing Robust Standard Errors for Within-Groups Estimators," Mimeo RAND.

BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2003): "How Much Should We Trust Differences-in-Differences Estimates?," forthcoming QJE.

BHARGAVA, A., L. FRANZINI, AND W. NARENDRANATHAN (1982): "Serial Correlation and the Fixed Effects Model," *Review of Economic Studies*, 49, 533–549.

BILLINGSLEY, P. (1995): *Probability and Measure*. New York: Wiley, third edn.

DICKENS, W. T. (1990): "Error Components in Grouped Data: Is It Ever Worth Weighting?," *Review of Economics and Statistics*, 72(2), 328–333.

DONALD, S., AND K. LANG (2001): "Inference with Difference in Differences and Other Panel Data," Mimeo.

DOUKHAN, P. (1994): *Mixing: Properties and Examples*, vol. 85 of *Lecture Notes in Statistics (Springer-Verlag)*. New York: Springer-Verlag.

FITZPATRICK, P. M. (1996): *Advanced Calculus: A Course in Mathematical Analysis.* PWS Publishing Company: Boston, Massachussetts.

HAHN, J., AND G. M. KUERSTEINER (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both N and T Are Large," *Econometrica*, 70(4), 1639–1657.

HAHN, J., AND W. NEWEY (2002): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," Mimeo.

HAMILTON, J. D. (1994): *Time Series Analysis.* Princeton University Press: Princeton, New Jersey.

HAUSMAN, J., AND G. KUERSTEINER (2003): "Differences in Differences Meets Generalized Least Squares: Higher Order Properties of Hypothesis Tests," Mimeo.

KEZDI, G. (2002): "Robust Standard Errors Estimation in Fixed-Effects Panel Models," Mimeo.

KIEFER, N. M. (1980): "Estimation of Fixed-Effects Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance," *Journal of Econometrics*, 14, 195–202.

LANCASTER, T. (2002): "Orthogonal Parameters and Panel Data," *Review of Economic Studies*, 69, 647–666.

LIANG, K.-Y., AND S. ZEGER (1986): "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73(1), 13–22.

MACURDY, T. E. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 18(1), 83–114.

MOULTON, B. R. (1986): "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 32(3), 385–397.

NICKELL, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–1426.

PHILLIPS, P. C. B., AND H. R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67(5), 1057–1111.

Solon, G. (1984): "Estimating Autocorrelations in Fixed Effects Models," NBER Technical Working Paper No. 32.

White, H. (2001): *Asymptotic Theory for Econometricians.* San Diego: Academic Press, revised edn.

Wooldridge, J. M. (2003): "Cluster-Sample Methods in Applied Econometrics," *American Economic Review*, 93(2), 133–188.

TABLE 1. Bias and MSE of $\widehat{\alpha}$ and $\widehat{\alpha}_{BC}$ with Fixed Effects

			AR(1) Model $\alpha_1 = 0.8$				AR(2) Model $\alpha_1 = 0.80$, $\alpha_2 = 0$			
S	T	$\widehat{\alpha}$	$\widehat{\alpha}^{(1)}$	$\widehat{\alpha}^{(\infty)}$	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_1^{(1)}$	$\widehat{\alpha}_2^{(1)}$	$\widehat{\alpha}_1^{(\infty)}$	$\widehat{\alpha}_2^{(\infty)}$
51	23	-0.099	-0.013	-0.003	-0.061	-0.058	-0.009	-0.008	-0.002	-0.002
		(0.010)	(0.0008)	(0.0007)	(0.005)	(0.004)	(0.0012)	(0.0011)	(0.0012)	(0.0011)
				[0]						[0]
204	23	-0.098	-0.011	-0.001	-0.060	-0.056	-0.008	-0.007	-0.001	-0.00003
		(0.010)	(0.0003)	(0.0002)	(0.004)	(0.003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
				[0]						[0]
1020	23	-0.096	-0.010	-0.0001	-0.059	-0.056	-0.007	-0.007	0.0001	-0.0001
		(0.009)	(0.0001)	(0.00003)	(0.004)	(0.003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
				[0]						[0]
51	12	-0.216	-0.052	-0.010	-0.149	-0.129	-0.045	-0.033	-0.009	-0.002
		(0.048)	(0.005)	(0.003)	(0.025)	(0.018)	(0.005)	(0.003)	(0.003)	(0.003)
				[0]						[0]
204	12	-0.210	-0.044	-0.002	-0.142	-0.127	-0.039	-0.031	-0.002	-0.0003
		(0.044)	(0.002)	(0.001)	(0.021)	(0.017)	(0.002)	(0.002)	(0.001)	(0.001)
				[0]						[0]
1020	12	-0.208	-0.043	-0.0003	-0.141	-0.127	-0.037	-0.031	-0.0002	-0.0001
		(0.044)	(0.002)	(0.0001)	(0.020)	(0.016)	(0.002)	(0.001)	(0.0001)	(0.0001)
				[0]						[0]
51	6	-0.491	-0.201	-0.023	-0.424	-0.325	-0.246	-0.164	-0.036	-0.020
		(0.244)	(0.046)	(0.010)	(0.187)	(0.109)	(0.068)	(0.033)	(0.012)	(0.009)
				[30]						[67]
204	6	-0.484	-0.192	-0.005	-0.418	-0.321	-0.238	-0.158	-0.005	0.0007
		(0.235)	(0.038)	(0.003)	(0.176)	(0.104)	(0.058)	(0.027)	(0.003)	(0.003)
				[0]						[4]
1020	6	-0.481	-0.189	-0.0007	-0.415	-0.322	-0.234	-0.159	-0.0006	0.0001
		(0.232)	(0.036)	(0.0006)	(0.172)	(0.104)	(0.055)	(0.026)	(0.0005)	(0.0005)
				[0]						[0]

Monte Carlo results for simulation model based on aggregate CPS-MORG data with fixed effects. Data are manufactured so that variances match those in the CPS data. S is the number of aggregate cross-sectional observations, and T denotes the number of time series observations. $S = 51$ and $T = 23$ were chosen to match CPS data, and remaining sample sizes were chosen to explore behavior for small T and large S. $\widehat{\alpha}$, $\widehat{\alpha}^{(1)}$, and $\widehat{\alpha}^{(\infty)}$ are, respectively, least squares, one-step bias-corrected least squares, and iteratively bias-corrected least squares estimates of the autocorrelation coefficients. MSE is given in parentheses, and the number of times $\alpha_T^{-1}(\widehat{\alpha})$ failed to exist is given in brackets. The number of simulations is 1000.

TABLE 2. Bias and MSE of $\widehat{\alpha}$ and $\widehat{\alpha}_{BC}$ with Fixed Effects and State-Specific Trends

		AR(1) Model $\alpha_1 = 0.8$			AR(2) Model $\alpha_1 = 0.80$, $\alpha_2 = 0$					
S	T	$\widehat{\alpha}$	$\widehat{\alpha}^{(1)}$	$\widehat{\alpha}^{(\infty)}$	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_1^{(1)}$	$\widehat{\alpha}_2^{(1)}$	$\widehat{\alpha}_1^{(\infty)}$	$\widehat{\alpha}_2^{(\infty)}$
51	23	-0.219	-0.054	-0.005	-0.147	-0.119	-0.047	-0.029	-0.004	-0.0005
		(0.049)	(0.004)	(0.002)	(0.023)	(0.015)	(0.004)	(0.002)	(0.002)	(0.001)
				[0]						[3]
204	23	-0.215	-0.050	0.0003	-0.142	-0.119	-0.042	-0.029	0.0006	-0.0004
		(0.046)	(0.003)	(0.0004)	(0.021)	(0.014)	(0.002)	(0.001)	(0.0004)	(0.0003)
				[0]						[0]
1020	23	-0.215	-0.050	-0.0005	-0.143	-0.119	-0.043	-0.029	-0.0004	-0.0002
		(0.046)	(0.003)	(0.0001)	(0.021)	(0.014)	(0.002)	(0.001)	(0.0001)	(0.0001)
				[0]						[0]
51	12	-0.467	-0.200	-0.030	-0.369	-0.248	-0.203	-0.109	-0.054	-0.025
		(0.219)	(0.043)	(0.009)	(0.139)	(0.063)	(0.045)	(0.014)	(0.012)	(0.004)
				[87]						[173]
204	12	-0.460	-0.192	0.002	-0.361	-0.247	-0.195	-0.108	-0.007	-0.006
		(0.212)	(0.038)	(0.003)	(0.131)	(0.061)	(0.039)	(0.012)	(0.004)	(0.001)
				[4]						[45]
1020	12	-0.460	-0.192	-0.0001	-0.360	-0.247	-0.194	-0.108	0.001	-0.001
		(0.212)	(0.037)	(0.001)	(0.130)	(0.061)	(0.038)	(0.012)	(0.001)	(0.0002)
				[0]						[0]
51	6	-0.955	-0.660	-0.262	-0.919	-0.493	-0.803	-0.369	-0.403	-0.162
		(0.914)	(0.441)	(0.126)	(0.850)	(0.246)	(0.653)	(0.141)	(0.253)	(0.046)
				[270]						[318]
204	6	-0.950	-0.653	-0.135	-0.914	-0.492	-0.795	-0.366	-0.268	-0.106
		(0.903)	(0.428)	(0.068)	(0.836)	(0.242)	(0.634)	(0.135)	(0.162)	(0.029)
				[152]						[230]
1020	6	-0.949	-0.652	-0.020	-0.914	-0.492	-0.796	-0.366	-0.083	-0.023
		(0.902)	(0.426)	(0.009)	(0.835)	(0.242)	(0.634)	(0.134)	(0.028)	(0.005)
				[8]						[28]

Monte Carlo results for simulation model based on aggregate CPS-MORG data with fixed effects. Data are manufactured so that variances match those in the CPS data. S is the number of aggregate cross-sectional observations, and T denotes the number of time series observations. $S = 51$ and $T = 23$ were chosen to match CPS data, and remaining sample sizes were chosen to explore behavior for small T and large S. $\widehat{\alpha}$, $\widehat{\alpha}^{(1)}$, and $\widehat{\alpha}^{(\infty)}$ are, respectively, least squares, one-step bias-corrected least squares, and iteratively bias-corrected least squares estimates of the autocorrelation coefficients. MSE is given in parentheses, and the number of times $\alpha_T^{-1}(\widehat{\alpha})$ failed to exist is given in brackets. The number of simulations is 1000.

TABLE 3. Estimated Variance of Treatment Parameter in Simulated AR(1) Model with $\alpha = .8$.

	$S = 51, T = 23$			$S = 51, T = 12$			$S = 51, T = 6$		
	$\widehat{\sigma}^2$	σ_a^2	σ_s^2	$\widehat{\sigma}^2$	σ_a^2	σ_s^2	$\widehat{\sigma}^2$	σ_a^2	σ_s^2
A. Conventional Inference Methods									
OLS	0.024	0.108	0.107	0.035	0.101	0.109	0.046	0.078	0.081
OLS - Cluster by state	0.102	0.108	0.107	0.096	0.101	0.109	0.075	0.078	0.081
FGLS-U	0.020	0.037	0.062	0.030	0.040	0.049	0.038	0.044	0.051
AR(1)	0.034	0.037	0.039	0.038	0.040	0.046	0.045	0.044	0.060
AR(1) - Cluster by state	0.037	0.037	0.039	0.044	0.040	0.046	0.054	0.044	0.060
AR(2)	0.034	0.037	0.040	0.038	0.040	0.046	0.045	0.044	0.060
AR(2) - Cluster by state	0.038	0.037	0.040	0.045	0.040	0.046	0.055	0.044	0.060
B. FGLS with Bias-Corrected AR Coefficients									
$AR(1)_{bc}$	0.037	0.037	0.037	0.040	0.040	0.038	0.044	0.044	0.046
$AR(1)_{bc}$ - Cluster by state	0.035	0.037	0.037	0.039	0.040	0.038	0.041	0.044	0.046
$AR(2)_{bc}$	0.037	0.037	0.037	0.040	0.040	0.038	0.044	0.044	0.046
$AR(2)_{bc}$ - Cluster by state	0.035	0.037	0.037	0.038	0.040	0.038	0.041	0.044	0.046

Monte Carlo results for simulation model based on aggregate CPS-MORG data. Data are manufactured so that variances correspond to those in the CPS data, and the error process is assumed to follow an AR(1) with $\alpha = .8$. S is the number of aggregate cross-sectional observations, and T denotes the number of time series observations. $S = 51$ and $T = 23$ correspond to the full sample of CPS data, and remaining sample sizes were chosen to explore behavior for small T. Results are for the variance of the treatment parameter, β_1, only. $\widehat{\sigma}^2$ is 1000 times the mean of the estimated variances, σ_a^2 is 1000 times the asymptotic variance, and σ_s^2 is 1000 times the variance of the β_1's estimated in the simulation. The number of simulations is 1000.

TABLE 4. Monte Carlo Results from CPS-MORG Microdata

	$S = 51, T = 23$			$S = 51, T = 12$			$S = 51, T = 6$		
	Size	Power	Length	Size	Power	Length	Size	Power	Length
A. Conventional Inference Methods									
OLS	0.594	0.860	0.006	0.398	0.866	0.009	0.072	0.850	0.012
	(0.022)	(0.016)		(0.022)	(0.015)		(0.012)	(0.016)	
OLS - Cluster by state	0.078	0.354	0.024	0.078	0.548	0.019	0.064	0.798	0.013
	(0.012)	(0.021)		(0.012)	(0.022)		(0.011)	(0.018)	
Random Effects - state × year	0.392	0.778	0.010	0.222	0.870	0.011	0.050	0.808	0.013
	(0.022)	(0.019)		(0.019)	(0.015)		(0.009)	(0.018)	
OLS - Cluster by state × year	0.398	0.766	0.011	0.280	0.804	0.011	0.076	0.844	0.012
	(0.022)	(0.019)		(0.020)	(0.018)		(0.012)	(0.016)	
B. FGLS with Bias-Corrected AR Coefficients									
$AR(2)_{bc}$	0.078	0.714	0.015	0.152	0.880	0.012	0.050	0.804	0.014
	(0.012)	(0.020)		(0.016)	(0.015)		(0.010)	(0.018)	
$AR(2)_{bc}$ - Cluster by state	0.078	0.728	0.016	0.070	0.880	0.015	0.058	0.792	0.014
	(0.012)	(0.020)		(0.011)	(0.015)		(0.010)	(0.018)	
$AR(3)_{bc}$	0.076	0.738	0.015	0.142	0.888	0.012			
	(0.012)	(0.020)		(0.016)	(0.014)				
$AR(3)_{bc}$ - Cluster by state	0.076	0.744	0.015	0.072	0.774	0.015			
	(0.012)	(0.020)		(0.012)	(0.019)				

Monte Carlo results for simulation model using individual CPS-MORG data. Data are manufactured by resampling states from actual CPS-MORG data. S is the number of aggregate cross-sectional observations, and T denotes the number of time series observations. $S = 51$ and $T = 23$ correspond to the full sample of CPS data, and remaining sample sizes were chosen to explore behavior for small T. Results are for the treatment parameter which enters the model with a true coefficient of $\beta_1 = 0$. Size and power are for 5% level tests, and power is versus the alternative that $\beta_1 = .02$. Length is the 95% confidence interval length divided by two. Interval lengths are based on a t_{50} for tests with standard errors clustered at the state level, a t_{ST-1} for tests with standard errors clustered at the state × year level, and a $N(0, 1)$ for the remaining tests. Simulation standard errors are reported in parentheses. (For interval length, the simulation standard error is negligible, so it is not reported.) The number of simulations is 500.

TABLE 5. Monte Carlo Results from Aggregate CPS-MORG Data

	$S = 51, T = 23$			$S = 51, T = 12$			$S = 51, T = 6$		
	Size	Power	Length	Size	Power	Length	Size	Power	Length
A. Conventional Inference Methods									
OLS	0.374	0.762	0.011	0.158	0.846	0.012	0.052	0.730	0.015
	(0.022)	(0.019)		(0.016)	(0.016)		(0.010)	(0.020)	
OLS - Cluster by state	0.066	0.344	0.025	0.062	0.656	0.017	0.052	0.744	0.015
	(0.011)	(0.021)		(0.011)	(0.021)		(0.010)	(0.020)	
FGLS-U	0.362	0.962	0.007	0.112	0.932	0.010	0.086	0.804	0.013
	(0.021)	(0.009)		(0.014)	(0.011)		(0.013)	(0.018)	
AR(2)	0.104	0.772	0.014	0.116	0.844	0.014	0.080	0.774	0.014
	(0.014)	(0.019)		(0.014)	(0.016)		(0.012)	(0.019)	
AR(2) - Cluster by state	0.064	0.662	0.016	0.062	0.726	0.016	0.048	0.716	0.015
	(0.011)	(0.021)		(0.011)	(0.020)		(0.010)	(0.020)	
AR(3)	0.106	0.770	0.014	0.132	0.834	0.013			
	(0.014)	(0.019)		(0.015)	(0.017)				
AR(3) - Cluster by state	0.072	0.688	0.016	0.068	0.708	0.016			
	(0.012)	(0.021)		(0.011)	(0.020)				
B. FGLS with Bias-Corrected AR Coefficients									
$AR(2)_{bc}$	0.080	0.750	0.015	0.064	0.814	0.014	0.042	0.720	0.015
	(0.012)	(0.019)		(0.011)	(0.017)		(0.009)	(0.020)	
$AR(2)_{bc}$ - Cluster by state	0.062	0.754	0.015	0.062	0.800	0.014	0.060	0.734	0.015
	(0.011)	(0.019)		(0.011)	(0.018)		(0.011)	(0.020)	
$AR(3)_{bc}$	0.064	0.766	0.014	0.066	0.810	0.014			
	(0.011)	(0.019)		(0.011)	(0.018)				
$AR(3)_{bc}$ - Cluster by state	0.076	0.788	0.014	0.060	0.808	0.014			
	(0.012)	(0.018)		(0.011)	(0.018)				

Monte Carlo results for simulation model using aggregate CPS-MORG data. Data are manufactured by resampling states from actual CPS-MORG data and aggregating data to the state-year level using the method of Amemiya (1978) outlined in the text. S is the number of aggregate cross-sectional observations, and T denotes the number of time series observations. $S = 51$ and $T = 23$ correspond to the full sample of CPS data, and remaining sample sizes were chosen to explore behavior for small T. Results are for the treatment parameter which enters the model with a true coefficient of $\beta_1 = 0$. Size and power are for 5% level tests, and power is versus the alternative that $\beta_1 = .02$. Length is the 95% confidence interval length divided by two. Interval lengths are based on a t_{50} for tests with standard errors clustered at the state level and a t_{ST-S-T} for the remaining tests. Simulation standard errors are reported in parentheses. (For interval length, the simulation standard error is negligible, so it is not reported.) The number of simulations is 500.

TABLE 6. Monte Carlo Results from Simulated AR(1) Model with $\alpha = .8$.

	$S = 51, T = 23$			$S = 51, T = 12$			$S = 51, T = 6$		
	Size	Power	Length	Size	Power	Length	Size	Power	Length
A. Conventional Inference Methods									
OLS	0.356	0.842	0.010	0.285	0.794	0.012	0.141	0.730	0.013
	(0.015)	(0.011)		(0.014)	(0.013)		(0.011)	(0.014)	
OLS - Cluster by state	0.056	0.490	0.020	0.065	0.520	0.020	0.058	0.574	0.017
	(0.007)	(0.016)		(0.008)	(0.016)		(0.007)	(0.016)	
FGLS-U	0.269	0.915	0.009	0.121	0.900	0.011	0.091	0.834	0.012
	(0.014)	(0.008)		(0.011)	(0.009)		(0.009)	(0.012)	
AR(1)	0.078	0.914	0.011	0.077	0.878	0.012	0.083	0.786	0.013
	(0.008)	(0.009)		(0.008)	(0.011)		(0.008)	(0.013)	
AR(1) - Cluster by state	0.057	0.884	0.012	0.059	0.830	0.013	0.064	0.728	0.015
	(0.007)	(0.010)		(0.008)	(0.012)		(0.008)	(0.014)	
AR(2)	0.080	0.912	0.011	0.081	0.874	0.012	0.093	0.778	0.013
	(0.008)	(0.009)		(0.008)	(0.011)		(0.009)	(0.013)	
AR(2) - Cluster by state	0.062	0.882	0.012	0.060	0.828	0.013	0.064	0.724	0.015
	(0.008)	(0.010)		(0.008)	(0.012)		(0.008)	(0.014)	
B. FGLS with Bias-Corrected AR Coefficients									
$AR(1)_{bc}$	0.056	0.908	0.012	0.044	0.880	0.012	0.062	0.818	0.013
	(0.007)	(0.009)		(0.006)	(0.011)		(0.008)	(0.012)	
$AR(1)_{bc}$ - Cluster by state	0.061	0.904	0.012	0.050	0.874	0.012	0.069	0.822	0.013
	(0.008)	(0.009)		(0.007)	(0.011)		(0.008)	(0.012)	
$AR(2)_{bc}$	0.057	0.904	0.012	0.044	0.882	0.012	0.066	0.822	0.013
	(0.007)	(0.009)		(0.006)	(0.010)		(0.008)	(0.012)	
$AR(2)_{bc}$ - Cluster by state	0.065	0.900	0.012	0.049	0.880	0.012	0.070	0.824	0.013
	(0.008)	(0.009)		(0.007)	(0.011)		(0.008)	(0.012)	

Monte Carlo results for simulation model based on aggregate CPS-MORG data. Data are manufactured so that variances correspond to those in the CPS data, and the error process is assumed to follow an AR(1) with $\alpha = .8$. S is the number of aggregate cross-sectional observations, and T denotes the number of time series observations. $S = 51$ and $T = 23$ correspond to the full sample of CPS data, and remaining sample sizes were chosen to explore behavior for small T. Results are for the treatment parameter which enters the model with a true coefficient of $\beta_1 = 0$. Size and power are for 5% level tests, and power is versus the alternative that $\beta_1 = .02$. Length is the 95% confidence interval length divided by two. Interval lengths are based on a t_{50} for tests with standard errors clustered at the state level and a t_{ST-S-T} for the remaining tests. Simulation standard errors are reported in parentheses. (For interval length, the simulation standard error is negligible, so it is not reported.) The number of simulations is 1000.

TABLE 7. Monte Carlo Results from Simulated AR(2) Model with $\alpha_1 = .43$ and $\alpha_2 = .30$.

	$S = 51, T = 23$			$S = 51, T = 12$			$S = 51, T = 6$		
	Size	Power	Length	Size	Power	Length	Size	Power	Length
A. Conventional Inference Methods									
OLS	0.324	0.856	0.010	0.210	0.764	0.012	0.091	0.702	0.015
	(0.015)	(0.011)		(0.013)	(0.013)		(0.009)	(0.014)	
OLS - Cluster by state	0.054	0.536	0.019	0.063	0.508	0.019	0.058	0.612	0.017
	(0.007)	(0.016)		(0.008)	(0.016)		(0.007)	(0.016)	
FGLS-U	0.233	0.890	0.010	0.132	0.813	0.013	0.078	0.746	0.015
	(0.013)	(0.010)		(0.011)	(0.012)		(0.008)	(0.013)	
AR(1)	0.101	0.816	0.013	0.120	0.744	0.014	0.095	0.708	0.015
	(0.009)	(0.012)		(0.011)	(0.014)		(0.009)	(0.014)	
AR(1) - Cluster by state	0.054	0.744	0.015	0.065	0.658	0.016	0.058	0.614	0.017
	(0.007)	(0.014)		(0.008)	(0.015)		(0.007)	(0.016)	
AR(2)	0.065	0.850	0.013	0.115	0.750	0.014	0.106	0.698	0.015
	(0.008)	(0.011)		(0.010)	(0.013)		(0.010)	(0.015)	
AR(2) - Cluster by state	0.051	0.806	0.014	0.065	0.680	0.016	0.059	0.582	0.018
	(0.007)	(0.013)		(0.008)	(0.015)		(0.008)	(0.016)	
B. FGLS with Bias-Corrected AR Coefficients									
$AR(1)_{bc}$	0.070	0.816	0.013	0.077	0.708	0.015	0.061	0.686	0.016
	(0.008)	(0.012)		(0.008)	(0.014)		(0.008)	(0.015)	
$AR(1)_{bc}$ - Cluster by state	0.059	0.766	0.015	0.070	0.704	0.015	0.054	0.676	0.016
	(0.008)	(0.013)		(0.008)	(0.014)		(0.007)	(0.015)	
$AR(2)_{bc}$	0.043	0.838	0.013	0.053	0.746	0.014	0.055	0.706	0.016
	(0.006)	(0.011)		(0.007)	(0.013)		(0.007)	(0.014)	
$AR(2)_{bc}$ - Cluster by state	0.060	0.828	0.013	0.070	0.746	0.014	0.058	0.700	0.016
	(0.008)	(0.012)		(0.008)	(0.013)		(0.007)	(0.014)	

Monte Carlo results for simulation model based on aggregate CPS-MORG data. Data are manufactured so that variances correspond to those in the CPS data, and the error process is assumed to follow an AR(2) with $\alpha_1 = .43$ and $\alpha_2 = .3$. S is the number of aggregate cross-sectional observations, and T denotes the number of time series observations. $S = 51$ and $T = 23$ correspond to the full sample of CPS data, and remaining sample sizes were chosen to explore behavior for small T. Results are for the treatment parameter which enters the model with a true coefficient of $\beta_1 = 0$. Size and power are for 5% level tests, and power is versus the alternative that $\beta_1 = .02$. Length is the 95% confidence interval length divided by two. Interval lengths are based on a t_{50} for tests with standard errors clustered at the state level and a t_{ST-S-T} for the remaining tests. Simulation standard errors are reported in parentheses. (For interval length, the simulation standard error is negligible, so it is not reported.) The number of simulations is 1000.

Chapter 2

Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large

2.1 Introduction

The use of heteroskedasticity robust covariance matrix estimators, cf. White (1980), in cross-sectional settings and of heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators, cf. Andrews (1991), in time series contexts is extremely common in applied econometrics. The popularity of these robust covariance matrix estimators is due to their consistency under weak functional form assumptions. In particular, their use allows the researcher to form valid confidence regions about a set of parameters from a model of interest without specifying an exact process for the disturbances in the model.

With the increasing use of panel data methods and the increasing availability of panel data, it is natural that extensions of existing robust covariance matrix estimators for panel data settings that allow for arbitrary within individual correlation are becoming more common. A recent paper by Bertrand, Duflo, and Mullainathan (2003) illustrated the pitfalls of ignoring serial correlation in panel data, finding through a simulation study that inference procedures which fail to account for the potential within individual serial correlation may be severely size distorted. As a potential resolution of this problem, Bertrand, Duflo, and

Mullainathan (2003) suggest the use of a robust covariance matrix estimator proposed by Arellano (1987) and explored in Kezdi (2002) which allows arbitrary within individual correlation and find in a simulation study that tests based on this estimator of the covariance parameters have correct size.

One drawback of the estimator of Arellano (1987), hereafter referred to as the "clustered" covariance matrix (CCM) estimator, is that its properties are only known in conventional panel asymptotics as the cross section dimension, n, increases with the time dimension, T, fixed. While many panel data sets are indeed characterized by large n and relatively small T, this is not necessarily the case. For example, in many differences-in-differences and policy evaluation studies, the cross-section is composed of states and the time dimension of yearly or quarterly (or occasionally monthly) observations on each state. In this case, it is not uncommon to have 20 or more annual observations, in which case $n \approx T$.

In this paper, I address this issue by exploring the theoretical properties of the CCM estimator in asymptotics that allow n and T to go to infinity jointly and in asymptotics where T goes to infinity with n fixed. Perhaps surprisingly, I find that the CCM estimator, appropriately normalized, is consistent without imposing any conditions on the rate of growth of T relative to n even when the time series dependence between the observations within each individual is left unrestricted. In this case, both the OLS estimator and the CCM estimator converge at only the $\sqrt{n}$-rate, essentially because the only information is coming from cross-sectional variation. If the time series process is restricted to be strongly mixing, I show that the OLS estimator is $\sqrt{nT}$-consistent but that, because high lags are not down weighted, the robust covariance matrix estimator still converges at only the $\sqrt{n}-$rate. This behavior suggests, as indicated in the simulations found in Kezdi (2002), that it is the n dimension and not the size of n relative to T that matters for determining the properties of the CCM estimator.

It is interesting to note that the limiting behavior of $\widehat{\beta}$ changes "discontinuously" as the amount of dependence is limited. In particular, the rate of convergence of $\widehat{\beta}$ changes from $\sqrt{n}$ in the "no-mixing case" to $\sqrt{nT}$ when mixing is imposed. However, despite the difference in the limiting behavior of $\widehat{\beta}$, there is no difference in the behavior of standard inference procedures based on the CCM estimator between the two cases. In particular, the same t and F statistics will be valid in either case (and in the $n \to \infty$ with T fixed case) without reference to the asymptotics or degree of dependence in the data.

I also derive the behavior of the CCM estimator as $T \to \infty$ with n fixed, where I find the estimator is not consistent but does have a limiting distribution. This result corresponds to asymptotic results for HAC estimators without truncation found in recent work by Kiefer and Vogelsang (2002), Kiefer and Vogelsang (2003), Phillips, Sun, and Jin (2003), and Vogelsang (2002). While the limiting distribution is not proportional to the true covariance matrix in general, it is proportional to the covariance matrix in the important special case of iid data across individuals,[1] allowing construction of asymptotically pivotal statistics in this case. In fact, in this case, the standard t-statistic is not asymptotically normal but converges in distribution to a random variable which is exactly proportional to a t_{n-1} distribution. This behavior suggests the use of the t_{n-1} for constructing confidence intervals and tests when the CCM estimator is used as a general rule, as this will provide asymptotically correct critical values under any asymptotic sequence.

I then explore the finite sample behavior of the CCM estimator and tests based upon it through a short simulation study. The simulation results indicate that tests based on the robust standard error estimates generally have approximately correct size in serially correlated panel data even in small samples. However, the standard error estimates themselves are considerably more variable than their counterparts based on simple parametric models. The bias of the simple parametric estimators is also typically smaller in the cases where the parametric model is correct, suggesting that these standard error estimates are likely preferable when the researcher is confident in the form of the error process. In the simulation, I also explore the behavior of an analog of White's (1980) direct test for heteroskedasticity proposed by Kezdi (2002).[2] The results indicate the performance of the test is fairly good for moderate n, though it is quite poor when n is small. This simulation behavior suggests that this test may be useful for choosing between the use of robust standard error estimates and standard errors estimated from a more parsimonious model when n is reasonably large.

The remainder of this paper is organized as follows. In Section 2, I present the basic framework and the estimator and test statistics that will be considered. The asymptotic properties of these estimators are collected in Section 3, and Section 4 contains a discussion of a Monte Carlo study assessing the finite sample performance of the estimators in simple

[1] Note that this still allows arbitrary correlation and heteroskedasticity within individuals, but restricts that the pattern is the same across individuals.

[2] Solon and Inoue (2004) offers a different testing procedure for detecting serial correlation in fixed effects panel models.

models. Section 5 concludes.

2.2 A Heteroskedasticity-Autocorrelation Consistent Covariance Matrix Estimator for Panel Data

Consider a regression model defined by

$$y_{it} = x_{it}'\beta + \epsilon_{it} \tag{2.1}$$

where $i = 1, \ldots, n$ indexes individuals, $t = 1, \ldots, T$ indexes time, x_{it} is a $k \times 1$ vector of observable covariates, and ϵ_{it} is an unobservable error component. Note that this formulation incorporates the standard fixed effects model as well as models that include other covariates which enter the model with individual specific coefficients, such as individual specific time trends, where these covariates have been partialed out. In these cases, the variables x_{it}, y_{it}, and ϵ_{it} should be interpreted as residuals from regressions of x_{it}^*, y_{it}^*, and ϵ_{it}^* on an auxiliary set of covariates z_{it}^* from the underlying model $y_{it}^* = x_{it}^{*'}\beta + z_{it}^{*'}\gamma + \epsilon_{it}^*$. For example, in the fixed effects model, Z^* is a matrix of dummy variables for each individual and γ is a vector of individual specific fixed effects. In this case, $x_{it} = x_{it}^* - \frac{1}{T}\sum_{t=1}^{T} x_{it}^*$, and y_{it} and ϵ_{it} are defined similarly. Alternatively, x_{it}, y_{it}, and ϵ_{it} could be interpreted as variables resulting from other transformations which remove the nuisance parameters from the equation, such as first-differencing to remove the fixed effects. In what follows, all properties are given in terms of the transformed variables for convenience. Alternatively, conditions could be imposed on the underlying variables and the properties derived as $T \to \infty$ as in Hansen (2004).

Within each individual, the equations defined by (2.1) may be stacked and represented in matrix form as

$$y_i = x_i\beta + \epsilon_i \tag{2.2}$$

where y_i is a $T \times 1$ vector of individual outcomes, x_i is a $T \times k$ vector of observed covariates, and ϵ_i is a $T \times 1$ vector of unobservables affecting the outcomes y_i with $\mathrm{E}[\epsilon_i\epsilon_i'|x_i] = \Omega_i$. The OLS estimator of β from equation (2.2) may then be defined as $\widehat{\beta} = (\sum_{i=1}^{n} x_i'x_i)^{-1}\sum_{i=1}^{n} x_i'y_i$. The properties of $\widehat{\beta}$ as $n \to \infty$ with T fixed are well known. In particular, under regularity conditions, $\sqrt{n}(\widehat{\beta} - \beta)$ is asymptotically normal with covariance matrix $M^{-1}WM^{-1}$ where $M = \lim_n \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[x_i'x_i]$ and $W = \lim_n \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[x_i'\Omega_i x_i]$.

The problem of robust covariance matrix estimation is then estimating W without imposing a parametric structure on the Ω_i. In this paper, I consider the estimator suggested by Arellano (1987) which may be defined as

$$\widehat{W} = \frac{1}{nT}\sum_{i=1}^{n} x_i' \widehat{\epsilon}_i \widehat{\epsilon}_i' x_i \tag{2.3}$$

where $\widehat{\epsilon}_i = y_i - x_i\widehat{\beta}$ are OLS residuals from equation (2.2). This estimator is an appealing generalization of White's (1980) heteroskedasticity consistent covariance matrix estimator that allows for arbitrary intertemporal correlation patterns and heteroskedasticity across individuals.[3] The estimator is also appealing in that, unlike HAC estimators for time series data, its implementation does not require the selection of a kernel or bandwidth parameter. The properties of $\widehat{W}$ under conventional panel asymptotics where $n \to \infty$ with T fixed are well-established. In the remainder of this paper, I extend this analysis by considering the properties of $\widehat{W}$ under asymptotic sequences where $T \to \infty$ as well.

The chief reason for interest in the CCM estimator is for performing inference about $\widehat{\beta}$. Suppose $\sqrt{d_{nT}}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, B)$ and define an estimator of the asymptotic variance of $\widehat{\beta}$ as $\frac{1}{d_{nT}}\widehat{B}$ where $\widehat{B} \xrightarrow{p} B$. The following estimator of the asymptotic variance of $\widehat{\beta}$ based on $\widehat{W}$ is used throughout the remainder of the paper:

$$\begin{aligned}\widehat{Avar}(\widehat{\beta}) &= (\sum_{i=1}^{n} x_i' x_i)^{-1}(nT\widehat{W})(\sum_{i=1}^{n} x_i' x_i)^{-1} \\ &= (\sum_{i=1}^{n} x_i' x_i)^{-1}(\sum_{i=1}^{n} x_i' \widehat{\epsilon}_i \widehat{\epsilon}_i' x_i)(\sum_{i=1}^{n} x_i' x_i)^{-1}.\end{aligned} \tag{2.4}$$

In addition, for testing the hypothesis $R\beta = r$ for a $q \times k$ matrix R with rank q, the usual t (for R a $1 \times k$ vector) and Wald statistics can be defined as

$$t^* = \frac{\sqrt{nT}(R\widehat{\beta} - r)}{\sqrt{R\widehat{Q}^{-1}\widehat{W}\widehat{Q}^{-1}R'}} \tag{2.5}$$

and

$$F^* = nT(R\widehat{\beta} - r)'[R\widehat{Q}^{-1}\widehat{W}\widehat{Q}^{-1}R']^{-1}(R\widehat{\beta} - r) \tag{2.6}$$

[3]It does, however, ignore the possibility of cross-sectional correlation, and it will be assumed that there is no cross-sectional correlation for the remainder of the paper.

respectively, where $\widehat{W}$ is defined above and $\widehat{Q} = \frac{1}{nT}\sum_{i=1}^{n} x_i' x_i$. In Section 2.3, I verify that, despite differences in the limiting behavior of $\widehat{\beta}$, $t^* \xrightarrow{d} N(0,1)$, $F^* \xrightarrow{d} \chi_q^2$, and $\widehat{Avar}(\widehat{\beta})$ is valid for estimating the asymptotic variance of $\widehat{\beta}$ as $n \to \infty$ regardless of the behavior of T. I also consider the behavior of t^* and F^* as $T \to \infty$ with n fixed. In this case, $\widehat{W}$ is not consistent for W but does have a limiting distribution; and when the data are iid across i,[4] I show that $t^* \xrightarrow{d} (\frac{n}{n-1})^{1/2} t_{n-1}$ and that F^* is asymptotically pivotal and so can be used to construct valid tests. This behavior suggests that inference using $\frac{n}{n-1}\widehat{W}$ and forming critical values using a t_{n-1} distribution will be valid regardless of the asymptotic sequence considered.

It is worth noting that the estimator $\widehat{W}$ has also been used extensively in multilevel models to account for the presence of correlation between individuals within cells; cf. Liang and Zeger (1986) and Bell and McCaffrey (2002). For example, in a schooling study, one might have data on individual outcomes where the individuals are grouped into classes. In this case, the cross-sectional unit of observation could be defined as the class, and arbitrary correlation between all individuals within each class could be allowed. In this case, one would expect the presence of a classroom specific random effect resulting in equicorrelation between all individuals within a class. While this would clearly violate the mixing assumptions imposed in obtaining the asymptotic behavior as $T \to \infty$ with n fixed, it would not invalidate the use of $\widehat{W}$ for inference about β in cases where n and T go to infinity jointly.

In addition to being useful for performing inference about $\widehat{\beta}$, $\widehat{W}$ may also be used to test the specification of simple parametric models of the error process.[5] Such a test may be useful for a number of reasons. If a parametric model is correct, the estimates of the variance of $\widehat{\beta}$ based on this model will tend to behave better than the estimates obtained from $\widehat{W}$. In particular, parametric estimates of the variance of $\widehat{\beta}$ will often be considerably less variable and will typically converge faster than estimates made using $\widehat{W}$; and if the parametric model is deemed to be adequate, this model may be used to perform FGLS estimation. The FGLS estimator is asymptotically more efficient than the OLS estimator, and simulation evidence in Hansen (2004) suggests that the efficiency gain to using FGLS over OLS in serially correlated panel data may be substantial.

To define the specification test, called hereafter the heteroskedasticy-autocorrelation

[4]Note that this still allows arbitrary correlation and heteroskedasticity within individuals but restricts that the pattern is the same across individuals.

[5]The test considered is a straightforward generalization of the test proposed by White (1980) for heteroskedasticity and was suggested in the panel context by Kezdi (2002).

(HA) test, let $\widehat{W}(\widehat{\theta}) = \frac{1}{nT}\sum_{i=1}^{n} x_i'\Omega_i(\widehat{\theta})'x_i$ where $\widehat{\theta}$ are estimates of a finite set of parameters describing the disturbance process and $\Omega_i(\widehat{\theta})$ is the implied covariance matrix for individual i.[6] Define a test statistic

$$S^* = (nT)[\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))'\widehat{D}^{-}\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))] \tag{2.7}$$

where $\widehat{D}$ is a positive semi-definite weighting matrix that estimates the variance of $\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))$ and A^{-} is the generalized inverse of a matrix A.[7] In the following section, it will be shown that $S^* \xrightarrow{d} \chi^2_{k(k+1)/2}$ for $\widehat{D}$ defined below.

A natural choice for $\widehat{D}$ is

$$\widehat{D} = \frac{1}{nT}\sum_{i=1}^{n}[(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i - x_i'\Omega_i(\widehat{\theta})x_i))(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i - x_i'\Omega_i(\widehat{\theta})x_i))']. \tag{2.8}$$

Under asymptotics where $\{n,\ T\} \to \infty$ jointly, another potential choice for $\widehat{D}$ is

$$\widehat{V} = \frac{1}{nT}\sum_{i=1}^{n}[(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i - \widehat{W}))(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i - \widehat{W}))']. \tag{2.9}$$

That $\widehat{V}$ provides an estimator of the variance of $\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))$ follows from the fact that as $\{n,\ T\} \to \infty$, $\text{vec}(\widehat{W})$ is $\sqrt{n}$-consistent while $\text{vec}(\widehat{W}(\widehat{\theta}))$ will be $\sqrt{nT}$−consistent in many cases, so $\text{vec}(\widehat{W}(\widehat{\theta}))$ may be taken as a constant relative to $\text{vec}(\widehat{W})$. The difference in rates of convergence would arise, for example, in a fixed effects panel model where the errors follow an AR process with common AR coefficients across individuals. However, it is important

[6] Consistency and asymptotic normality of $\widehat{W}(\widehat{\theta})$ will generally follow from consistency and asymptotic normality of $\widehat{\theta}$. In particular, defining $W_i(\theta)$ as the derivative of W with respect to θ_i and letting θ be a $p \times 1$ vector, a Taylor series expansion of $W(\widehat{\theta})$ yields $W(\widehat{\theta}) = W(\theta) + \sum_{i=1}^{p} W_i(\bar{\theta})(\widehat{\theta} - \theta)$ where $\bar{\theta}$ is an intermediate value. As long as a uniform law of large numbers applies to $W_i(\theta)$, $W(\widehat{\theta}) - W(\theta)$ will inherit the properties of $\widehat{\theta} - \theta$. The problem is then reduced to finding an estimator of θ that is consistent and asymptotically normal with a mean zero asymptotic distribution. Finding such an estimator in fixed effects panel models with serial correlation and/or heteroskedasticity when $n \to \infty$ and $T/n \to \rho$ where $\rho < \infty$ is complicated, though there are estimators which exist. For example, Hansen (2004) provides an estimator of parameters of an AR(p) model which will have these properties under any asymptotic sequence, and Hahn and Kuersteiner (2002) and Hahn and Newey (2002) provide estimators that will be consistent, asymptotically normal, and asymptotically unbiased as long as $n/T^3 \to 0$. See also Nickell (1981), Macurdy (1982), Solon (1984), and Lancaster (2002).

[7] The test could alternatively be defined by only considering the $\frac{k(k+1)}{2}$ unique elements of $\widehat{W} - \widehat{W}(\widehat{\theta})$ and using the inverse of the implied covariance matrix. This test will be equivalent to the test outlined above.

to note that this will not always be the case. In particular, in random effects models, the estimator of the variance of the individual specific shock will converge at only a $\sqrt{n}$ rate, implying the same rate of convergence for both the robust and parametric estimators of the variance. In the following section, I outline the asymptotic properties of $\widehat{\beta}$, $\widehat{W}$, and $\widehat{V}$ from which the behavior of t^*, F^*, and S^* will follow. The properties of $\widehat{D}$, though not discussed, will generally be the same as those of $\widehat{V}$ under the different asymptotic sequences considered.

2.3 Asymptotic Properties of the Robust Covariance Matrix Estimator

To develop the asymptotic inference results, I impose the following conditions.

Assumption 4 $\{x_i, \epsilon_i\}$ *are independent across* i, *and* $\mathrm{E}[\epsilon_i \epsilon_i' | x_i] = \Omega_i$.

Assumption 5 $Q_{nT} = \mathrm{E}\left[\sum_{i=1}^{n} \frac{x_i' x_i}{nT}\right]$ *is uniformly positive definite with constant limit* Q.

Assumption 6 *Either (a)* $\mathrm{E}[\epsilon_i | x_i] = 0$ *or (b)* $\mathrm{E}[x_{it} \epsilon_{it}] = 0$.

Assumptions 4-6 are quite standard for panel data models. Assumption 4 imposes independence across individuals, ruling out cross-sectional correlation, but leaves the time series correlation unconstrained and allows general heterogeneity across individuals. Assumption 5 is a standard full rank condition, and the restriction that M_{nT} has a constant limit could be relaxed at the cost of more complicated notation. Assumption 6 imposes that one of two orthogonality conditions is satisfied. Assumption 6.b imposes that x_{it} and ϵ_{it} are uncorrelated and is weaker than the strict exogeneity imposed in Assumption 6.a. Assumption 6.a is stronger than necessary, but it simplifies the proof of asymptotic normality of $\widehat{W}$ and consistency of $\widehat{V}$. In addition, Assumption 6.a would typically be imposed in fixed effects models.

The first theorem, which is stated here for completeness, collects the properties of $\widehat{\beta}$ and $\widehat{W}$ in asymptotics where $n \to \infty$ with T fixed.

Theorem 1 *Suppose the data are generated by model (2.1), that Assumptions 4 and 5 are satisfied, and that $n \to \infty$ with T fixed.*

i. If Assumption 6.b holds and $\mathrm{E}|x_{ith}|^{4+\delta} < \Delta < \infty$ *and* $\mathrm{E}|\epsilon_{it}|^{4+\delta} < \Delta < \infty$ *for some* $\delta > 0$*, then*

$$\sqrt{nT}(\widehat{\beta} - \beta) \xrightarrow{d} Q^{-1}N(0, W = \lim_n \frac{1}{nT}\sum_{i=1}^{n}\mathrm{E}[x_i'\Omega_i x_i]),$$

and

$$\widehat{W} \xrightarrow{p} W.$$

ii. In addition, if Assumption 6.a holds and $\mathrm{E}|x_{ith}|^{8+\delta} < \Delta < \infty$ *and* $\mathrm{E}|\epsilon_{it}|^{8+\delta} < \Delta < \infty$ *for some* $\delta > 0$*, then*

$$\sqrt{nT}[\mathrm{vec}(\widehat{W} - W)] \xrightarrow{d}$$
$$N(0, V = \lim_n \frac{1}{nT}\sum_{i=1}^{n}\mathrm{E}[(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))']),$$

and

$$\widehat{V} \xrightarrow{p} V.$$

Remark 2.3.1 It follows from Theorem 1.i that the asymptotic variance of $\widehat{\beta}$ can be estimated using (2.4) since

$$\begin{aligned}\widehat{Avar}(\widehat{\beta}) &= (\sum_{i=1}^{n} x_i'x_i)^{-1}\sum_{i=1}^{n} x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i(\sum_{i=1}^{n} x_i'x_i)^{-1}\\ &= \frac{1}{nT}(\frac{1}{nT}\sum_{i=1}^{n} x_i'x_i)^{-1}\widehat{W}(\frac{1}{nT}\sum_{i=1}^{n} x_i'x_i)^{-1} = \frac{1}{nT}\widehat{Q}^{-1}\widehat{W}\widehat{Q}^{-1}\end{aligned}$$

where $\widehat{Q}^{-1}\widehat{W}\widehat{Q}^{-1} \xrightarrow{p} Q^{-1}WQ^{-1}$. It also follows immediately from the definitions of t^* and F^* in equations (2.5) and (2.6) and Theorem 1.i that, under the null hypothesis, $t^* \xrightarrow{d} N(0,1)$ and $F^* \xrightarrow{d} \chi_q^2$. Similarly, using Theorem 1.ii and assuming $\widehat{W}(\widehat{\theta})$ has properties similar to those of $\widehat{W}$, it will follow that the HA test statistic, S^*, formed using $\widehat{D}$ defined above converges in distribution to a $\chi^2_{k(k+1)/2}$ under the null hypothesis.

Theorem 1 verifies that $\widehat{\beta}$ and $\widehat{W}$ are consistent and asymptotically normal as $n \to \infty$ with T fixed without imposing any restrictions on the time series dimension. In the following results, I consider alternate asymptotic approximations under the assumption that both n and T are going to infinity. In these cases, consistency and asymptotic normality of suitably normalized versions of $\widehat{W}$ are established under weak conditions.

Theorem 2 *Suppose the data are generated by model (2.1), that Assumptions 4 and 5 are satisfied, and that* $\{n,\ T\} \to \infty$ *jointly.*

i. If Assumption 6.b holds and $\mathrm{E}|x_{ith}|^{4+\delta} < \Delta < \infty$ *and* $\mathrm{E}|\epsilon_{it}|^{4+\delta} < \Delta < \infty$ *for some* $\delta > 0$*, then*

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} Q^{-1}N(0, W = \lim_{n,T} \frac{1}{nT^2} \sum_{i=1}^{n} \mathrm{E}[x_i'\Omega_i x_i]),$$

and

$$\widehat{W}/T \xrightarrow{p} W.$$

ii. In addition, if Assumption 6.a holds and $\mathrm{E}|x_{ith}|^{8+\delta} < \Delta < \infty$ *and* $\mathrm{E}|\epsilon_{it}|^{8+\delta} < \Delta < \infty$ *for some* $\delta > 0$*, then*

$$\sqrt{n}[\mathrm{vec}(\widehat{W}/T - W)] \xrightarrow{d}$$

$$N(0, V = \lim_{n,T} \frac{1}{nT^4} \sum_{i=1}^{n} \mathrm{E}[(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))']),$$

and

$$\widehat{V}/T^3 \xrightarrow{p} V.$$

Remark 2.3.2 Theorem 2 verifies consistency and asymptotic normality of both $\widehat{\beta}$ and $\widehat{W}$ while imposing essentially no constraints on the time series dependence in the data. The large cross-section effectively allows the time series dimension to be ignored even when T is large. However, without constraints on the time series, $\widehat{\beta}$ is $\sqrt{n}$-consistent, not $\sqrt{nT}$-consistent. Intuitively, the slower rate of convergence is due to the fact that there may be little information contained in the time series since it is allowed to be arbitrarily dependent.

Remark 2.3.3 The fact that $\widehat{\beta}$ and $\widehat{W}$ are not $\sqrt{nT}-$consistent will not affect practical implementation of inference about $\widehat{\beta}$. In particular, the estimate of the asymptotic variance of $\widehat{\beta}$ based on equation (2.4) is

$$\begin{aligned}
\widehat{Avar}(\widehat{\beta}) &= (\sum_{i=1}^{n} x_i'x_i)^{-1} \sum_{i=1}^{n} x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i(\sum_{i=1}^{n} x_i'x_i)^{-1} \\
&= \frac{1}{n}(\frac{1}{nT}\sum_{i=1}^{n} x_i'x_i)^{-1}(\widehat{W}/T)(\frac{1}{nT}\sum_{i=1}^{n} x_i'x_i)^{-1} = \frac{1}{n}\widehat{Q}^{-1}(\widehat{W}/T)\widehat{Q}^{-1}
\end{aligned}$$

where $\widehat{Q}^{-1}(\widehat{W}/T)\widehat{Q}^{-1} \xrightarrow{p} Q^{-1}WQ^{-1}$. The t-statistic defined in equation (2.5) above may also be expressed as

$$\begin{aligned}
t^* &= \frac{\sqrt{nT}(R\widehat{\beta} - r)}{\sqrt{R\widehat{Q}^{-1}\widehat{W}\widehat{Q}^{-1}R'}} \\
&= \frac{\sqrt{n}(R\widehat{\beta} - r)}{\sqrt{R\widehat{Q}^{-1}(\widehat{W}/T)\widehat{Q}^{-1}R'}}
\end{aligned}$$

which converges in distribution to a $N(0,1)$ random variable under the null hypothesis, $R\beta = r$, by Theorem 2.i. Similarly, it follows that $F^* \xrightarrow{d} \chi_q^2$ under the null. Finally, the HA test statistic, S^*, defined above also satisfies

$$\begin{aligned}
S^* &= (nT)[\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))'\widehat{D}^{-}\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))] \\
&= n[\text{vec}(\widehat{W}/T - \widehat{W}(\widehat{\theta})/T)'(\widehat{D}/T^3)^{-}\text{vec}(\widehat{W}/T - \widehat{W}(\widehat{\theta})/T)],
\end{aligned}$$

which converges in distribution to a $\chi^2_{k(k+1)/2}$ under the conditions of the theorem and the additional assumption that $\widehat{W}(\widehat{\theta})$ behaves similarly to $\widehat{V}$.

Remark 2.3.4 It is important to note that the results presented in Theorem 2 are not interesting in the setting where the $\{j,k\}$ element of Ω_i becomes small when $|j-k|$ is large since in these circumstances $\frac{1}{nT^2}\sum_{i=1}^{n} \text{E}[x_i'\Omega_i x_i] \to 0$. Theorem 3 below presents results which are relevant in this case.

The previous theorem establishes the properties of $\widehat{\beta}$ and the robust variance matrix estimator as n and T go to infinity jointly without imposing restrictions on the time series dependence. While the result is interesting, there are many cases in which one might expect the time series dependence to diminish over time. In the following theorem, the properties of $\widehat{\beta}$ and $\widehat{W}$ are established under the assumption that the data are strong mixing in the time series dimension.

Theorem 3 *Suppose the data are generated by model (2.1), that Assumptions 4 and 5 are satisfied, and that* $\{n,\ T\} \to \infty$ *jointly.*

(i) If Assumption 6.b is satisfied, $\mathrm{E}|x_{ith}|^{r+\delta} < \Delta$ *and* $\mathrm{E}|\epsilon_{it}|^{r+\delta} < \Delta$ *for some* $\delta > 0$, *and* $\{x_{it}, \epsilon_{it}\}$ *is a strong mixing sequence in t with* α *of size* $-3r/(r-4)$ *for* $r > 4$,

$$\sqrt{nT}(\widehat{\beta} - \beta) \xrightarrow{d} Q^{-1}N(0, W = \lim_{n,T} \frac{1}{nT} \sum_{i=1}^{n} \mathrm{E}[x_i'\Omega_i x_i])$$

and

$$\widehat{W} - W \xrightarrow{p} 0.$$

(ii) In addition, if Assumption 6.a is satisfied, $\mathrm{E}|x_{ith}|^{r+\delta} < \Delta$ *and* $\mathrm{E}|\epsilon_{it}|^{r+\delta} < \Delta$ *for some* $\delta > 0$, *and* $\{x_{it}, \epsilon_{it}\}$ *is a strong mixing sequence in t with* α *of size* $-7r/(r-8)$ *for* $r > 8$,

$$\sqrt{n}[\mathrm{vec}(\widehat{W} - W)] \xrightarrow{d}$$
$$N(0, V = \lim_{n,T} \frac{1}{nT^2} \sum_{i=1}^{n} \mathrm{E}[(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))']),$$

and

$$\widehat{V}/T \xrightarrow{p} V.$$

Remark 2.3.5 Theorem 3 verifies consistency and asymptotic normality of both $\widehat{\beta}$ and $\widehat{W}$ under fairly conventional conditions on the time series dependence of the variables. The added restriction on the time series dependence allows estimation of β at the $\sqrt{nT}$-rate, which differs from the case above where $\widehat{\beta}$ is only $\sqrt{n}$-consistent. Intuitively, the increase in the rate of convergence is due to the fact that under the mixing conditions, the time series is more informative than in the case analyzed in Theorem 2.

Remark 2.3.6 It follows immediately from the conclusions of Theorem 3 and the definitions of $\widehat{Avar}(\widehat{\beta})$, t^*, and F^* in equations (2.4), (2.5), and (2.6) that $\widehat{Avar}(\widehat{\beta})$ is valid for estimating the asymptotic variance of $\widehat{\beta}$ and that $t^* \xrightarrow{d} N(0,1)$ and $F^* \xrightarrow{d} \chi^2_q$ under the null hypothesis. The HA test statistic, S^*, also satisfies

$$\begin{aligned} S^* &= (nT)[\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))'\widehat{D}^{-}\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))] \\ &= n[\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))'(\widehat{D}/T)^{-}\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))], \end{aligned}$$

which converges in distribution to a $\chi^2_{k(k+1)/2}$ under the conditions of the theorem and the assumption that $\widehat{D}$ behaves similarly to $\widehat{V}$. In this case, $\widehat{V}$ could also typically be used as the weighting matrix in forming S^* since it will often be the case that $\widehat{W}(\widehat{\theta})$ will be $\sqrt{nT}$-consistent while $\widehat{W}$ is $\sqrt{n}$-consistent.

Theorems 1-3 establish that conventional estimators of the asymptotic variance of $\widehat{\beta}$ and t and F statistics formed using $\widehat{W}$ have their usual properties as long as $n \to \infty$ regardless of the behavior of T. In addition, the results indicate that it is essentially only the size of n that matters for the asymptotic behavior of the estimators under these sequences. To complete the theoretical analysis, I present the asymptotic properties of $\widehat{W}$ as $T \to \infty$ with n fixed below.

Theorem 4 *Suppose the data are generated by model (2.1), that Assumptions 4, 5, and 6.b are satisfied, and that $T \to \infty$ with n fixed. If $\mathrm{E}|x_{ith}|^{r+\delta} < \Delta$, $\mathrm{E}|\epsilon_{it}|^{r+\delta} < \Delta$, and $\{x_{it}, \epsilon_{it}\}$ is a strong mixing sequence in t with α of size $-3r/(r-4)$ for $r > 4$, then*

$$\sqrt{nT}(\widehat{\beta} - \beta) \xrightarrow{d} Q^{-1}N(0, W), \quad x_i'x_i/nT - Q_i/n \xrightarrow{p} 0, \quad x_i'\epsilon_i/\sqrt{nT} \xrightarrow{d} N(0, W_i/n),$$

and

$$\begin{aligned} \widehat{W} \xrightarrow{d} U = \frac{1}{n}\sum_{i=1}^{n}(\Lambda_i B_i B_i' \Lambda_i - \Lambda_i B_i (\sum_{j=1}^{n} B_j' \Lambda_j)(\sum_{j=1}^{n} Q_j)^{-1} Q_i \\ - Q_i (\sum_{j=1}^{n} Q_j)^{-1} (\sum_{j=1}^{n} \Lambda_j B_j) B_i' \Lambda_i \\ + Q_i (\sum_{j=1}^{n} Q_j)^{-1} (\sum_{j=1}^{n} \Lambda_j B_j)(\sum_{j=1}^{n} B_j' \Lambda_j)(\sum_{j=1}^{n} Q_j)^{-1} Q_i, \end{aligned}$$

where $W_i = \lim_T \frac{1}{T}\mathrm{E}[x_i'\Omega_i x_i]$, $W = \lim_T \frac{1}{nT}\sum_i \mathrm{E}[x_i'\Omega_i x_i]$, $B_i \sim N(0, I_k)$ is a $k-$dimensional normal vector with $\mathrm{E}[B_i B_j'] = 0$ and $\Lambda_i = W_i^{1/2}$.

Remark 2.3.7 Theorem 4 verifies that $\widehat{W}$ is not consistent but does have a limiting distribution as $T \to \infty$ with n fixed. Unfortunately, the result here differs from results obtained in Phillips, Sun, and Jin (2003), Kiefer and Vogelsang (2002), Kiefer and Vogelsang (2003), and Vogelsang (2002) who consider HAC estimation in time series data without truncation in that how to construct asymptotically pivotal statistics from U is not immediately obvious. However, in one important special case, U is proportional to the true covariance matrix allowing construction of asymptotically pivotal tests.

Corollary 4.1 *Suppose the conditions of Theorem 4 are satisfied and that $Q_i = Q$ and $W_i = W$ for all i. Then*

$$\widehat{W} \xrightarrow{d} U = \frac{1}{n}\Lambda(\sum_{i=1}^{n} B_i B_i' - \frac{1}{n}\sum_{i=1}^{n} B_i \sum_{i=1}^{n} B_i')\Lambda$$

for B_i defined in Theorem 4 and $\Lambda = W^{1/2}$. Then, for testing the null hypothesis $H_0 : R\beta = r$ against the alternative $H_1 : R\beta \neq r$ for a $q \times k$ matrix R with rank q, the limiting distributions of the conventional Wald (F^*) and t-type (t^*) tests under H_0 are

$$\begin{aligned} F^* &= (nT)(R\widehat{\beta} - r)'[R\widehat{Q}^{-1}\widehat{W}\widehat{Q}^{-1}R']^{-1}(R\widehat{\beta} - r)) \\ &\xrightarrow{d} \widetilde{B}_{q,n}' \left[\frac{1}{n}(\sum_i B_{q,i}B_{q,i}' - \widetilde{B}_{q,n}\widetilde{B}_{q,n}')\right]^{-1} \widetilde{B}_{q,n}, \end{aligned} \tag{2.10}$$

and

$$\begin{aligned} t^* &= \frac{\sqrt{nT}(R\widehat{\beta} - r)}{\sqrt{R\widehat{Q}^{-1}\widehat{W}\widehat{Q}^{-1}R'}} \\ &\xrightarrow{d} \frac{\widetilde{B}_{1,n}}{\sqrt{\frac{1}{n}(\sum_i B_{1,i}^2 - \widetilde{B}_{1,n}^2)}} = \sqrt{\frac{n}{n-1}}t_{n-1}, \end{aligned} \tag{2.11}$$

where $B_{q,i} \sim N(0, I_q)$, $\widetilde{B}_{q,n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} B_{q,i}$, and t_{n-1} is a t distribution with $n-1$ degrees of freedom.

Corollary 4.1 gives the limiting distribution of $\widehat{W}$ as $T \to \infty$ under the additional restriction that $Q_i = Q$ and $W_i = W$ for all i. These restrictions would be satisfied when the data vectors for each individual $\{x_i, y_i\}$ are iid across i. While this is more restrictive than the condition imposed in Assumption 4, it still allows for quite general forms of conditional heteroskedasticity and does not impose any structure on the time series process within individuals.

The most interesting feature about the result in Corollary 4.1 is that under the conditions imposed, the limiting distribution of $\widehat{W}$ is proportional to the actual covariance matrix in the data. This allows construction of asymptotically pivotal statistics based on standard t and Wald tests as in Phillips, Sun, and Jin (2003), Kiefer and Vogelsang (2002), Kiefer and Vogelsang (2003), and Vogelsang (2002). This is particularly convenient in the panel case since the limiting distribution of the t-statistic is exactly $\sqrt{\frac{n}{n-1}}t_{n-1}$ where t_{n-1} denotes the t distribution with $n-1$ degrees of freedom.[8] It is also interesting that $\mathrm{E}U = (1 - \frac{1}{n})W$. This suggests normalizing the estimator $\widehat{W}$ by $\frac{n}{n-1}$ will result in an asymptotically unbiased estimator in asymptotics where $T \to \infty$ with n fixed and will likely reduce the finite-sample bias under asymptotics where $n \to \infty$. In addition, the t-statistic constructed based on the estimator defined by $\frac{n}{n-1}\widehat{W}$ will be asymptotically distributed as a t_{n-1} for which critical values are readily available.[9]

The conclusions of Corollary 4.1 suggest a simple procedure for testing hypotheses regarding regression coefficients which will be valid under any of the asymptotics considered. Using $\frac{n}{n-1}\widehat{W}$ and obtaining critical values from a t_{n-1} distribution will yield tests which are asymptotically valid regardless of the asymptotic sequence since the $t_{n-1} \to N(0,1)$ and $\frac{n}{n-1} \to 1$ as $n \to \infty$. Thus, this approach will yield valid tests under any of the asymptotics considered in the presence of quite general heteroskedasticity and serial correlation.[10]

Finally, it is worth noting that the maximum rank of $\widehat{W}$ will generally be $n-1$, which suggests that $\widehat{W}$ will be rank deficient when $k > n-1$. Since $\widehat{W}$ is supposed to estimate a full rank matrix, it seems likely that inference based on $\widehat{W}$ will perform poorly in these cases. Also, the above development ignores time effects, which will often be included in panel data models. Under T fixed, $n \to \infty$ asymptotics, the time effects can be included in the

[8]If $n = 1$, $\widehat{W}$ is identically equal to 0. In this case, it is easy to verify that U equals 0, though the results of Theorem 4 and Corollary 4.1 are obviously uninteresting in this case.

[9]This is essentially the normalization used in Stata's *cluster* command, which normalizes $\widehat{W}$ by $\frac{nT-1}{nT-k}\frac{n}{n-1}$, where the normalization is motivated as a finite-sample adjustment under the usual $n \to \infty$, T fixed asymptotics; see *Stata User's Guide Release 8* p. 275.

[10]This argument also applies to testing multiple parameters using F^*, though in that case, the n fixed asymptotic approximation is less convenient to work with.

covariate vector x_{it} and pose no additional complications. However, as $T \to \infty$, they also need to be considered separately from x and partialed out with the individual fixed effects. This partialing out will generally result in the presence of an $O(1/n)$ correlation between individuals. When n is large, this correlation should not matter, but in the fixed n, $T \to \infty$ case, it will invalidate the results. The effect of the presence of time effects was explored in a simulation study with the same design as that reported in the following section where each model was estimated including a full set of time fixed effects. The results, which are not reported below but are available upon request, show that tests based on $\widehat{W}$ are somewhat more size distorted than when no time effects are included for small n, but that this size distortion diminishes quickly as n increases.

2.4 Monte Carlo Evidence

The asymptotic results presented above suggest that tests based on the robust standard error estimates should have good properties regardless of the relative sizes of n and T. I report results from a simple simulation study used to assess the finite sample effectiveness of the robust covariance matrix estimator and tests based upon it below. Specifically, the simulation focuses on t-tests for regression coefficients and the HA test discussed above.

The Monte Carlo simulations are based on two different specifications: a "fixed effect" specification and a random effects specification. The fixed effect specification is

$$y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it},$$

where x_{it} is a scalar and α_i is an individual specific effect. The data generating process for the fixed effect specification allows for serial correlation in both x_{it} and ε_{it} and heteroskedasticity:

$$\begin{aligned}
x_{it} &= .5x_{it-1} + v_{it}, \quad v_{it} \sim N(0, .75), \\
\varepsilon_{it} &= \rho\varepsilon_{it-1} + \sqrt{a_0 + a_1 x_{it}^2} u_{it}, \quad u_{it} \sim N(0, 1 - \rho^2), \\
\alpha_i &\sim N(0, .5).
\end{aligned}$$

Data are simulated using four different values of ρ, $\rho \in \{0, .3, .6, .9\}$, in both the homoskedastic ($a_0 = 1, a_1 = 0$) and heteroskedastic ($a_0 = a_1 = .5$) cases, resulting in a total of eight distinct parameter settings. The models are estimated including x_{it} and a full set of individual specific fixed effects.[11]

The random effects specifications is

$$y_{it} = x_{it}'\beta + \epsilon_{it}$$

where x_{it} is a normally distributed scalar with $\mathrm{E}[x_{it}^2] = 1$ and $\mathrm{E}[x_{it_1}x_{it_2}] = .8$ for all $t_1 \neq t_2$. ϵ_{it} contains an individual specific random component and a random error term:

$$\begin{aligned} \epsilon_{it} &= \alpha_i + u_{it} \\ \alpha_i &\sim N(0, \rho) \\ u_{it} &\sim N(0, 1-\rho) \end{aligned}$$

Note that the random effects data generating process implies that $\mathrm{E}[\epsilon_{it_1}\epsilon_{it_2}] = \rho$ for $t_1 \neq t_2$. Three values of ρ are employed for the random effects specification: .3, .6, and .9. The model is estimated by regressing y_{it} on x_{it} and a constant.

The fixed effects model is commonly used in empirical work when panel data are available. The random effects specification is also widely used in the policy evaluation literature. In many policy evaluation studies, the covariate of interest is a policy variable that is highly correlated within aggregate cells, often with a correlation of one, which has led to the dominance of the random effects estimator in this context. For example, a researcher may desire to estimate the effect of classroom level policies on student-level micro data containing observations from multiple classrooms. In this setting, T indexes the number of students within each class, n indexes the number of classrooms, and α_i is a classroom specific random effect. The CCM estimator has been widely utilized in such situations in order to consistently estimate standard errors.[12]

The simulations are performed for various values of the cross-sectional (n) and time (T) dimensions. For example, to explore the properties of the tests and estimators for large

[11]Since α_i is uncorrelated with x_i, this model could be estimated using random effects. I chose to consider a different specification for the random effects estimates where the x_{it} were generated to more closely resemble covariates which appear in policy analysis studies.

[12]This is, in fact, one of the original motivations for the development of the CCM estimator, cf. Liang and Zeger (1986).

n and small T, $n = 200$ and $T = 10$ is used. To mimic a typical situation in state-level data from the CPS, $n = 50$, $T = 20$ is used. In total, nine combinations of n and T are simulated. For each of these nine combinations the eleven parameter settings (eight for the fixed effects specification and three for the random effects specification) are used for a total of 99 distinct 'cells'. 1,000 Monte Carlo simulations are performed for each cell. Each simulation estimates three types of standard errors for $\widehat{\beta}$: unadjusted OLS standard errors, $\hat{s}_{OLS}$, CCM standard errors, $\hat{s}_{CLUS}$, and standard errors consistent with an AR(1) process, $\hat{s}_{AR(1)}$.[13] For the random effects specification, standard errors consistent with random effects, $\hat{s}_{RE}$, are substituted for $\hat{s}_{AR(1)}$.[14] $\hat{s}_{CLUS}$ is consistent for all parameter settings. $\hat{s}_{OLS}$ is consistent only in the iid case (the homoskedastic data generating process with $\rho = 0$). $\hat{s}_{AR(1)}$ is consistent in all homoscedastic data generating processes, and $\hat{s}_{RE}$ is consistent in all models for which it is reported. In all cases, the CCM estimator is computed using the normalization implied by $T \to \infty$ with n fixed asymptotics; that is, the CCM estimator is computed as $\frac{n}{n-1}\widehat{W}$ for $\widehat{W}$ defined in equation (2.3).

Tables 1 - 9 present the results of the Monte Carlo study, where each table corresponds to a different $\{n,\ T\}$ combination.[15] In each table, Panel A presents the fixed effects results for the homoskedastic and heteroskedastic cases, while Panel B presents the random effects results. Column (1) presents t-test rejection rates for 5% level tests based on OLS, CCM, and AR(1) standard errors. The critical values for tests based on OLS and AR(1) errors are taken from a t_{nT-n-1} distribution, and the critical values for tests based on clustered standard errors are taken from a t_{n-1} distribution. Columns (2) and (3) present the mean and standard deviation of the estimated standard errors respectively. Column (4) presents the standard deviation of the $\widehat{\beta}$'s. The difference between columns (2) and (4) is therefore the bias of the estimated standard errors. Finally, column (5) presents the rejection rates for the HA test described above which tests the null hypothesis that both the CCM estimator

[13] $\hat{s}_{AR(1)}$ imposes the parametric structure implied by an AR(1) process. The ρ parameter is estimated from the OLS residuals using the procedure described in Hansen (2004) which consistently estimates AR parameters in fixed effects panel models. The standard errors are then computed as $(X'X)^{-1}X'\ \Omega(\widehat{\rho})X(X'X)^{-1}$ where $\Omega(\widehat{\rho})$ is the covariance matrix implied by an AR(1) process.

[14] $\hat{s}_{RE}$ is estimated in a manner analogous to $\hat{s}_{AR(1)}$ where the covariance parameters are estimated in the usual manner from the OLS and within residuals.

[15] Tables 1-9 correspond to $\{n,\ T\} = \{10, 10\}$, $\{n,\ T\} = \{10, 50\}$, $\{n,\ T\} = \{10, 200\}$, $\{n,\ T\} = \{50, 10\}$, $\{n,\ T\} = \{50, 20\}$, $\{n,\ T\} = \{50, 50\}$, $\{n,\ T\} = \{50, 200\}$, $\{n,\ T\} = \{200, 10\}$, and $\{n,\ T\} = \{200, 50\}$ respectively.

and the parametric estimator are consistent.

As expected, tests based on $\hat{s}_{OLS}$ and $\hat{s}_{AR(1)}$ perform well in the cases where the assumed model is consistent with the data across the full range of n and T combinations. The results are also consistent with the asymptotic theory, clearly illustrating the $\sqrt{nT}$-consistency of $\widehat{\beta}$ and $\widehat{W}$ with the bias of $\widehat{W}$ and the variance of both $\widehat{\beta}$ and $\widehat{W}$ decreasing as either n or T increases. Of course, when the assumed parametric model is inconsistent with the data, tests based on parametric standard errors suffer from size distortions and the standard error estimates are biased. The RE tests have the correct size for moderate and large n, but not for small n (i.e. $n = 10$); and as indicated by the asymptotic theory, the T dimension has no apparent impact on the size of RE based tests or the overall performance of the RE estimates.

Tests based on the CCM estimator have approximately correct size across all combinations of n and T and all models of the disturbances considered in the fixed effect specification. The estimator does, however, display a moderate bias in the small n case; this bias does not translate into a large size distortion due to the use of the t_{n-1} distribution to obtain the critical values. While the clustered standard errors perform well in terms of size of tests and reasonably well in terms of bias, the simulations reveal that a potential weakness of the clustered estimator is a relatively high variance. The CCM estimates have a substantially higher standard deviation than the other estimators and this difference, in percentage terms, increases with T. This behavior is consistent with the $\sqrt{n}$-consistency of the estimator and does suggest that if a parametric estimator is available, it may have better properties for estimating the variance of $\widehat{\beta}$.

The clustered estimator performs less well in the random effects specification. For small n, tests based on the CCM estimator suffer from a substantial size distortion for all values of T. For moderate to large values of n, the tests have the correct size, and the overall performance does not appear to depend on T. In addition, the variance of $\widehat{\beta}$ does not appear to decrease as T increases. These results are consistent with the lack of $\sqrt{nT}$-consistency in this case.[16]

The performance of the HA test is much less robust than that of t-tests based on clustered standard errors. For small n, the tests are badly size distorted and have essentially no power

[16]The inconsistency of $\widehat{\beta}$ when T increases with n fixed in differences-in-differences and policy evaluation studies has also been discussed in Donald and Lang (2001).

against any alternative hypotheses. As n and T grow, the test performance improves. With $n = 50$, the test remains size distorted, but it does have some power against alternatives that increases as T increases. Table 9, which displays results for $n = 200$ and $T = 50$ suggests the test has good properties in cases of both large n and T. The HA test also performs poorly for the random effects specification for small n. However, for moderate or large n, the test has both the correct size and good power.

Overall, the simulation results support the use of clustered standard errors for performing inference on regression coefficient estimates in serially correlated panel data, though they also suggest care should be taken if n is small and one suspects a "random effects" structure. The poor performance of $\widehat{W}$ in "random effects" models with small n is already well-known; see for example Bell and McCaffrey (2002) who also suggest a bias reduction for $\widehat{W}$ in this case. However, that the estimator does quite well even for small n in the serially correlated case where the errors are mixing is somewhat surprising and is a new result which is suggested by the asymptotic analysis of the previous section. The simulation results confirm the asymptotic results, suggesting that the clustered standard errors are consistent as long as $n \to \infty$ and that they are not sensitive to the size of n relative to T. The chief drawback of the CCM estimates is that the robustness comes at the cost of increasing the variance of the standard error estimate relative to that of standard errors estimated through more parsimonious models.

The HA test offers one simple information based criterion for choosing between the CCM estimator and a simple parametric model of the error process. However, the simulation evidence regarding its usefulness is mixed. In particular, the properties of the test are poor in small sample settings where there is likely to be the largest gain to using a parsimonious model. However, in moderate sized samples, the test performs reasonably well, and there still may be gains to using a simple parametric model in these cases.

2.5 Conclusion

This paper explores the asymptotic behavior of the robust covariance matrix estimator of Arellano (1987). It extends the usual analysis performed under asymptotics where $n \to \infty$ with T fixed to cases where n and T go to infinity jointly, considering both non-mixing and mixing cases, and to the case where $T \to \infty$ with n fixed. The limiting behavior of the OLS estimator, $\widehat{\beta}$, in each case is different. However, the analysis shows that the conventional estimator of the asymptotic variance and the usual t and F statistics have the same properties regardless of the behavior of the time series as long as $n \to \infty$. In addition, when $T \to \infty$ with n fixed and the data satisfy mixing conditions and an iid assumption across individuals, the usual t and F statistics can be used for inference despite the fact that the robust covariance matrix estimator is not consistent but converges in distribution to a limiting random variable. In this case, it is shown that the t statistic constructed using $\frac{n}{n-1}$ times the estimator of Arellano (1987) is asymptotically t_{n-1}, suggesting the use of $\frac{n}{n-1}$ times the estimator of Arellano (1987) and critical values obtained from a t_{n-1} in all cases. The use of this procedure is also supported in a short simulation experiment, which verifies that it produces tests with approximately correct size regardless of the relative size of n and T in cases where the time series correlation between observations diminishes as the distance between observations increases. The simulations also verify that tests based on the robust standard errors are consistent as n increases regardless of the relative size of n and T even in cases when the data are equicorrelated.

2.6 Appendix 1. Preliminaries

Throughout the appendix, let $\|A\| = [\text{trace}(A'A)]^{1/2}$ be the Euclidean norm of a matrix A, and let $\sum_i = \sum_{i=1}^n$, $\sum_t = \sum_{t=1}^T$, and $\sum_h = \sum_{h=1}^k$. Repeated use will be made of the following simple results, which are stated here for convenience.

Lemma 2.6.1 *For matrices A and B,* $\mathrm{E}\|A \otimes B\|^r \leq (\mathrm{E}\|A\|^{2r}\mathrm{E}\|B\|^{2r})^{1/2}$.

Proof. $\mathrm{E}\|A \otimes B\|^r = \mathrm{E}\{[\text{trace}(AA' \otimes BB')]^{r/2}\} = \mathrm{E}\{[\text{trace}(AA')]^{r/2}[\text{trace}(BB')]^{r/2}\} = \mathrm{E}(\|A\|^r\|B\|^r)$
$\leq (\mathrm{E}\|A\|^{2r}\mathrm{E}\|B\|^{2r})^{1/2}$ where the equalities follow from the definition of $\|A\|$ and properties of the Kronecker product and the inequality results from the Cauchy-Schwarz inequality. ■

Lemma 2.6.2 *Suppose $\{Z_{i,T}\}$ are independent across i for all T with $\mathrm{E}[Z_{i,T}] = \mu_{i,T}$ and $\mathrm{E}|Z_{i,T}|^{1+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i, T. Then $\frac{1}{n}\sum_{i=1}^n(Z_{i,T} - \mu_{i,T}) \xrightarrow{p} 0$ as $\{n,\ T\} \to \infty$ jointly.*

Proof. The proof follows from standard arguments, cf. Chung (2001) Chapter 5. Details are given in Appendix 2.12. ■

Lemma 2.6.3 *For $k \times 1$ vectors $Z_{i,T}$, suppose $\{Z_{i,T}\}$ are independent across i for all T with $\mathrm{E}[Z_{i,T}] = 0$, $\mathrm{E}[Z_{i,T}Z_{i,T}'] = \Omega_{i,T}$, and $\mathrm{E}\|Z_{i,T}\|^{2+\delta} < \Delta < \infty$ for some $\delta > 0$. Assume $\Omega = \lim_{n,T} \frac{1}{n}\sum_{i=1}^{n} \Omega_{i,T}$ is positive definite with minimum eigenvalue $\lambda_{\min} > 0$. Then $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,T} \xrightarrow{d} N(0, W)$ as $\{n,\ T\} \to \infty$ jointly.*

Proof. The result follows from verifying the Lindeberg condition of Theorem 2 in Phillips and Moon (1999) using an argument similar to that used in the proof of Theorem 3 in Phillips and Moon (1999). Details are given in Appendix 2.12. ■

The final lemma simply restates Doukhan (1994) Theorem 2 with a slight change of notation. Its proof may be found in Doukhan (1994) p. 25-30.

Lemma 2.6.4 *Let $\{z_t\}$ be a strong mixing sequence with $\mathrm{E}[z_t] = 0$, $\mathrm{E}\|z_t\|^{\tau+\epsilon} < \Delta < \infty$, and mixing coefficient $\alpha(m)$ of size $\frac{(1-c)r}{r-c}$ where $c \in 2\mathbb{N}$, $c \geq \tau$, and $r > c$. Then there is a constant C depending only on τ and $\alpha(m)$ such that $\mathrm{E}|\sum_{t=1}^{T} y_t|^{\tau} \leq CD(\tau, \epsilon, T)$ with $D(\tau, \epsilon, T)$ defined in Doukhan (1994) and satisfying $D(\tau, \epsilon, T) = O(T)$ if $\tau \leq 2$ and $D(\tau, \epsilon, T) = O(T^{\tau/2})$ if $\tau > 2$.*

Finally note

$$
\begin{aligned}
\widehat{W} &= \frac{1}{nT}\sum_i x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i \\
&= \frac{1}{nT}\sum_i x_i'\epsilon_i\epsilon_i'x_i && \text{(O.1)} \\
&\quad - \frac{1}{nT}\sum_i x_i'x_i(\widehat{\beta} - \beta)\epsilon_i'x_i && \text{(O.2)} \\
&\quad - \frac{1}{nT}\sum_i x_i'\epsilon_i(\widehat{\beta} - \beta)'x_i'x_i && \text{(O.3)} \\
&\quad + \frac{1}{nT}\sum_i x_i'x_i(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)'x_i'x_i, && \text{(O.4)}
\end{aligned}
$$

and

$$\begin{aligned}\widehat{V} &= \frac{1}{nT}\sum_i[(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i - \widehat{W}))(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i - \widehat{W}))'] \\ &= \frac{1}{nT}\sum_i(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i))(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i))' - \frac{1}{nT}\sum_i(\text{vec}(\widehat{W}))\frac{1}{nT}\sum_i(\text{vec}(\widehat{W}))'\end{aligned}$$

where

$$\frac{1}{nT}\sum_i(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i))(\text{vec}(x_i'\widehat{\epsilon}_i\widehat{\epsilon}_i'x_i))'$$

$$= \frac{1}{nT}\sum_i(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))' \tag{V.1}$$

$$- \frac{1}{nT}\sum_i(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))(\text{vec}(x_i'\epsilon_i(\widehat{\beta}-\beta)'x_i'x_i))' \tag{V.2}$$

$$- \frac{1}{nT}\sum_i(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)\epsilon_i'x_i))' \tag{V.3}$$

$$+ \frac{1}{nT}\sum_i(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'x_i'x_i))' \tag{V.4}$$

$$- \frac{1}{nT}\sum_i(\text{vec}(x_i'\epsilon_i(\widehat{\beta}-\beta)'x_i'x_i))(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))' \tag{V.5}$$

$$+ \frac{1}{nT}\sum_i(\text{vec}(x_i'\epsilon_i(\widehat{\beta}-\beta)'x_i'x_i))(\text{vec}(x_i'\epsilon_i(\widehat{\beta}-\beta)'x_i'x_i))' \tag{V.6}$$

$$+ \frac{1}{nT}\sum_i(\text{vec}(x_i'\epsilon_i(\widehat{\beta}-\beta)'x_i'x_i))(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)\epsilon_i'x_i))' \tag{V.7}$$

$$- \frac{1}{nT}\sum_i(\text{vec}(x_i'\epsilon_i(\widehat{\beta}-\beta)'x_i'x_i))(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'x_i'x_i))' \tag{V.8}$$

$$- \frac{1}{nT}\sum_i(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)\epsilon_i'x_i))(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))' \tag{V.9}$$

$$+ \frac{1}{nT}\sum_i(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)\epsilon_i'x_i))(\text{vec}(x_i'\epsilon_i(\widehat{\beta}-\beta)'x_i'x_i))' \tag{V.10}$$

$$+ \frac{1}{nT}\sum_i(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)\epsilon_i'x_i))(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)\epsilon_i'x_i))' \tag{V.11}$$

$$- \frac{1}{nT}\sum_i(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)\epsilon_i'x_i))(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'x_i'x_i))' \tag{V.12}$$

$$+ \frac{1}{nT}\sum_i(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'x_i'x_i))(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))' \tag{V.13}$$

$$-\frac{1}{nT}\sum_i(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'x_i'x_i))(\text{vec}(x_i'\epsilon_i(\widehat{\beta}-\beta)'x_i'x_i))' \qquad (V.14)$$

$$-\frac{1}{nT}\sum_i(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'x_i'x_i))(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)\epsilon_i'x_i))' \qquad (V.15)$$

$$+\frac{1}{nT}\sum_i(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'x_i'x_i))(\text{vec}(x_i'x_i(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'x_i'x_i))'. \qquad (V.16)$$

Theorems 1-4 then follow by examining the properties of equations (O.1)-(O.4) and (V.1)-(V.16).

2.7 Appendix 2. Proof of Theorem 1

(i) $\widehat{\beta}-\beta \xrightarrow{p} 0$ and $\sqrt{nT}(\widehat{\beta}-\beta) \xrightarrow{d} Q^{-1}N(0, W = \lim_n \frac{1}{nT}\sum_{i=1}^n \text{E}[x_i'\Omega_i x_i])$ follow immediately under the conditions of Theorem 1 from the Markov LLN and the Liapounov CLT.

We also have that $\text{E}\|x_i'x_i/T\|^{2+2\delta} \overset{(a)}{\leq} \text{E}\|x_i/\sqrt{T}\|^{4+4\delta} \overset{(b)}{\leq} \frac{[\sum_t\sum_h(\text{E}|x_{ith}^2|^{2+2\delta})^{1/(2+2\delta)}]^{2+2\delta}}{T^{2+2\delta}} \overset{(c)}{<} \frac{(kT)^{2+2\delta}\Delta}{T^{2+2\delta}}$, where (a) is by the Cauchy-Schwarz inequality, (b) follows from the definition of $\|A\|$ and Minkowski's inequality, and (c) is by $\text{E}|x_{ith}|^{4+\delta} < \Delta$. Making use of $\text{E}|\epsilon_{ith}|^{4+\delta} < \Delta$, $\text{E}\|x_i'\epsilon_i/T\|^{2+2\delta} < k^{2+2\delta}\Delta$ follows similarly. Noting that $\text{E}[x_i'x_i \otimes \epsilon_i'x_i] = 0$ by Assumption 6.a, it follows that terms (O.2)-(O.4) of $\widehat{W}$ are $o_p(n^{-1/2})$ by the Markov LLN.[17] The Markov LLN also yields $\frac{1}{nT}\sum_i x_i'\epsilon_i\epsilon_i'x_i \xrightarrow{p} W$, which implies $\widehat{W} \xrightarrow{p} W$.

(ii) In addition, under $\text{E}|x_{ith}|^{8+\delta} < \Delta$ and $\text{E}|\epsilon_{ith}|^{8+\delta} < \Delta$, $\text{E}\|x_i'\epsilon_i/T\|^{4+4\delta} < k^{4+4\delta}\Delta$ and $\text{E}\|x_i'x_i/T\|^{4+4\delta} < k^{4+4\delta}\Delta$ follow by an argument similar to that used to show $\text{E}\|x_i'x_i/T\|^{2+2\delta} < k^{2+2\delta}\Delta$. Then, using that (O.2)-(O.4) of $\widehat{W}$ are $o_p(n^{-1/2})$, it follows that

$$\sqrt{nT}[\text{vec}(\widehat{W}-W)] = \sqrt{nT}[\text{vec}(\frac{1}{nT}\sum_i x_i'\epsilon_i\epsilon_i'x_i - W)] + o_p(1) \xrightarrow{d} N(0,V)$$

where $V = \lim_n \frac{1}{nT}\sum_{i=1}^n \text{E}[(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))']$ by the Liapounov CLT.

It is also straightforward to show that $\widehat{\beta}-\beta \xrightarrow{p} 0$, $\text{E}\|x_i'\epsilon_i/T\|^{4+4\delta} < k^{4+4\delta}\Delta$, and $\text{E}\|x_i'x_i/T\|^{4+4\delta} < k^{4+4\delta}\Delta$ imply that terms (V.2)-(V.16) of $\widehat{V}$ are $o_p(1)$ and that $\text{E}\|x_i'\epsilon_i/T\|^{4+4\delta}$

[17]Under Assumption 6.b, (O.2)-(O.4) will be $o_p(1)$ and consistency of $\widehat{W}$ follows.

$< k^{4+4\delta}\Delta$ implies

$$\frac{1}{nT}\sum_i[(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))' - \text{E}[(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i))']] \xrightarrow{p} 0.$$

We then have $\widehat{V} - V \xrightarrow{p} 0$. ■

2.8 Appendix 3. Proof of Theorem 2

(i) $\text{E}\|x_i'x_i/T\|^{2+2\delta} \overset{(a)}{\leq} \text{E}\|x_i/\sqrt{T}\|^{4+4\delta} \overset{(b)}{\leq} \frac{[\sum_t\sum_h(\text{E}|x_{ith}^2|^{2+2\delta})^{1/(2+2\delta)}]^{2+2\delta}}{T^{2+2\delta}} \overset{(c)}{<} \frac{(kT)^{2+2\delta}\Delta}{T^{2+2\delta}}$, where (a) is by the Cauchy-Schwarz inequality, (b) follows from the definition of $\|A\|$ and Minkowski's inequality, and (c) is by $\text{E}|x_{ith}|^{4+\delta} < \Delta$. Making use of $\text{E}|\epsilon_{ith}|^{4+\delta} < \Delta$, $\text{E}\|x_i'\epsilon_i/T\|^{2+2\delta} < k^{2+2\delta}\Delta$ follows similarly. Then Lemma 2.6.2 gives $\frac{1}{n}\sum_i x_i'x_i/T \xrightarrow{p} Q$ and $\frac{1}{n}\sum_i x_i'\epsilon/T \xrightarrow{p} 0$ as $\{n,\ T\} \to \infty$ jointly, so $\widehat{\beta} - \beta \xrightarrow{p} 0$. In addition, since $\text{E}\|x_i'\epsilon_i/T\|^{2+2\delta} < k^{2+2\delta}\Delta$, we have that $\frac{1}{\sqrt{n}}\sum_i x_i'\epsilon/T \xrightarrow{d} N(0, W = \lim_{n,T}\frac{1}{nT^2}\sum_{i=1}^n \text{E}[x_i'\Omega_i x_i])$ by Lemma 2.6.3, so $\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} Q^{-1}N(0, W)$.

Now consider term (O.2) of $\widehat{W}/T$:

$$\text{vec}(\frac{1}{nT^2}\sum_i x_i'x_i(\widehat{\beta} - \beta)\epsilon_i'x_i) = \frac{1}{\sqrt{n}}(\frac{1}{nT^2}\sum_i x_i'\epsilon_i \otimes x_i'x_i)\sqrt{n}(\widehat{\beta} - \beta) = \frac{1}{\sqrt{n}}O_p(1)O_p(1)$$

by Lemma 2.6.2 since $\text{E}\|x_i'\epsilon_i/T \otimes x_i'x_i/T\|^{1+\delta} \leq (\text{E}\|x_i'\epsilon_i/T\|^{2+2\delta}\text{E}\|x_i'x_i/T\|^{2+2\delta})^{1/2} < C$ by Lemma 2.6.1 and the argument in the preceding paragraph. That (O.3) is $O_p(1/\sqrt{n})$ and (O.4) is $O_p(1/n)$ follow similarly. Finally, $\text{E}\|x_i'\epsilon_i\epsilon_i'x_i/T^2\|^{1+\delta} \leq \text{E}\|x_i'\epsilon_i/T\|^{2+2\delta} < C$ where the first inequality follows from the Cauchy-Schwarz inequality and the second inequality is proven above. Hence, by Lemma 2.6.2 $\frac{1}{nT^2}\sum_i \text{vec}(x_i'\epsilon_i\epsilon_i'x_i - \text{E}[x_i'\epsilon_i\epsilon_i'x_i]) = o_p(1)$. It follows immediately that $\widehat{W}/T - W \xrightarrow{p} 0$.

(ii) Under Assumption 6.a, we also have that

$$\frac{1}{\sqrt{n}}(\frac{1}{nT^2}\sum_i x_i'\epsilon_i \otimes x_i'x_i)\sqrt{n}(\widehat{\beta} - \beta) = \frac{1}{\sqrt{n}}o_p(1)O_p(1) = o_p(\frac{1}{\sqrt{n}})$$

since $E[x_i'\epsilon_i \otimes x_i'x_i] = 0$. Similarly, $\frac{1}{\sqrt{n}}(\frac{1}{nT^2}\sum_i x_i'x_i \otimes \epsilon_i'x_i) = o_p(\frac{1}{\sqrt{n}})$. It follows from Lemma 2.6.3 that $\sqrt{n}(\widehat{W}/T - W) = \sqrt{n}(\frac{1}{nT^2}\sum_i \text{vec}(x_i'\epsilon_i\epsilon_i'x_i - \text{E}[x_i'\epsilon_i\epsilon_i'x_i])) + o_p(1) \xrightarrow{d} N(0, V)$ where

$$V = \lim_{n,T}\frac{1}{nT^4}\sum_{i=1}^n \text{E}[(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))(\text{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))']$$

since $\mathrm{E}\|x_i'\epsilon_i\epsilon_i'x_i/T^2\|^{2+\delta} \le \mathrm{E}\|x_i'\epsilon_i/T\|^{4+2\delta} < C$ by an argument similar to that used to show $\mathrm{E}\|x_i'\epsilon_i/T\|^{2+2\delta} < k^{2+2\delta}\Delta$ as long as $\mathrm{E}|x_{ith}|^{8+\delta} < \Delta$ and $\mathrm{E}|\epsilon_{ith}|^{8+\delta} < \Delta$.

To show $\widehat{V}/T^3 - V \xrightarrow{p} 0$, consider

$$\mathrm{E}\|\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i/T^2)\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i/T^2)'\|^{1+\delta} \le \mathrm{E}\|x_i'\epsilon_i/T\|^{4+4\delta} < C$$

where the first inequality is by repeated application of the Cauchy-Schwarz inequality and the second is by an argument similar to that used to show $\mathrm{E}\|x_i'\epsilon_i/T\|^{2+2\delta} < k^{2+2\delta}\Delta$ which holds if $\mathrm{E}|x_{ith}|^{8+\delta} < \Delta$ and $\mathrm{E}|\epsilon_{ith}|^{8+\delta} < \Delta$. It then follows by Lemma 2.6.2 that

$$\frac{1}{nT^4}\sum_i[\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i)\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i)' - \mathrm{E}[\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i)\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i)']] \xrightarrow{p} 0.$$

Turning to (V.2), we have

$$\begin{aligned}\mathrm{vec}[(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i/T^2))(\mathrm{vec}(x_i'\epsilon_i(\widehat{\beta}-\beta)'x_i'x_i/T^2))'] \\ = [(x_i'\epsilon_i/T \otimes x_i'\epsilon_i/T) \otimes (x_i'x_i/T \otimes \epsilon_i'x_i/T)]\mathrm{vec}(\widehat{\beta}-\beta).\end{aligned}$$

$\mathrm{E}\|(x_i'\epsilon_i/T \otimes x_i'\epsilon_i/T) \otimes (x_i'x_i/T \otimes \epsilon_i'x_i/T)\|^{1+\delta} \le [(\mathrm{E}\|x_i'\epsilon_i/T\|^{4+4\delta})^3\mathrm{E}\|x_i'x_i\|^{4+4\delta}]^{1/4} < C$ where the first inequality is by repeated application of Lemma 2.6.1 and the second inequality is by the moment conditions. It then follows from Lemma 2.6.2 that $\frac{1}{n}\sum_i(x_i'\epsilon_i/T \otimes x_i'\epsilon_i/T) \otimes (x_i'x_i/T \otimes \epsilon_i'x_i/T) = O_p(1)$, so (V.2) is $o_p(1)$. Using similar arguments, it is also straightforward to show that terms (V.3)-(V.16) are $o_p(1)$, and the conclusion follows. ■

2.9 Appendix 4. Proof of Theorem 3

(i) $\mathrm{E}\|x_i'x_i/T\|^{2+2\delta} \overset{(a)}{\le} \mathrm{E}\|x_i/\sqrt{T}\|^{4+4\delta} \overset{(b)}{\le} \frac{[\sum_t\sum_h(\mathrm{E}|x_{ith}^2|^{2+2\delta})^{1/(2+2\delta)}]^{2+2\delta}}{T^{2+2\delta}} \overset{(c)}{<} \frac{(kT)^{2+2\delta}\Delta}{T^{2+2\delta}}$, where (a) is by the Cauchy-Schwarz inequality, (b) by definition of $\|A\|$ and Minkowski's inequality, and (c) by $\mathrm{E}|x_{ith}|^{r+\delta} < \Delta$. Also, $\mathrm{E}\|x_i'\epsilon_i/\sqrt{T}\|^{2+2\delta} < C$ by Lemma 2.6.4, $\mathrm{E}|\epsilon_{it}|^{r+\delta} < \Delta$, and the mixing condition that α is of size $-3r/(r-4)$ for $r > 4$. It follows by Lemmas 2.6.2 and 2.6.3 that $\sqrt{nT}(\widehat{\beta}-\beta) \xrightarrow{d} Q^{-1}N(0, W = \lim_{n,T}\frac{1}{nT}\sum_{i=1}^n \mathrm{E}[x_i'\Xi_i x_i])$ as $\{n,T\} \to \infty$.

Now consider term (O.2) of $\widehat{W}$:

$$\mathrm{vec}(\frac{1}{nT}\sum_i x_i'x_i(\widehat{\beta}-\beta)\epsilon_i'x_i) = (\frac{1}{(nT)^{3/2}}\sum_i x_i'\epsilon_i \otimes x_i'x_i)\sqrt{nT}(\widehat{\beta}-\beta) = \frac{1}{\sqrt{n}}o_p(1)O_p(1)$$

by Lemma 2.6.2 since $\mathrm{E}[x_i'\epsilon_i \otimes x_i'x_i] = 0$ by Assumption 6.a and $\mathrm{E}\|x_i'\epsilon_i/\sqrt{T} \otimes x_i'x_i/T\|^{1+\delta} \leq (\mathrm{E}\|x_i'\epsilon_i/\sqrt{T}\|^{2+2\delta}\mathrm{E}\|x_i'x_i/T\|^{2+2\delta})^{1/2} < C$ by Lemma 2.6.1 and the argument in the preceding paragraph. That (O.3) is $o_p(1/\sqrt{n})$ and (O.4) is $O_p(1/n)$ follow similarly.[18] Finally, $\mathrm{E}\|x_i'\epsilon_i\epsilon_i'x_i/T\|^{1+\delta}\| \leq \mathrm{E}\|x_i'\epsilon_i/\sqrt{T}\|^{2+2\delta} < C$ by the Cauchy-Schwarz inequality and the preceding argument. It then follows that $\widehat{W} - W \xrightarrow{p} 0$ by Lemma 2.6.2.

(ii) In addition, under the mixing condition that α is of size $-7r/(r-8)$ for $r > 8$, $\mathrm{E}|x_{ith}|^{r+\delta} < \Delta$, and $\mathrm{E}|\epsilon_{ith}|^{r+\delta} < \Delta$, $\mathrm{E}\|x_i'\epsilon_i/\sqrt{T}\|^{4+4\delta} < C$ follows by an argument similar to that used to show $\mathrm{E}\|x_i'\epsilon_i/\sqrt{T}\|^{2+2\delta} < C$. Then, since $\mathrm{E}\|x_i'\epsilon_i\epsilon_i'x_i/T\|^{2+\delta} \leq \mathrm{E}\|x_i'\epsilon_i/\sqrt{T}\|^{4+4\delta} < C$ and using that (O.2)-(O.4) of $\widehat{W}$ are $o_p(n^{-1/2})$, the second conclusion follows from Lemma 2.6.3 since

$$\sqrt{n}[\mathrm{vec}(\widehat{W} - W)] = \sqrt{n}[\mathrm{vec}(\frac{1}{nT}\sum_i x_i'\epsilon_i\epsilon_i'x_i - W)] + o_p(1) \xrightarrow{d} N(0, V)$$

where

$$V = \lim_{n,T} \frac{1}{nT^2}\sum_{i=1}^{n} \mathrm{E}[(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i - W))']$$

.

To show $\widehat{V}/T - V \xrightarrow{p} 0$, consider

$$\mathrm{E}\|\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i/T)\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i/T)'\|^{1+\delta} \leq \mathrm{E}\|x_i'\epsilon_i/\sqrt{T}\|^{4+4\delta} < C$$

where the first inequality is by repeated application of the Cauchy-Schwarz inequality and the second is by an argument similar to that used above which holds if $\mathrm{E}|\epsilon_{it}|^{8+\delta} < \Delta < \infty$, $\mathrm{E}|x_{ith}|^{8+\delta} < \Delta < \infty$, and the strong mixing coefficient α is of size $-7r/(r-8)$ for $r > 8$. It then follows from Lemma 2.6.2 that

$$\frac{1}{nT^2}\sum_i [\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i)\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i)' - \mathrm{E}[\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i)\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i)']] \xrightarrow{p} 0.$$

Turning to (V.2), we have

$$\begin{aligned}\frac{1}{n^{3/2}}\sum_i &\mathrm{vec}[(\mathrm{vec}(x_i'\epsilon_i\epsilon_i'x_i/T))(\mathrm{vec}(x_i'\epsilon_i(\sqrt{nT}(\widehat{\beta} - \beta))'x_i'x_i/T^{3/2}))'] \\ &= \frac{1}{n^{3/2}}\sum_i [(x_i'\epsilon_i/\sqrt{T} \otimes x_i'\epsilon_i/\sqrt{T}) \otimes (x_i'x_i/T \otimes \epsilon_i'x_i/\sqrt{T})]\mathrm{vec}(\sqrt{nT}(\widehat{\beta} - \beta)).\end{aligned}$$

[18]Under Assumption 6.b, (O.2)-(O.4) will be $o_p(1)$ and consistency of $\widehat{W}$ follows.

By Lemma 2.6.2, we have that $\frac{1}{n}\sum_i(x_i'\epsilon_i/\sqrt{T}\otimes x_i'\epsilon_i/\sqrt{T})\otimes(x_i'x_i/T\otimes\epsilon_i'x_i/\sqrt{T}) = O_p(1)$ since

$$\mathrm{E}\|(x_i'\epsilon_i/\sqrt{T}\otimes x_i'\epsilon_i/\sqrt{T})\otimes(x_i'x_i/T\otimes\epsilon_i'x_i/\sqrt{T})\|^{1+\delta} \leq [(\mathrm{E}\|x_i'\epsilon_i/\sqrt{T}\|^{4+4\delta})^3\mathrm{E}\|x_i'x_i\|^{4+4\delta}]^{1/4} < C$$

where the first inequality is by repeated application of Lemma 2.6.1 and the second inequality is by the moment and mixing conditions. It then follows that (V.2) is $O_p(1/\sqrt{n})$. It can similarly be shown that terms (V.3)-(V.16) are $O_p(1/\sqrt{n})$, and the conclusion follows. ■

2.10 Appendix 5. Proof of Theorem 4

Under the hypotheses of the theorem, $\sqrt{n}(\widehat{\beta}-\beta)\xrightarrow{d}Q^{-1}N(0,W)$, $x_i'x_i/T - Q_i \xrightarrow{p} 0$, and $x_i'\epsilon_i/\sqrt{T}\xrightarrow{d}N(0,W_i)$ are immediate from a LLN and CLT for mixing sequences, cf. White (2001) Theorems 3.47 and 5.20. The conclusion then follows from the definition of $\widehat{W}$ and $\widehat{\epsilon}_i$.

2.11 Appendix 6. Proof of Corollary 4.1

Consider $t^* = \frac{\sqrt{nT}(R\widehat{\beta}-r)}{\sqrt{R\widehat{Q}^{-1}\widehat{W}\widehat{Q}^{-1}R'}}$. Under the null hypothesis, $R\beta = r$, so the numerator of $t*$ is $\sqrt{nT}R(\widehat{\beta}-\beta) = R(\frac{1}{nT}\sum_i x_i'x_i)^{-1}(\frac{1}{\sqrt{nT}}\sum_i x_i'\epsilon_i)\xrightarrow{d}RQ^{-1}\Lambda\sum_i B_i/\sqrt{n}$. From Theorem 4 and the hypotheses of the Corollary, the denominator of t^* converges in distribution to $\sqrt{RQ^{-1}\frac{1}{n}\Lambda(\sum_{i=1}^n B_iB_i' - \frac{1}{n}\sum_{i=1}^n B_i\sum_{i=1}^n B_i')\Lambda Q^{-1}R'}$. It follows from the Continuous Mapping Theorem that

$$t^*\xrightarrow{d}\frac{RQ^{-1}\Lambda\sum_i B_i/\sqrt{n}}{\sqrt{\frac{1}{n}RQ^{-1}\Lambda(\sum_{i=1}^n B_iB_i' - \frac{1}{n}\sum_{i=1}^n B_i\sum_{i=1}^n B_i')\Lambda Q^{-1}R'}}.$$

Define $\delta = (RQ^{-1}\Lambda\Lambda Q^{-1}R')^{1/2}$, so

$$\begin{aligned} t^*\xrightarrow{d}U &= \frac{\delta\sum_i B_{1,i}/\sqrt{n}}{\sqrt{\frac{\delta^2}{n}(\sum_{i=1}^n B_{1,i}B_{1,i}' - \frac{1}{n}\sum_{i=1}^n B_{1,i}\sum_{i=1}^n B_{1,i}')}} \\ &= \frac{\widetilde{B}_{1,n}}{\sqrt{\frac{1}{n}(\sum_i B_{1,i}^2 - \widetilde{B}_{1,n}^2)}}. \end{aligned}$$

It is straightforward to show that $\widetilde{B}_{1,n} \sim N(0,1)$, that $\sum_i B_{1,i}^2 - \widetilde{B}_{1,n}^2 \sim \chi_{n-1}^2$, and that $\sum_i B_{1,i}^2 - \widetilde{B}_{1,n}^2$ and $\widetilde{B}_{1,n}$ are independent, from which it follows that

$$U = \left(\frac{n}{n-1}\right)^{1/2} \frac{\widetilde{B}_{1,n}}{\sqrt{(\sum_i B_{1,i}^2 - \widetilde{B}_{1,n}^2)/(n-1)}} \sim \left(\frac{n}{n-1}\right)^{1/2} t_{n-1}.$$

The result for F^* is obtained through a similar argument. ■

2.12 Appendix 7. Proof of Lemma 2.6.2 and Lemma 2.6.3

Proof of Lemma 2.6.2 Instead of proving Lemma 2.6.2 directly, we prove the following: If $\{Z_{i,T}\}$ are independent across i for all T with $\mathrm{E}[Z_{i,T}] = 0$ and $\mathrm{E}|Z_{i,T}|^{1+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i, T, $\frac{1}{n}\sum_i Z_{i,T} \xrightarrow{p} 0$ as $\{n, T\} \to \infty$ jointly. Lemma 2.6.2 then follows immediately from an argument similar to that used to prove White (2001) Corollary 3.9.

Define $Y_{i,T} = Z_{i,T}1(|Z_{i,T}| \leq n)$. Then $Var(\sum_i Y_{i,T}/n) = \sum_i Var(Y_{i,T}/n)$ by independence of $Z_{i,T}$ across i. Now $\sum_i Var(Y_{i,T}/n) \leq \mathrm{E}(Y_{i,T}/n)^2 = \sum_i \int_{|Z|\leq n}(Z^2/n^2)dF_{i,T}(Z)$ where $F_{i,T}$ is the distribution function of $Z_{i,T}$. $Z^2/n^2 \leq Z^{1+\delta}/n^{1+\delta}$ for $|Z| \leq n$ implies

$$\begin{aligned}\sum_i \int_{|Z|\leq n}(Z^2/n^2)dF_{i,T}(Z) &\leq \sum_i \int_{|Z|\leq n}(Z^{1+\delta}/n^{1+\delta})dF_{i,T}(Z) \\ &\leq \sum_i \int (Z^{1+\delta}/n^{1+\delta})dF_{i,T}(Z) < \Delta/n^\delta\end{aligned}$$

where the last inequality results from $\mathrm{E}|Z_{i,T}|^{1+\delta} < \Delta$. It follows that

$$Var(\sum_i Y_{i,T}/n) < \Delta/n^\delta \to 0 \tag{2.12}$$

as $\{n,\ T\} \to \infty$ jointly.

Now consider

$$\begin{aligned}|\mathrm{E}\frac{1}{n}\sum_i Y_{i,T}| &= |\sum_i \int_{|Z|\leq n}(Z/n)dF_{i,T}(Z)| \\ &= |\sum_i \int (Z/n)dF_{i,T}(Z) - \sum_i \int_{|Z|>n}(Z/n)dF_{i,T}(Z)|\end{aligned}$$

$$
\begin{aligned}
&= |\sum_i \int_{|Z|>n} (Z/n) dF_{i,T}(Z)| \\
&\leq \sum_i \int_{|Z|>n} (|Z|/n) dF_{i,T}(Z)
\end{aligned}
$$

by the Triangle inequality and Jensen's inequality. For $|Z| > n$, $|Z|/n \leq |Z|^{1+\delta}/n^{1+\delta}$, so

$$
\sum_i \int_{|Z|>n} \frac{|Z|}{n} dF_{i,T}(Z) \leq \sum_i \int_{|Z|>n} \frac{|Z|^{1+\delta}}{n^{1+\delta}} dF_{i,T}(Z) \leq \sum_i \int \frac{|Z|^{1+\delta}}{n^{1+\delta}} dF_{i,T}(Z) \leq \Delta/n^{\delta} \to 0 \tag{2.13}
$$

which yields

$$
|\mathrm{E}\frac{1}{n}\sum_i Y_{i,T}| \to 0. \tag{2.14}
$$

By Chebyshev's inequality and (2.12),

$$
\lim_{n,T} P(|\sum_i (Y_{i,T} - \mathrm{E}[Y_{i,T}])/n| \geq \epsilon) \leq \lim_{n,T} Var(\sum_i Y_{i,T}/n)/\epsilon^2 = 0,
$$

so $\frac{1}{n}\sum_i Y_{i,T} - \frac{1}{n}\sum_i \mathrm{E}[Y_{i,T}] \xrightarrow{p} 0$, which implies, using (2.14), that $\frac{1}{n}\sum_i Y_{i,T} \xrightarrow{p} 0$.

Finally, consider

$$
\begin{aligned}
P(|\frac{1}{n}\sum_i Z_{i,T} - \frac{1}{n}\sum_i Y_{i,T}| \geq \epsilon) &= P(|\frac{1}{n}\sum_i (1 - 1(|Z_{i,T}| \leq n)) Z_{i,T}| \geq \epsilon) \\
&\leq \mathrm{E}|\frac{1}{n}\sum_i (1 - 1(|Z_{i,T}| \leq n)) Z_{i,T}|/\epsilon
\end{aligned}
$$

by the Markov inequality. $\mathrm{E}|\frac{1}{n}\sum_i (1 - 1(|Z_{i,T}| \leq n)) Z_{i,T}| \leq \frac{1}{n}\sum_i \mathrm{E}|(1 - 1(|Z_{i,T}| \leq n)) Z_{i,T}|$ by the Triangle inequality, and

$$
\begin{aligned}
\frac{1}{n}\sum_i \mathrm{E}|(1 - 1(|Z_{i,T}| \leq n)) Z_{i,T}| &= \frac{1}{n}\sum_i [\int |Z| dF_{i,T}(Z) - \int_{|Z|\leq n} |Z| dF_{i,T}(Z) \\
&= \sum_i \int_{|Z|>n} \frac{|Z|}{n} dF_{i,T}(Z) \to 0
\end{aligned}
$$

by (2.13). It then follows that $\frac{1}{n}\sum_i Z_{i,T} - \frac{1}{n}\sum_i Y_{i,T} \xrightarrow{p} 0$ which, with $\frac{1}{n}\sum_i Y_{i,T} \xrightarrow{p} 0$, implies $\frac{1}{n}\sum_i Z_{i,T} \xrightarrow{p} 0$. ■

Proof of Lemma 2.6.3 Define $\xi_{i,n,T} = \Omega_{n,T}^{-1/2} Z_{i,T}$ where $\Omega_{n,T} = \sum_i \Omega_{i,T}$. By the Cramer-Wold device, $\sum_i \xi_{i,n,T} \xrightarrow{d} N(0, I_k)$ as $\{n,\ T\} \to \infty$ jointly if $\forall\ c \in \mathbb{R}^k$ with $\|c\| = 1$, $c' \sum_i \xi_{i,n,T} \xrightarrow{d} N(0,1)$ as $\{n,\ T\} \to \infty$ jointly. Then, by $\Omega = \lim_{n,T} \frac{1}{n} \sum_{i=1}^n \Omega_{i,T} > 0$, $\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,T} \xrightarrow{d} N(0, \Omega)$ as $\{n,\ T\} \to \infty$ jointly.

To establish $c' \sum_i \xi_{i,n,T} \xrightarrow{d} N(0,1)$, it is sufficient to verify

$$\lim_{n,T} \sum_i \mathrm{E}[\xi_{i,n,T}^2 1(|\xi_{i,n,T}^2 > \epsilon)] = 0. \tag{2.15}$$

For a given $\epsilon > 0$ and $c \in \mathbb{R}^k$ with $\|c\| = 1$,

$$\begin{aligned} & c' \sum_i \mathrm{E}[\xi_{i,n,T} \xi_{i,n,T}' 1(|c' \xi_{i,n,T} \xi_{i,n,T}' c| > \epsilon)] c \\ &\qquad = c' \Omega_{n,T}^{-1/2} \sum_i \mathrm{E}[Z_{i,T} Z_{i,T}' 1(|c' \Omega_{n,T}^{-1/2} Z_{i,T} Z_{i,T}' \Omega_{n,T}^{-1/2} c| > \epsilon)] \Omega_{n,T}^{-1/2} c. \end{aligned}$$

Considering first the indicator function, we have

$$\begin{aligned} 1(|c' \Omega_{n,T}^{-1/2} Z_{i,T} Z_{i,T}' \Omega_{n,T}^{-1/2} c| > \epsilon) &\le 1(\|c\| \|\Omega_{n,T}^{-1/2} Z_{i,T} Z_{i,T}' \Omega_{n,T}^{-1/2}\| > \epsilon) \\ &\le 1(\lambda_{\max}(\Omega_{n,T}^{-1}) \|Z_{i,T}\|^2 > \epsilon) \\ &= 1(\|Z_{i,T}\|^2 > \epsilon \lambda_{\min}(\Omega_{n,T})) \end{aligned}$$

Then

$$\begin{aligned} & c' \Omega_{n,T}^{-1/2} \sum_i \mathrm{E}[Z_{i,T} Z_{i,T}' 1(|c' \Omega_{n,T}^{-1/2} Z_{i,T} Z_{i,T}' \Omega_{n,T}^{-1/2} c| > \epsilon)] \Omega_{n,T}^{-1/2} c \\ &\qquad \le \|c\|^2 \|\Omega_{n,T}^{-1/2} \sum_i \mathrm{E}[Z_{i,T} Z_{i,T}' 1(|c' \Omega_{n,T}^{-1/2} Z_{i,T} Z_{i,T}' \Omega_{n,T}^{-1/2} c| > \epsilon)] \Omega_{n,T}^{-1/2}\| \\ &\qquad \le \lambda_{\max}(\Omega_{n,T}^{-1}) \|\sum_i \mathrm{E}[Z_{i,T} Z_{i,T}' 1(\|Z_{i,T}\|^2 > \epsilon \lambda_{\min}(\Omega_{n,T}))]\| \\ &\qquad \le \frac{1}{\lambda_{\min}(\Omega_{n,T})} \sum_i \mathrm{E}[\|Z_{i,T}\|^2 1(\|Z_{i,T}\|^2 > \epsilon \lambda_{\min}(\Omega_{n,T}))] \\ &\qquad \le \frac{1}{\lambda_{\min}(\Omega_{n,T})} \sum_i \frac{\mathrm{E}\|Z_{i,T}\|^{2+\delta}}{(\epsilon \lambda_{\min}(\Omega_{n,T}))^\delta} \\ &\qquad \le \frac{n\Delta}{\epsilon^\delta [n \lambda_{\min}(\frac{1}{n} \sum_i \Omega_{i,T})]^{1+\delta}} = \frac{\Delta}{\epsilon^\delta n^\delta \lambda_{\min}^{1+\delta}} \to 0 \end{aligned}$$

as $\{n,\ T\} \to \infty$ jointly, and it follows that (2.15) $\to 0$ as $\{n,\ T\} \to \infty$ jointly.

Bibliography

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

ARELLANO, M. (1987): "Computing Robust Standard Errors for Within-Groups Estimators," *Oxford Bulletin of Economics and Statistics*, 49(4), 431–434.

BELL, R. M., AND D. F. MCCAFFREY (2002): "Computing Robust Standard Errors for Within-Groups Estimators," Mimeo RAND.

BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2003): "How Much Should We Trust Differences-in-Differences Estimates?," forthcoming QJE.

CHUNG, K. L. (2001): *A Course in Probability Theory.* San Diego, California: Academic Press, third edn.

DONALD, S., AND K. LANG (2001): "Inference with Difference in Differences and Other Panel Data," Mimeo.

DOUKHAN, P. (1994): *Mixing: Properties and Examples*, vol. 85 of *Lecture Notes in Statistics (Springer-Verlag).* New York: Springer-Verlag.

HAHN, J., AND G. M. KUERSTEINER (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both N and T Are Large," *Econometrica*, 70(4), 1639–1657.

HAHN, J., AND W. NEWEY (2002): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," Mimeo.

HANSEN, C. (2004): "Generalized Least Squares Inference in Multilevel Models with Serial Correlation and Fixed Effects," Mimeo.

KEZDI, G. (2002): "Robust Standard Errors Estimation in Fixed-Effects Panel Models," Mimeo.

KIEFER, N. M., AND T. J. VOGELSANG (2002): "Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size," *Econometric Theory*, 18, 1350–1366.

——— (2003): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," Mimeo.

LANCASTER, T. (2002): "Orthogonal Parameters and Panel Data," *Review of Economic Studies*, 69, 647–666.

LIANG, K.-Y., AND S. ZEGER (1986): "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73(1), 13–22.

MACURDY, T. E. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 18(1), 83–114.

NICKELL, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–1426.

PHILLIPS, P. C. B., AND H. R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67(5), 1057–1111.

PHILLIPS, P. C. B., Y. SUN, AND S. JIN (2003): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," Mimeo.

SOLON, G. (1984): "Estimating Autocorrelations in Fixed Effects Models," NBER Technical Working Paper No. 32.

SOLON, G., AND A. INOUE (2004): "A Portmanteau for Serially Correlated Errors in Fixed Effects Models," Mimeo.

STATA CORPORATION (2003): *Stata User's Guide Release 8*. College Station, Texas: Stata Press.

VOGELSANG, T. J. (2002): "Testing in GMM Models without Truncation," Mimeo.

WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

——— (2001): *Asymptotic Theory for Econometricians*. San Diego: Academic Press, revised edn.

Table 1
N = 10, T = 10

Data Generating Process	t Test Rejection Rate (1)	Mean(s.e.) (2)	Std(s.e.) (3)	Std(β) (4)	HA Test Rejection Rate (5)
			A. Fixed Effects		
homoskedastic, r = 0					
OLS	0.038	0.1180	0.0133	0.1152	0.152
Cluster	0.043	0.1149	0.0330	0.1152	
AR1	0.041	0.1170	0.0141	0.1152	0.135
homoskedastic, r = .3					
OLS	0.082	0.1130	0.0136	0.1269	0.095
Cluster	0.054	0.1212	0.0357	0.1269	
AR1	0.055	0.1240	0.0161	0.1269	0.133
homoskedastic, r = .6					
OLS	0.093	0.1005	0.0133	0.1231	0.074
Cluster	0.060	0.1167	0.0352	0.1231	
AR1	0.051	0.1219	0.0181	0.1231	0.123
homoskedastic, r = .9					
OLS	0.145	0.0609	0.0090	0.0818	0.038
Cluster	0.053	0.0772	0.0249	0.0818	
AR1	0.054	0.0795	0.0136	0.0818	0.085
heteroskedastic, r = 0					
OLS	0.126	0.1150	0.0126	0.1502	0.051
Cluster	0.057	0.1410	0.0458	0.1502	
AR1	0.126	0.1140	0.0137	0.1502	0.042
heteroskedastic, r = .3					
OLS	0.171	0.1165	0.0137	0.1708	0.036
Cluster	0.068	0.1538	0.0500	0.1708	
AR1	0.143	0.1284	0.0172	0.1708	0.044
heteroskedastic, r = .6					
OLS	0.187	0.1238	0.0153	0.1853	0.027
Cluster	0.074	0.1717	0.0572	0.1853	
AR1	0.117	0.1503	0.0219	0.1853	0.049
heteroskedastic, r = .9					
OLS	0.198	0.1406	0.0209	0.2181	0.031
Cluster	0.087	0.1872	0.0641	0.2181	
AR1	0.097	0.1830	0.0336	0.2181	0.074
			B. Random Effects		
ρ = .3					
OLS	0.295	0.1063	0.0231	0.1926	0.017
Cluster	0.115	0.1561	0.0609	0.1926	
RE	0.097	0.1693	0.0460	0.1926	0.027
ρ = .6					
OLS	0.399	0.1030	0.0248	0.2438	0.054
Cluster	0.118	0.2024	0.0788	0.2438	
RE	0.094	0.2180	0.0600	0.2438	0.023
ρ = .9					
OLS	0.482	0.0987	0.0293	0.2925	0.093
Cluster	0.108	0.2346	0.0909	0.2925	
RE	0.095	0.2546	0.0723	0.2925	0.018

Table 2
N = 10, T = 50

Data Generating Process	t Test Rejection Rate (1)	Mean(s.e.) (2)	Std(s.e.) (3)	Std(β) (4)	HA Test Rejection Rate (5)
			A. Fixed Effects		
homoskedastic, r = 0					
OLS	0.054	0.0462	0.0024	0.0472	0.184
Cluster	0.050	0.0449	0.0117	0.0472	
AR1	0.057	0.0460	0.0026	0.0472	0.185
homoskedastic, r = .3					
OLS	0.088	0.0459	0.0024	0.0519	0.077
Cluster	0.043	0.0520	0.0133	0.0519	
AR1	0.050	0.0529	0.0031	0.0519	0.159
homoskedastic, r = .6					
OLS	0.155	0.0447	0.0028	0.0590	0.049
Cluster	0.042	0.0574	0.0150	0.0590	
AR1	0.047	0.0598	0.0044	0.0590	0.184
homoskedastic, r = .9					
OLS	0.225	0.0372	0.0034	0.0600	0.046
Cluster	0.046	0.0562	0.0159	0.0600	
AR1	0.049	0.0583	0.0072	0.0600	0.150
heteroskedastic, r = 0					
OLS	0.158	0.0459	0.0021	0.0637	0.052
Cluster	0.051	0.0606	0.0169	0.0637	
AR1	0.162	0.0458	0.0023	0.0637	0.057
heteroskedastic, r = .3					
OLS	0.199	0.0479	0.0022	0.0724	0.046
Cluster	0.041	0.0735	0.0198	0.0724	
AR1	0.142	0.0553	0.0032	0.0724	0.047
heteroskedastic, r = .6					
OLS	0.229	0.0558	0.0031	0.0934	0.067
Cluster	0.043	0.0928	0.0260	0.0934	
AR1	0.112	0.0748	0.0054	0.0934	0.059
heteroskedastic, r = .9					
OLS	0.239	0.0857	0.0079	0.1490	0.059
Cluster	0.046	0.1428	0.0451	0.1490	
AR1	0.076	0.1338	0.0163	0.1490	0.099
			B. Random Effects		
ρ = .3					
OLS	0.568	0.0471	0.0092	0.1636	0.147
Cluster	0.104	0.1356	0.0547	0.1636	
RE	0.097	0.1475	0.0413	0.1626	0.014
ρ = .6					
OLS	0.703	0.0466	0.0105	0.2331	0.212
Cluster	0.104	0.1897	0.0727	0.2331	
RE	0.095	0.2079	0.0567	0.2331	0.007
ρ = .9					
OLS	0.744	0.0450	0.0130	0.2785	0.245
Cluster	0.106	0.2310	0.0920	0.2785	
RE	0.103	0.2539	0.0701	0.2785	0.014

Table 3
N = 10, T = 200

Data Generating Process	t Test Rejection Rate (1)	Mean(s.e.) (2)	Std(s.e.) (3)	Std(β) (4)	HA Test Rejection Rate (5)
			A. Fixed Effects		
homoskedastic, r = 0					
OLS	0.050	0.0226	0.0006	0.0223	0.177
Cluster	0.048	0.0219	0.0052	0.0223	
AR1	0.049	0.0226	0.0006	0.0223	0.175
homoskedastic, r = .3					
OLS	0.077	0.0225	0.0006	0.0256	0.098
Cluster	0.039	0.0255	0.0066	0.0256	
AR1	0.043	0.0261	0.0008	0.0256	0.186
homoskedastic, r = .6					
OLS	0.163	0.0224	0.0007	0.0314	0.044
Cluster	0.051	0.0297	0.0077	0.0314	
AR1	0.054	0.0304	0.0011	0.0314	0.169
homoskedastic, r = .9					
OLS	0.252	0.0214	0.0011	0.0353	0.075
Cluster	0.057	0.0339	0.0091	0.0353	
AR1	0.054	0.0345	0.0025	0.0353	0.165
heteroskedastic, r = 0					
OLS	0.139	0.0225	0.0005	0.0309	0.048
Cluster	0.053	0.0308	0.0081	0.0309	
AR1	0.141	0.0225	0.0006	0.0309	0.052
heteroskedastic, r = .3					
OLS	0.217	0.0236	0.0006	0.0380	0.065
Cluster	0.056	0.0367	0.0096	0.0380	
AR1	0.158	0.0274	0.0008	0.0380	0.049
heteroskedastic, r = .6					
OLS	0.279	0.0279	0.0008	0.0497	0.100
Cluster	0.040	0.0479	0.0122	0.0497	
AR1	0.086	0.0379	0.0014	0.0497	0.046
heteroskedastic, r = .9					
OLS	0.270	0.0490	0.0025	0.0861	0.089
Cluster	0.045	0.0837	0.0224	0.0861	
AR1	0.056	0.0791	0.0056	0.0861	0.124
			B. Random Effects		
ρ = .3					
OLS	0.773	0.0236	0.0047	0.1593	0.222
Cluster	0.115	0.1333	0.0523	0.1593	
RE	0.093	0.1447	0.0392	0.1593	0.014
ρ = .6					
OLS	0.837	0.0227	0.0050	0.2230	0.269
Cluster	0.116	0.1808	0.0704	0.2230	
RE	0.093	0.2007	0.0533	0.2230	0.016
ρ = .9					
OLS	0.874	0.0222	0.0066	0.2900	0.290
Cluster	0.110	0.2310	0.0933	0.2900	
RE	0.104	0.2497	0.0710	0.2900	0.022

Table 4
N = 50, T = 10

Data Generating Process	t Test Rejection Rate (1)	Mean(s.e.) (2)	Std(s.e.) (3)	Std(β) (4)	HA Test Rejection Rate (5)
			A. Fixed Effects		
homoskedastic, r = 0					
OLS	0.049	0.0522	0.0026	0.0526	0.106
Cluster	0.057	0.0515	0.0062	0.0526	
AR1	0.047	0.0522	0.0028	0.0526	0.099
homoskedastic, r = .3					
OLS	0.080	0.0500	0.0027	0.0569	0.053
Cluster	0.059	0.0552	0.0072	0.0569	
AR1	0.055	0.0556	0.0033	0.0569	0.092
homoskedastic, r = .6					
OLS	0.102	0.0447	0.0026	0.0539	0.132
Cluster	0.048	0.0549	0.0071	0.0539	
AR1	0.049	0.0553	0.0037	0.0539	0.072
homoskedastic, r = .9					
OLS	0.156	0.0273	0.0273	0.0387	0.220
Cluster	0.075	0.0364	0.0367	0.0387	
AR1	0.067	0.0367	0.0367	0.0387	0.078
heteroskedastic, r = 0					
OLS	0.119	0.0517	0.0025	0.0659	0.213
Cluster	0.047	0.0673	0.0093	0.0659	
AR1	0.116	0.0516	0.0028	0.0659	0.210
heteroskedastic, r = .3					
OLS	0.197	0.0521	0.0026	0.0768	0.369
Cluster	0.062	0.0741	0.0114	0.0768	
AR1	0.139	0.0581	0.0033	0.0768	0.140
heteroskedastic, r = .6					
OLS	0.214	0.0558	0.0031	0.0840	0.451
Cluster	0.048	0.0820	0.0126	0.0840	
AR1	0.108	0.0688	0.0045	0.0840	0.056
heteroskedastic, r = .9					
OLS	0.152	0.0623	0.0043	0.0883	0.324
Cluster	0.038	0.0899	0.0144	0.0883	
AR1	0.057	0.0834	0.0070	0.0883	0.023
			B. Random Effects		
ρ = .3					
OLS	0.291	0.0451	0.0041	0.0822	0.673
Cluster	0.062	0.0776	0.0135	0.0822	
RE	0.059	0.0788	0.0091	0.0822	0.058
ρ = .6					
OLS	0.357	0.0452	0.0049	0.1034	0.892
Cluster	0.073	0.1004	0.0183	0.1034	
RE	0.068	0.1028	0.0127	0.1034	0.054
ρ = .9					
OLS	0.497	0.0447	0.0056	0.1246	0.943
Cluster	0.062	0.1192	0.0212	0.1246	
RE	0.063	0.1210	0.0147	0.1246	0.048

Table 5
N = 50, T = 20

Data Generating Process	t Test Rejection Rate (1)	Mean(s.e.) (2)	Std(s.e.) (3)	Std(β) (4)	HA Test Rejection Rate (5)
			A. Fixed Effects		
homoskedastic, r = 0					
OLS	0.050	0.0342	0.0013	0.0341	0.097
Cluster	0.490	0.0341	0.0040	0.0341	
AR1	0.052	0.0342	0.0014	0.0341	0.088
homoskedastic, r = .3					
OLS	0.094	0.0334	0.0013	0.0393	0.077
Cluster	0.051	0.0379	0.0045	0.0393	
AR1	0.056	0.0382	0.0016	0.0393	0.086
homoskedastic, r = .6					
OLS	0.120	0.0315	0.0014	0.0414	0.300
Cluster	0.059	0.0407	0.0052	0.0414	
AR1	0.050	0.0412	0.0021	0.0414	0.092
homoskedastic, r = .9					
OLS	0.200	0.0222	0.0013	0.0336	0.580
Cluster	0.059	0.0327	0.0047	0.0336	
AR1	0.060	0.0329	0.0024	0.0336	0.094
heteroskedastic, r = 0					
OLS	0.168	0.0340	0.0011	0.0479	0.408
Cluster	0.063	0.0458	0.0056	0.0479	
AR1	0.171	0.0340	0.0012	0.0479	0.406
heteroskedastic, r = .3					
OLS	0.209	0.0350	0.0012	0.0536	0.675
Cluster	0.051	0.0527	0.0068	0.0536	
AR1	0.145	0.0399	0.0016	0.0536	0.294
heteroskedastic, r = .6					
OLS	0.228	0.0394	0.0017	0.0653	0.802
Cluster	0.050	0.0636	0.0084	0.0653	
AR1	0.119	0.0514	0.0027	0.0653	0.123
heteroskedastic, r = .9					
OLS	0.196	0.0507	0.0028	0.0775	0.681
Cluster	0.036	0.0809	0.0131	0.0775	
AR1	0.058	0.0751	0.0056	0.0775	0.034
			B. Random Effects		
ρ = .3					
OLS	0.405	0.0320	0.0029	0.0756	0.915
Cluster	0.069	0.0726	0.0131	0.0756	
RE	0.063	0.0738	0.0085	0.0756	0.064
ρ = .6					
OLS	0.515	0.0318	0.0033	0.1012	0.944
Cluster	0.066	0.0976	0.0169	0.1012	
RE	0.055	0.0996	0.0118	0.1012	0.055
ρ = .9					
OLS	0.614	0.0314	0.0038	0.1203	0.948
Cluster	0.054	0.1166	0.0204	0.1203	
RE	0.051	0.1194	0.0140	0.1203	0.053

Table 6
N = 50, T = 50

Data Generating Process	t Test Rejection Rate (1)	Mean(s.e.) (2)	Std(s.e.) (3)	Std(β) (4)	HA Test Rejection Rate (5)
			A. Fixed Effects		
homoskedastic, r = 0					
OLS	0.057	0.0206	0.0005	0.0212	0.112
Cluster	0.058	0.0205	0.0022	0.0212	
AR1	0.053	0.0206	0.0005	0.0212	0.104
homoskedastic, r = .3					
OLS	0.079	0.0205	0.0005	0.0232	0.114
Cluster	0.043	0.0236	0.0027	0.0232	
AR1	0.051	0.0237	0.0007	0.0232	0.094
homoskedastic, r = .6					
OLS	0.128	0.0200	0.0006	0.0254	0.529
Cluster	0.051	0.0267	0.0030	0.0254	
AR1	0.050	0.0269	0.0009	0.0254	0.089
homoskedastic, r = .9					
OLS	0.220	0.0168	0.0007	0.0270	0.834
Cluster	0.055	0.0264	0.0033	0.0270	
AR1	0.059	0.0265	0.0015	0.0270	0.082
heteroskedastic, r = 0					
OLS	0.158	0.0206	0.0004	0.0279	0.610
Cluster	0.044	0.0285	0.0031	0.0279	
AR1	0.162	0.0206	0.0005	0.0279	0.617
heteroskedastic, r = .3					
OLS	0.225	0.0214	0.0005	0.0346	0.865
Cluster	0.053	0.0333	0.0040	0.0346	
AR1	0.153	0.0247	0.0006	0.0346	0.515
heteroskedastic, r = .6					
OLS	0.256	0.0249	0.0006	0.0429	0.942
Cluster	0.051	0.0425	0.0050	0.0429	
AR1	0.122	0.0335	0.0011	0.0429	0.287
heteroskedastic, r = .9					
OLS	0.275	0.0384	0.0016	0.0670	0.902
Cluster	0.048	0.0662	0.0090	0.0670	
AR1	0.062	0.0605	0.0034	0.0670	0.042
			B. Random Effects		
ρ = .3					
OLS	0.582	0.0202	0.0018	0.0719	0.949
Cluster	0.061	0.0685	0.0126	0.0719	
RE	0.058	0.0704	0.0084	0.0719	0.062
ρ = .6					
OLS	0.693	0.0202	0.0021	0.1001	0.961
Cluster	0.063	0.0957	0.0170	0.1001	
RE	0.051	0.0982	0.0116	0.1001	0.047
ρ = .9					
OLS	0.763	0.0199	0.0024	0.1242	0.959
Cluster	0.078	0.1162	0.0200	0.1242	
RE	0.070	0.1188	0.0141	0.1242	0.055

Table 7
N = 50, T = 200

Data Generating Process	t Test Rejection Rate (1)	Mean(s.e.) (2)	Std(s.e.) (3)	Std(β) (4)	HA Test Rejection Rate (5)
			A. Fixed Effects		
homoskedastic, r = 0					
OLS	0.037	0.0101	0.0001	0.0100	0.104
Cluster	0.047	0.0100	0.0011	0.0100	
AR1	0.038	0.0101	0.0001	0.0100	0.105
homoskedastic, r = .3					
OLS	0.083	0.0101	0.0001	0.0113	0.141
Cluster	0.044	0.0116	0.0012	0.0113	
AR1	0.042	0.0117	0.0002	0.0113	0.094
homoskedastic, r = .6					
OLS	0.151	0.0100	0.0001	0.0136	0.608
Cluster	0.055	0.0134	0.0014	0.0136	
AR1	0.050	0.0136	0.0002	0.0136	0.108
homoskedastic, r = .9					
OLS	0.243	0.0096	0.0002	0.0161	0.934
Cluster	0.055	0.0155	0.0017	0.0161	
AR1	0.059	0.0154	0.0005	0.0161	0.072
heteroskedastic, r = 0					
OLS	0.156	0.0101	0.0001	0.0139	0.720
Cluster	0.042	0.0141	0.0015	0.0139	
AR1	0.156	0.0101	0.0001	0.0139	0.723
heteroskedastic, r = .3					
OLS	0.244	0.0105	0.0001	0.0174	0.943
Cluster	0.057	0.0169	0.0018	0.0174	
AR1	0.185	0.0122	0.0002	0.0174	0.651
heteroskedastic, r = .6					
OLS	0.259	0.0125	0.0002	0.0219	0.982
Cluster	0.047	0.0220	0.0024	0.0219	
AR1	0.120	0.0170	0.0003	0.0219	0.432
heteroskedastic, r = .9					
OLS	0.276	0.0219	0.0005	0.0397	0.977
Cluster	0.049	0.0389	0.0044	0.0397	
AR1	0.080	0.0353	0.0011	0.0397	0.065
			B. Random Effects		
ρ = .3					
OLS	0.791	0.0101	0.0009	0.0692	0.969
Cluster	0.059	0.0674	0.0120	0.0692	
RE	0.055	0.0686	0.0086	0.0692	0.063
ρ = .6					
OLS	0.855	0.0100	0.0010	0.0974	0.972
Cluster	0.052	0.0951	0.0170	0.0974	
RE	0.051	0.0969	0.0114	0.0974	0.049
ρ = .9					
OLS	0.870	0.0100	0.0012	0.1184	0.972
Cluster	0.065	0.1166	0.0206	0.1184	
RE	0.058	0.1184	0.0143	0.1184	0.056

Table 8

N = 200, T = 10

Data Generating Process	t Test Rejection Rate (1)	Mean(s.e.) (2)	Std(s.e.) (3)	Std(β) (4)	HA Test Rejection Rate (5)
	A. Fixed Effects				
homoscedastic, ρ = 0					
OLS	0.061	0.0261	0.0007	0.0263	0.050
Cluster	0.063	0.0261	0.0015	0.0263	
AR1	0.061	0.0261	0.0007	0.0263	0.042
homoscedastic, ρ = .3					
OLS	0.061	0.0250	0.0007	0.0266	0.300
Cluster	0.026	0.0278	0.0018	0.0266	
AR1	0.028	0.0279	0.0008	0.0266	0.062
homoscedastic, ρ = .6					
OLS	0.104	0.0224	0.0007	0.0275	0.860
Cluster	0.050	0.0277	0.0018	0.0275	
AR1	0.046	0.0278	0.0009	0.0275	0.041
homoscedastic, ρ = .9					
OLS	0.131	0.0136	0.0005	0.0181	0.981
Cluster	0.043	0.0182	0.0013	0.0181	
AR1	0.051	0.0183	0.0007	0.0181	0.044
heteroscedastic, ρ = 0					
OLS	0.136	0.0258	0.0006	0.0339	0.961
Cluster	0.055	0.0341	0.0025	0.0339	
AR1	0.136	0.0258	0.0007	0.0339	0.960
heteroscedastic, ρ = .3					
OLS	0.159	0.0261	0.0007	0.0369	0.994
Cluster	0.047	0.0377	0.0029	0.0369	
AR1	0.124	0.0292	0.0009	0.0369	0.921
heteroscedastic, ρ = .6					
OLS	0.188	0.0279	0.0008	0.0426	0.997
Cluster	0.058	0.0420	0.0033	0.0426	
AR1	0.120	0.0345	0.0012	0.0426	0.616
heteroscedastic, ρ = .9					
OLS	0.176	0.0313	0.0011	0.0454	0.983
Cluster	0.050	0.0459	0.0040	0.0454	
AR1	0.069	0.0421	0.0018	0.0454	0.048
	B. Random Effects				
ρ = .3					
OLS	0.264	0.0224	0.0011	0.0393	1.000
Cluster	0.052	0.0395	0.0037	0.0393	
RE	0.053	0.0396	0.0024	0.0393	0.038
ρ = .6					
OLS	0.430	0.0223	0.0012	0.0533	1.000
Cluster	0.061	0.0511	0.0048	0.0533	
RE	0.060	0.0513	0.0031	0.0533	0.045
ρ = .9					
OLS	0.456	0.0224	0.0014	0.0617	1.000
Cluster	0.064	0.0610	0.0056	0.0617	
RE	0.060	0.0612	0.0037	0.0617	0.046

Table 9
N = 200, T = 50

Data Generating Process	t Test Rejection Rate (1)	Mean(s.e.) (2)	Std(s.e.) (3)	Std(β) (4)	HA Test Rejection Rate (5)
			A. Fixed Effects		
homoskedastic, r = 0					
OLS	0.050	0.0103	0.0001	0.0101	0.074
Cluster	0.047	0.0103	0.0005	0.0101	
AR1	0.050	0.0103	0.0001	0.0101	0.073
homoskedastic, r = .3					
OLS	0.095	0.0102	0.0001	0.0121	0.677
Cluster	0.055	0.0117	0.0006	0.0121	
AR1	0.054	0.0118	0.0002	0.0121	0.065
homoskedastic, r = .6					
OLS	0.138	0.0100	0.0001	0.0131	1.000
Cluster	0.050	0.0134	0.0007	0.0134	
AR1	0.052	0.0134	0.0002	0.0134	0.061
homoskedastic, r = .9					
OLS	0.201	0.0084	0.0002	0.0130	1.000
Cluster	0.047	0.0131	0.0008	0.0130	
AR1	0.048	0.0132	0.0004	0.0130	0.058
heteroskedastic, r = 0					
OLS	0.148	0.0103	0.0001	0.0139	1.000
Cluster	0.045	0.0143	0.0008	0.0139	
AR1	0.146	0.0103	0.0001	0.0139	1.000
heteroskedastic, r = .3					
OLS	0.210	0.0107	0.0001	0.0168	1.000
Cluster	0.051	0.0168	0.0010	0.0168	
AR1	0.148	0.0124	0.0002	0.0168	0.999
heteroskedastic, r = .6					
OLS	0.264	0.0125	0.0002	0.0220	1.000
Cluster	0.061	0.0216	0.0013	0.0220	
AR1	0.138	0.0168	0.0003	0.0220	0.976
heteroskedastic, r = .9					
OLS	0.261	0.0192	0.0004	0.0333	1.000
Cluster	0.050	0.0333	0.0023	0.0333	
AR1	0.078	0.0303	0.0009	0.0333	0.181
			B. Random Effects		
ρ = .3					
OLS	0.593	0.0100	0.0004	0.0353	1.000
Cluster	0.052	0.0354	0.0032	0.0353	
RE	0.048	0.0356	0.0020	0.0353	0.041
ρ = .6					
OLS	0.703	0.0100	0.0005	0.0497	1.000
Cluster	0.057	0.0491	0.0045	0.0497	
RE	0.049	0.0495	0.0030	0.0497	0.045
ρ = .9					
OLS	0.741	0.0100	0.0006	0.0584	1.000
Cluster	0.053	0.0596	0.0054	0.0584	
RE	0.051	0.0601	0.0035	0.0584	0.036

Chapter 3

The Impact of 401(k) Participation on the Wealth Distribution: An Instrumental Quantile Regression Analysis

Co-Authored with Victor Chernozhukov (MIT)

3.1 Introduction

In the early 1980s, the United States introduced several tax deferred savings options in an effort to increase individual saving for retirement. The two options which have generated the most interest are Individual Retirement Accounts (IRAs) and 401(k) plans. Tax deferred IRAs and 401(k) plans are similar in that both allow the individual to deduct contributions from taxable income and allow tax-free accrual of interest on assets held within the plan. The key differences between the two savings options are that employers provide 401(k) plans, and employers may also match a certain percentage of an employee's contribution. Since 401(k) plans are provided by employers, only workers in firms offering plans are eligible for participation, while participation in IRAs is open to everyone.[1]

While it is clear that 401(k) plans and, to a lesser extent, IRAs are widely used as vehicles for retirement saving, their impact on assets is less clear. The key problem in determining the effect of participation in IRA and 401(k) plans on accumulated assets is saver heterogeneity coupled with non-random selection into participation states. In particular, it is generally recognized that some people have a higher preference for saving than others. Thus, it seems likely that those individuals with the highest unobserved preference for saving would be most likely to choose to participate in tax-advantaged retirement savings plans and would also have higher savings in other assets than individuals with lower unobserved saving propensity. This implies that conventional estimates that do not account for saver heterogeneity and selection of the participation state will be biased upward, tending to overstate the actual savings effects of 401(k) and IRA participation.

This problem has long been recognized in the savings literature and has led to numerous important studies which attempt to overcome this problem. In a series of articles, Poterba, Venti, and Wise (1994, 1995, 1996) use comparisons between groups based on eligibility for 401(k) participation. They argue that 401(k) eligibility can be taken as exogenous given income. The argument is motivated by the fact that eligibility is determined by the employer, and so may be taken as exogenous conditional on covariates. Poterba, Venti, and Wise (1996) contains an overview of suggestive evidence based on pre-program savings used to substantiate this claim and reports mean and median regression estimates of the impact of 401(k) eligibility on household net financial assets. The results show that 401(k) eligibility has significant and positive effects on net financial assets. Based on the assumed exogeneity of 401(k) eligibility, they attribute this difference to the causal effect of 401(k) eligibility on savings.[2] Recent work by Benjamin (2003) examines the effects of eligibility on savings using matching based on the propensity score and finds positive, although much more modest, effects of 401(k) eligibility on assets.[3]

A similar approach, which we follow in this paper, is that of Abadie (2003). Abadie,

[1] A detailed description of regulations regarding retirement programs can be found in a recent publication of the Employee Benefit Research Institute (1997).

[2] For a differing viewpoint, see Engen, Gale, and Scholz (1996), which contends that eligibility should not be treated as exogenous.

[3] Benjamin uses a more inclusive definition of assets and makes adjustments to account for replacement or substitution of an existing defined contribution or defined benefit plan by a 401(k).

assuming that eligibility for a 401(k) is exogenous given income (and other covariates), uses 401(k) eligibility as an instrument for 401(k) participation in order to estimate the effect of 401(k) participation, not eligibility, on net financial assets. Abadie uses a novel semiparametric estimator which estimates the average effect for compliers.[4] Since only individuals eligible for a 401(k) can participate, the average effect for compliers also corresponds to the average effect for the treated. Abadie's results suggest that the average effect for the treated of 401(k) participation is significant and positive.

One drawback of all of these studies is that they focus the analysis on measures of central tendency: the mean or the median. While the mean and median impacts are interesting and important measures in determining a program's impact, they are not sufficient to fully characterize the impact of the treatment except under very restrictive conditions. In particular, they are uninformative about the impact of the treatment on other, perhaps more interesting, points in the outcome distribution when the treatment effect is heterogeneous. Understanding the distributional impact of 401(k) plans is especially interesting from a policy perspective since policy makers may be particularly concerned about the impact of 401(k) plans on the lower part of the wealth distribution. In addition, knowledge of the distributional impact of a program provides a clearer picture of what is driving the mean results.

As with estimates of the mean effect, the analysis of the distributional effect is complicated by the possibility that individuals choose whether or not to participate in a 401(k) based on their unobserved preferences for saving. One estimator which would allow a more full characterization of the effect of a heterogeneous treatment given treatment exogeneity is the quantile regression estimator of Koenker and Bassett (1978). However, the self-selection of the participation state makes the conventional quantile regression estimator inappropriate.[5]

In this paper, we contribute to the extensive set of existing literature of the impact of 401(k) plans on wealth by analyzing the impact of 401(k) participation on the entire wealth

[4] In the context of 401(k) participation, the group of compliers is the group of individuals who would participate in a 401(k) if eligible but would not if ineligible. Non-compliers in this example are people who would not participate in the 401(k) regardless of their eligibility status.

[5] Also, treatment heterogeneity renders the two stage least absolute deviation estimator of Amemiya (1982) and its extension to quantile regression by Chen and Portnoy (1996) inconsistent. The inconsistency was first demonstrated in Chernozhukov and Hansen (2001).

distribution. Using the reasoning of Poterba, Venti, and Wise (1994, 1995, 1996) and Abadie (2003) outlined above, we use 401(k) eligibility as an instrument for 401(k) participation in order to estimate the effect of participating in a 401(k) on various measures of wealth. To do this, we employ a model and an estimator developed in Chernozhukov and Hansen (2001). The model provides a set of assumptions under which the conditional quantiles of the outcome distribution may be recovered from a set of statistical moment equations through the use of instrumental variables. The estimator we use is computationally convenient for linear quantile models and can be computed through a series of conventional linear quantile regressions. Chernozhukov and Hansen (2001) demonstrate that the estimator is consistent under endogeneity and treatment effect heterogeneity. Thus, this paper provides an important complement to the work discussed above which focuses on estimating the impact of 401(k) plans on the center of the outcome distribution. Also, due to the binary nature of both the participation decision and the eligibility instrument, the approach developed by Abadie, Angrist, and Imbens (2002) to estimate quantile effects for binary treatments under endogeneity also applies. We present estimates obtained through both procedures to provide both a robustness check and a comparison of the two approaches. We find that the results are very similar using either estimation procedure.

The instrumental quantile regression estimates indicate that there is considerable heterogeneity in the effect of 401(k) participation on net financial assets, with the treatment effect increasing monotonically as one moves from the lower to the upper tail of the asset distribution. The results are also uniformly positive and significant, suggesting that 401(k) participation positively impacts net financial assets across the entire distribution. The effect of participation on total wealth is positive and approximately constant for all quantiles. In addition, it is of the same magnitude as the effect of participation on net financial assets for low quantiles, but is substantially smaller than the effect of participation on the upper quantiles of net financial assets. These results suggest that the increase in net financial assets observed in the lower tail of the conditional assets distribution can be interpreted as an increase in wealth, while the increase in the upper tail of the distribution is mitigated by substitution with some other component of wealth. The effect of participation on net non-401(k) financial assets is uniformly insignificant, which suggests there is little substitution for 401(k) assets along this dimension of wealth.[6]

[6]Net non-401(k) financial assets are net financial assets minus 401(k) balances. More details about the wealth measures are found in the description of the data in Section 3.

The remainder of the paper is organized as follows. Section 2 reviews the model of quantile treatment effects of Chernozhukov and Hansen (2001) and demonstrates how an empirical model for assets may be embedded in the model. In Section 3, the data used in the empirical analysis are described. Section 4 presents the empirical results and compares the results from the estimator of Chernozhukov and Hansen (2001) to those obtained with the estimator of Abadie, Angrist, and Imbens (2002), and Section 5 concludes.

3.2 An Instrumental Variable Model for Quantile Treatment Effects

In the following, we briefly present the assumptions and main implications of the instrumental variables model of quantile treatment effects developed in Chernozhukov and Hansen (2001). We then show how an empirical model of savings decisions may be embedded in this framework. This discussion helps illustrate the interpretation of the estimates of the model, especially the interpretation of the quantile index τ, and isolates the key identifying assumptions.

3.2.1 Potential Outcomes and the QTE

The model is developed within the conventional potential (latent) outcome framework. Potential real-valued outcomes are indexed against treatment d and denoted Y_d. For example, Y_d is an individual's outcome when $D = d$. Treatments d take values in a subset $\mathcal{D}$ of $\mathbb{R}^l$. The potential outcomes $\{Y_d\}$ are latent because, given the selected treatment D, the observed outcome for each individual or state of the world is $Y \equiv Y_D$. That is, only one component of potential outcomes vector $\{Y_d\}$ is observed for each observational unit.

While there are many features of the distributions of potential outcomes that may be interesting, we focus on the quantiles of potential outcomes conditional on covariates X,[7]

$$\left\{Q_{Y_d}(\tau|x),\ \tau \in (0,1)\right\},$$

[7]We use $Q_Y(\tau|x)$ and $f_Y(y|x)$ to denote the conditional τ-quantile and density of Y given $X = x$. Capitals such as Y denote random variables, and lower case letters such as y denote the values they take.

and the quantile treatment effects (QTE) that summarize the difference between the quantiles under different treatments (e.g. Doksum (1974)):

$$Q_{Y_d}(\tau|x) - Q_{Y_{d'}}(\tau|x) \quad \text{or, if defined,} \quad \frac{\partial}{\partial d} Q_{Y_d}(\tau|x).$$

Quantile treatment effects represent a useful way of describing the effect of treatment d on different points of the marginal distribution of potential outcomes.

Typically D is selected in relation to $\{Y_d\}$ inducing endogeneity, so that the conditional quantile of Y given the selected treatment $D = d$, denoted $Q_Y(\tau|d, x)$, is generally not equal to the quantile of potential or latent outcome $Q_{Y_d}(\tau|x)$. This makes the conventional quantile regression inappropriate for the estimation of $Q_{Y_d}(\tau|x)$. The model of Chernozhukov and Hansen (2001), briefly presented below, states the conditions under which we can recover the quantiles of latent outcomes through a set of conditional moment restrictions.

3.2.2 The Instrumental Quantile Treatment Model.

We build the model from the basic Skorohod representation of latent outcomes Y_d, which yields for each d given $X = x$

$$Y_d = q(d, x, U_d), \text{ where } U_d \overset{d}{\sim} U(0, 1), \tag{3.1}$$

and $q(d, x, \tau) = Q_{Y_d}(\tau|x)$ is the conditional τ-quantile of latent outcome Y_d.[8] This representation is essential to the rest of the analysis.

The variable U_d is responsible for heterogeneity of outcomes for individuals with the same observed characteristics x and treatment d. It also determines their relative ranking in terms of potential outcomes. Hence we will call U_d the rank variable, and may think of it as representing some innate ability or level of preference. This allows interpretation of the quantile treatment effect as the treatment effect for people with a given rank in the distribution of U_d, making quantile analysis an interesting tool for describing and learning the structure of heterogeneous treatment effects.

[8] The basic Skorohod representation states that, given a collection of variables $\{\zeta_j\}$, each variable ζ_j can be represented as, a.s.

$$\zeta_j = Q_{\zeta_j}(U_j), \text{ for some } U_j \overset{d}{\sim} U(0, 1).$$

Recall that $Q_{\zeta_j}(\tau)$ denotes the τ-quantile of variable ζ_j.

The model consists of five main conditions (some are representations) that hold jointly.

The IVQT Model: Given a common probability space (Ω, F, P), for P-almost every value of X, Z, where X represents covariates and Z represents excluded instruments, the following conditions A1-A5 hold jointly:

A1 POTENTIAL OUTCOMES. Given $X = x$, for each d, for some $U_d \overset{d}{\sim} U(0,1)$,

$$Y_d = q(d, x, U_d),$$

where $q(d, x, \tau)$ is strictly increasing and left-continuous in τ.

A2 INDEPENDENCE. Given $X = x$, $\{U_d\}$ is independent of Z.

A3 SELECTION. Given $X = x, Z = z$, for unknown function δ and random vector V,

$$D \equiv \delta(z, x, V).$$

A4 RANK SIMILARITY. For each d and d', given (V, X, Z)

$$U_d \text{ is equal in distribution to } U_{d'}.$$

A5 OBSERVED variables consist of (for $U_D \equiv \sum_{d \in \mathcal{D}} I(D = d) \cdot U_d$)

$$\begin{cases} Y \equiv q(D, X, U_D), \\ D \equiv \delta(Z, X, V), \\ X, Z. \end{cases}$$

Chernozhukov and Hansen (2001) demonstrate that the following result is an implication of the IVQT model.

Theorem 5 ***(Main Statistical Implication)*** *Suppose conditions A1-A5 hold. Then, for any $\tau \in (0,1)$, a.s.*

$$P[Y \leq q(D, X, \tau) | X, Z] = \tau \text{ and } P[Y < q(D, X, \tau) | X, Z] = \tau. \tag{3.2}$$

This result provides an important link of the parameters of the IVQT model to a set of conditional moment equations which are used in Chernozhukov and Hansen (2001) to develop identification conditions for the IVQT model as well as for estimation and inference. In addition, Chernozhukov and Hansen (2001) give an extensive discussion of the IVQT model, its assumptions, and its identification. While this discussion will not be repeated here, it is important to note that the assumptions of the IVQT model differ from those in other models with endogeneity and heterogeneous treatment effects in two key respects.[9] First, the IVQT model imposes a different set of independence conditions; in particular, it does not require that the instruments, Z, are independent of the errors in the selection equation V. The independence of Z and V may be violated when Z is measured with error or related to V in other ways. Second, the IVQT model imposes rank similarity, Assumption A4, which will be discussed in the context of saving decisions below.

3.2.3 The Instrumental Quantile Regression Model and Saving Decisions

Assumptions A1-A5 represent a plausible framework within which to analyze the effects of participating in a 401(k) plan on an individuals accumulated wealth. First, wealth, Y_d, in the participation state $d \in \{0, 1\}$ can be represented as

$$Y_d = q(d, X, U_d), \quad U_d \sim U(0, 1)$$

by the Skorohod representation of random variables, where $\tau \mapsto q(d, X, \tau)$ is the conditional quantile function of Y_d and U_d is an unobserved random variable. Following the discussion in Section 2.2, we will refer to U_d as the preference for saving and thus interpret the quantile index τ as indexing rank in the preference for saving distribution.[10] The individual selects

[9] See, for example, Amemiya (1982), Heckman and Robb (1986), Imbens and Angrist (1994), and Vytlacil (2000).

[10] Since the outcomes of interest are all measures of accumulated wealth, perhaps more appropriate, but more cumbersome, terminology would be preference for accumulated assets. In addition, if there are unobservable factors besides preferences then this interpretation of U_d and τ is incorrect, and τ should be only interpreted as indexing rank in the conditional distribution of Y_d given x. For simplicity and clarity, we will refer to U_d and τ as relating to preference for saving throughout the rest of the paper.

the 401(k) participation state to maximize expected utility:

$$D = \arg\max_{d \in \mathcal{D}} \; E\left[W\{Y_d, d\} \Big| X, Z, V \right] = \arg\max_{d \in \mathcal{D}} \; E\left[W\{q(d, x, U_d), d\} \Big| X, Z, V \right], \quad (3.3)$$

where $W\{y, d\}$ is the unobserved Bernoulli utility function. As a result, the participation decision is represented by

$$D = \delta(Z, X, V)$$

where Z and X are observed, V is an unobserved information component that depends on rank U_d, and includes other unobserved variables that affect the participation state, and function δ is unknown. Thus this model is a special case of the IVQT model. In this model, the independence condition A2 only requires that U_d is independent of Z, conditional on X.

The simplest form of rank similarity is rank invariance, under which the preference for saving vector U_d may be collapsed to a single random variable:

$$U = U_0 = U_1.$$

In this case, a single preference for saving is responsible for an individual's ranking across all treatment states. It is important to note that U is defined relative to observationally identical people (individuals with the same X and Z). Rank invariance has been used in many interesting models without endogeneity,[11] and traditional simultaneous equations models are built assuming rank invariance. However, as noted in Heckman and Smith (1997), rank invariance may be implausible on logical grounds since it implies that the potential outcomes Y_d are not truly multivariate, but have a jointly degenerate distribution.

The similarity condition A4 is a more general form of rank invariance – it relaxes the exact invariance of ranks U_d across d by allowing noisy, unsystematic variations of U_d across d, conditional on (V, X, Z). This relaxation allows for variation in the ranks across the treatment states, requiring only a "rank invariance in expectation". Therefore, similarity accommodates general multivariate models of outcomes. It states that given the information in (V, X, Z) employed to make the selection of treatment D, the expectations of any function of rank U_d does not vary across the treatment states. That is, ex-ante, conditional on (V, X, Z), the ranks may be considered to be the same across potential treatments, but the realized, ex-post, rank may be different across treatment states.

[11]For example, Doksum (1974) and Heckman and Smith (1997).

From an econometric perspective, the similarity assumption is nothing but a restriction on the evolution on the unobserved heterogeneity component which precludes systematic variation of U_d across the treatment states. Similarity allows interpretation of the quantile treatment effect as the treatment effect holding the level of unobserved heterogeneity constant across the treatment states:

$$q(d, x, \tau) - q(d', x, \tau) = q(d, x, U_d) - q(d', x, U_{d'})\Big|_{U_d = U_{d'} = \tau}.$$

Since changes in U_d across d are assumed to be asystematic, the quantile treatment effect not only summarizes the distributional impact but also the actual likely treatment effect.

To be more concrete, consider the following simple example where

$$U_d = F_{V+\eta_d}(V + \eta_d),$$

where $F_{V+\eta_d}(\cdot)$ is the distribution function of $V + \eta_d$ and $\{\eta_d\}$ are mutually iid conditional on V, X, and Z. The variable V represents an individual's "mean" saving preference, while η_d is a noisy adjustment.[12] This more general assumption leaves the individual optimization problem (3.3) unaffected, while allowing variation in an individual's rank across different potential outcomes.

While we feel that similarity may be a reasonable assumption in many contexts, imposing similarity is not innocuous. In the context of 401(k) participation, matching practices of employers could jeopardize the validity of the similarity assumption. This is because individuals in firms with high match rates may be expected to have a higher rank in the asset distribution than workers in firms with less generous match rates. This suggests that the distribution of U_d may be different across the treatment states.

Similarity may still hold in the presence of the employer match if the rank, U_d, in the asset distribution is insensitive to the match rate. The rank may be insensitive if, for example, individuals follow simple rules of thumb such as target saving when they make their savings decisions. Also, if the variation of match rates is small relative to the variation of individual heterogeneity or if the covariates capture most of the variation in match rates, then similarity will be satisfied approximately. Since the model is just-identified in our data, specification tests based on the implications of Theorem 1 may not be used to perform

[12] Clearly similarity holds in this case, $U_d \stackrel{d}{=} U_{d'}$ given V, X, and Z.

overidentifying tests. However, the quantile treatment effects model and estimator of Abadie, Angrist, and Imbens (2002), which apply only to binary treatment variables, provide a useful robustness check. While the approach of Abadie, Angrist, and Imbens (2002) and the approach presented in this paper generally identify and estimate different quantities, they will estimate the same thing when the assumptions of both models, including similarity, are satisfied and the set of compliers is representative of the population. If these conditions are not met, then the two estimators will in general have different probability limits, suggesting that a comparison of results based on the two models will provide evidence on the plausibility of these assumptions. Comparisons between the two estimators are presented with the empirical results in Section 4. The results from the two estimators are very similar, suggesting that employer matching does not result in a serious violation of rank similarity.

3.3 The Data

To estimate the QTE, we use data on a sample of households from Wave 4 of the 1990 Survey of Income and Program Participation (SIPP).[13] The sample is limited to households in which the reference person is 25-64 years old, in which at least one person is employed, and in which no one is self-employed.[14] The sample consists of 9,915 households and all dollar amounts are in 1991 dollars.

The 1991 SIPP reports household financial data across a range of asset categories. These data include a variable for whether a person works for a firm that offers a 401(k) plan. Households in which a member works for such a firm are classified as eligible for a 401(k). In addition, the survey also records the amount of 401(k) assets. Households with a positive 401(k) balance are classified as participants, while eligible households with a zero balance are considered non-participants.

[13]This sample has been used extensively to study the effect of 401(k) plans on wealth. See, for example, Benjamin (2003), Abadie (2003), Engen and Gale (2000), Engen, Gale, and Scholz (1996), and Poterba, Venti, and Wise (1994, 1995, 1996). The sample is often referred to as the 1991 SIPP because data were collected between February and May of 1991.

[14]Analyses are restricted to this sample because the SIPP only asks 401(k) questions to people 25 and older, because retirement and saving behavior of people over 65 would complicate the analysis, and because the self-employed and unemployed do not have access to 401(k)s. The household reference person is the person in whose name the family's home is owned or rented.

While there are several potential measures of wealth in the 1991 SIPP, we choose to focus our analysis on total wealth, net financial assets, and net non-401(k) financial assets. Net non-401(k) assets are defined as the sum of checking accounts, U.S. saving bonds, other interest-earning accounts in banks and other financial institutions, other interest-earning assets (such as bonds held personally), stocks and mutual funds less nonmortgage debt, and IRA balances. Net financial assets are net non-401(k) financial assets plus 401(k) balances, and total wealth is net financial assets plus housing equity and the value of business, property, and motor vehicles.[15]

We use the same set of covariates as Benjamin (2003). Specifically, we use age, income, family size, education, marital status, two-earner status, defined benefit (DB) pension status, IRA participation status, and home ownership status. Marital status, two-earner status, DB pension status, IRA participation status, and home ownership status are binary variables, where two-earner status indicates whether both household heads, where present, contribute to household income and DB pension status indicates whether the household's employer offers a DB pension plan. The education variable measures the number of years of school completed by the household reference person, and for the analysis we have categorized this variable into four groups: less than 12 years of education, 12 years of education, 13-15 years of education, and 16 or more years of education. Households are classified as IRA participants if they have positive IRA asset balances, and households are classified as home owners if the household has a positive home value. In addition, in the estimates reported below, we control for age using categorical variables: less than 30 years old, 30-35 years old, 36-44 years old, 45-54 years old, and 55 years old or older. Following Poterba, Venti, and Wise (1995), we control for income through the use of seven categorical variables. The income intervals are as follows: <\$10K, \$10-20K, \$20-30K, \$30-40K, \$40-50K, \$50-75K, and \$75K+.

Table 1 contains descriptive statistics for the full sample as well as by eligibility and participation status. 37% of the sample is eligible for a 401(k) plan while 26% choose to participate. Among those eligible for a 401(k) account, the participation rate is 70%. The descriptive statistics indicate that participants have larger holdings of all measures of wealth that we consider. As expected, the means of all of the wealth variables are substantially larger than their medians, indicating the high degree of skewness in wealth. The means also show that 401(k) participants have more income, are more likely to be married, are more

[15]Housing equity is defined as housing value less mortgage.

likely to have IRAs and defined benefit pensions, are more likely to be home-owners, and are more educated than non-participants. Average age and family size are similar between the two groups. Descriptive statistics for the dependent variables by income category are also provided in Table 2.

3.4 Empirical Results

3.4.1 Estimation and Inference Procedures

To capture the effects of 401(k) participation on net financial assets, we estimate linear quantile models of the form

$$Q_{Y_d|X}(\tau) = d\alpha(\tau) + X'\beta(\tau)$$

where d indicates 401(k) participation status and is instrumented for by 401(k) eligibility, following Abadie (2003) and Poterba, Venti, and Wise (1994, 1995, 1996).[16] The outcomes Y are the three previously mentioned measures of wealth (total wealth, net financial assets, and net non-401(k) financial assets), and X consists of dummies for income category, dummies for age category, dummies for education category, a marital status indicator, family size, two-earner status, DB pension status, IRA participation status, home ownership status, and a constant.[17] To more fully control for income, we also consider estimates obtained within each income category. In these cases, the income category dummies are omitted and a linear term in income is included to account for any remaining variation within income category.

The main results reported below are for the standard quantile regression (QR) estimator and the instrumental quantile regression (IQR) estimator of Chernozhukov and Hansen (2001) which corrects for the endogeneity of 401(k) participation under the assumptions of the model presented in Section 2 of this paper. The IQR estimator may be viewed as a

[16]The OLS and 2SLS estimates are based on analogous specifications.

[17]We also considered alternate specifications of the covariate vector. However, the estimate of the treatment effect was found to be largely insensitive to the specification. The most substantial difference is that when the home ownership dummy was excluded the results for total wealth closely tracked those of net financial assets across the entire distribution, indicating little or no substitution between 401(k) assets and other forms of wealth. All other results were very similar.

convenient method of approximately solving the sample analog of moment equations (3.2):[18]

$$\frac{1}{n}\sum_{i=1}^{n}(1(Y_i \leq D_i'\widehat{\alpha} + X_i'\widehat{\beta}) - \tau)(X_i', Z_i')' = o_p\left(\frac{1}{\sqrt{n}}\right). \quad (3.4)$$

When there is only one endogenous regressor and the model is just identified, the IQR estimator for a given quantile may be computed as follows:

1. Run a series of standard quantile regressions of $Y - D\alpha_j$ on covariates X and instrument Z where $\{\alpha_j\}$ is a grid over α.

2. Take the α_j that minimizes the absolute value of the coefficient on Z as the estimate of α, $\widehat{\alpha}$. Estimates of β, $\widehat{\beta}$, are then the corresponding coefficients on X.

In Chernozhukov and Hansen (2001), we show that, under regularity conditions and for $\theta = [\alpha, \beta']'$,

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, J^{-1}\Omega(J^{-1})'),$$

where, for $\Psi = [Z, X']'$ and $\epsilon = Y - D\alpha - X'\beta$,

$$\Omega = \tau(1-\tau)E\Psi\Psi' \text{ and } J = E[f_\epsilon(0|D, X, Z)\Psi[D, X']].$$

Chernozhukov and Hansen (2001) also provides further details covering estimation and asymptotic theory in the general, potentially overidentified model.

Estimates of the QTE, $\alpha(\tau)$, for many different points τ also provide an estimate of the QTE process $\alpha(\cdot)$ which treats α as a function of τ.[19] Knowledge of the QTE process allows formal testing of a number of interesting hypotheses. These include the constant effect hypothesis ($\alpha(\cdot) = \alpha$), of which the hypothesis of no effect ($\alpha(\cdot) = 0$) is a special case, and the hypothesis of no endogeneity ($\alpha(\cdot) = \alpha_{QR}(\cdot)$ where α_{QR} denotes the ordinary quantile regression estimate). If the constant effect hypothesis is not rejected, the distributional impact of the treatment may be captured by a single statistic, such as the mean or the median treatment effect. Also, failure to reject the hypothesis of no endogeneity suggests that the endogeneity bias is not statistically important and that standard QR estimates

[18]Estimation using a similar set of moment equations was considered by Abadie (1997) who noted the computational difficulty in obtaining their solution.

[19]The following discussion also applies to the coefficients of the covariates, $\beta(\tau)$.

may be used. Chernozhukov and Hansen (2002) provides asymptotic theory for the IQR process and suggests a computationally attractive method for performing inference on the IQR process.

3.4.2 OLS and 2SLS Results

Table 3 provides OLS and 2SLS results of the participation effect. These estimates serve as a benchmark for the quantile and instrumental quantile regression estimates presented later. In addition, they are interesting in their own right. Indeed, in the case of a constant treatment effect, these estimates would be sufficient to fully characterize the distributional impact of the treatment.[20]

The first stage estimates, reported in the third column of Table 3, confirm that eligibility for a 401(k) is highly correlated with participation. In the full sample and within each income category, the first stage estimate is large, positive, and highly significant. Indeed, conditional on eligibility, the rest of the covariates have very small effects on 401(k) participation.

In the full sample, the 2SLS estimates are uniformly smaller than the OLS estimates, confirming the intuition that the OLS estimates should be upward biased. However, the biases appear to be modest, especially when compared to the standard errors of the estimates. After accounting for endogeneity, the impact of 401(k) participation on both total wealth and net financial assets remains large and significant. Relative to the means, 401(k) participation increases net financial assets by approximately 70% and total wealth by approximately 14%. The magnitude of both effects is also quite similar, though slightly larger for net financial assets, suggesting little substitution between 401(k) assets and other forms of wealth. On the other hand, 401(k) participation has relatively little impact on net non-401(k) financial assets. Neither the OLS nor the 2SLS estimate of the effect of participation on net non-401(k) financial assets is significantly different from zero, and both are quite small in magnitude. Overall, these results suggest that the majority of the increase in net financial assets may be attributed to new saving due to 401(k) plans and not to substitution from other forms of wealth.

The results by income category provide additional evidence on substitution patterns.

[20]The process tests reported below suggest that this is the case when the dependent variable is total wealth or net non-401(k) financial assets.

The loss of precision resulting from estimating the treatment effect within income categories makes drawing any firm conclusion difficult, but the patterns of the estimates are still quite interesting.[21] The impact of 401(k) participation on net financial assets is uniformly positive and significant and tends to increase as one moves from lower to higher income categories. This result appears to be consistent with the resource constraints of the different income groups. The results for net non-401(k) financial assets are never significantly different from zero. However, in all cases but one, the point estimate is negative and non-negligible, which provides weak evidence that there is financial asset substitution that was obscured in the results obtained in the full sample. While the results for total wealth show much less of a pattern as one looks across income categories, it can be seen that in no case is the effect significantly different from zero. The point estimates are uniformly positive and, in the majority of cases, are reasonably large. This again provides weak evidence that 401(k) participation increases total wealth by a modest amount, but that this increase is smaller than the increase to net financial assets, indicating substitution between 401(k) and other assets.

3.4.3 Quantile Regression and Instrumental Quantile Regression Results: Full Sample

While the OLS and 2SLS results presented above provide a summary statistic for the impact of the treatment, they fail to capture the distributional impact of 401(k) participation on wealth. To further explore the effect of 401(k) participation on wealth, we report results obtained from both standard quantile regression and the instrumental quantile regression of Chernozhukov and Hansen (2001) in Figure 1.

The left column of Figure 1 contains QR estimates of the effect of 401(k) participation on the wealth measures, and the right column of Figure 1 presents the IQR estimates of the QTE. The shaded region in the first six panels represents the 95% confidence interval.[22] The last two panels plot the estimated effects for each of the dependent variables together

[21]In the following, we ignore estimates in the lowest income category which are greatly influenced by outliers in the upper tail of the distribution and the small sample size. The influence of the upper tail is seen clearly in the quantile regression results presented below.

[22]Standard errors were estimated using heteroskedasticity consistent standard errors as in Powell (1984, 1986) and Buchinsky (1995) using the methods outlined in Chernozhukov and Hansen (2001).

to provide a comparison of the magnitudes and to facilitate the discussion of substitution between the different wealth measures.

The results exhibit a number of striking features. First, the difference between the QR and IQR estimators is not dramatic. Both exhibit the same pattern of results, though there is some upward bias evident in the QR estimates. This bias is most evident in the estimates for net financial assets and net non-401(k) financial assets, but is hardly noticeable in the total wealth results.

Another interesting feature of the results is that the effect of participation on net financial assets is highly non-constant, appearing to increase monotonically in the quantile index. This result suggests that, conditional on income and other observables, people who rank higher in the conditional wealth distribution are impacted far more than those ranking lower in the conditional distribution. In addition, the effect is strongly positive across the entire distribution. While these results correspond to our intuition, there is actually no other a priori reason to believe that net financial assets must react in this way. In particular, if people were simply substituting financial assets held in 401(k)s for other forms of financial assets, the effect of 401(k) participation on net financial assets would be zero. These results provide strong evidence against this hypothesis at all quantiles.

The impact of 401(k) participation on total wealth relative to its impact on net financial assets also provides interesting insights. As with net non-401(k) financial assets, the effect of participating in a 401(k) on total wealth is roughly constant, though in this case it is uniformly positive. The most interesting feature of the effect on total wealth is that for low quantiles it is of almost the same magnitude as the effect on net financial assets, while it is substantially smaller than the effect on net financial assets in the upper tail of the distribution. Taken together, these findings suggest that the increase in net financial assets observed in the lower tail of the conditional assets distribution can be interpreted as an increase in wealth, while the increase in the upper tail of the distribution is being mitigated by substitution with some other (non-financial) component of wealth. However, even for the highest quantiles, the substitution does not appear to be complete.

A final outstanding feature of the results is the indication that 2SLS estimates substantially overstate the treatment effect across a large range of the net financial asset distribution. In fact, the 2SLS estimates of the treatment effect on net financial assets correspond much more closely to the treatment effect at the 75^{th} percentile of the distribution than to that of the median.

In order to strengthen and further develop our conclusions, we present test results based on the empirical instrumental quantile regression process computed using the methods of Chernozhukov and Hansen (2002). Kolmogorov-Smirnov (KS) test statistics and 95% critical values are given in Table 4. The test results lend further support to the conclusions already drawn. The tests strongly reject the hypothesis that the impact of 401(k) participation on net financial assets is constant and confirm that the impact is significantly different from zero. In addition, we see that the hypothesis of exogeneity of treatment is rejected for net financial assets. However, the tests fail to reject both the hypothesis of a constant treatment effect (equal to the median effect) and the hypothesis of exogeneity for total wealth and net non-401(k) financial assets. That the treatment effect for both total wealth and net non-401(k) financial assets is statistically constant adds further credibility to the conclusion that there is little substitution between 401(k) assets and other forms of wealth in the low tail of the assets distribution but that there is substantial substitution in the upper tail. In addition, the results of the exogeneity tests provide some evidence that there is endogeneity bias in the conventional QR estimates of the treatment effects.

3.4.4 Quantile Regression and Instrumental Quantile Regression Results: By Income Interval

As with analysis of the mean effect presented above, additional insights about the QTE may be gained by examining the effect of 401(k) participation on our chosen wealth measures within income interval. The independence assumption, A2, may also be more plausible within income categories due to the finer conditioning on income, since the arguments of Poterba, Venti, and Wise (1995) suggest that 401(k) eligibility is as good as randomly assigned once income is conditioned upon. Of course, the estimates within income category do suffer from a loss of precision relative to estimates obtained with a coarser income control, which makes drawing firm inferences more difficult.

IQR estimation results by income category are reported in Figures 2-5.[23] The figures are arranged by dependent variable, with Figure 2 corresponding to net financial assets, Figure 3 to net non-401(k) financial assets, and Figure 4 to total wealth. In all cases, the shaded region

[23]QR results are not reported but are quite similar to the IQR results.

represents the 95% confidence interval.[24] Figure 5 contains plots of the estimated effects for each of the dependent variables together to facilitate comparison of the magnitudes. Table 5 reports process test results.

Within income categories, the results for net financial assets follow roughly the same pattern as the results in the full sample. In all categories, the results are generally increasing in the quantile index, and in all but the first income category, the process tests reveal that the treatment effect is different from zero. In addition, the hypothesis of a constant effect is rejected in all but the first and last income categories. As would be expected, the magnitudes of the results increases as income increases. The point estimates in the first category are close to zero across the majority of the quantiles, suggesting that participation in a 401(k) has little effect on those with incomes of less than $10,000. Also, in each income interval, the results are fairly constant and quite modest for quantiles below the median. Overall, these results indicate that 401(k) participation increases accumulated net financial assets in all, excepting possibly the first, income categories, but that these effects may be quite modest through much of the distribution.

As with the results in the full sample, the estimated treatment effect of 401(k) participation on net non-401(k) financial assets is not significantly different from zero in any case. The point estimates are also generally quite small, though they do exhibit some tendency to be negative more often than positive. This negative tendency provides weak evidence for some substitution between financial assets held in 401(k)s and other forms of financial assets. That this negative tendency appears to be most pronounced for low quantiles also suggests that those with low preferences for saving, who probably have relatively little in the form of financial assets, are choosing to accumulate assets within 401(k)s instead of elsewhere, whereas those with higher preferences for saving are saving in both locations.

The results for the effect of 401(k) participation on total wealth are the most varied across income categories, though the lack of precision makes comparison difficult. One result which is quite interesting is that, within the lowest income category, there appear to be extreme outliers in the upper tail of the distribution. Examining the quantile results within the first income category suggests there is little effect of 401(k) participation on wealth across the majority of the wealth distribution. However, at approximately the 60^{th} percentile, the

[24]Standard errors were estimated using heteroskedasticity consistent standard errors as in Powell (1984, 1986) and Buchinsky (1995) using the methods outlined in Chernozhukov and Hansen (2001).

effects increase dramatically. These large effects in the upper tail also explain the anomalous OLS and 2SLS results within the first income category illustrated in Table 2. The process test of no effect does not reject within the first income category, which seems to be a plausible conclusion given the small effect for most quantiles. It is also interesting that in the highest income category the estimated participation effect on total wealth is close to zero in the upper quantiles of the wealth distribution while the estimated effect on net financial assets is quite large, suggesting a large amount of substitution in these quantiles. Overall, it is difficult to draw any firm conclusions due to large estimated standard errors of the effects. However, one robust finding seems to be that the estimated effect of participation on total wealth and the estimated effect of participation on net financial assets are quite similar in the lower tail of the wealth distribution, which suggests that participation in 401(k) plans stimulates asset accumulation of those with low preferences for saving.

A final interesting note is that, within income categories, the hypothesis of the exogeneity of 401(k) participation is never rejected. This could be because, conditional on income and other covariates, 401(k) participation is as good as randomly assigned, or it could be driven by small sample size and the lack of precision of the estimates. We choose to focus on the IQR estimates because they are robust to endogeneity, but there is no statistical evidence that endogeneity is present.

3.4.5 Comparison with Abadie, Angrist, and Imbens (2002)

One key criticism of the approach pursued thus far in this paper is that employer matching practices may invalidate the similarity assumption required in the model in Section 2. However, since both the instrument and endogenous variable are binary, the model and approach of Abadie, Angrist, and Imbens (2002) apply. A comparison between the results from the two approaches then provides a specification check of the developed results.

The estimator of Abadie, Angrist, and Imbens (2002) is developed within the LATE framework of Imbens and Angrist (1994). In particular, Abadie, Angrist, and Imbens (2002) show that if

1. the instrument Z is independent of the outcome error, U_d in our notation, and the error in the selection equation, V in our notation,

2. monotonicity, $P(D_1 \geq D_0|X) = 1$ where D_1 is the treatment state of an individual when $Z = 1$ and D_0 is defined similarly, holds,

3. and other standard conditions are met,

then the QTE for compliers, those individuals with $D_1 > D_0$, is identified and develop an estimator for the QTE for compliers. Since only individuals eligible for a 401(k) can participate, monotonicity holds trivially, and the QTE for compliers corresponds to the QTE for the treated, which will correspond to the quantity identified by the IVQT model of Section 2 if the treated are representative of the population and the assumptions of the IVQT model are satisfied.

Given that the two models are mutually compatible under the conditions outlined above and the monotonicity assumption of Abadie, Angrist, and Imbens (2002) holds in the case of 401(k) participation, a comparison of the previous results obtained via IQR and results obtained via the estimator of Abadie, Angrist, and Imbens (2002) provides a useful robustness check of the previous results and the assumptions that underlie their interpretation. Figure 6 reports results from the estimator of Abadie, Angrist, and Imbens (2002) in the full sample and comparisons with corresponding IQR estimates.[25] From this exercise, we see that the pattern of results obtained from the two estimators are quite similar, with the major differences being that the Abadie, Angrist, and Imbens (2002) estimates of the effects of 401(k) participation on total wealth and on net non-401(k) financial assets appear to be even more constant than those obtained through IQR.[26]

It appears that the difference in the estimates is small relative to sampling variation and that one would not draw substantively different conclusions from either set of estimates. The striking similarity between the estimates provides further support for the IQR results discussed above and strongly suggests that employer matching of 401(k) contributions does not result in failure of rank similarity.

[25]The Abadie, Angrist, and Imbens (2002) estimator may be computed by running weighted quantile regression, where the weights are nonparametrically estimated. In our analysis, we used series methods to estimate the weights. The exact parameterization used to estimate the weights is available upon request. We also found that the overall results were not sensitive to the exact specification used to estimate the weights.

[26]Estimates using the estimator of Abadie, Angrist, and Imbens (2002) within income categories were also very similar to the IQR estimates previously reported.

3.4.6 Overall Conclusions and Cautions.

Overall, the results indicate that 401(k) participation has a positive impact on the accumulation of net financial assets. The results suggest that the effect on net financial assets is increasing as one approaches the upper tail of the net financial asset distribution. Estimates for the effect of 401(k) participation on total wealth and net non-401(k) financial assets are approximately constant and indicate that 401(k) participation generally increases total wealth but has little effect on net non-401(k) financial assets. We interpret these results as indicating that participation in 401(k)s increases total wealth and that there is little substitution between financial assets in 401(k)s and other financial assets. In addition, the results suggest that there is substitution between assets held in 401(k)s and other components of wealth in the upper tail of the wealth distribution, but that most financial assets held in 401(k)s in the lower tail of the distribution represent new savings. This has important policy implications, as the people in the low tail of the net financial asset distribution are also likely to be the people with the lowest retirement savings.

The estimates also clearly indicate the inability of a single summary statistic, such as the 2SLS regression estimate of the treatment impact, to provide a clear picture of the impact of a program on the distribution of the outcomes of interest. The 2SLS estimate for the effect of 401(k) participation on net financial assets appears to overstate the actual treatment across much of the distribution, corresponding most closely to the estimates for the upper tail of the asset distribution. In addition, the single summary statistic provided by 2SLS or OLS obscures the regions where divergences between the effect of 401(k) participation on the different wealth measures occur and thus do not provide as full a description of the program impact as the quantile-based methods.

While we feel that this paper provides insight into the effect of 401(k) participation on wealth, it does suffer from limitations. First, all of the dependent variables used in this analysis represent stocks of assets rather than the flow of savings. The accumulated level of assets is interesting because it provides a summary of a person's wealth and the resources that are available to the individual. However, they are not sufficient to capture the effect of the program on savings. In particular, given employer matching and the tax-advantaged nature of 401(k) saving, it may be possible to have a large increase in accumulated assets with little change in the individual's flow of savings. Second, the data available in the SIPP do not report all sources of pension wealth. In particular, the SIPP does not contain information on assets held in DB plans or defined contribution plans other than IRAs and 401(k)s. The lack of these data could potentially bias the results upward if 401(k) assets are substituting for these other forms of assets. While evidence from Poterba, Venti, and Wise (2001) and Papke, Peterson, and Poterba (1996) is consistent with the view that 401(k)s rarely cause DB termination, it does not preclude substantial substitution between the different forms of pensions.

3.5 Conclusion

In this paper, we apply the instrumental quantile regression model and estimators developed in Chernozhukov and Hansen (2001) and Chernozhukov and Hansen (2002) to data from the SIPP, which has previously been used by Poterba, Venti, and Wise (1996), Abadie (2003), Benjamin (2003), and Engen, Gale, and Scholz (1996), to examine the effects of 401(k) plans on savings. Following Poterba, Venti, and Wise (1996), Abadie (2003), and Benjamin (2003), we use 401(k) eligibility as an instrument for 401(k) participation to estimate the QTE of participation in a 401(k) plan on various wealth measures. The QTE provide a more full characterization of the effect of 401(k) participation on savings than do conventional IV methods and supplement these methods by providing a more detailed description of the distributional impact of 401(k) program participation.

The IQR estimates suggest that the effect of 401(k) participation on net financial assets is quite heterogeneous, with the largest returns accruing to those who are in the upper tail of the assets distribution. The results also indicate that the effect of 401(k) participation on net financial assets is positive and significant over the entire range of the asset distribution and that the effect is monotonically increasing in the quantile index. Effects on total wealth and net non-401(k) financial assets, on the other hand, appear to be constant, and the effect on net non-401(k) financial assets is not significantly different from zero while the effect on total wealth is positive and significant. Overall, the results suggest that participation in 401(k)s increases net financial assets across the asset distribution, but that this effect is mitigated by substitution with other forms of wealth in the upper tail of the distribution. They also demonstrate that estimates of treatment effects which focus on a single feature of the outcome distribution may fail to capture the full impact of the treatment and that examining additional features may enhance our understanding of the economic relationships involved.

Bibliography

ABADIE, A. (1997): "Changes in Spanish Labor Income Structure During the 1980s: A Quantile Regression Approach," *Investigaciones Económicas*, 21(2), 253–272.

——— (2003): "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, 113(2), 231–263.

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

AMEMIYA, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711.

BENJAMIN, D. J. (2003): "Does 401(k) eligibility increase saving? Evidence from propensity score subclassification," *Journal of Public Economics*, 87(5-6), 1259–1290.

BUCHINSKY, M. (1995): "Estimating the asymptotic covariance matrix for quantile regression models: A Monte Carlo study," *Journal of Econometrics*, 68, 303–338.

CHEN, L., AND S. PORTNOY (1996): "Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models," *Comm. Statist. Theory Methods*, 25(5), 1005–1032.

CHERNOZHUKOV, V., AND C. HANSEN (2001): "An IV Model of Quantile Treatment Effects," Sepetmber, Department of Economics Working Paper (download at WWW.SSRN.COM).

——— (2002): "Inference for Distributional Effects using Instrumental Quantile Regression," March, MIT Department of Economics Working Paper (download at WWW.SSRN.COM).

DOKSUM, K. (1974): "Empirical probability plots and statistical inference for nonlinear models in the two-sample case," *Ann. Statist.*, 2, 267–277.

Employee Benefit Research Institute (1997): *Fundamentals of Employee Benefit Programs.* EBRI, Washington, DC.

Engen, E. M., and W. G. Gale (2000): "The effects of 401(k) plans on household wealth: differences across earnings groups," NBER Working Paper 8032.

Engen, E. M., W. G. Gale, and J. K. Scholz (1996): "The illusory effects of saving incentives on saving," *Journal of Economic Perspectives*, 10(4), 113–138.

Heckman, J., and R. Robb (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in *Drawing Inference from Self-Selected Samples*, ed. by H. Wainer, pp. 63–107. Springer-Verlag, New York.

Heckman, J. J., and J. Smith (1997): "Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts," *Rev. Econom. Stud.*, 64(4), 487–535.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Koenker, R., and G. S. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Papke, L. E., M. Peterson, and J. M. Poterba (1996): "401(k) plans and tax-deferred savings," in *Advances in the Economics of Aging*, ed. by D. A. Wise. University of Chicago Press: Chicago.

Poterba, J. M., S. F. Venti, and D. A. Wise (1994): "401(k) plans and tax-deferred savings," in *Studies in the Economics of Aging*, ed. by D. A. Wise. University of Chicago Press: Chicago.

——— (1995): "Do 401(k) contributions crowd out other personal saving?," *Journal of Public Economics*, 58, 1–32.

——— (1996): "Personal retirement saving programs and asset accumulation: Reconciling the evidence," NBER Working Paper 5599.

——— (2001): "The transition to personal accounts and increasing retirement wealth: Macro and micro evidence," NBER Working Paper 8610.

POWELL, J. L. (1986a): "Censored regression quantiles," *Journal of Econometrics*, 32, 143–155.

——— (1986b): "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics*, 25, 303–325.

VYTLACIL, E. (2000): "Semiparametric Identification of the Average Treatment Effect in Nonseparable Models," mimeo.

TABLE 1. MEANS, STANDARD DEVIATIONS, AND MEDIANS

	Entire Sample	By 401(k) Participation: Participants	By 401(k) Participation: Non-Participants	By 401(k) Eligibility: Eligibles	By 401(k) Eligibility: Non-Eligibles
Treatment:					
Participation in 401(k)	0.26			0.70	0.00
	(0.44)			(0.46)	(0.00)
Instrument:					
Eligibility for 401(k)	0.37	1.00	0.15		
	(0.48)	(0.00)	(0.36)		
Outcome Variables:					
Net Financial Assets	18,051	38,262	10,890	30,347	10,788
	(63,523)	(79,088)	(55,257)	(74,800)	(54,518)
	[1,499]	[15,249]	[200]	[9,122]	[145]
Net Non-401(k) Assets	13,877	22,775	10,724	19,396	10,617
	(59,605)	(70,415)	(54,930)	(67,439)	(54,192)
	[542]	[3,830]	[200]	[2,711]	[130]
Total Wealth	63,817	96,920	52,088	86,240	50,571
	(111,530)	(127,790)	(102,646)	(124,006)	(101,155)
	[25,100]	[53,441]	[16,645]	[45,356]	[14,640]
Covariates:					
Income	37,201	49,367	32,890	46,862	31,494
	(24,774)	(27,208)	(22,316)	(25,958)	(22,151)
Age	41.06	41.51	40.90	41.48	40.81
	(10.34)	(9.66)	(10.57)	(9.61)	(10.75)
Family Size	2.87	2.92	2.85	2.90	2.84
	(1.54)	(1.47)	(1.56)	(1.48)	(1.57)
Married	0.60	0.69	0.57	0.67	0.56
	(0.49)	(0.46)	(0.49)	(0.47)	(0.50)
Participation in IRA	0.24	0.36	0.20	0.32	0.20
	(0.43)	(0.48)	(0.40)	(0.47)	(0.40)
Defined Benefit Pension	0.27	0.39	0.23	0.42	0.19
	(0.44)	(0.49)	(0.42)	(0.49)	(0.39)
Home Owner	0.64	0.77	0.59	0.74	0.57
	(0.48)	(0.42)	(0.49)	(0.44)	(0.49)
Years Education:					
< 12	0.13	0.07	0.15	0.07	0.16
	(0.33)	(0.25)	(0.36)	(0.26)	(0.37)
12	0.38	0.35	0.39	0.35	0.39
	(0.48)	(0.48)	(0.49)	(0.48)	(0.49)
> 12 and < 16	0.24	0.26	0.24	0.26	0.24
	(0.43)	(0.44)	(0.43)	(0.44)	(0.42)
> 16	0.25	0.33	0.22	0.32	0.21
	(0.43)	(0.47)	(0.42)	(0.47)	(0.41)

Note: The sample is drawn from the 1991 SIPP and consists of 9,915 observations. The observational units are household reference persons aged 25-64 and spouse if present. Households are included in the sample if at least one person is employed and no one is self-employed. Standard deviations are in parentheses, and medians are in brackets.

TABLE 2. MEANS, STANDARD DEVIATIONS, AND MEDIANS OF ASSET MEASURES BY INCOME INTERVAL

	Income						
	<$10K	$10-20K	$20-30K	$30-40K	$40-50K	$50-75K	$75K+
Net Financial Assets	735	2,308	6,311	11,938	19,348	33,708	83,709
	(10,827)	(15,498)	(30,615)	(43,519)	(54,773)	(66,894)	(157,168)
	[0]	[0]	[400]	[2,053]	[5,761]	[14,500]	[43,779]
Net Non-401(k) Financial Assets	431	1,543	4,979	8,775	14,942	25,179	66,999
	(9,143)	(14,699)	(29,525)	(40,991)	(52,718)	(62,438)	(151,627)
	[0]	[0]	[110]	[651]	[3,437]	[8,676]	[29,800]
Total Wealth	16,235	21,620	36,730	55,119	74,006	105,285	202,240
	(40,772)	(43,631)	(67,659)	(83,203)	(97,913)	(119,531)	(226,077)
	[1,258]	[4,225]	[12,500]	[29,224]	[44,197]	[71,025]	[152,500]
N	638	1,948	2,074	1,712	1,204	1,572	767

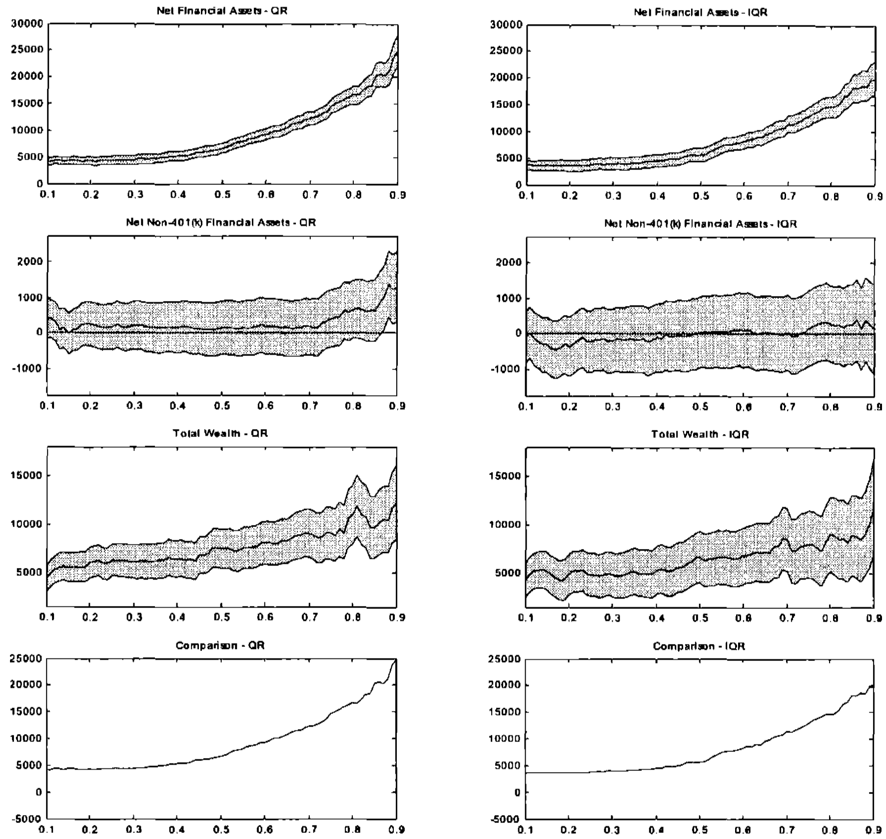
Note: The sample is drawn from the 1991 SIPP and consists of 9,915 observations. The observational units are household reference persons aged 25-64 and spouse if present. Households are included in the sample if at least one person is employed and no one is self-employed. Standard deviations are in parentheses, and medians are in brackets.

TABLE 3. OLS AND 2SLS ESTIMATES OF EFFECT OF 401(K) PARTICIPATION

Sample	N	First Stage	Net Financial Assets OLS	Net Financial Assets 2SLS	Net Non-401(k) Financial Assets OLS	Net Non-401(k) Financial Assets 2SLS	Total Wealth OLS	Total Wealth 2SLS
A. Full Sample								
Full Sample	9,915	0.697	14,250	13,087	778	-355	10,694	9,259
		(0.006)	(1,551)	(1,922)	(1,477)	(1,855)	(2,388)	(3,035)
B. By Income Interval								
<$10K	638	0.711	9,843	9,149	4,093	3,443	20,464	17,224
		(0.020)	(4,921)	(4,914)	(3,447)	(3,527)	(11,311)	(11,518)
$10-20K	1,948	0.650	5,591	5,352	-759	-917	4,729	6,138
		(0.013)	(1,463)	(1,629)	(1,227)	(1,427)	(2,265)	(3,218)
$20-30K	2,074	0.627	7,083	4,143	448	-2,518	5,462	0.183
		(0.013)	(1,315)	(2,268)	(1,124)	(2,152)	(3,119)	(4,502)
$30-40K	1,712	0.672	12,136	10,273	1,077	-909	10,683	4,881
		(0.015)	(2,513)	(2,880)	(2,297)	(2,677)	(3,891)	(5,103)
$40-50K	1,204	0.723	12,858	9,980	500	-2,479	13,470	13,205
		(0.018)	(2,470)	(3,741)	(2,330)	(3,646)	(4,905)	(6,675)
$50-75K	1,572	0.744	20,800	21,920	1,803	2,985	12,881	12,202
		(0.017)	(3,010)	(3,444)	(2,876)	(3,310)	(5,132)	(6,718)
$75K+	767	0.831	23,103	24,013	-6,735	-5,252	5,514	10,470
		(0.022)	(10,417)	(12,895)	(10,228)	(12,713)	(13,645)	(17,174)

Note: The table reports OLS and 2SLS estimates of the effect of 401(k) participation on wealth. The second column contains the sample size used for the estimates in each row. The third column reports the first stage coefficient estimate from a regression of 401(k) participation on 401(k) eligibility and covariates. Covariates are as described in the main text. Heteroskedasticity robust standard errors are given in parentheses.

FIGURE 1. QR AND IQR ESTIMATES OF EFFECT OF 401(K) PARTICIPATION



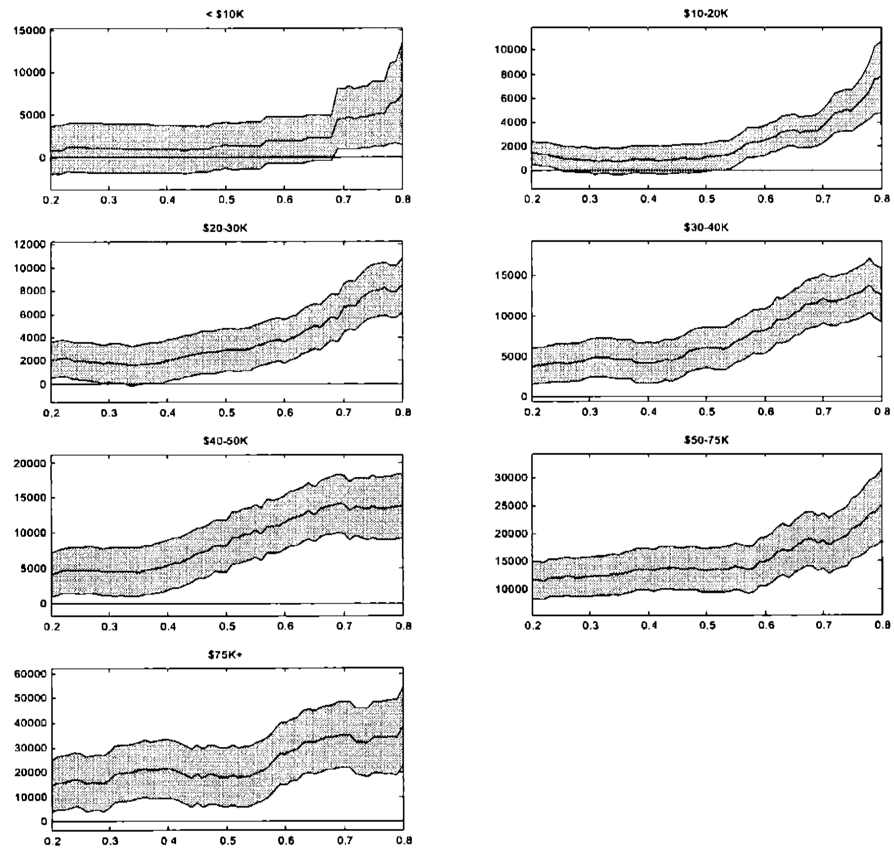
Note: The sample size is 9,915. The left column contains standard quantile regression estimates, and the right column contains instrumental quantile regression. Each panel is labeled with the dependent variable used in estimation of the presented results. The bottom panel in each column compares the point estimates for each wealth measure. The solid line corresponds to net financial assets, the dashed line to net non-401(k) financial assets, and the dash-dot line to total wealth. The vertical axis measures the dollar increase in the wealth measure due to 401(k) participation. The quantile of the conditional wealth distribution is on the horizontal axis. Covariates are as described in the main text. The shaded region is the 95% confidence band using robust standard errors. Estimates are reported for $\tau \in [.10, .90]$ at .01 unit intervals.

TABLE 4. TESTS ON THE INSTRUMENTAL QUANTILE REGRESSION PROCESS IN THE FULL SAMPLE

Null Hypothesis	Net Financial Assets		Net Non-401(k) Financial Assets		Total Wealth	
	Statistic	$c_{.95}$	Statistic	$c_{.95}$	Statistic	$c_{.95}$
No Effect	12.875	3.009	0.921	2.882	4.538	3.003
Constant Effect	9.093	3.321	0.843	3.452	1.850	3.213
Exogeneity	3.851	3.209	2.287	3.056	1.899	3.086

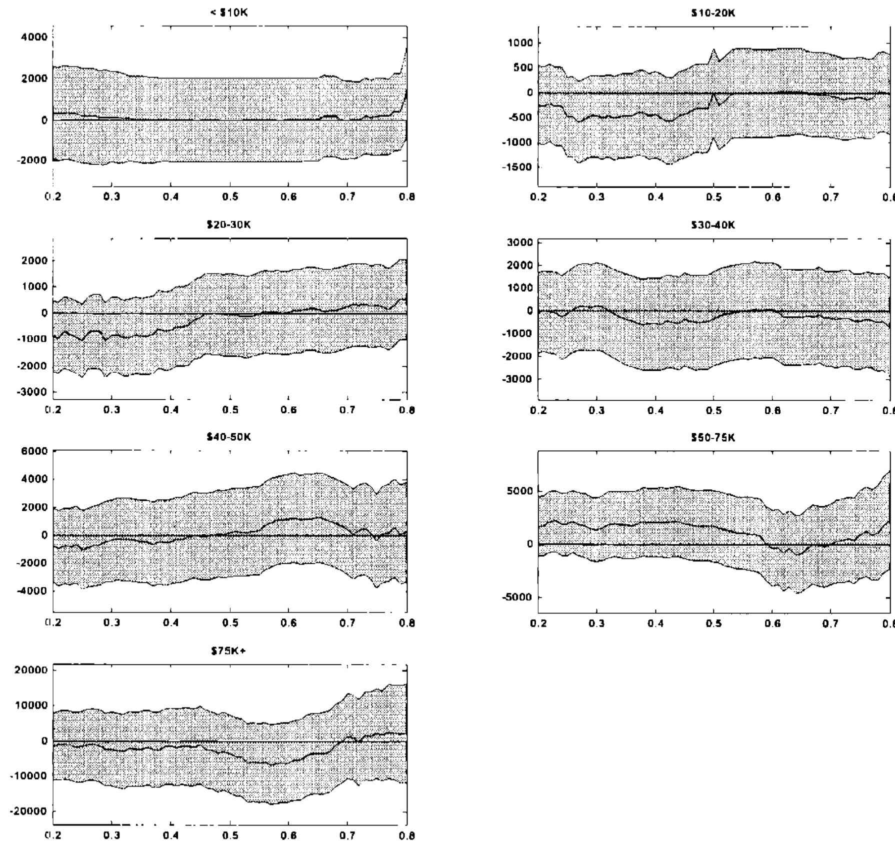
Note: The table report inference results on the inverse quantile regression process. Reported are Kolmogorov-Smirnov statistics and 95% critical values. The statistics and critical values are computed using the methods in Chernozhukov and Hansen (2002). The null hypotheses tested are as follows: no effect - $\alpha(\cdot) = 0$, constant effect - $\alpha(\cdot) = \alpha(.5)$, exogeneity - $\alpha(\cdot) = \alpha_{QR}(\cdot)$, where $\alpha(\cdot)$ denotes the instrumental quantile regression process and $\alpha_{QR}(\cdot)$ denotes the quantile regression process.

FIGURE 2. EFFECT OF 401(K) PARTICIPATION ON NET FINANCIAL ASSETS BY INCOME INTERVAL



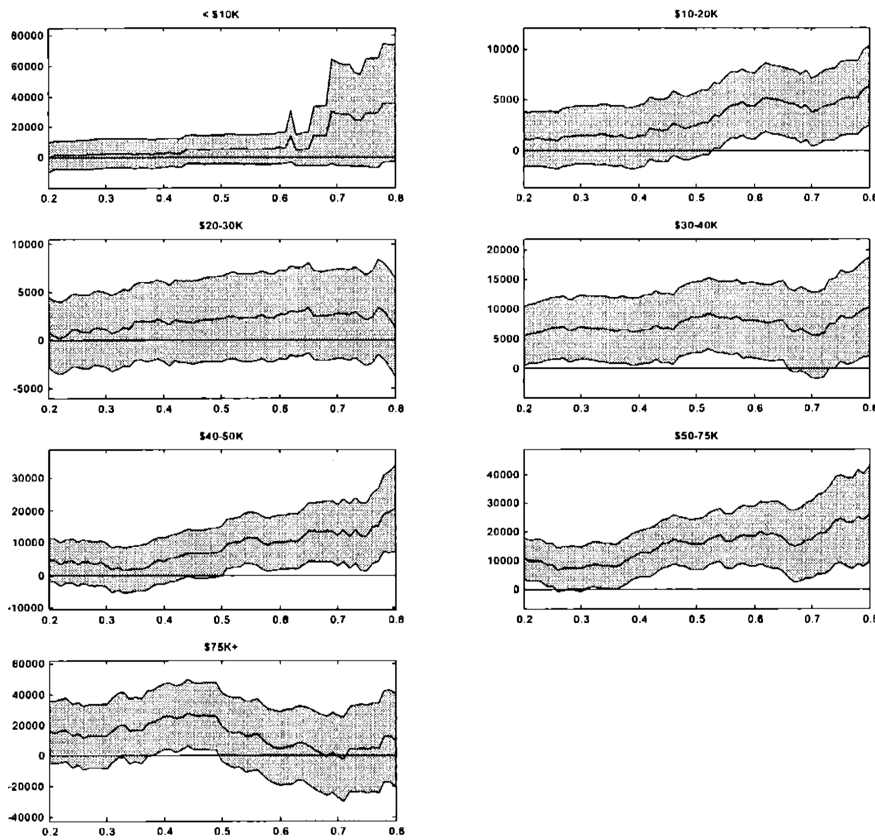
Note: The figure reports the effect of 401(k) participation on net financial assets by income interval. Each panel is labeled with the income interval to which it corresponds. The vertical axis measures the dollar increase in net financial assets due to 401(k) participation. The quantile of the conditional net financial assets distribution is on the horizontal axis. Covariates are as described in the main text. The shaded region is the 95% confidence band using robust standard errors. Estimates are reported for $\tau \in [.20, .80]$ at .01 unit intervals.

Figure 3. Effect of 401(k) Participation on Non-401(k) Financial Assets by Income Interval



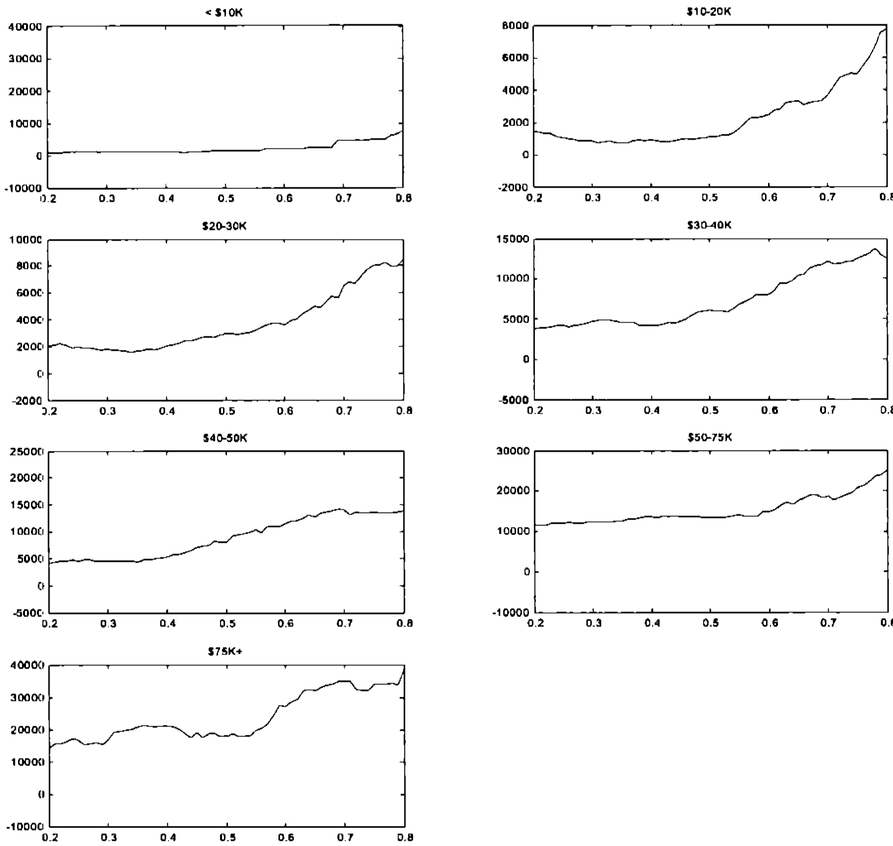
Note: The figure reports the effect of 401(k) participation on net non-401(k) financial assets by income interval. Each panel is labeled with the income interval to which it corresponds. The vertical axis measures the dollar increase in net non-401(k) financial assets due to 401(k) participation. The quantile of the conditional net non-401(k) financial assets distribution is on the horizontal axis. Covariates are as described in the main text. The shaded region is the 95% confidence band using robust standard errors. Estimates are reported for $\tau \in [.20, .80]$ at .01 unit intervals.

FIGURE 4. EFFECT OF 401(K) PARTICIPATION ON TOTAL WEALTH BY INCOME INTERVAL



Note: The figure reports the effect of 401(k) participation on total wealth by income interval. Each panel is labeled with the income interval to which it corresponds. The vertical axis measures the dollar increase in total wealth due to 401(k) participation. The quantile of the conditional total wealth distribution is on the horizontal axis. Covariates are as described in the main text. The shaded region is the 95% confidence band using robust standard errors. Estimates are reported for $\tau \in [.20, .80]$ at .01 unit intervals.

Figure 5. Effect of 401(k) Participation on Wealth Measures by Income Interval



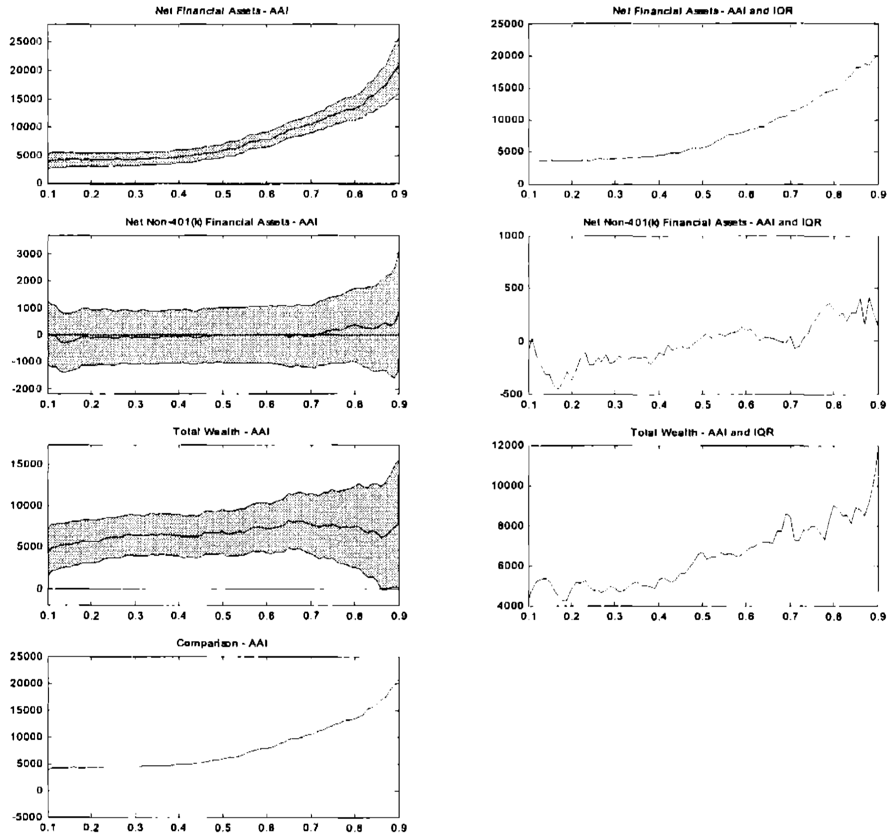
Note: The figure compares the effect of 401(k) participation on net financial assets, net non-401(k) financial assets, and total wealth by income interval. Each panel is labeled with the income interval to which it corresponds. The vertical axis measures the dollar increase in the wealth measures due to 401(k) participation. The horizontal axis corresponds to the quantiles of the conditional distributions. Covariates are as described in the main text. The solid line corresponds to net financial assets, the dashed line to net non-401(k) financial assets, and the dash-dot line to total wealth. Estimates are reported for $\tau \in [.20, .80]$ at .01 unit intervals.

TABLE 5. TESTS ON THE INSTRUMENTAL QUANTILE REGRESSION PROCESS BY INCOME INTERVAL

Null Hypothesis	Net Financial Assets		Net Non-401(k) Financial Assets		Total Wealth	
	Statistic	$c_{.95}$	Statistic	$c_{.95}$	Statistic	$c_{.95}$
A. <$10K						
No Effect	2.231	2.909	0.982	3.209	1.744	2.955
Constant Effect	2.021	3.786	0.982	3.947	1.636	3.894
Exogeneity	1.981	3.801	1.469	4.032	1.243	3.722
B. $10-20K						
No Effect	4.811	2.754	1.177	2.725	2.723	2.758
Constant Effect	4.649	3.169	1.633	3.361	2.114	3.455
Exogeneity	1.927	3.012	1.756	3.073	0.783	3.058
C. $20-30K						
No Effect	6.133	2.603	1.217	2.817	1.204	2.786
Constant Effect	4.418	3.140	1.260	3.262	0.807	3.145
Exogeneity	1.367	2.930	1.292	2.909	2.472	3.132
D. $30-40K						
No Effect	6.746	2.895	0.550	2.738	2.538	3.001
Constant Effect	4.243	3.186	0.707	3.127	1.367	3.340
Exogeneity	1.451	2.967	0.964	3.087	1.863	2.831
E. $40-50K						
No Effect	5.678	2.760	0.673	3.000	2.648	2.724
Constant Effect	3.139	3.124	0.937	3.298	2.286	3.349
Exogeneity	0.980	2.972	1.127	2.959	1.909	3.134
F. $50-75K						
No Effect	6.475	2.817	1.253	2.612	3.402	2.831
Constant Effect	3.410	3.202	1.820	3.135	2.112	3.109
Exogeneity	1.009	2.880	0.947	2.878	1.674	2.859
G. $75K+						
No Effect	4.396	2.829	0.990	2.853	2.112	2.964
Constant Effect	2.721	3.266	1.017	3.339	1.827	3.410
Exogeneity	1.359	3.321	0.823	2.945	1.098	3.089

Note: The table report inference results on the inverse quantile regression process computed by income interval. Reported are Kolmogorov-Smirnov statistics and 95% critical values. The statistics and critical values are computed using the methods in Chernozhukov and Hansen (2002). The null hypotheses tested are as follows: no effect - $\alpha(\cdot) = 0$, constant effect - $\alpha(\cdot) = \alpha(.5)$, exogeneity - $\alpha(\cdot) = \alpha_{QR}(\cdot)$, where $\alpha(\cdot)$ denotes the instrumental quantile regression process and $\alpha_{QR}(\cdot)$ denotes the quantile regression process.

FIGURE 6. COMPARISON OF AAI AND IQR



Note: The sample size is 9,915. The left column contains estimates obtained using the estimator of Abadie, Angrist, and Imbens (2002) (AAI), and the right column compares them with the corresponding estimates obtained through the IQR estimator of Chernozhukov and Hansen (2001). The solid line corresponds to the IQR estimator, and the dashed line corresponds to the AAI estimator. Each panel is labeled with the dependent variable used in estimation of the presented results. The bottom panel in the left column compares the AAI point estimates for each wealth measure. The solid line corresponds to net financial assets, the dashed line to net non-401(k) financial assets, and the dash-dot line to total wealth. The vertical axis measures the dollar increase in the wealth measure due to 401(k) participation. The quantile of the conditional wealth distribution is on the horizontal axis. Covariates are as described in the main text. The shaded region is the 95% confidence band using robust standard errors. Estimates are reported for $\tau \in [.10, .90]$ at .01 unit intervals.