Three Essays on Nonlinear Panel Data Models and Quantile Regression Analysis

by

Iván Fernández-Val

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2005 [June 2005]

Author ....
Department of Economics
April 15, 2005

Certified by ....
Joshua D. Angrist
Professor of Economics
Thesis Supervisor

Certified by ....
Victor Chernozhukov
Castle Krob Career Development Assistant Professor of Economics
Thesis Supervisor

Certified by ....
Whitney K. Newey
Jane Berkowitz Carlton and Dennis William Carton Professor of Economics
Thesis Supervisor

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

Three Essays on Nonlinear Panel Data Models and Quantile Regression Analysis

by

Iván Fernández-Val

Submitted to the Department of Economics
on April 15, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This dissertation is a collection of three independent essays in theoretical and applied econometrics, organized in the form of three chapters. In the first two chapters, I investigate the properties of parametric and semiparametric fixed effects estimators for nonlinear panel data models. The first chapter focuses on fixed effects maximum likelihood estimators for binary choice models, such as probit, logit, and linear probability model. These models are widely used in economics to analyze decisions such as labor force participation, union membership, migration, purchase of durable goods, marital status, or fertility. The second chapter looks at generalized method of moments estimation in panel data models with individual-specific parameters. An important example of these models is a random coefficients linear model with endogenous regressors. The third chapter (co-authored with Joshua Angrist and Victor Chernozhukov) studies the interpretation of quantile regression estimators when the linear model for the underlying conditional quantile function is possibly misspecified.

Thesis Supervisor: Joshua D. Angrist
Title: Professor of Economics

Thesis Supervisor: Victor Chernozhukov
Title: Castle Krob Career Development Assistant Professor of Economics

Thesis Supervisor: Whitney K. Newey
Title: Jane Berkowitz Carlton and Dennis William Carton Professor of Economics

Acknowledgments

This thesis was written with the help and support of a long list of people, to whom I am most grateful. First and foremost, I owe an immense debt of gratitude to my advisors, Joshua Angrist, Victor Chernozhukov, and Whitney Newey. I thank Whitney for always knowing the answers to my questions, helping me graduate in finite time. Josh has been a constant source of guidance and encouragement. Without his wise advice this dissertation would be much less interesting and relevant. Victor is a friend and excellent advisor, incredibly generous with his knowledge and research ideas. I am also indebted to Josh and Victor for letting me be their coauthor. Working with them was an extraordinary experience, and I look forward to continue learning from them in the years to follow.

I would further like to acknowledge the financial support of the Fundación Caja Madrid and Fundación Ramón Areces. I am also very grateful to my professors at CEMFI for their help in finding my way to MIT. A special thank you is due to Alberto Abadie for having shared with me his experiences in the job market process.

I was honored to share these years at MIT with my classmates and friends, Raphael Auer, Karna Basu, Matilde Bombardini, Thomas Chaney, Chris Hansen, Geraint Jones, Ashley Lester, Gerard Padró-i-Miquel, Daniel Paravisini, Nancy Qian, Verónica Rappoport, Tali Regev, Rubén Segura-Cayuela, and Petia Topalova. From them I learned plenty. Thank you also to the rest of Ibbetsonians and regulars, Paula Bustos, Valeria Carou, Sylvain Chassang, Patricia Cortés, Pascaline Dupas, Antara Dutta, Merçe Gené, Gero Jung, Isabel Neto, and Karine Sefarty, for making my time at MIT much more pleasant and enjoyable.

Finally I would like to thank my brother and sister, Juanjo and María, for supporting me from far away during my years in graduate school. Most specially, I would like to thank my parents for continuous help, tireless support and encouragement. It is to them that I dedicate this thesis.

A mis padres,
Gloria y Juan José

Contents

Introduction **13**

1 Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects **17**

1.1 Introduction 17

1.2 The Model and Estimators 20

1.2.1 The Model 20

1.2.2 Fixed Effects MLE 21

1.2.3 Incidental Parameters Problem 22

1.2.4 Large-T Approximation to the Bias 23

1.3 Bias Corrections for Discrete Choice Panel Data Models 25

1.3.1 Bias Correction of the Estimator 25

1.3.2 Bias Correction of the Estimating Equation 27

1.3.3 Modified (Profile) Maximum Likelihood (MML) 27

1.3.4 Example: Andersen (1973) Two-Period Logit Model 28

1.4 Bias for Static Panel Probit Model 29

1.5 Marginal Effects: Small Bias Property 32

1.5.1 Parameters of Interest 32

1.5.2 Bias Correction of Marginal Effects 34

1.5.3 Panel Probit: Small Bias Property 35

1.6 Extension: Dynamic Discrete Choice Models 38

1.6.1 The Model 38

1.6.2 Large-T Approximation to the Bias 39

1.6.3 Marginal Effects 40

1.7 Monte Carlo Experiments 41

1.7.1 Static Model 42

1.7.2 Dynamic Model . . . 43
1.8 Empirical Application: Female Labor Force Participation . . . 44
1.9 Summary and conclusions . . . 47
Appendix . . . 48
1.A Bias Formulas for Binary Choice Models . . . 48
1.A.1 Static Case . . . 48
1.A.2 Dynamic Case . . . 48
1.B Bias Corrections in Dynamic models . . . 49
1.C Proofs . . . 50
1.C.1 Lemmas . . . 50
1.C.2 Proof of Proposition 1 . . . 51
1.C.3 Proof of Proposition 2 . . . 51
1.D Relationship between Bias Correction of the Estimating Equation and Modified Maximum Likelihood . . . 53

2 Bias Correction in Panel Data Models with Individual-Specific Parameters 71
2.1 Introduction . . . 71
2.2 Example: A Linear IV Panel Model with Individual-Specific Coefficients . . . 74
2.2.1 The Model . . . 74
2.2.2 Example . . . 74
2.2.3 The Problem . . . 75
2.2.4 Incidental Parameters Bias . . . 77
2.2.5 Bias Corrections . . . 78
2.3 The Model and Estimators . . . 79
2.3.1 Some Notation . . . 80
2.3.2 Fixed Effects GMM Estimator (FE-GMM) . . . 80
2.3.3 First Order Conditions for FE-GMM . . . 81
2.4 Asymptotic Properties of Panel Data GMM Estimators . . . 82
2.5 Bias Corrections . . . 89
2.5.1 Bias Correction of the Estimator . . . 90
2.5.2 Bias Correction of the Plug-in Score . . . 91
2.5.3 Asymptotic Properties of bias-corrected FE-GMM Estimators . . . 93
2.6 Monte Carlo Experiment . . . 95
2.7 Empirical Application . . . 96

2.8 Conclusion . . . 99
Appendix . . . 100
2.A Consistency of the First Stage GMM Estimator . . . 100
2.A.1 Some Lemmas . . . 100
2.A.2 Proof of Theorem 1 . . . 104
2.A.3 Proof of Theorem 2 . . . 104
2.B Asymptotic Distribution of the First Stage GMM Estimator . . . 106
2.B.1 Some Lemmas . . . 106
2.B.2 Proof of Theorem 3 . . . 111
2.C Consistency of Second Stage GMM Estimator . . . 113
2.C.1 Some Lemmas . . . 113
2.C.2 Proof of Theorem 4 . . . 115
2.C.3 Proof of Theorem 5 . . . 115
2.D Asymptotic Distribution of the Second Stage GMM Estimator . . . 117
2.D.1 Some Lemmas . . . 117
2.D.2 Proof of Theorem 6 . . . 122
2.E Asymptotic Distribution of the bias-corrected Second Stage GMM Estimator . . . 124
2.E.1 Some Lemmas . . . 124
2.E.2 Proof of Theorem 7 . . . 131
2.F Stochastic Expansion for $\tilde{\gamma}_{i0} = \tilde{\gamma}_i(\theta_0)$ . . . 131
2.G Stochastic Expansion for $\hat{s}_i^W(\theta_0, \tilde{\gamma}_{i0})$ . . . 135
2.H Stochastic Expansion for $\hat{\gamma}_{i0} = \hat{\gamma}_i(\theta_0)$ . . . 139
2.I Stochastic Expansion for $\hat{s}_i(\theta_0, \hat{\gamma}_{i0})$ . . . 145
2.J V - Statistics . . . 152
2.J.1 Properties of Normalized V-Statistics . . . 152
2.J.2 Properties of Modified Normalized V-Statistics . . . 157
2.K First Stage Score and Derivatives: Fixed Effects . . . 161
2.K.1 Score . . . 161
2.K.2 Derivatives with respect to the fixed effects . . . 162
2.K.3 Derivatives with respect to the common parameter . . . 165
2.L First Stage Score and Derivatives: Common Parameters . . . 165
2.L.1 Score . . . 165
2.L.2 Derivatives with respect to the fixed effects . . . 165

2.L.3 Derivatives with respect to the common parameters 167
2.M Second Stage Score and Derivatives: Fixed Effects 167
2.M.1 Score 167
2.M.2 Derivatives with respect to the fixed effects 167
2.M.3 Derivatives with respect to the common parameters 167
2.N Second Stage Score and Derivatives: Common Parameters 168
2.N.1 Score 168

3 Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure 177

3.1 Introduction 177
3.2 Interpreting QR Under Misspecification 179
3.2.1 Notation and Framework 179
3.2.2 QR Approximation Properties 180
3.2.3 Partial Quantile Regression and Omitted Variable Bias 184
3.3 Sampling Properties of QR Under Misspecification 186
3.4 Application to U.S. Wage Data 189
3.5 Summary and conclusions 192
Appendix 194
3.A Proof of Theorems 3 and its Corollaries 194
3.A.1 Uniform consistency of $\hat{\beta}(\cdot)$ 194
3.A.2 Asymptotic Gaussianity of $\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$ 194
3.A.3 Proof of Corollaries 196
3.A.4 Uniform Consistency of $\widehat{\Sigma}(\cdot,\cdot)$ and $\widehat{J}(\cdot)$. 196

Introduction

This dissertation is a collection of three independent essays in theoretical and applied econometrics, organized in the form of three chapters. In the first two chapters, I investigate the properties of parametric and semiparametric fixed effects estimators for nonlinear panel data models. The first chapter focuses on fixed effects maximum likelihood estimators for binary choice models, such as probit, logit, and linear probability model. These models are widely used in economics to analyze decisions such as labor force participation, union membership, migration, purchase of durable goods, marital status, or fertility. The second chapter looks at generalized method of moments estimation in panel data models with individual-specific parameters. An important example of these models is a random coefficients linear model with endogenous regressors. The third chapter (co-authored with Joshua Angrist and Victor Chernozhukov) studies the interpretation of quantile regression estimators when the linear model for the underlying conditional quantile function is possibly misspecified.

Chapter 1, "Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Models with Fixed Effects," analyzes the properties of maximum likelihood estimators for index coefficients and marginal effects in binary choice models with individual effects. The inclusion of individual effects in these models helps identify causal effects of regressors on the outcome of interest because these effects provide control for unobserved time-invariant characteristics. However, it also poses important technical challenges in estimation. In particular, if these individual effects are treated as parameters to be estimated (fixed-effects approach), then the resulting estimators suffer from the well-known incidental parameters problem (Neyman and Scott, 1948).

This chapter derives the incidental parameter bias for fixed effects estimators of index coefficients and introduces analytical bias correction methods for these estimators. In particular, the first term of a large-T expansion of the bias is characterized using higher-order asymptotics. For the important case of the probit, this bias is a positive definite matrix-weighted average of the true parameter value for general distributions of regressors and individual effects. This implies, for instance, that probit estimates are biased away from zero when the regressors are scalar, which helps explain previous Monte Carlo evidence. The expression of the bias is also used to derive the bias of other quantities of interest, such as ratios of index coefficients and marginal effects. In the absence of heterogeneity, fixed effect estimates of these quantities do not suffer from the incidental parameters problem. Moreover, numerical computations and Monte Carlo examples show that the small bias property for fixed effects estimates of marginal effects holds for a wide range of distributions of regressors and individual effects. The methods and estimators introduced in this chapter are applied to the analysis of the effect of fertility on female labor force participation.

Chapter 2, "Bias Correction in Panel Data Models with Individual-Specific Parameters," introduces a new class of semiparametric estimators for panel models where the response to the regressors can be individual-specific in an unrestricted way. These estimators are based on moment conditions that can be nonlinear functions in parameters and variables, accommodating both linear and nonlinear models and allowing for the presence of endogenous regressors. In models with individual-specific parameters and endogenous regressors, these estimators are generally biased in short panels because of the finite-sample bias of GMM estimators. This chapter derives bias correction methods for fixed effects GMM estimators of model parameters and other quantities of interest, such as means or standard deviations of the individual-specific parameters. These methods are illustrated by estimating earnings equations for young men allowing the effect of the union status to be different for each individual. The results suggest that there is large heterogeneity in the union premium. Moreover, fixed coefficients estimators overestimate the average effect of union status on earnings.

Chapter 3, "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," studies the approximation properties of quantile regression (QR) estimators to the conditional quantile function. The analysis here is motivated by the minimum mean square error linear approximation property of traditional mean regression in the estimation of conditional expectation functions. Empirical research using quantile regression with discrete covariates suggests that QR may have a similar property, but the exact nature of the linear approximation has remained elusive. This chapter shows that QR can be interpreted as minimizing a weighted mean-squared error loss function for specification error. The weighting function is an average density of the dependent variable near the true conditional quantile. The weighted least squares interpretation of QR is used to derive an omitted variables bias formula and a partial quantile correlation concept, similar to the relationship between partial correlation and

OLS. General asymptotic results for QR processes allowing for misspecification of the conditional quantile function are also derived, extending earlier results from a single quantile to the entire process. The approximation properties of QR are illustrated through an analysis of the wage structure and residual inequality in US census data for 1980, 1990, and 2000. The results suggest continued residual inequality growth in the 1990s, primarily in the upper half of the wage distribution and for college graduates.

Chapter 1

Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects

1.1 Introduction

Panel data models are widely used in empirical economics because they allow researchers to control for unobserved individual time-invariant characteristics. However, these models pose important technical challenges. In particular, if individual heterogeneity is left completely unrestricted, then estimates of model parameters in nonlinear and/or dynamic models suffer from the incidental parameters problem, first noted by Neyman and Scott (1948). This problem arises because the unobserved individual characteristics are replaced by inconsistent sample estimates, which, in turn, bias estimates of model parameters. Examples include probit with fixed effects, and linear and nonlinear models with lagged dependent variables and fixed effects (see, e.g., Nerlove, 1967; Nerlove, 1971; Heckman, 1981; Nickell, 1981; Greene, 2002; Katz, 2001; and Hahn and Newey, 2004).

Incidental parameters bias is a longstanding problem in econometrics, but general bias correction methods have been developed only recently. Efforts in this direction include Lancaster (2000), Hahn and Kuersteiner (2001), Woutersen (2002), Arellano (2002), Alvarez and Arellano (2003), Carro (2003), Hahn and Kuersteiner (2003), and Hahn and Newey (2004). I refer to the approaches taken in these papers as providing large-T-consistent estimates because they rely on an asymptotic approximation to the behavior of the estimator that lets both the number of individuals, n, and the time dimension, T, grow with the sample size.[1] The idea behind these methods is to expand the incidental parameters bias of the estimator in orders of magnitude of T, and to remove an estimate of the leading term of the bias from the estimator.[2] As a result, the adjusted estimator has a bias of order T^{-2}, whereas the bias of the initial estimator is of order T^{-1}. This approach aims to approximate the properties of estimators in applications that use panels of moderate length, such as the PSID or the Penn World Table, where the most important part of the bias is captured by the first term of the expansion.

The first contribution of this chapter is to provide new correction methods for parametric binary choice models that attain the semiparametric efficiency bound of the bias estimation problem. The improvement comes from using the parametric structure of the model more intensively than in previous studies by taking conditional moments of the bias, given the regressors and individual effects. The correction is then constructed based on the new formulas. This approach is similar to the use of the conditional information matrix in the estimation of asymptotic variances in maximum likelihood, instead of other alternatives, such as the sample average of the outer product of the scores or the sample average of the negative Hessian (Porter, 2002).[3] The adjustment presented here not only simplifies the correction by removing terms with zero conditional expectation, but also reduces incidental parameter bias more effectively than other large-T corrections.

The second contribution of the chapter is to derive a lower bound and a proportionality result for the bias of probit fixed effects estimators of model parameters. The lower bound depends uniquely upon the number of time periods of the panel, and is valid for general distributions of regressors and individual effects. According to this bound, for instance, the incidental parameters bias is at least 20 % for 4-period panels and 10 % for 8-period panels. Proportionality, on the other hand, establishes that probit fixed effect estimators of model parameters are biased away from zero when the regressor is scalar, providing a theoretical explanation for the numerical evidence found in previous studies (see, for e.g., Greene, 2002). It also implies that fixed effects

[1]Fixed-T-consistent estimators have been also derived for panel logit models (see Cox, 1958, Andersen, 1973, Chamberlain, 1980, for the static case; and Cox, 1958, Chamberlain, 1985, and Honoré and Kyriazidou, 2000, for the dynamic case), and other semiparametric index models (see Manski, 1987, for the static case; and Honoré and Kyriazidou, 2000, for the dynamic case). These methods, however, do not provide estimates for individual effects, precluding estimation of other quantities of interest, such as marginal effects.

[2]To avoid complicated terminology, in the future I will generally refer to the first term of the large-T expansion of the bias simply as the bias.

[3]Porter (2002) also shows that the conditional information matrix estimator attains the semiparametric efficiency bound for the variance estimation problem.

estimators of ratios of coefficients do not suffer from the incidental parameters bias in probit models in the absence of heterogeneity. These ratios are often structural parameters of interest because they can be interpreted as marginal rates of substitution in many economic applications.

Finally, the bias of fixed effects estimators of marginal effects in probit models is explored.[4] The motivation for this analysis comes from a question posed by Wooldridge: "How does treating the *individual effects* as parameters to estimate - in a "fixed effects probit" analysis - affect estimation of the APEs *(average partial effects)*?"[5] Wooldridge conjectures that the estimators of the marginal effects have reasonable properties. Here, using the expansion of the bias for the fixed effects estimators of model parameters, I characterize the analytical expression for the bias of these average marginal effects. As Wooldridge anticipated, this bias is negligible relative to the true average effect for a wide range of distributions of regressors and individual effects, and is identically zero in the absence of heterogeneity. This helps explain the small biases in the marginal effects estimates that Hahn and Newey (2004) (HN henceforth) find in their Monte Carlo example.

The results presented in this chapter are also consistent with Angrist's (2001) argument for cross-sectional limited dependent variable (LDV) models. Angrist argues that much of the difficulty with LDV models comes from a focus on structural parameters, such as latent index coefficients in probit models, instead of directly interpretable causal effects, such as average treatment effects (see also Wooldridge, 2002; Wooldridge, 2003; and Hahn, 2001). He recommends the use of simple linear models, where the structural parameters are directly linked to the effects of interest, just as if the outcomes were continuous. Here, I show that the same approach of focusing directly on causal effects rather than structural parameters also pays off in panel data models. However, unlike Angrist (2001), I use nonlinear models that incorporate the restrictions on the data support explicitly. These models are better suited for LDVs in cases where some regressors are continuous or the model is not fully saturated.

Monte Carlo examples show that adjusted logit and probit estimators of model parameters based on the new bias formulas have improved finite sample properties. In particular, these corrections remove more effectively the incidental parameters bias and provide estimators with smaller dispersion than previous methods. Accurate finite sample inference for model parameters and marginal effects is obtained from distributions derived under asymptotic sequences where

[4]Marginal effects are defined either as the change in the outcome conditional probability as a response to an one-unit increase in a regressor, or as a local approximation based on the slope of the outcome conditional probability. For example, in the probit the marginal effects can be defined either as $\Phi((x+1)\theta) - \Phi(x\theta)$ or as $\theta\phi(x\theta)$, where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cdf and pdf of the standard normal distribution, respectively.

[5]C.f., Wooldridge (2002), p. 489 (italics mine).

$T/n^{1/3} \to \infty$ for static panels with 4 periods and dynamic panels with 8 periods. Results are also consistent with the small bias property of marginal effects for static models; they suggest that the property holds for the effects of exogenous variables in dynamic models, but not for the effects of lagged dependent variables. Simple linear probability models, in the spirit of Angrist (2001), also perform well in estimating average marginal effects.

The properties of probit and logit fixed effects estimators of model parameters and marginal effects are illustrated with an analysis of female labor force participation using 10 waves from the Panel Survey of Income Dynamics (PSID). The analysis here is motivated by similar studies in labor economics, where panel binary choice processes have been widely used to model female labor force participation decisions (see, e.g., Hyslop, 1999; Chay and Hyslop, 2000; and Carro, 2003). In particular, I find that fixed effects estimators, while biased for index coefficients, give very similar estimates to their bias corrected counterparts for marginal effects in static models. On the other hand, uncorrected fixed effects estimators are biased away from zero for both index coefficients and marginal effects of the fertility variables in dynamic models that account for true state dependence. In this case, the bias corrections presented here are effective reducing the incidental parameters problem.

The chapter is organized as follows. Section 1.2 describes the panel binary choice model and its maximum likelihood estimator. Section 1.3 reviews existing solutions to the incidental parameters problem and proposes improved correction methods for binary choice models. Section 1.4 derives the proportionality result of the bias in static probit models. Section 1.5 analyzes the properties of probit fixed effects estimators of marginal effects. Section 1.6 extends the previous results to dynamic models. Monte Carlo results and the empirical application are given in Sections 1.7 and 1.8, respectively. Section 1.9 concludes with a summary of the main results.

1.2 The Model and Estimators

1.2.1 The Model

Given a binary response Y and a $p \times 1$ regressor vector X, consider the following data generating process

$$Y = \mathbf{1}\left\{X'\theta_0 + \alpha - \epsilon \geq 0\right\}, \tag{1.2.1}$$

where $\mathbf{1}\{C\}$ is an indicator function that takes on value one if condition C is satisfied and zero otherwise; θ_0 denotes a $p \times 1$ vector of parameters; α is a scalar unobserved individual effect; and

ϵ is a time-individual specific random shock. This is an error-components model where the error term is decomposed into a permanent individual-specific component α and a transitory shock ϵ. Examples of economic decisions that can be modeled within this framework include labor force participation, union membership, migration, purchase of durable goods, marital status, or fertility (see Amemiya, 1981, for a survey).

1.2.2 Fixed Effects MLE

In economic applications, regressors and individual heterogeneity are correlated because regressors are decision variables and individual heterogeneity usually represents variation in tastes or technology. To avoid imposing any structure on this relationship, I adopt a fixed-effects approach and treat the sample realization of the individual effects $\{\alpha_i\}_{i=1,\dots,n}$ as parameters to be estimated, see Mundlak (1978), Lancaster (2000), Arellano and Honoré (2000), and Arellano (2003) for a similar interpretation of fixed effects estimators.[6]

To estimate the model parameters, a sample of the observable variables for individuals followed in subsequent periods of time $\{y_{it}, x_{it}\}_{t=1,\dots,T;\ i=1,\dots,n}$ is available, where i and t usually index individuals and time periods, respectively.[7] Then, assuming that ϵ follows a known distribution conditional on regressors and individual effects, typically normal or logistic, a natural way of estimating this model is by maximum likelihood.[8] Thus, if ϵ_{it} are i.i.d. conditional on $\bar{x}_i$ and α_i, with cdf $F_\epsilon(\cdot|\bar{X}, \alpha)$, the conditional log-likelihood for observation i at time t is[9]

$$l_{it}(\theta, \alpha_i) \equiv y_{it} \log F_{it}(\theta, \alpha_i) + (1 - y_{it}) \log(1 - F_{it}(\theta, \alpha_i)), \tag{1.2.2}$$

where $F_{it}(\theta, \alpha_i)$ denotes $F_\epsilon(x_{it}'\theta + \alpha_i|\bar{X} = \bar{x}_i, \alpha = \alpha_i)$, and the MLE of θ, concentrating out the α_i's, is the solution to

$$\hat{\theta} \equiv \arg\max_{\theta} \sum_{i=1}^{n} \sum_{t=1}^{T} l_{it}(\theta, \hat{\alpha}_i(\theta))/nT, \quad \hat{\alpha}_i(\theta) \equiv \arg\max_{\alpha} \sum_{t=1}^{T} l_{it}(\theta, \alpha)/T. \tag{1.2.3}$$

[6] Note that Kiefer and Wolfowitz's (1956) consistency result does not apply to here, since no assumption is imposed on the distribution of the individual effects conditional on regressors.

[7] In the sequel, for any random variable Z, z_{it} denotes observation at period t for individual i; $\bar{Z}$ denotes a random vector with T copies of Z; and $\bar{z}_i$ denotes an observation of $\bar{Z}$, i.e. $\{z_{i1}, ..., z_{iT}\}$.

[8] Since the inference is conditional on the realization of the regressors and individual effects, all the probability statements should be qualified with a.s. I omit this qualifier for notational convenience.

[9] Following the common practice in fixed effects panel models, I will assume that the regressor vector $\bar{X}$ is strictly exogenous. See Arellano and Carrasco (2003) for an example of random effects estimator with predetermined regressors.

1.2.3 Incidental Parameters Problem

Fixed effects MLEs generally suffer from the incidental parameters problem noted by Neyman and Scott (1948). This problem arises because the unobserved individual effects are replaced by sample estimates. In nonlinear model, estimation of the model parameters cannot generally be separated from the estimation of the individual effects.[10] Then, the estimation error of the individual effects introduces bias in the estimates of model parameters. To see this, for $\bar{z}_i \equiv (\bar{y}_i, \bar{x}_i)$ and any function $m(\bar{z}_i, \alpha_i)$, let $\bar{E}[m(\bar{z}_i, \alpha_i)] \equiv E_{\bar{X}\alpha}\left\{E_{\bar{Y}}\left[m(\bar{Z}, \alpha)|\bar{X}, \alpha\right]\right\}$, where the first expectation is taken with respect to the unknown joint distribution of $(\bar{X}, \alpha)$ and the second with respect to the known distribution of $\bar{Y}|\bar{X}, \alpha$.[11] Then, from the usual maximum likelihood properties, for $n \to \infty$ with T fixed,

$$\hat{\theta} \xrightarrow{p} \theta_T, \ \theta_T \equiv \arg\max_{\theta} \bar{E}\left[\sum_{t=1}^{T} l_{it}(\theta, \hat{\alpha}_i(\theta))/T\right]. \tag{1.2.4}$$

When the true conditional log-likelihood of Y is $l_{it}(\theta_0, \alpha_i)$ generally $\theta_T \neq \theta_0$, but $\theta_T \to \theta_0$ as $T \to \infty$. For the smooth likelihoods considered here, $\theta_T = \theta_0 + \frac{\mathcal{B}}{T} + O\left(\frac{1}{T^2}\right)$ for some $\mathcal{B}$.[12] By asymptotic normality of the MLE, $\sqrt{nT}(\hat{\theta} - \theta_T) \xrightarrow{d} N(0, -\mathcal{J}^{-1})$ as $n \to \infty$, and therefore

$$\sqrt{nT}(\hat{\theta} - \theta_0) = \sqrt{nT}(\hat{\theta} - \theta_T) + \sqrt{nT}\frac{\mathcal{B}}{T} + O\left(\sqrt{\frac{n}{T^3}}\right). \tag{1.2.5}$$

Here we can see that even if we let T grow at the same rate as n, that is $T = O(n)$, the MLE, while consistent, has a limiting distribution not centered around the true parameter value. Under asymptotic sequences where T grows large no faster than n, the estimates of the individual effects converge to their true value at a slower rate than the sample size nT, since only observations for each individual convey information about the corresponding individual effect. This slower rate translates into bias in the asymptotic distribution of the estimators of the model parameters.

[10] In static linear models the individual effects can be removed by taking differences with respect to the individual means without affecting the consistency of the estimator of model parameters. There is no general procedure, however, to remove the individual effects in nonlinear models. An exception is the panel logit, where the individual effects can be eliminated by conditioning in their sufficient statistics.

[11] When the observations are independent across time periods the conditional cdf of $\bar{Y}$ given $(\bar{X}, \alpha)$ can be factorized as $F(Y|\bar{X}, \alpha) \times ... \times F(Y|\bar{X}, \alpha)$. For dependent data, if $\bar{Y} = (Y_T, ..., Y_1)$, then the conditional cdf of $\bar{Y}$ given $(\bar{X}, Y_0, \alpha)$ factorizes as $F(Y_T|\bar{X}, Y_{T-1}, ..., Y_0, \alpha) \times ... \times F(Y_1|\bar{X}, Y_0, \alpha)$. The conditional cdf of Y can then be obtained from the conditional cdf of ϵ, which is assumed to be known.

[12] To see this intuitively, note that MLEs are implicit smooth functions of sample means and use the following result (Lehmann and Casella, 1998). Let $X_1, ..., X_T$ be i.i.d. with $E[X_1] = \mu_X$, $Var[X_1] = \sigma_X^2$, and finite fourth moment; suppose h is a function of real variable whose first four derivatives exist for all $x \in I$, where I is an interval with $\Pr\{X_1 \in I\} = 1$; and $h^{iv}(x) \leq M$ for all $x \in I$, for some $M < \infty$. Then $E\left[h(\sum_{t=1}^{T} X_t/T)\right] = h(\mu_X) + \sigma_X^2 h''(\mu_X)/2T + O(T^{-2})$.

1.2.4 Large-T Approximation to the Bias

In moderate-length panels the most important part of the bias is captured by the first term of the expansion, $\mathcal{B}$. A natural way to reduce bias is therefore to remove a consistent estimate of this term from the MLE. To implement this procedure, however, we need an analytical expression for $\mathcal{B}$.[13] This expression can be characterized using a stochastic expansion of the fixed effects estimator in orders of T.

A little more notation is useful for describing this expansion. Let

$$u_{it}(\theta, \alpha) \equiv \frac{\partial}{\partial \theta} l_{it}(\theta, \alpha), \quad v_{it}(\theta, \alpha) \equiv \frac{\partial}{\partial \alpha} l_{it}(\theta, \alpha), \tag{1.2.6}$$

and additional subscripts denote partial derivatives, e.g. $u_{it\theta}(\theta, \alpha) \equiv \partial u_{it}(\theta, \alpha)/\partial \theta'$. Then, the first order condition for the concentrated problem can be expressed as

$$0 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it}(\hat{\theta}, \hat{\alpha}_i(\hat{\theta})). \tag{1.2.7}$$

Expanding this expression around the true parameter value θ_0 yields

$$0 = \hat{u}(\theta_0) + \hat{\mathcal{J}}(\bar{\theta})(\hat{\theta} - \theta_0) \quad \Rightarrow \quad \hat{\theta} - \theta_0 = -\hat{\mathcal{J}}(\bar{\theta})^{-1} \hat{u}(\theta_0), \tag{1.2.8}$$

where $\bar{\theta}$ lies between $\hat{\theta}$ and θ_0;

$$\hat{u}(\theta_0) \equiv \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it}(\theta_0, \hat{\alpha}_i(\theta_0)) \tag{1.2.9}$$

is the fixed effects estimating equation evaluated at the true parameter value, the expectation of which is generally different from zero because of the randomness of $\hat{\alpha}(\theta_0)$ and determines the bias; and

$$\hat{\mathcal{J}}(\theta) \equiv \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ u_{it\theta}(\theta, \hat{\alpha}_i(\theta)) + u_{it\alpha}(\theta, \hat{\alpha}_i(\theta)) \frac{\partial \hat{\alpha}_i(\theta)}{\partial \theta'} \right\} \tag{1.2.10}$$

is the Jacobian of $\hat{u}(\theta)$.

For the estimators of the individual effects, the first order condition is $\sum_{t=1}^{T} v_{it}(\theta, \hat{\alpha}_i(\theta))/T =$

[13] Jackknife is an alternative bias correction method that does not require an analytical form for $\mathcal{B}$, see HN.

0. Differentiating this expressions with respect to θ and $\hat{\alpha}_i$ yields

$$\frac{\partial \hat{\alpha}_i(\theta)}{\partial \theta'} = -\frac{\hat{E}_T\left[v_{it\theta}(\theta, \hat{\alpha}_i(\theta))\right]}{\hat{E}_T\left[v_{it\alpha}(\theta, \hat{\alpha}_i(\theta)\right]}, \tag{1.2.11}$$

where $\hat{E}_T\left[f_{it}\right] \equiv \sum_{t=1}^{T} f_{it}/T$, for any function $f_{it} \equiv f(z_{it})$. Plugging this expression into (1.2.10) and taking (probability) limits as $n, T \to \infty$

$$\hat{\mathcal{J}}(\bar{\theta}) \xrightarrow{p} \bar{E}\left[E_T\left[u_{it\theta}\right] - E_T\left[u_{it\alpha}\right]\frac{E_T\left[v_{it\theta}\right]}{E_T\left[v_{it\alpha}\right]}\right] \equiv \mathcal{J}, \tag{1.2.12}$$

where $E_T\left[f_{it}\right] \equiv \lim_{T\to\infty}\sum_{t=1}^{T} f_{it}/T = E_Z\left[f_{it}|\alpha\right]$, by the Law of Large Numbers for i.i.d. sequences. For notational convenience the arguments are omitted when the expressions are evaluated at the true parameter value, i.e. $v_{it\theta} = v_{it\theta}(\theta_0, \alpha_i)$. Here, $-\mathcal{J}^{-1}$ gives the asymptotic variance of the fixed effect estimator of θ_0 under correct specification.

For the estimating equation, note that using independence across t, standard higher-order asymptotics for the estimator of the individual effects give (e.g., Ferguson, 1992, or Rilstone et al., 1996), as $T \to \infty$

$$\hat{\alpha}_i = \alpha_i + \psi_i/\sqrt{T} + \beta_i/T + o_p(1/T), \; \psi_i = \sum_{t=1}^{T}\psi_{it}/\sqrt{T} \xrightarrow{d} \mathcal{N}(0, \sigma_i^2), \tag{1.2.13}$$

$$\psi_{it} = \sigma_i^2 v_{it}, \; \sigma_i^2 = -E_T\left[v_{it\alpha}\right]^{-1}, \; \beta_i = \sigma_i^2\left\{E_T\left[v_{it\alpha}\psi_{it}\right] + \frac{1}{2}\sigma_i^2 E_T\left[v_{it\alpha\alpha}\right]\right\}. \tag{1.2.14}$$

Then, expanding $\hat{u}(\theta_0)$ around the α_i's, and assuming that orders in probability correspond to orders in expectation, we have as $n, T \to \infty$

$$\begin{aligned} T\hat{u}(\theta_0) &= T\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{E}_T\left[u_{it}\right] + \hat{E}_T\left[u_{it\alpha}(\hat{\alpha}_i - \alpha)\right] + \hat{E}_T\left[u_{it\alpha\alpha}(\hat{\alpha}_i - \alpha)^2\right]/2 + o_p(1/T)\right\} \\ &\xrightarrow{p} \bar{E}\left\{0 + E_T\left[u_{it\alpha}\right]\beta_i + E_T\left[u_{it\alpha}\psi_{it}\right] + \frac{1}{2}\sigma_i^2 E_T\left[u_{it\alpha\alpha}\right]\right\} \equiv b. \end{aligned} \tag{1.2.15}$$

Finally, the (first term of the large-T expansion of the) asymptotic bias is[14]

$$T(\hat{\theta} - \theta_0) \xrightarrow{p} T(\theta_T - \theta_0) = -p\lim \hat{\mathcal{J}}(\bar{\theta})^{-1}\; p\lim T\hat{u}(\theta_0) = -\mathcal{J}^{-1}b \equiv \mathcal{B}. \tag{1.2.16}$$

[14] This expansion also provides an alternative explanation for the absence of incidental parameters bias in the static panel linear model. In this case $v_{it} = y_{it} - x_{it}'\theta - \alpha_i$, $v_{it\alpha} = -1$, $v_{it\alpha\alpha} = 0$, and $u_{it} = v_{it}x_{it}$. Then, $\beta_i = b = 0$ since $E[v_{it\alpha}\psi_{it}] = E[u_{it\alpha}\psi_{it}] = 0$ and $E[v_{it\alpha\alpha}] = E[u_{it\alpha\alpha}] = 0$. Moreover, the bias terms of higher order are also zero because the second order expansions are exact.

1.3 Bias Corrections for Discrete Choice Panel Data Models

Large-T correction methods remove the bias of fixed effects estimators up to a certain order of magnitude in T. In particular, the incidental parameters bias is reduced from $O(T^{-1})$ to $O(T^{-2})$ as $T \to \infty$, and the asymptotic distribution is centered at the true parameter value if $T/n^{1/3} \to \infty$. To see this, note that if $\hat{\theta}^c$ is a bias corrected estimator of θ_0 with probability limit $\theta_T^c = \theta_0 + O\left(T^{-2}\right)$, then

$$\sqrt{nT}(\hat{\theta}^c - \theta_0) = \sqrt{nT}(\hat{\theta}^c - \theta_T^c) + O\left(\sqrt{\frac{n}{T^3}}\right). \tag{1.3.1}$$

These methods can take different forms depending on whether the adjustment is made in the estimator, estimating equation, or objective function (log-likelihood). The first purpose of this section is to review the existing methods of bias correction, focusing on how these methods can be applied to panel binary choice models. The second purpose is to modify the corrections in order to improve their asymptotic and finite sample properties. Finally, I compare the different alternatives in a simple example.

1.3.1 Bias Correction of the Estimator

HN propose the following correction

$$\hat{\theta}^1 \equiv \hat{\theta} - \frac{\hat{\mathcal{B}}}{T}, \tag{1.3.2}$$

where $\hat{\mathcal{B}}$ is an estimator of $\mathcal{B}$. Since $\hat{\mathcal{B}}$ generally depends on $\hat{\theta}$, i.e., $\hat{\mathcal{B}} = \hat{\mathcal{B}}(\hat{\theta})$, they also suggest to iterate the correction by solving $\hat{\theta}^\infty = \hat{\theta} - \hat{\mathcal{B}}(\hat{\theta}^\infty)$. To estimate $\mathcal{B}$, HN give two alternatives. The first alternative, only valid for the likelihood setting, is based on replacing derivatives for outer products in the bias formulas using Bartlett identities, and then estimate expectations using sample means. The second possibility replaces expectations for sample means using directly the bias formulas for general estimating equations derived in Section 1.2. These expressions rely only upon the unbiasedness of the estimating equation at the true value of the parameters and individual effects, and therefore are more robust to misspecification.

I argue here that the previous distinction is not very important for parametric discrete choice models, since the estimating equations are only valid under correct specification of the conditional distribution of ϵ. In other words, these estimating equations do not have a quasi-likelihood interpretation under misspecification. Following the same idea as in the estimation of asymptotic variances in MLE, I propose to take conditional expectations of the bias formulas (conditioning on the regressors and individual effects), and use the resulting expressions to construct the corrections.[15] These new corrections have optimality asymptotic properties. In particular, using Brown and Newey (1998) results for efficient estimation of expectations, it follows that the estimator of the bias proposed here attains the semiparametric efficiency bound for the bias estimation problem.[16] Intuitively, taking conditional expectations removes zero-mean terms of the bias formula that only add noise to the analog estimators.

To describe how to construct the correction from the new bias formulas, it is convenient to introduce some more notation. Let

$$
\begin{aligned}
F_{it}(\theta) &\equiv F_\epsilon(x_{it}'\theta + \hat{\alpha}_i(\theta)|\bar{X}, \alpha), \quad f_{it}(\theta) \equiv f_\epsilon(x_{it}'\theta + \hat{\alpha}_i(\theta)|\bar{X}, \alpha), && (1.3.3)\\
g_{it}(\theta) &\equiv f_\epsilon'(x_{it}'\theta + \hat{\alpha}_i(\theta)|\bar{X}, \alpha), \quad H_{it}(\theta) \equiv \frac{f_{it}(\theta)}{F_{it}(\theta)\left(1 - F_{it}(\theta)\right)}, && (1.3.4)
\end{aligned}
$$

where f is the pdf associated with F, and f' is the derivative of f. Also, define

$$
\hat{\sigma}_i^2(\theta) \equiv \hat{E}_T \left[H_{it}(\theta) f_{it}(\theta)\right]^{-1}, \quad \hat{\psi}_{it}(\theta) \equiv \hat{\sigma}_i^2(\theta) H_{it}(\theta) \left[y_{it} - F_{it}(\theta)\right]. \tag{1.3.5}
$$

Here, $\hat{\sigma}_i^2(\theta)$ and $\hat{\psi}_{it}(\theta)$ are estimators of the asymptotic variance and influence function, respectively, obtained from a expansion of $\hat{\alpha}_i(\theta)$ as T grows after taking conditional expectations, see (1.2.13) and the expressions in Appendix 1.A. Let

$$
\begin{aligned}
\hat{\beta}_i(\theta) &= -\hat{\sigma}_i^4(\theta) \hat{E}_T \left[H_{it}(\theta) g_{it}(\theta)\right]/2, && (1.3.6)\\
\hat{\mathcal{J}}(\theta) &= -\frac{1}{n}\sum_{i=1}^n \left\{\hat{E}_T\left[H_{it}(\theta) f_{it}(\theta) x_{it} x_{it}'\right] - \hat{\sigma}_i^2(\theta)\hat{E}_T\left[H_{it}(\theta) f_{it}(\theta) x_{it}\right]\hat{E}_T\left[H_{it}(\theta) f_{it}(\theta) x_{it}'\right]\right\}, && (1.3.7)
\end{aligned}
$$

where $\hat{\beta}_i(\theta)$ is an estimator of the higher-order asymptotic bias of $\hat{\alpha}_i(\theta)$ from a stochastic expansion as T grows, and $\hat{\mathcal{J}}(\theta)$ is an estimator of the Jacobian of the estimating equation for θ. Then, the estimator of $\mathcal{B}$ is

$$
\hat{\mathcal{B}}(\theta) = -\hat{\mathcal{J}}(\theta)^{-1}\hat{b}(\theta), \quad \hat{b}(\theta) = -\frac{1}{n}\sum_{i=1}^n \left\{\hat{E}_T\left[H_{it}(\theta) f_{it}(\theta) x_{it}\right]\hat{\beta}_i(\theta) + \hat{E}_T\left[H_{it}(\theta) g_{it}(\theta) x_{it}\right]\hat{\sigma}_i^2(\theta)/2\right\}, \tag{1.3.8}
$$

[15] Appendix 1.A gives the expressions of the bias for discrete choice models after taking conditional expectations.

[16] Brown and Newey (1998) results apply here by noting that the bias formula can be decomposed in unconditional expectation terms. The efficient estimators for each of these terms is the corresponding sample analog of the conditional expectation, given the regressors and individual effects. Then, the argument follows by delta method. See also Porter (2002).

where $\hat{b}(\theta)$ is an estimator of the bias of the estimating equation of θ.

One step bias corrected estimators can then be formed by evaluating the previous expression at the MLE, that is $\hat{\mathcal{B}} = \hat{\mathcal{B}}(\hat{\theta})$, and the iterated bias corrected estimator is the solution to $\hat{\theta}^{\infty} = \hat{\theta} - \hat{\mathcal{B}}(\hat{\theta}^{\infty})$. Monte Carlo experiments in Section 1.7 show that the previous higher-order refinements also improve the finite sample performance of the corrections.

1.3.2 Bias Correction of the Estimating Equation

The source of incidental parameters bias is the non-zero expectation of the estimating equation (first order condition) for $\hat{\theta}$ at the true parameter value θ_0, see (1.2.15). This suggests an alternative correction consisting of a modified estimating equation that has no bias at θ_0, up to a certain order in T (see, for e.g., Woutersen, 2002; HN; and Fernández-Val, 2004).[17] For the discrete choice model, the score-corrected estimator is the solution to

$$0 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}(\tilde{\theta}, \hat{\alpha}_i(\tilde{\theta})) - \frac{1}{T}\hat{b}(\tilde{\theta}). \tag{1.3.9}$$

HN and Fernández-Val (2004) show that this method is equivalent to the iterated bias correction of the estimator when the initial estimating equation is linear in θ. In general, the iterated estimator is the solution to an approximation to the unbiased estimating equation.

1.3.3 Modified (Profile) Maximum Likelihood (MML)

Cox and Reid (1987), in the context of robust inference with nuisance parameters, develop a method for reducing the sensitivity of MLEs of structural parameters to the presence of incidental parameters.[18] This method consists of adjusting the likelihood function to reduce the order of the bias of the corresponding estimating equation (see Liang, 1987; McCullagh and Tibsharani, 1990; and Ferguson, Reid and Cox, 1991). Lancaster (2000) and Arellano (2003) show that the modified profile likelihood, i.e. concentrating out the α_i's, takes the following form for panel discrete choice models

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} l_{it}(\theta, \hat{\alpha}_i(\theta)) - \frac{1}{2T}\frac{1}{n}\sum_{i=1}^{n}\log\frac{-\hat{E}_T\left[v_{it\alpha}(\theta, \hat{\alpha}_i(\theta))\right]}{\hat{E}_T\left[H_{it}(\theta) f_{it}(\theta)\right]^2} + \frac{1}{2}\log T. \tag{1.3.10}$$

[17] Neyman and Scott (1948) suggest this method, but do not give the general expression for the bias of the estimating equation.

[18] "Roughly speaking 'nuisance' parameters are those which are not of primary interest; 'incidental' parameters are nuisance parameters whose number increases with the sample size." C.f., Lancaster (2000), footnote 10.

Appendix 1.D shows that the estimating equation of the modified profile likelihood is equivalent to (1.3.9), up to order $o_p(1/T)$. The difference with (1.3.9) is that the MML estimating equation does not use conditional expectations of all the terms. For the logit, however, the two methods exactly coincide since the correction factor of the likelihood does not depend on ϵ. In this case $\hat{E}_T\left[v_{it\alpha}(\theta, \hat{\alpha}_i(\theta))\right] = -\hat{E}_T\left[H_{it}(\theta) f_{it}(\theta)\right]$, $H_{it}(\theta) = 1$, and the modified likelihood takes the simple form

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} l_{it}(\theta, \hat{\alpha}_i(\theta)) + \frac{1}{2T}\frac{1}{n}\sum_{i=1}^{n}\log \hat{E}_T\left[f_{it}(\theta)\right] + \frac{1}{2}\log T. \tag{1.3.11}$$

1.3.4 Example: Andersen (1973) Two-Period Logit Model

The previous discussion suggests that the most important distinction between the correction methods is whether to adjust the estimator or the estimating equation, and in the former case whether to use a one-step or an iterative procedure. Here, I compare the asymptotic properties of these alternative procedures in a simple example from Andersen (1973). This example is convenient analytically because the fixed-T probability limit of the MLE has a closed-form expression. This expression allows me to derive the probability limits of the bias-corrected estimators, and to compare them to the true parameter value.

The model considered is

$$y_{it} = 1\left\{x_{it}\theta_0 + \alpha_i - \epsilon_{it} \geq 0\right\}, \epsilon_{it} \sim \mathcal{L}(0, \pi^2/3) \quad t = 1, 2; \; i = 1, ..., n, \tag{1.3.12}$$

where $x_{i1} = 0$ and $x_{i2} = 1$ for all i, and $\mathcal{L}$ denotes the standardized logistic distribution. Andersen (1973) shows that the MLE, $\hat{\theta}$, converges to $2\theta_0 \equiv \theta^{ML}$, as $n \to \infty$, and derives a fixed-T consistent estimator for this model, the conditional logit estimator. Using the probability limit of the MLE and the expression of the bias in (1.3.8), the limit of the one-step bias-corrected estimator is

$$\hat{\theta}^1 = \hat{\theta} - \left(e^{\hat{\theta}/2} - e^{-\hat{\theta}/2}\right)/2 \xrightarrow{p} 2\theta - \left(e^{\theta} - e^{-\theta}\right)/2 \equiv \theta^1. \tag{1.3.13}$$

For the iterated bias-corrected estimator, the limit of the estimator (θ^∞) is the solution to the following nonlinear equation

$$\theta^\infty = 2\theta - \left(e^{\theta^\infty/2} - e^{-\theta^\infty/2}\right)/2. \tag{1.3.14}$$

Finally, Arellano (2003) derives the limit for the score-corrected estimator[19]

$$\tilde{\theta} \xrightarrow{p} 2\log\left(\frac{5e^{\theta}+1}{5+e^{\theta}}\right) \equiv \theta^{s}. \tag{1.3.15}$$

Figures 1 and 2 plot the limit of the corrected estimators as functions of the true parameter value θ_0.[20] Here, we can see that the large-T adjustments produce significant improvements over the MLE for a wide range of parameter values, even for $T = 2$. Among the corrected estimators considered, no estimator uniformly dominates the rest in terms of having smaller bias for all the parameter values. Thus, the one-step bias correction out-performs the other alternatives for low values of θ, but its performance deteriorates very quickly as the true parameter increases; the score correction dominates for medium range parameter values; and the iterated correction becomes the best for high values of θ.

1.4 Bias for Static Panel Probit Model

The expression for the bias takes a simple form for the static panel probit model, which helps explain the results of previous Monte Carlo studies (Greene, 2002; and HN). In particular, the bias can be expressed as a matrix-weighted average of the true parameter value, where the weighting matrices are positive definite. This implies that probit fixed effects estimators are biased away from zero if the regressor is scalar (as in the studies aforementioned). This property also holds regardless of the dimension of the regressor vector in the absence of heterogeneity, because in this case the weighting matrix is a scalar multiple of the identity matrix (Nelson, 1995). In general, however, matrix-weighted averages are difficult to interpret and sign except in special cases (see Chamberlain and Leamer, 1976). These results are stated in the following proposition:

Proposition 1 (Bias for Model Parameters) *Assume that (i)* $\epsilon_{it}|\bar{X}_i, \alpha_i \sim$ *i.i.d.* $\mathcal{N}(0,1)$, *(ii)* $E[XX'|\alpha]$ *exists and is nonsingular for almost all* α, *(iii)* $X_{it}|\alpha$ *is stationary and strongly missing with missing coefficients such that* $\sum_{m=1}^{\infty} \alpha_m^{59/60}/m < \infty$, *for almost all* α, *(iv)* α_i *are independent, (v)* $E\left[\|(X,\alpha)\|^{120}\right] < \infty$, *and (vi)* $n = o(T^3)$.[21] *Then,*

1.

$$\mathcal{B} = \frac{1}{2}\bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2 \mathcal{J}_i\right]\theta_0, \tag{1.4.1}$$

[19] He obtains this result for the MML estimator, but MML is the same as score correction for the logit case.

[20] See Arellano (2002) for a similar exercise comparing the probability limits of MML and ML estimators.

[21] $\|\cdot\|$ denotes the Euclidean norm.

where

$$\mathcal{J}_i = E_T\left[H_{it}f_{it}x_{it}x_{it}'\right] - \sigma_i^2 E_T\left[H_{it}f_{it}x_{it}\right] E_T\left[H_{it}f_{it}x_{it}'\right], \tag{1.4.2}$$

$$\sigma_i^2 = E_T\left[H_{it}f_{it}\right]^{-1} = E_T\left\{\phi(x_{it}'\theta_0+\alpha_i)^2 / \left[\Phi(x_{it}'\theta_0+\alpha_i)\left(1-\Phi(x_{it}'\theta_0+\alpha_i)\right)\right]\right\}^{-1}. \tag{1.4.3}$$

2. $\bar{E}\left[\mathcal{J}_i\right]^{-1}\sigma_i^2\mathcal{J}_i$ *is positive definite for almost all* α_i.

3. If $\alpha_i = a\ \forall i$, *then*

$$\mathcal{B} = \frac{1}{2}\sigma^2\theta_0, \tag{1.4.4}$$

where $\sigma^2 = E_T\left\{\phi(x_{it}'\theta_0+a)^2 / \left[\Phi(x_{it}'\theta_0+a)\left(1-\Phi(x_{it}'\theta_0+a)\right)\right]\right\}^{-1}$.

Proof. See Appendix 1.C. ■

Condition (i) is the probit modelling assumption; condition (ii) is standard for MLE (Newey and McFadden, 1994), and guarantees identification and asymptotic normality for MLEs of model parameters and individual effects; assumptions (iii), together with the moment condition (v), and (iv) are imposed in order to apply a Law of Large Numbers;[22] and assumptions (v) and (vi) guarantee the existence of, and uniform convergence of remainder terms in, the higher-order expansion of the bias (HN, and Fernández-Val, 2004). Note that the second result follows because σ_i^2 is the asymptotic variance of the estimator of the individual effect α_i, and $\mathcal{J}_i$ corresponds to the contribution of individual i to the inverse of the asymptotic variance of the estimator of the model parameter θ_0. Moreover, since $\sigma_i^2 \geq \Phi(0)\left[1-\Phi(0)\right]/\phi(0)^2 = \pi/2$, $\mathcal{B}$ can be bounded from below.

Corollary 1 *Under the conditions of Proposition 1*

$$\|\mathcal{B}\| \geq \frac{\pi}{4}\ \|\theta_0\|. \tag{1.4.5}$$

When the regressor is scalar or there is no heterogeneity, this lower bound establishes that the first order bias for each index coefficient is at least $\pi/8 \approx 40\%$, $\pi/16 \approx 20\%$ and $\pi/32 \approx 10\%$ for panels with 2, 4 and 8 periods, respectively.[23] In general, these bounds apply to the norm of the coefficient vector. Tighter bounds can be also established for the proportionate bias,

[22] The stationarity condition can be relaxed to accommodate deterministic regressors, such as time dummies or linear trends.

[23] For two-period panels, the incidental parameters bias of the probit estimator is 100 % (Heckman, 1981). Part of the difference between the bias and the lower bound in this case can be explained by the importance of higher order terms, which have growing influence as the number of periods decreases.

$||\mathcal{B}||/||\theta_0||$, as a function of the true parameter value θ_0. These bounds, however, depend on the joint distribution of regressor and individual effects, and are therefore application specific. Thus, using standard matrix algebra results (see, e.g., Rao, 1973, p. 74), the proportionate bias can be bounded from below and above by the minimum and maximum eigenvalues of the matrix $\bar{E}\left[\mathcal{J}_i\right]^{-1} \bar{E}\left[\sigma_i^2 \mathcal{J}_i\right]/2$, for any value of the parameter vector θ_0.[24]

The third result of the proposition establishes that the bias is proportional to the true parameter value in the absence of heterogeneity. The intuition for this result can be obtained by looking at the linear model. Specifically, suppose that $y_{it} = x_{it}'\beta_0 + \alpha_i + \epsilon_{it}$, where $\epsilon_{it} \sim i.i.d.(0, \sigma_\epsilon^2)$. Next, note that in the probit the index coefficients are identified only up to scale, that is $\theta_0 = \beta_0/\sigma_\epsilon$. The probability limit of the fixed effects estimator of this quantity in the linear model, as $n \to \infty$, is

$$\hat{\theta} = \frac{\hat{\beta}}{\hat{\sigma}_\epsilon} \xrightarrow{p} \frac{\beta_0}{\sqrt{1 - 1/T}\sigma_\epsilon} = \left[1 + \frac{1}{2T}\right]\theta_0 + O(T^{-2}), \tag{1.4.6}$$

where the last equality follows from a standard Taylor expansion of $(1 - 1/T)^{-1/2}$ around $1/T = 0$. Here we can see the parallel with the probit, where $\theta_T = \left[1 + \sigma^2/2T\right]\theta_0 + O(T^{-2})$. Hence, we can think of the bias as coming from the estimation of σ_ϵ, which in the probit case cannot be separated from the estimation of β_0. In other words, the over-fitting due to the fixed effects biases upwards the estimates of the model parameters because the standard deviation is implicitly in the denominator of the model parameter estimated by the probit.

Proportionality implies, in turn, zero bias for fixed effects estimators of ratios of index coefficients. These ratios are often structural parameters of interest because they are direct measures of the relative effect of the regressors, and can be interpreted as marginal rates of substitution in many economic applications.

Corollary 2 *Assume that the conditions of the Proposition 1 hold and $\alpha_i = a \;\; \forall i$. Then, for any $j \neq k \in \{1, ..., p\}$ and $\theta = (\theta_1, ..., \theta_p)$*

$$\frac{\hat{\theta}_j}{\hat{\theta}_k} \xrightarrow{p} \frac{\theta_{0,j}}{\theta_{0,k}} + O(T^{-2}). \tag{1.4.7}$$

In general, the first term of the bias is different for each coefficient depending on the distribution of the individual effects, and the relationship between regressors and individual effects; and shrinks to zero with the inverse of the variance of the underlying distribution of individual

[24] Chesher and Jewitt (1987) use a similar argument to bound the bias of the Eicker-White heteroskedasticity consistent covariance matrix estimator.

effects.

1.5 Marginal Effects: Small Bias Property

1.5.1 Parameters of Interest

In discrete choice models the ultimate quantities of interest are often the marginal effects of specific changes in the regressors on the response conditional probability (see, e.g., Angrist, 2001; Ruud 2001; Greene, 2002; Wooldridge, 2002; and Wooldridge, 2003). However, unlike in linear models, structural parameters in nonlinear index models are only informative about the sign and relative magnitude of the effects. In addition, an attractive feature of these models is that marginal effects are heterogeneous across individuals. This allows, for instance, the marginal effects to be decreasing in the propensity (measured by the individual effect) to experience the event. Thus, individuals more prone to work are arguably less sensitive to marginal changes on other observable characteristics when deciding labor force participation.

For a model with two regressors, say X_1 and X_2, and corresponding parameters θ_1 and θ_2, the marginal effect of a one-unit increase in X_1 on the conditional probability of Y is defined as

$$F_\epsilon((x_1+1)\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) - F_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha). \tag{1.5.1}$$

When X_1 is continuous, the previous expression is usually approximated by a local version based on the derivative of the conditional probability with respect to x_1, that is

$$\frac{\partial}{\partial x_1} F_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) = \theta_1 f_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha), \tag{1.5.2}$$

where $f_\epsilon(\cdot|\bar{X}, \alpha)$ is the conditional pdf associated with $F_\epsilon(\cdot|\bar{X}, \alpha)$. These measures are heterogenous in the individual effect α and the level chosen for evaluating the regressors.

What are the relevant effects to report? A common practice is to give some summary measure, for example, the average effect or the effect for some interesting value of the regressors. Chamberlain (1984) suggests reporting the average effect for an individual randomly drawn from the population, that is

$$\mu(x_1) = \int \left[F_\epsilon((x_1+1)\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) - F_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha)\right] dG_{\bar{X}_2,\alpha}(\bar{x}_2, a), \tag{1.5.3}$$

or

$$\mu = \int \theta_1 f_\epsilon(x_1\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) dH_{\bar{X}_1,\bar{X}_2,\alpha}(\bar{x}_1, \bar{x}_2, a), \tag{1.5.4}$$

where G and H are the joint distribution of $(\bar{X}_2, \alpha)$ and $(\bar{X}, \alpha)$, respectively, and x_1 is some interesting value of X_1. The previous measures correspond to different thought experiments. The first measure, commonly used for discrete variables, corresponds to the counterfactual experiment where the change on the outcome probability is evaluated as if all the individuals would have chosen x_1 initially and receive an additional unit of X_1. The second measure, usually employed for continuous variables, is the average derivative of the response probabilities with respect to x_1, i.e., the average effect of giving one additional unit of X_1. The fixed effects estimators for these measures are

$$\hat{\mu}(x_1) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[F_\epsilon((x_1+1)\hat{\theta}_1 + x_{2it}\hat{\theta}_2 + \hat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha) - F_\epsilon(x_1\hat{\theta}_1 + x_{2it}\hat{\theta}_2 + \hat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha)\right], \tag{1.5.5}$$

and

$$\hat{\mu} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\theta}_1 f_\epsilon(x_{1it}\hat{\theta}_1 + x_{2it}\hat{\theta}_2 + \hat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha), \tag{1.5.6}$$

respectively. Note that the first measure corresponds to the average treatment effect if X_1 is a treatment indicator.

These effects can be calculated also for subpopulations of interest by conditioning on the relevant values of the covariates. For example, if X_1 is binary (treatment indicator), the average treatment effect on the treated (ATT) is

$$\mu_{ATT} = \int \left[F_\epsilon(\theta_1 + x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha) - F_\epsilon(x_2\theta_2 + a|\bar{X}_1, \bar{X}_2, \alpha)\right] dG_{\bar{X}_2,\alpha}(x_2, a|X_1 = 1), \tag{1.5.7}$$

and can be estimated by

$$\hat{\mu}_{ATT} = \frac{1}{N_1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[F_\epsilon(\hat{\theta}_1 + x_{2it}\hat{\theta}_2 + \hat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha) - F(x_{2it}\hat{\theta}_2 + \hat{\alpha}_i|\bar{X}_1, \bar{X}_2, \alpha)\right]\mathbf{1}\{x_{1it} = 1\}, \tag{1.5.8}$$

where $N_1 = \sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{1}\{x_{1it} = 1\}$. Other alternative measures used in cross-section models, such as the effect evaluated for an individual with average characteristics, are less attractive for panel data models because they raise conceptual and implementation problems (see Carro, 2003, for a related discussion about other measures of marginal effects).[25]

[25]On the conceptual side, Chamberlain (1984) and Ruud (2000) argue that this effect may not be relevant for

1.5.2 Bias Correction of Marginal Effects

HN develop analytical and jackknife bias correction methods for fixed effect averages, which include marginal effects. Let $m(r, \theta, \alpha)$ denote the change in the outcome conditional probability as a response to a one-unit increase in the first regressor $F_\epsilon\left((r_1 + 1)\theta_1 + r_{-1}'\theta_{-1} + \alpha|\bar{X}, \alpha\right) - F_\epsilon(r_1\theta_1 + r_{-1}'\theta_{-1} + \alpha|\bar{X}, \alpha)$, or its local approximation $\theta_1 f_\epsilon(r'\theta + \alpha|\bar{X}, \alpha)$ if X_1 is continuous.[26] The object of interest is then

$$\mu = \bar{E}\left[m(r, \theta_0, \alpha)\right], \tag{1.5.9}$$

where $r = (x_1, X_2)$ for discrete X_1 and $r = X$ for continuous X_1.[27] The fixed effects MLE of μ is then given by

$$\hat{\mu} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m\left(r_{it}, \hat{\theta}, \hat{\alpha}_i(\hat{\theta})\right), \tag{1.5.10}$$

where $r_{it} = (x_1, x_{2it})$ or $r_{it} = x_{it}$. For the bias corrections, let $\tilde{\theta}$ be a bias-corrected estimator (either one-step, iterated or derived from a bias-corrected estimating equation) of θ_0 and $\tilde{\alpha}_i = \hat{\alpha}_i(\tilde{\theta})$, $i = 1, ..., n$, the corresponding estimators of the individual effects.[28] Then, a bias-corrected estimator of μ is given by

$$\tilde{\mu} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m\left(r_{it}, \tilde{\theta}, \tilde{\alpha}_i\right) - \frac{1}{T}\tilde{\Delta}, \tag{1.5.11}$$

$$\tilde{\Delta} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{m_\alpha\left(r_{it}, \tilde{\theta}, \tilde{\alpha}_i\right)\left[\hat{\beta}_i\left(\tilde{\theta}\right) + \hat{\psi}_{it}\left(\tilde{\theta}\right)\right] + \frac{1}{2}m_{\alpha\alpha}\left(r_{it}, \tilde{\theta}, \tilde{\alpha}_i\right)\hat{\sigma}_i^2\left(\tilde{\theta}\right)\right\}, \tag{1.5.12}$$

where subscripts on m denote partial derivatives. Note that the $\hat{\psi}_{it}\left(\tilde{\theta}\right)$ term can be dropped since r_{it} does not depend on ϵ_{it}.

most of the population. The practical obstacle relates to the difficulty of estimating average characteristics in panel models. Thus, replacing population expectations for sample analogs does not always work in binary choice models estimated using a fixed-effects approach. The problem here is that the MLEs of the individual effects are unbounded for individuals that do not change status in the sample, and therefore the sample average of the estimated individual effects is generally not well defined.

[26] For a $p \times 1$ vector v, v_i denotes the $i-th$ component and v_{-i} denotes $(v_1, ..., v_{i-1}, v_{i+1}, ..., v_p)'$.

[27] If X_1 is binary r_1 is usually set to 0.

[28] Bias-corrected estimators for marginal effects can also be constructed from fixed-T consistent estimators. Thus, the conditional logit estimator can be used as $\tilde{\theta}$ in the logit model. This possibility is explored in the Monte Carlo experiments and the empirical application.

1.5.3 Panel Probit: Small Bias Property

Bias-corrected estimators of marginal effects are consistent up to order $O\left(T^{-2}\right)$ and have asymptotic distributions centered at the true parameter value if $T/n^{1/3} \to \infty$. This can be shown using a large T-expansion of the estimator, just as for model parameters. The question addressed here is whether these corrections are indeed needed. In other words, how important is the bias that the corrections aim to remove? This question is motivated by Monte Carlo evidence in HN, which shows negligible biases for uncorrected fixed effects estimators of marginal effects in a specific example. The following proposition gives the analytical expression for the bias of probit fixed effects estimators of marginal effects.

Proposition 2 (Bias for Marginal Effects) *Let* $\hat{\mu} = \hat{\theta}\ \sum_{i=1}^{n}\sum_{t=1}^{T}\phi\left(x_{it}'\hat{\theta} + \hat{\alpha}_i(\hat{\theta})\right)/nT$ *and* $\mu = \theta_0\ \bar{E}\left[\phi\left(X'\theta_0 + \alpha\right)\right]$, *then under the conditions of Proposition 1, as* $n, T \to \infty$

$$\hat{\mu} \xrightarrow{p} \mu + \frac{1}{T}\mathcal{B}_\mu + O(T^{-2}), \tag{1.5.13}$$

where

$$\mathcal{B}_\mu = \frac{1}{2}\bar{E}\left\{\phi\left(\xi_{it}\right)\left[\xi_{it}\theta_0\left(x_{it} - \sigma_i^2 E_T\left[H_{it}f_{it}x_{it}\right]\right)' - \mathcal{I}_p\right]\left(\sigma_i^2\mathcal{I}_p - \bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\right)\right\}\theta_0, \tag{1.5.14}$$

$$\xi_{it} = x_{it}'\theta_0 + \alpha_i, \tag{1.5.15}$$

$$\sigma_i^2 = E_T\left[H_{it}f_{it}\right]^{-1} = E_T\left\{\phi(\xi_{it})^2/\left[\Phi(\xi_{it})\Phi(-\xi_{it})\right]\right\}^{-1}, \tag{1.5.16}$$

$$\mathcal{J}_i = -\left\{E_T\left[H_{it}f_{it}x_{it}x_{it}'\right] - \sigma_i^2 E_T\left[H_{it}f_{it}x_{it}\right]E_T\left[H_{it}f_{it}x_{it}'\right]\right\}, \tag{1.5.17}$$

and $\mathcal{I}_p$ *denotes a* $p \times p$ *identity matrix.*

Proof. See Appendix 1.C. ■

When x_{it} is scalar all the formulas can be expressed as functions uniquely of the index ξ_{it} as

$$\mathcal{B}_\mu = E_\alpha\left[\mathcal{B}_{\mu i}\right] = E_\alpha\left[\delta_i\pi_i\right]\theta_0/2, \tag{1.5.18}$$

with $\delta_i = E_X\left\{\phi(\xi_{it})\left[\xi_{it}\left(\xi_{it} - E_T\left[H_{it}f_{it}\right]^{-1}E_T\left[H_{it}f_{it}\xi_{it}\right]\right) - 1\right]|\alpha\right\}$, $\pi_i = \sigma_i^2 - \bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]$, and $\mathcal{J}_i = -\left\{E_T\left[\omega_{it}\xi_{it}^2\right] - \sigma_i^2 E_T\left[\omega_{it}\xi_{it}\right]^2\right\}$. Here, we can see that the bias is an even function of the index, ξ_{it}, since all the terms are centered around weighted means that are symmetric around the origin.[29] The terms δ_i and π_i have U shapes and take the same sign for $\alpha_i = 0$ (when

[29] A function $f : \mathbb{R}^n \to \mathbb{R}$ is *even* if $f(-x) = f(x)\ \forall x \in \mathbb{R}^n$.

the mean of the regressors is absorbed in the individual effects); the term $\phi(\xi_{it})$ in δ_i acts to reduce the weights in the tails, where the other components are large. This is shown in Figure 3, which plots the components of the bias for independent normally distributed regressor and individual effect. In this case the bias, as a function of α, is positive at zero and takes negative values as we move away from the origin. Then, positive and negative values compensate each other when they are integrated using the distribution of the individual effect to obtain $\mathcal{B}_\mu$.

The following example illustrates the argument of the proof (in Appendix 1.C) and shows a case where the bias is exactly zero.

Example 1 (Panel probit without heterogeneity) *Consider a probit model where the individual effect is the same for all the individuals, that is $\alpha_i = a \;\; \forall i$. In this case, as $n, T \to \infty$*

$$\hat{\mu} \xrightarrow{p} \mu + O_p(T^{-2}). \tag{1.5.19}$$

First, note that as $n \to \infty$

$$\hat{\mu} \xrightarrow{p} \bar{E}\left\{\theta_T E_T\left[\phi\left(x_{it}'\theta_T + \hat{\alpha}_i(\theta_T)\right)\right]\right\}. \tag{1.5.20}$$

Next, in the absence of heterogeneity (see Proposition 1, and proof of Proposition 2 in Appendix 1.C)

$$\theta_T = \theta_0 + \frac{1}{2T}\sigma^2\theta_0 + O\left(T^{-2}\right), \;\; \hat{\alpha}_i(\theta_T) = a + \tilde{\psi}_i/\sqrt{T} + a\sigma^2/2T + R_i/T^{3/2}, \tag{1.5.21}$$

where under the conditions of Proposition 2

$$\tilde{\psi}_i \overset{a}{\sim} N(0, \sigma^2), \;\; \sigma^2 = E_T\left[H_{it} f_{it}\right]^{-1}, \quad \bar{E}\left[R_i\right] = O(T^{-2}). \tag{1.5.22}$$

Combining these results, the limit of the index, $\hat{\xi}_{it} = x_{it}'\theta_T + \hat{\alpha}_i(\theta_T)$, has the following expansion

$$\hat{\xi}_{it} = \left(1 + \sigma^2/2T\right)\xi_{it} + \psi_i/\sqrt{T} + R_i/\sqrt{T} + O(T^{-2}), \quad \xi_{it} = x_{it}'\theta_0 + a. \tag{1.5.23}$$

Finally, replacing this expression in (1.5.20), using the convolution properties of the normal distribution, see Lemma 1 in Appendix 1.C, and assuming that orders in probability correspond *to orders in expectation,*

$$\hat{\mu} \xrightarrow{p} \bar{E}\left\{\theta_T\ E\left[\phi(\hat{\xi}_{it})|\bar{X},\alpha\right]\right\} = \bar{E}\left[\theta_T\ \int \phi\left(\left(1+\frac{\sigma^2}{2T}\right)\xi_{it}+v\right)\frac{\sqrt{T}}{\sigma}\ \phi\left(\frac{\sqrt{T}v}{\sigma}\right)dv\right]$$
$$+\quad O(T^{-2}) = \bar{E}\left[\frac{(1+\sigma^2/2T)}{\sqrt{1+\sigma^2/T}}\ \theta_0\ \phi\left(\frac{(1+\sigma^2/2T)\,\xi}{\sqrt{1+\sigma^2/T}}\right)\right] + O(T^{-2}) = \mu + O(T^{-2}), \tag{1.5.24}$$

since by a standard Taylor expansion

$$\left(1+\sigma^2/T\right)^{-1/2}\left(1+\sigma^2/2T\right) \;=\; \left(1-\frac{\sigma^2}{2T}+O(T^{-2})\right)\left(1+\frac{\sigma^2}{2T}\right) = 1+O(T^{-2}). \tag{1.5.25}$$

In other words, the standard deviation of the random part of the limit index exactly compensates for the first term of the bias in the conditional expectation of the nonlinear function $\phi(\cdot)$.

The intuition for this result is the equivalent for panel probit of the consistency of average survivor probabilities in the linear Gaussian panel model. Thus, HN show that $\hat{S} = \sum_{i=1}^{n}\Phi\left(\frac{x'\hat{\theta}+\hat{\alpha}_i(\hat{\theta})}{\hat{\sigma}}\right)/n$ is a consistent estimator for $S = \bar{E}\left\{\Phi\left(\frac{x'\theta_0+\alpha_i}{\sigma}\right)\right\}$ for fixed T, because averaging across individuals exactly compensates the bias of the estimator of σ. In the nonlinear model, however, the result holds only approximately, since averaging reduces the bias of the MLE of the average effects by one order of magnitude from $O\left(T^{-1}\right)$ to $O\left(T^{-2}\right)$.

This example shows that, as in linear models, the inclusion of irrelevant variables, while reducing efficiency, does not affect the consistency of the probit estimates of marginal effects. Moreover, this example also complements Wooldridge's (2002, Ch. 15.7.1) result about neglected heterogeneity in panel probit models. Wooldridge shows that estimates of average effects that do not account for unobserved heterogeneity are consistent, if the omitted heterogeneity is normally distributed and independent of the included regressors. Here, on the other hand, I find that estimates of marginal effects that account for heterogeneity are consistent in the absence of such heterogeneity.

In general, the bias depends upon the degree of heterogeneity and the joint distribution of regressors and individual effects. Table 1 reports numerical values for the bias of fixed effects estimators of model parameters and marginal effects (in percent of the true value) for several distributions of regressors and individual effects. These examples correspond to an 8-period model with one regressor, and the model parameter θ_0 equal to 1. All the distributions, except for the Nerlove process for the regressor, are normalized to have zero mean and unit variance.[30] The numerical results show that the first term of the bias for the marginal effect is below 2% for all the configurations considered, and is always lower than the bias for the model parameter, which is about 15% (larger than the lower bound of 10%). The values of the bias for Nerlove regressor and normal individual effect are close to their Monte Carlo estimates in Section 1.7. Thus, the theoretical bias for the model parameter and marginal effect are 15% and -0.23%, and their Monte Carlo estimates are 18% and -1% (see Tables 2 and 3).

When can we use uncorrected fixed effects estimators of marginal effects in practice? The expression of the bias derived in Proposition 2 is also useful to answer this question. Thus, since the bias is a fixed effects average, its value can be estimated in the sample using the procedure described in HN. Moreover, a standard Wald test can be constructed to determine whether the bias is significantly different from zero.

1.6 Extension: Dynamic Discrete Choice Models

1.6.1 The Model

Consider now the following dynamic version of the panel discrete choice model

$$Y = \mathbf{1}\left\{\theta_{y,0}Y_{-1} + X'\theta_{x,0} + \alpha - \epsilon \geq 0\right\}, \tag{1.6.1}$$

where Y_{-1} is a binary random variable that takes on value one if the outcome occurred in the previous period and zero otherwise. The rest of the variables are defined as in the static case. In this model, persistence in the outcome can be a consequence of higher unobserved individual propensity to experience the event in all the periods, as measured by α, or to alterations in the individual behavior for having experienced the event in the previous period, as measured by $\theta_{y,0}Y_{-1}$. Heckman (1981) refers to these sources of persistence as heterogeneity and true state dependence, respectively. Examples of empirical studies that use this type of specification include Card and Sullivan (1988), Moon and Stotsky (1993), Roberts and Tybout (1997), Hyslop (1999), Chay and Hyslop (2000), and Carro (2003).

To estimate the model parameters, I adopt a fixed-effects estimation approach. This approach has the additional advantage in dynamic models of not imposing restrictions on the initial conditions of the process (Heckman, 1981). Then, given a sample of the observable vari-

[30]The Nerlove process is not normalized to help me compare with the results of the Monte Carlo in Section 1.7.

ables and assuming a distribution for ϵ conditional on $(Y_{-1}, \bar{X}, \alpha)$, the model parameters can be estimated by maximum likelihood conditioning on the initial observation of the sample.

1.6.2 Large-T Approximation to the Bias

In the presence of dynamics, fixed effects MLEs of structural parameters suffer from the incidental parameters problem even when the model is linear; see, for e.g., Nerlove (1967), Nerlove (1971), and Nickell (1981). Formulas for the bias can be obtained using large-T asymptotic expansions of the estimators, which in this case include Hurwicz-type terms due to the correlations between the observations. Thus, let $\bar{Z}_t$ denote $(\bar{X}, Y_{t-1}, ..., Y_0)$, and $E_{T-j}[z_{t-k}]$ denote $\lim_{T\to\infty} \sum_{t=k+1}^{T} E\left[z_{t-k}|\bar{Z}_{t-j}, \alpha\right]/(T-k)$ for $k \leq j$, then a large-T expansion for the estimators of the individual effects can be constructed as

$$\hat{\alpha}_i = \alpha_i + \psi_i^d/\sqrt{T} + \beta_i^d/T + o_p(1/T), \quad \psi_i^d = \sum_{t=1}^{T} \psi_{it}^d/\sqrt{T} \xrightarrow{d} \mathcal{N}(0, \sigma^{2,d}), \tag{1.6.2}$$

$$\psi_{it}^d = -E_T\left[v_{it\alpha}\right]^{-1} v_{it}, \quad \sigma_i^{d,2} = -E_T\left[v_{it\alpha}\right]^{-1} + 2\lim_{T\to\infty}\sum_{j=1}^{T-1} E_{T-j}\left[\psi_{it}^d \psi_{i,t-j}^d\right], \tag{1.6.3}$$

$$\beta_i^d = -E_T\left[v_{it\alpha}\right]^{-1}\left\{\lim_{T\to\infty}\sum_{j=0}^{T-1} E_{T-j}\left[v_{it\alpha}\psi_{i,t-j}^d\right] + \frac{1}{2}\sigma_i^{d,2} E_T\left[v_{it\alpha\alpha}\right]\right\}. \tag{1.6.4}$$

As in the static model, ψ_{it}^d, $\sigma_i^{d,2}$ and β_i^d are the influence function, first-order variance and higher-order bias of $\hat{\alpha}_i$ as $T \to \infty$. For the common parameter, the expressions for the Jacobian and the first term of the bias of the estimating equation are

$$\mathcal{J}^d = \bar{E}\left\{E_T\left[u_{it\theta}\right] - E_T\left[u_{it\alpha}\right]\frac{E_T\left[v_{it\theta}\right]}{E_T\left[v_{it\alpha}\right]}\right\}, \tag{1.6.5}$$

$$b^d = \bar{E}\left\{E_T\left[u_{it\alpha}\right]\beta_i^d + \lim_{T\to\infty}\sum_{j=0}^{T-1} E_{T-j}\left[u_{it\alpha}\psi_{i,t-j}^d\right] + \frac{1}{2}\sigma_i^{d,2} E_T\left[u_{it\alpha\alpha}\right]\right\}. \tag{1.6.6}$$

The first term of the bias of the fixed effects estimator of θ_0 is then $\mathcal{B}^d = -\left(\mathcal{J}^d\right)^{-1} b^d$. This expression corresponds to the bias formula for general nonlinear panel models derived in Hahn and Kuersteiner (2003), where all the terms that depend on Y have been replaced for their conditional expectation given $(\bar{X}, Y_{-1}, \alpha)$. This adjustment removes zero conditional mean terms without affecting the asymptotic properties of the correction.[31] Monte Carlo results in Sec-

[31] Appendix 1.A derives the bias formulas for dynamic discrete choice models, and Appendix 1.B describes the corrections.

tion 1.7 show that this higher-order refinement improves the finite sample performance of the correction for a dynamic logit model.

1.6.3 Marginal Effects

Marginal effects can be defined in the same manner as for the static model. Bias corrections for fixed effects estimators also extend naturally to the dynamic case by adding some correlation terms. Thus, using the notation of Section 1.5, the estimator of the bias can be formed as

$$
\begin{aligned}
\tilde{\Delta}^d &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{m_\alpha\left(r_{it},\tilde{\theta},\tilde{\alpha}_i\right)\left[\hat{\beta}_i^d\left(\tilde{\theta}\right)+\hat{\psi}_{it}^d\left(\tilde{\theta}\right)\right]+\frac{1}{2}m_{\alpha\alpha}\left(r_{it},\tilde{\theta},\tilde{\alpha}_i\right)\hat{\sigma}_i^{d,2}\left(\tilde{\theta}\right)\right\} \\
&+ \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\sum_{j=1}^{J}m_\alpha\left(r_{it},\tilde{\theta},\tilde{\alpha}_i\right)\hat{\psi}_{i,t-j}^d\left(\tilde{\theta}\right)\right\}. \qquad (1.6.7)
\end{aligned}
$$

Here, J is a bandwidth parameter that needs to be chosen such that $J/T^{1/2} \to 0$ as $T \to \infty$, see Hahn and Kuersteiner (2003).

The small bias property for fixed effects estimators of marginal effects does not generally hold in dynamic models. The reason is that averaging across individuals does not remove the additional bias components due to the dynamics. To understand this result, we can look at the survivor probabilities at zero in a dynamic Gaussian linear model. Specifically, suppose that $y_{it} = \theta_0 y_{i,t-1}+\alpha_i+\epsilon_{it}$, where $\epsilon_{it}|y_{i,t-1},...,y_{i,0},\alpha_i \sim N(0,\sigma^2)$, $y_{i0}|\alpha_i \sim \mathcal{N}(\alpha_i/(1-\theta_0),\sigma^2/(1-\theta_0^2))$, and $0 \leq \theta_0 < 1$. The survivor probability evaluated at $y_{i,t-1} = r$ and its fixed effects estimator are

$$
S = \bar{E}\left\{\Phi\left(\frac{\theta_0\, r+\alpha_i}{\sigma}\right)\right\},\quad \hat{S} = \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{\hat{\theta}\, r+\hat{\alpha}_i(\hat{\theta})}{\hat{\sigma}}\right), \qquad (1.6.8)
$$

where $\hat{\theta}$ and $\hat{\sigma}^2$ are the fixed effects MLEs of θ_0 and σ^2. It can be shown that $\hat{\theta}$ converges to $\theta_T = \theta_0 - (1+\theta_0)/T + O(T^{-2})$, and $\hat{\sigma}^2$ converges to $\sigma_T^2 = \sigma^2 - \sigma^2/T + O(T^{-2})$, as $n \to \infty$ (Nickell, 1981). For the estimator of the individual effects, a large-T expansion gives

$$
\hat{\alpha}_i(\theta_T) = \alpha_i + v_i - (\theta_T - \theta_0)\frac{\alpha_i}{1-\theta_0} + o_p(1/T),\quad v_i \sim \mathcal{N}(0,\sigma^2/T). \qquad (1.6.9)
$$

Then, as $n, T \to \infty$

$$\begin{aligned}
\hat{S} & \xrightarrow{p} \bar{E}\left\{E\left[\Phi\left(\frac{\theta_T\ u + \hat{\alpha}_i(\theta_T)}{\sigma_T}\right)\right]\right\} \\
& = \bar{E}\left\{E\left[\Phi\left(\frac{\theta_0 u + \alpha_i + v_i + (\theta_T - \theta_0)\left(u - \frac{\alpha_i}{1-\theta_0}\right) + o_p(T^{-1})}{\sigma_T}\right)\right]\right\} \\
& = \bar{E}\left\{\Phi\left(\frac{\theta_0 u + \alpha_i - \frac{1}{T}(1+\theta_0)\left(u - \frac{\alpha_i}{1-\theta_0}\right) + o(T^{-1})}{\sigma}\right)\right\}, \qquad (1.6.10)
\end{aligned}$$

by the convolution properties of the normal distribution, see Lemma 1 in Appendix 1.C.

In (1.6.10) we can see that averaging across individuals eliminates the bias of $\hat{\sigma}^2$, but does not affect the bias of $\hat{\theta}$. The sign of the bias of $\hat{S}$ generally depends on the distribution of the individual effects. When there is no heterogeneity ($\alpha_i = a\ \forall i$), for example, $\hat{S}$ underestimates (overestimates) the underlying survivor probability when evaluated at values above (below) the unconditional mean of the response, $a/(1-\theta_0)$. This means that if the marginal effects are thought of as differences in survivor probabilities evaluated at two different values u_1 and u_0, fixed effects estimates of marginal effects would be biased downward if the values chosen are $u_1 = 1$ and $u_0 = 0$. For exogenous variables, Monte Carlo results suggest that the bias problem is less severe (see Section 1.7). Intuitively, it seems that the part of the bias due to the dynamics is less important for the exogenous regressors.[32]

1.7 Monte Carlo Experiments

This section reports evidence on the finite sample behavior of fixed effects estimators of model parameters and marginal effects for static and dynamic models. In particular, I analyze the finite sample properties of uncorrected and bias-corrected fixed effects estimators in terms of bias and

[32]For example, assume that in the previous dynamic linear model we add an exogenous regressor X with coefficient β_0, such that $x_{it} \sim$ *i.i.d.* $(0, \sigma_x^2)$ (the individual means are absorbed in the individual effect). Then, it can be shown that the fixed effects estimator of β_0 is fixed-T consistent, and the estimators of θ_0 and σ^2 have the same probability limits as before. The fixed effect estimator of the survivor probability, evaluated at (u_y, u_x), converges to, as $n, T \to \infty$,

$$\hat{S} = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{\hat{\theta}u_y + \hat{\beta}u_x + \hat{\alpha}_i}{\hat{\sigma}}\right) \xrightarrow{p} \bar{E}\left\{\Phi\left(\frac{\theta_0 u_y + \beta_0 u_x + \alpha_i - \frac{1}{T}(1+\theta_0)\left(u_y - \frac{\alpha_i}{1-\theta_0}\right) + o(T^{-1})}{\sigma}\right)\right\}. \qquad (1.6.11)$$

Hence, if there is no individual heterogeneity ($\alpha_i = a\ \forall i$) and the probability is evaluated at the unconditional mean of the lagged endogenous variable, i.e. $u_y = a/(1-\theta_0)$, then the fixed effects estimator of the survivor probability is large-T-consistent for every value of u_x. As a result, the derivative with respect to u_x, which is the analog of the marginal effect of X, is also large-T-consistent.

inference accuracy of the asymptotic distribution. The small bias property for marginal effects is illustrated for several lengths of the panel. Robustness of the estimators to small deviations from correct specification is also considered. Thus, the performance of probit and logit estimators is evaluated when the error term is logistic and normal, respectively. All the results presented are based on 1000 replications, and the designs are as in Heckman (1981), Greene (2002), and HN for the static probit model, and as in Honoré and Kyriazidou (2000), Carro (2003), and Hahn and Kuersteiner (2003) for the dynamic logit model.

1.7.1 Static Model

The model design is

$$y_{it} = \mathbf{1}\{x_{it}\theta_0 + \alpha_i - \epsilon_{it} \geq 0\}, \ \epsilon_{it} \sim \mathcal{N}(0,1), \ \alpha_i \sim \mathcal{N}(0,1), \tag{1.7.1}$$

$$x_{it} = t/10 + x_{i,t-1}/2 + u_{it}, \ \ x_{i0} = u_{i0}, \ \ u_{it} \sim \mathcal{U}(-1/2, 1/2), \tag{1.7.2}$$

$$n = 100, \ T = 4, \ 8, \ 12; \ \theta_0 = 1, \tag{1.7.3}$$

where $\mathcal{N}$ and $\mathcal{U}$ denote normal and uniform distribution, respectively. Throughout the tables reported, SD is the standard deviation of the estimator; $\hat{p}$; # denotes a rejection frequency with # specifying the nominal value; SE/SD is the ratio of the average standard error to standard deviation; and MAE denotes median absolute error.[33] $BC1$ and $BC2$ correspond to the one-step analytical bias-corrected estimators of HN based on maximum likelihood setting and general estimating equations, respectively. JK is the bias correction based on the leave-one-period-out Jackknife-type estimator, see HN. $BC3$ is the one-step bias-corrected estimator proposed here. $CLOGIT$ denotes Andersen's (1973) conditional logit estimator, which is fixed-T consistent when the disturbances are logistically distributed. Iterated and score-corrected estimators are not considered because they are much more cumbersome computationally.[34]

Table 2 gives the Monte Carlo results for the estimators of θ_0 when ϵ_{it} is normally distributed. Both probit and logit estimators are considered and logit estimates are normalized to help compare to the probit.[35] The results here are similar to previous studies (Greene, 2002; and HN) and show that the probit MLE is severely biased, even when $T = 12$, and has important distortions in rejection probabilities. $BC3$ has negligible bias, relative to standard deviation,

[33] I use median absolute error instead of root mean squared error as an overall measure of goodness of fit because it is less sensitive to outliers.

[34] HN find that iterating the bias correction does not matter much in this example.

[35] Estimates and standard errors for logit are multiplied by $\sqrt{3}/\pi$ in order to have the same scale as for probit.

and improves in terms of bias and rejection probabilities over HN's analytical and jackknife bias-corrected estimators for small sample sizes.[36] It is also remarkable that all the bias corrections and the conditional logit estimator are robust to the type of misspecification considered, even for a sample size as small as $T = 4$. This resembles the well-known similarity between probit and logit estimates in cross sectional data, see Amemiya (1981), but it is more surprising here since the bias correction formulas and conditional logit estimator rely heavily on the form of the true likelihood.

Table 3 reports the ratio of estimators to the truth for marginal effects. Here, I include also two estimators of the average effect based on linear probability models. $LPM - FS$ is the standard linear probability model that uses all the observations; LPM is an adjusted version that calculates the slope from individuals that change status during the sample, i.e., excluding individuals with $y_{it} = 1\ \forall t$ or $y_{it} = 0\ \forall t$, and assigns zero effect to the rest. The results are similar to HN and show small bias in uncorrected fixed effects estimators of marginal effects. Rejection frequencies are higher than their nominal levels, due to underestimation of dispersion. As in cross-section models (Angrist, 2001), both linear models work fairly well in estimating the average effect.[37]

1.7.2 Dynamic Model

The model design is

$$\begin{aligned}
y_{i0} &= \mathbf{1}\left\{\theta_{X,0} x_{i0} + \alpha_i - \epsilon_{i0} \geq 0\right\}, && (1.7.4)\\
y_{it} &= \mathbf{1}\left\{\theta_{Y,0} y_{i,t-1} + \theta_{X,0} x_{it} + \alpha_i - \epsilon_{it} \geq 0\right\}, \quad t = 1, ..., T-1, && (1.7.5)\\
\epsilon_{it} &\sim \mathcal{L}(0, \pi^2/3), \quad x_{it} \sim \mathcal{N}\left(0, \pi^2/3\right), && (1.7.6)\\
n &= 250;\ T = 8,\ 12,\ 16;\ \theta_{Y,0} = .5;\ \theta_{X,0} = 1, && (1.7.7)
\end{aligned}$$

where $\mathcal{L}$ denotes the standardized logistic distribution. Here, the individual effects are correlated with the regressor. In particular, to facilitate the comparison with other studies, I follow

[36] In this case the bias correction reduces the dispersion of the fixed effects estimator. This can be explained by the proportionality result of the bias in Proposition 1. Thus, the bias corrected estimator takes the form $\left[\mathcal{I}_p - \hat{A}/T\right]\hat{\theta}$, which reduces to $\left[1 - \hat{\sigma}^2/2T\right]\hat{\theta}$ in the absence of heterogeneity.

[37] Stoker (1986) shows that linear probability models estimate consistently average effects in index models (e.g., probit and logit) under normality of regressors and individual effects. In general, however, the bias of the linear probability model depends on the covariance between the conditional probability of the index and the deviations from normality of the index, and on the covariance between the conditional probability of the index and the deviations from linearity of the conditional expectation of regressors and individual effects given the index (see equation (6.1) in Stoker, 1986).

Honoré and Kyriazidou (2000) and generate $\alpha_i = \sum_{t=0}^{3} x_{it}/4$. The measures reported are the same as for the static case, and logit and probit estimators are considered.[38] $BC1$ denotes the bias-corrected estimator of Hahn and Kuersteiner (2003); HK is the dynamic version of the conditional logit of Honoré and Kyriazidou (2000), which is fixed-T consistent; MML is the Modified MLE for dynamic models of Carro (2003); and $BC3$ is the bias-corrected estimator that uses conditional expectations in the derivation of the bias formulas.[39] For the number of lags, I choose a bandwidth parameter $J = 1$, as in Hahn and Kuersteiner (2003).

Tables 4 and 5 present the Monte Carlo results for the structural parameters $\theta_{Y,0}$ and $\theta_{X,0}$. Overall, all the bias-corrected estimators have smaller finite sample bias and better inference properties than the uncorrected MLEs. Large-T-consistent estimators have median absolute error comparable to HK for $T = 8$.[40] Among them, $BC3$ and MML are slightly superior to $BC1$, but there is no clear ranking between them.[41] Thus, $BC3$ has smaller bias and MAE for $\theta_{Y,0}$, but has larger bias and MAE for $\theta_{X,0}$. As for the static model, the bias corrections are robust to the type of misspecification considered for moderate T.

Tables 6 and 7 report the Monte Carlo results for ratios of the estimator to the truth for average effects for the lagged dependent variable and exogenous regressor, respectively. These effects are calculated using expression (1.5.5) with $x_1 = 0$ for the lagged dependent variable, and expression (1.5.6) for the exogenous regressor. Here, I present results for MLE, $BC1$, $BC3$, linear probability models (LPM and $LPM-FS$), and bias-corrected linear models ($BC-LPM$ and $BC-LPM-FS$) constructed using Nickell's (1981) bias formulas. As in the example of the linear model in Section 1.6, uncorrected estimates of the effects of the lagged dependent variable are biased downward. Uncorrected estimates of the effect for the exogenous variable, however, have small biases. Large-T corrections are effective in reducing bias and fixing rejection probabilities for both linear and nonlinear estimators.

1.8 Empirical Application: Female Labor Force Participation

The relationship between fertility and female labor force participation is of longstanding interest in labor economics and demography. For a recent discussion and references to the literature, see

[38]Probit estimates and standard errors are multiplied by $\pi/\sqrt{3}$ to have the same scale as for logit.

[39]HK and MML results are extracted from the tables reported in their articles and therefore some of the measures are not available. HK results are based on a bandwidth parameter equal to 8.

[40]An aspect not explored here is that the performance of HK estimator deteriorates with the number of exogenous variables. Thus, Hahn and Kuersteiner (2003) find that their large-T-consistent estimator out-performs HK for $T = 8$ when the model includes two exogenous variables.

[41]Note, however, that $BC3$ is computationally much more simpler than MML.

Angrist and Evans (1998). Research on the causal effect of fertility on labor force participation is complicated because both variables are jointly determined. In other words, there exist multiple unobserved factors (to the econometrician) that affect both decisions. Here, I adopt an empirical strategy that aims to solve this omitted variables problem by controlling for unobserved individual time-invariant characteristics using panel data. Other studies that follow a similar approach include Heckman and MaCurdy (1980), Heckman and MaCurdy (1982), Hyslop (1999), Chay and Hyslop (2000), Carrasco (2001), and Carro (2003).

The empirical specification I use is similar to Hyslop (1999). In particular, I estimate the following equation

$$P_{it} = \mathbf{1}\left\{\delta_t + P_{i,t-1}\theta_P + X_{it}'\theta_X + \alpha_i - \epsilon_{it} \geq 0\right\}, \tag{1.8.1}$$

where P_{it} is the labor force participation indicator; δ_t is a period-specific intercept; $P_{i,t-1}$ is the participation indicator of the previous period; and X_{it} is a vector of time-variant covariates that includes three fertility variables - the numbers of children aged 0-2, 3-5, and 6-17 -, log of husband's earnings, and a quadratic function of age.[42]

The sample is selected from waves 13 to 22 of the Panel Study of Income Dynamics (PSID) and contains information of the ten calendar years 1979-1988. Only women aged 18-60 in 1985, continuously married, and whose husband is in the labor force in each of the sample periods are included in the sample. The final sample consists of 1,461 women, 664 of whom change labor force participation status during the sample period. The first year is excluded to use it as initial condition for the dynamic model.

Descriptive statistics for the sample are shown in Table 8. Twenty-one percent of the sample is black, and the average age in 1985 was 37. Roughly 72% of women participate in the labor force at some period, the average schooling is 12 years, and the average numbers of children are .2, .3 and 1.1 for the three categories 0-2 year-old, 3-5 year-old, and 6-17 year-old children, respectively.[43] Women that change participation status during the sample, in addition to be younger, less likely to be black, and less educated, have more dependent children and their husband's earnings are slightly higher than average. Interestingly, women who never participate do not have more children than women who are employed each year, though this can be explained in part by the non-participants being older. All the covariates included in the empirical

[42]Hyslop (1999) specification includes also the lag of the number of 0 to 2 year-old children as additional regressor. This regressor, however, is statistically nonsignificant at the 10% level.

[43]Years of schooling is imputed from the following categorical scheme: 1 = '0-5 grades' (2.5 years); 2 = '6-8 grades' (7 years); 3 = '9-11 grades' (10 years); 4 = '12 grades' (12 years); 5 = '12 grades plus nonacademic training' (13 years); 6 = 'some college' (14 years); 7 = 'college degree' (15 years); 7 = 'college degree, not advanced' (16 years); 8 = 'college and advanced degree' (18 years). See also Hyslop (1999).

specification display time variation over the period considered.

Table 9 reports fixed effects estimates of index coefficients and marginal effects obtained from a static specification, that is, excluding the lag of participation in equation (1.8.1). Estimators are labeled as in the Monte Carlo example. The results show that uncorrected estimates of index coefficients are about 15 percent larger than their bias-corrected counterparts; whereas the corresponding differences for marginal effects are less than 2 percent, and insignificant relative to standard errors. It is also remarkable that all the corrections considered give very similar estimates for both index coefficients and marginal effects (for example, bias-corrected logit estimates are the same as conditional logit estimates, up to two decimal points).[44] The adjusted linear probability model gives estimates of the marginal effects closer to logit and probit than the standard linear model. According to the static model estimates, an additional child aged less than 2 reduces the probability of participation by 9 percent, while each child aged 3-5 and 6-17 reduces the probability of participation by 5 percent and 2 percent, respectively.

In the presence of positive state dependence, estimates from a static model overstate the effect of fertility because additional children reduce the probability of participation and participation is positively serially correlated. This can be seen in Table 10, which reports fixed effects estimates of index coefficients and marginal effects using a dynamic specification. Here, as in the Monte Carlo example, uncorrected estimates of the index and effect of the lagged dependent variable are significantly smaller (relative to standard errors) than their bias-corrected counterparts for both linear and nonlinear models. Moreover, unlike in the Monte Carlo examples, uncorrected estimates of the effects of the regressors are biased away from zero. Bias-corrected probit gives estimates of index coefficients very similar to probit Modified Maximum Likelihood.[45] The adjusted linear probability model, again, gives estimates of the average effects closer to logit and probit than the standard linear model. Each child aged 0-2 and 3-5 reduces the probability of participation by 6 percent and 3 percent, respectively; while an additional child aged more than 6 years does not have a significant effect on the probability of participation (at the 5 percent level). Finally, a one percent increase in the income earned by the husband reduces a woman's probability of participation by about 0.03%. This elasticity is not sensitive to the omission of dynamics or to the bias corrections.

[44]Logit index coefficients are multiplied by $\sqrt{3}/\pi$ to have the same scale as probit index coefficients.

[45]Modified Maximum Likelihood estimates are taken from Carro (2003)

1.9 Summary and conclusions

This chapter derives bias-corrected fixed effects estimators for model parameters of panel discrete choice models that have better asymptotic and finite sample properties than other similar corrections. The idea behind these corrections is analogous to the use of the conditional information matrix in the variance estimation problem. Thus, the corrections presented here are based on bias formulas that use more intensively the parametric structure of the problem by taking conditional expectations given regressors and individual effects.

The new bias formulas are used to derive analytical expressions for the bias of fixed effects estimators of index coefficients and marginal effects in probit models. The expression for the index coefficients shows that the bias is proportional to the true value of the parameter and can be bounded from below. Moreover, fixed effects estimators of ratios of coefficients and marginal effects do not suffer from the incidental parameters problem in the absence of heterogeneity, and generally have smaller biases than fixed effects estimators of the index coefficients. These results are illustrated with Monte Carlo examples and an empirical application that analyzes female labor force participation using data from the PSID.

It would be useful to know if the small bias property of fixed effects estimators of average effects generalizes to other statistics of the distribution of effects in the population, like median effects or other quantile effects. However, such analysis is expected to be more complicated because these statistics are non-smooth functions of the data and therefore the standard expansions cannot be used. I leave this analysis for future research.

Appendix

1.A Bias Formulas for Binary Choice Models

1.A.1 Static Case

The conditional log-likelihood and the scores for observation i at time t are

$$
\begin{aligned}
l_{it}(\theta, \alpha_i) &= y_{it} \log F_{it}(\theta, \alpha_i) + (1 - y_{it}) \log(1 - F_{it}(\theta, \alpha_i)), && (1.A.1)\\
v_{it}(\theta, \alpha_i) &= H_{it}(\theta, \alpha_i)\left(y_{it} - F_{it}(\theta, \alpha_i)\right), \quad u_{it}(\theta, \alpha_i) = v_{it}(\theta, \alpha_i) x_{it}, && (1.A.2)
\end{aligned}
$$

where $F_{it}(\theta, \alpha_i)$ denotes $F_{\epsilon}(x_{it}'\theta + \alpha_i | \bar{X} = \bar{x}_i, \alpha = \alpha_i)$, f is the pdf associated to F, and $H_{it}(\theta, \alpha_i) = f_{it}(\theta, \alpha_i) / \left[F_{it}(\theta, \alpha_i)\left(1 - F_{it}(\theta, \alpha_i)\right)\right]$.

Next, since by the Law of Iterated Expectations $E_Z\left[h(z_{it})|\alpha\right] = E_X\left[E_Y\left[h(z_{it})|X, \alpha\right]|\alpha\right]$ for any function $h(z_{it})$, taking conditional expectations of the expressions for the components of the bias in Section 1.2 yields

$$
\begin{aligned}
\sigma_i^2 &= E_T\left[H_{it} f_{it}\right]^{-1}, \quad \beta_i = -\sigma_i^4 E_T\left[H_{it} g_{it}\right]/2, && (1.A.3)\\
b &= -\bar{E}\left\{E_T\left[H_{it} f_{it} x_{it}\right]\beta_i + E_T\left[H_{it} g_{it} x_{it}\right]\sigma_i^2/2\right\}, && (1.A.4)\\
\mathcal{J} &= -\bar{E}\left\{E_T\left[H_{it} f_{it} x_{it} x_{it}'\right] - \sigma_i^2 E_T\left[H_{it} f_{it} x_{it}\right] E_T\left[H_{it} f_{it} x_{it}'\right]\right\}, && (1.A.5)
\end{aligned}
$$

where g denotes the derivative of f and all the expressions are evaluated at the true parameter value (θ_0, α_i).

1.A.2 Dynamic Case

The conditional log-likelihood and the scores for observation i at time t are

$$
\begin{aligned}
l_{it}(\theta, \alpha_i) &= y_{it} \log F_{it}(\theta, \alpha_i) + (1 - y_{it}) \log(1 - F_{it}(\theta, \alpha_i)), && (1.A.6)\\
v_{it}(\theta, \alpha_i) &= H_{it}(\theta, \alpha_i)\left(y_{it} - F_{it}(\theta, \alpha_i)\right), \quad u_{it}(\theta, \alpha_i) = v_{it}(\theta, \alpha_i) x_{it}, && (1.A.7)
\end{aligned}
$$

where $F_{it}(\theta, \alpha_i)$ denotes $F_{\epsilon}(\theta_y y_{i,t-1} + x_{it}'\theta_x + \alpha_i | Y_{t-1} = y_{i,t-1}, ..., Y_0 = y_0, \bar{X} = \bar{x}_i, \alpha = \alpha_i)$, $H_{it}(\theta, \alpha_i) = f_{it}(\theta, \alpha_i) / \left[F_{it}(\theta, \alpha_i)\left(1 - F_{it}(\theta, \alpha_i)\right)\right]$, and f is the pdf associated to F.

Next, taking conditional expectations of the expressions for the components of the bias in Section 1.6

and using the formulas for the static case, yields

$$\psi_{it}^d = \sigma_i^2 v_{it}, \quad \sigma_i^{2,d} = \sigma_i^2 + 2 \lim_{T\to\infty} \sum_{j=1}^{T-1} E_{T-j}\left[\psi_{it}^d \psi_{i,t-j}^d\right], \tag{1.A.8}$$

$$\beta_i^d = \beta_i + \sigma_i^2 \lim_{T\to\infty} \sum_{j=1}^{T-1} \left\{E_{T-j}\left[H_{it} f_{it} \psi_{i,t-j}^d\right] + E_{T-j}\left[\psi_{it}^d \psi_{i,t-j}^d\right] E_T\left[H_{it} g_{it} + 2G_{it} f_{it}\right]\right\}, \tag{1.A.9}$$

$$\begin{aligned} b^d &= -\bar{E}\left\{E_T\left[H_{it} f_{it} x_{it}\right]\beta_i^d + E_T\left[H_{it} g_{it} x_{it}\right]\sigma_i^2/2 + \lim_{T\to\infty}\sum_{j=1}^{T-1} E_{T-j}\left[H_{it} f_{it} \psi_{i,t-j}^d x_{it}\right]\right\} \\ &- \bar{E}\left\{\lim_{T\to\infty}\sum_{j=1}^{T-1} E_{T-j}\left[\psi_{it}^d \psi_{i,t-j}^d\right] E_T\left[H_{it} g_{it} x_{it} + 2G_{it} f_{it} x_{it}\right]\right\}, \end{aligned} \tag{1.A.10}$$

$$\mathcal{J}^d = \mathcal{J}, \tag{1.A.11}$$

where g denotes the derivative of f, $G_{it} = \left(g_{it} F_{it}(1 - F_{it}) - f_{it}^2(1 - 2F_{it})\right) / \left[F_{it}(1 - F_{it})\right]^2$ is the derivative of H_{it}, and all the expressions are evaluated at the true parameter value (θ_0, α_i).

1.B Bias Corrections in Dynamic models

Here, I use the expressions in Appendix 1.A to construct bias-corrected estimators for the dynamic model. Let $\theta \equiv (\theta_y, \theta_x')'$, $z_{it} \equiv (y_{i,t-1}, x_{it}')'$ and

$$\begin{array}{ll} F_{it}(\theta) \equiv F_\epsilon(z_{it}'\theta + \hat{\alpha}_i(\theta)|\bar{Z}_t, \alpha), & f_{it}(\theta) \equiv f_\epsilon(z_{it}'\theta + \hat{\alpha}_i(\theta)|\bar{Z}_t, \alpha), \\ g_{it}(\theta) \equiv f_\epsilon'(z_{it}'\theta + \hat{\alpha}_i(\theta)|\bar{Z}_t, \alpha), & H_{it}(\theta) \equiv \frac{f_{it}(\theta)}{F_{it}(\theta)(1-F_{it}(\theta))}, \\ G_{it}(\theta) \equiv \frac{g_{it}(\theta)}{F_{it}(\theta)(1-F_{it}(\theta))} - \frac{f_{it}(\theta)^2(1-2F_{it}(\theta))}{[F_{it}(\theta)(1-F_{it}(\theta))]^2}. & \end{array} \tag{1.B.1}$$

Then, the components of the large-T expansion for the estimator of the individual effects can be estimated adding some terms to the analogous expressions for the static case. Thus,

$$\hat{\psi}_{it}^d(\theta) = H_{it}(\theta)\left[y_{it} - z_{it}'\theta - \hat{\alpha}_i(\theta)\right], \tag{1.B.2}$$

$$\hat{\sigma}_i^{d,2}(\theta) = \hat{\sigma}_i^2(\theta) + 2\sum_{j=1}^{J} \hat{E}_{T-j}\left[\hat{\psi}_{it}^d(\theta)\hat{\psi}_{i,t-j}^d(\theta)\right], \tag{1.B.3}$$

$$\begin{aligned} \hat{\beta}_i^d(\theta) &= \hat{\beta}_i(\theta) + \hat{\sigma}_i^2(\theta)\sum_{j=1}^{J} \hat{E}_{T-j}\left[H_{it}(\theta) f_{it}(\theta)\hat{\psi}_{i,t-j}^d(\theta)\right] \\ &+ \hat{\sigma}_i^2(\theta)\sum_{j=1}^{J} \hat{E}_{T-j}\left[\hat{\psi}_{i,t}^d(\theta)\hat{\psi}_{i,t-j}^d(\theta)\right] \hat{E}_T\left[H_{it}(\theta) g_{it}(\theta) + 2G_{it}(\theta) f_{it}(\theta)\right]. \end{aligned} \tag{1.B.4}$$

Similarly, for the estimator of the common parameter

$$
\begin{aligned}
\hat{\mathcal{J}}^d(\theta) &= \hat{\mathcal{J}}(\theta), &\text{(1.B.5)}\\
\hat{b}^d(\theta) &= -\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{E}_T\left[H_{it}(\theta)f_{it}(\theta)x_{it}\right]\hat{\beta}_i^d(\theta)+\sum_{j=1}^{J}\hat{E}_{T-j}\left[H_{it}(\theta)g_{it}(\theta)x_{it}\hat{\psi}^d_{i,t-j}(\theta)\right]\right\}\\
&- \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{J}\hat{E}_{T-j}\left[\hat{\psi}_{i,t}(\theta)\hat{\psi}_{i,t-j}(\theta)\right]\hat{E}_T\left[H_{it}\left(\theta\right)g_{it}(\theta)x_{it}+2G_{it}\left(\theta\right)f_{it}(\theta)x_{it}\right]\right\}\\
&- \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}^2(\theta)\hat{E}_T\left[H_{it}(\theta)g_{it}(\theta)x_{it}\right]/2. &\text{(1.B.6)}
\end{aligned}
$$

Here, J is a bandwidth parameter that needs to be chosen such that $J/T^{1/2}\to 0$ as $T\to\infty$, see Hahn and Kuersteiner (2003). For the first-order variance of the estimator of the individual effects, a kernel function can be used to guarantee that the estimates are positive, e.g., Newey and West (1987). From these formulas, all the bias-corrected estimators described in Section 1.3 can be formed.

For dynamic binary choice models, Lancaster (2000) and Woutersen (2002) derive score bias correction methods, and Carro (2003) extends the Modified Maximum Likelihood estimator of Cox and Reid (1987). The exact relationship between all these methods and the approach followed in Hahn and Kuersteiner (2003) is not yet known. The reason is that the equivalence results from the static model do not generalize directly to the dynamic case.

Finally, note that the asymptotic variance of the estimator of the common parameter needs to be adjusted to take into account the dependence across the observations, and is no longer the inverse of the Jacobian of the estimating equation. In this case, we have the standard sandwich formula

$$\mathcal{V}=\left(-\mathcal{J}^d\right)^{-1}\Omega\left(-\mathcal{J}^d\right)^{-1},\quad \Omega=\bar{V}\left\{T^{1/2}E_T\left[U_{it}\right]\right\},\quad U_{it}=u_{it}-v_{it}\frac{E_T\left[u_{it\alpha}\right]}{E_T\left[v_{it\alpha}\right]}, \tag{1.B.7}$$

where Ω can be estimated using a kernel function to guarantee positive definiteness, see Hahn and Kuersteiner (2003).

1.C Proofs

1.C.1 Lemmas

Lemma 1 *(**McFadden and Reid, 1975**) Let $Z\sim\mathcal{N}(\mu_Z,\sigma_Z^2)$, and $a,b\in\mathbb{R}$ with $b>0$. Then,*

$$\Phi\left(\frac{\mu_Z+a}{\sqrt{b^2+\sigma_Z^2}}\right)=\int\Phi\left(\frac{z+a}{b}\right)\frac{1}{\sigma_Z}\phi\left(\frac{z-\mu_Z}{\sigma_Z}\right)dz, \tag{1.C.1}$$

and

$$\frac{1}{\sqrt{b^2+\sigma_Z^2}}\phi\left(\frac{\mu_Z+a}{\sqrt{b^2+\sigma_Z^2}}\right)=\int\frac{1}{b}\phi\left(\frac{z+a}{b}\right)\frac{1}{\sigma_Z}\phi\left(\frac{z-\mu_Z}{\sigma_Z}\right)dz, \tag{1.C.2}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote cdf and pdf of the standard normal distribution, respectively.

Proof. First, take X independent of Z, with $X \sim \mathcal{N}(-a, b^2)$. Then,

$$\Pr\{X - Z \leq 0\} = \Phi\left(\frac{\mu_Z + a}{\sqrt{b^2 + \sigma_Z^2}}\right) \tag{1.C.3}$$

since $X - Z \sim \mathcal{N}(-a - \mu_Z, b^2 + \sigma_Z^2)$. Alternatively, using the law of iterated expectations and $X|Z \sim X$ by independence,

$$\Pr\{X - Z \leq 0\} = E_Z\left[\Pr\{X \leq Z|Z\}\right] = \int \Phi\left(\frac{z+a}{b}\right)\frac{1}{\sigma_Z}\phi\left(\frac{z - \mu_Z}{\sigma_Z}\right)dz. \tag{1.C.4}$$

The second statement follows immediately by deriving both sides of expression (1.C.1) with respect to a. ∎

1.C.2 Proof of Proposition 1

Proof. First, note that for the probit $g_{it} = -(x_{it}'\theta_0 + \alpha_i)f_{it}$. Then, substituting this expression for g_{it} in the bias formulas of the static model, see Appendix 1.A, yields

$$\begin{aligned}
\beta_i &= \left\{\sigma_i^4 E_T\left[H_{it} f_{it} x_{it}'\right]\theta_0 + \sigma_i^2 \alpha_i\right\}/2, &(1.C.5)\\
\mathcal{J} &= -\bar{E}\left\{E_T\left[H_{it} f_{it} x_{it} x_{it}'\right] - \sigma_i^2 E_T\left[H_{it} f_{it} x_{it}\right] E_T\left[H_{it} f_{it} x_{it}'\right]\right\} = \bar{E}\left[\mathcal{J}_i\right], &(1.C.6)\\
b &= -\bar{E}\left\{\sigma_i^4 E_T\left[H_{it} f_{it} x_{it}\right] E_T\left[H_{it} f_{it} x_{it}'\right]\theta_0 + \sigma_i^2 E_T\left[H_{it} f_{it} x_{it}\right]\alpha_i\right\}/2\\
&+ \bar{E}\left\{\sigma_i^2 E_T\left[H_{it} f_{it} x_{it} x_{it}'\right]\theta_0 + \sigma_i^2 E_T\left[H_{it} f_{it} x_{it}\right]\alpha_i\right\}/2\\
&= \bar{E}\left\{\sigma_i^2\left(E_T\left[H_{it} f_{it} x_{it} x_{it}'\right] - \sigma_i^2 E_T\left[H_{it} f_{it} x_{it}\right] E_T\left[H_{it} f_{it} x_{it}'\right]\right)\right\}\theta_0/2 = -\bar{E}\left[\sigma_i^2 \mathcal{J}_i\right]\theta_0/2. &(1.C.7)
\end{aligned}$$

Finally, we have for the bias

$$\mathcal{B} = -\mathcal{J}^{-1} b = \frac{1}{2}\bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2 \mathcal{J}_i\right]\theta_0. \tag{1.C.8}$$

The second and third results are immediate and are described in the text. ∎

1.C.3 Proof of Proposition 2

Proof. We want to find the probability limit of $\hat{\mu} = \sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\theta}\phi\left(x_{it}'\hat{\theta} + \hat{\alpha}_i(\hat{\theta})\right)/nT$, as $n, T \to \infty$, and compare it to the population parameter of interest $\mu = \bar{E}\left[\theta_0 \phi\left(x_{it}'\theta_0 + \alpha_i\right)\right]$.

First, note that by the Law of Large Number and Continuous Mapping Theorem, as $n \to \infty$

$$\hat{\mu} \xrightarrow{p} \bar{E}\left\{\theta_T E_T\left[\phi\left(x_{it}'\theta_T + \hat{\alpha}_i(\theta_T)\right)\right]\right\}. \tag{1.C.9}$$

Next, we have the following expansion for the limit index, $\hat{\xi}_{it}(\theta_T) \equiv x'_{it}\theta_T + \hat{\alpha}_i(\theta_T)$, around θ_0

$$\hat{\xi}_{it}(\theta_T) \quad = \quad x'_{it}\theta_0 + \hat{\alpha}_i(\theta_0) + \left[x'_{it} + \frac{\partial \hat{\alpha}_i(\bar{\theta})}{\partial \theta'}\right](\theta_T - \theta_0). \tag{1.C.10}$$

Using independence across t, standard higher-order asymptotics for $\hat{\alpha}_i(\theta_0)$ give (e.g., Ferguson, 1992, or Rilstone *et al.*, 1996), as $T \to \infty$

$$\hat{\alpha}_i(\theta_0) = \alpha_i + \psi_i/\sqrt{T} + \beta_i/T + R_{1i}/T^{3/2}, \quad \psi_i \xrightarrow{d} \mathcal{N}(0, \sigma_i^2 \equiv -E_T\left[v_{it\alpha}\right]^{-1}), \tag{1.C.11}$$

where $R_i = O_p(1)$ and $E_T\left[R_i/T^2\right] = O(1)$ uniformly in i by the conditions of the proposition (e.g., HN and Fernández-Val, 2004). From the first order conditions for $\hat{\alpha}_i(\theta)$, we have, as $T \to \infty$

$$\frac{\partial \hat{\alpha}_i(\bar{\theta})}{\partial \theta'} = -\frac{E_T\left[v_{it\theta}\right]}{E_T\left[v_{it\alpha}\right]} + R_{2i}/\sqrt{T} = \sigma_i^2 E_T\left[v_{it\theta}\right] + R_{2i}/\sqrt{T}, \tag{1.C.12}$$

where $R_{2i} = O_p(1)$ and $E_T\left[R_{2i}/T\right] = O(1)$ uniformly in i, again by the conditions of the proposition. Plugging (1.C.11) and (1.C.12) into the expansion for the index in (1.C.10) yields, for $\xi_{it} = x'_{it}\theta_0 + \alpha_i$,

$$\hat{\xi}_{it}(\theta_T) \quad = \quad \xi_{it} + \psi_i/\sqrt{T} + \beta_{\xi_i}/T + R_{3i}/T^{3/2}, \tag{1.C.13}$$

where $\beta_{\xi_i} = \beta_i + T\left(x'_{it} + \sigma_i^2 E_T\left[v_{it\theta}\right]\right)(\theta_T - \theta_0)$, $R_{3i} = O_p(1)$ and $E_T\left[R_{3i}/T^2\right] = O(1)$, uniformly in i by the properties of R_{1i} and R_{2i}.

Then, using the expressions for the bias for the static probit model, see proof of Proposition 1, and $E_T\left[v_{it\theta}\right] = -E_T\left[H_{it}f_{it}x'_{it}\right]$, we have

$$\begin{aligned}
\beta_{\xi_i} &= \left(\sigma_i^4 E_T\left[H_{it}f_{it}x'_{it}\right]\theta_0 + \sigma_i^2\alpha_i\right)/2 + \left(x'_{it} - \sigma_i^2 E_T\left[H_{it}f_{it}x'_{it}\right]\right)\mathcal{B} + O(T^{-2}) \\
&= \sigma_i^2\xi_{it}/2 - \left(x'_{it} - \sigma_i^2 E_T\left[H_{it}f_{it}x'_{it}\right]\right)\sigma_i^2\theta_0/2 + \left(x'_{it} - \sigma_i^2 E_T\left[H_{it}f_{it}x'_{it}\right]\right)\bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\theta_0 \\
&+ \ O(T^{-2}) = \sigma_i^2\xi_{it}/2 - \mathcal{D}_i + O(T^{-2}),
\end{aligned} \tag{1.C.14}$$

where $\mathcal{D}_i = \left(x'_{it} - \sigma_i^2 E_T\left[H_{it}f_{it}x'_{it}\right]\right)\left(\sigma_i^2\mathcal{I}_p - \bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\right)\theta_0/2$, and $\mathcal{I}_p$ denotes the $p \times p$ identity matrix. Substituting the expression for β_{ξ_i} in (1.C.13) gives

$$\hat{\xi}_{it}(\theta_T) = \left[1 + \sigma_i^2/2T\right]\xi_{it} + \psi_i/\sqrt{T} - \mathcal{D}_i/T + R_i/T^{3/2}, \quad \psi_i/\sqrt{T} \overset{a}{\sim} \mathcal{N}(0, \sigma_i^2/T) \tag{1.C.15}$$

where $R_i = O_p(1)$ and $E_T\left[R_i/T^2\right] = O(1)$ uniformly in i.

Finally, using Lemma 1 and expanding around θ_0, it follows that

$$\begin{aligned}
\hat{\mu} &\xrightarrow{p} \bar{E}\left\{\theta_T E_T\left[\phi\left(\hat{\xi}_i(\theta_T)\right)\right]\right\} \\
&= \bar{E}\left\{\theta_T \int \phi\left(\left[1+\sigma_i^2/2T\right]\xi_{it} + v - \mathcal{D}_i/T\right)\frac{\sqrt{T}}{\sigma_i}\,\phi\left(\frac{\sqrt{T}v}{\sigma_i}\right)dv + O(T^{-2})\right\} \\
&= \bar{E}\left\{\left(1+\sigma_i^2/T\right)^{-1/2}\theta_T\phi\left(\frac{\left[1+\sigma_i^2/2T\right]\xi_{it} - \mathcal{D}_i/T}{\sqrt{1+\sigma_i^2/T}}\right) + O(T^{-2})\right\} \\
&= \mu + \frac{1}{2T}\bar{E}\left\{\phi(\xi_{it})\left(\xi_{it}\theta_0\left(x_{it} - \sigma_i^2 E_T\left[H_{it}f_{it}x_{it}\right]\right)' - \mathcal{I}_p\right)\left(\sigma_i^2\mathcal{I}_p - \bar{E}\left[\mathcal{J}_i\right]^{-1}\bar{E}\left[\sigma_i^2\mathcal{J}_i\right]\right)\theta_0\right\} \\
&+ O(T^{-2}) = \mu + \frac{1}{T}\mathcal{B}_\mu + O\left(T^{-2}\right), \qquad (1.C.16)
\end{aligned}$$

as T grows, since

$$\frac{\sigma_{iT}^2}{\sqrt{1+\sigma_i^2/T}} = \left(1+\frac{\sigma_i^2}{2T}\right)\left(1-\frac{\sigma_i^2}{2T}+O(T^{-2})\right) = 1 + O(T^{-2}). \qquad (1.C.17)$$

■

1.D Relationship between Bias Correction of the Estimating Equation and Modified Maximum Likelihood

The first order condition for the MML estimator is

$$0 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}(\check{\theta}) - \frac{1}{T}\tilde{b}(\check{\theta}), \qquad (1.D.1)$$

where

$$\tilde{b}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{2}\frac{\hat{E}_T\left[v_{it\alpha\theta}(\theta) + v_{it\alpha\alpha}(\theta)\partial\hat{\alpha}_i(\theta)/\partial\theta'\right]}{\hat{E}_T\left[v_{it\alpha}(\theta)\right]} + \frac{\hat{E}_T\left[\left(H_{it}(\theta)g_{it}(\theta) + G_{it}(\theta)f_{it}(\theta)\right)\left(x_{it} + \partial\hat{\alpha}_i(\theta)/\partial\theta'\right)\right]}{\hat{E}_T\left[H_{it}(\theta)f_{it}(\theta)\right]}\right\}. \qquad (1.D.2)$$

From the first order condition for $\hat{\alpha}_i(\theta)$, note that as $T \to \infty$

$$\frac{\partial\hat{\alpha}_i(\theta)}{\partial\theta} = -\frac{E_T\left[v_{it\theta}(\theta)\right]}{E_T\left[v_{it\alpha}(\theta)\right]} + o_p(1). \qquad (1.D.3)$$

Replacing this expression in (1.D.1), and using the Bartlett identities $v_{it\theta} = u_{it\alpha}$ and $v_{it\alpha\theta} = u_{it\alpha\alpha}$, as $n, T \to \infty$

$$\begin{aligned}
\tilde{b}(\theta) &\xrightarrow{p} \bar{E}\left\{\frac{1}{2}\frac{E_T\left[H_{it}(\theta)g_{it}(\theta)\right]}{E_T\left[H_{it}(\theta)f_{it}(\theta)\right]^2}E_T\left[H_{it}(\theta)f_{it}(\theta)x_{it}\right] - \frac{1}{2}\frac{E_T\left[H_{it}(\theta)g_{it}(\theta)x_{it}\right]}{E_T\left[H_{it}(\theta)f_{it}(\theta)\right]}\right\} \\
&= -\bar{E}\left\{E_T\left[H_{it}(\theta)f_{it}(\theta)x_{it}\right]\beta_i(\theta) + E_T\left[H_{it}(\theta)g_{it}(\theta)x_{it}\right]\sigma_i^2(\theta)/2\right\} = b(\theta). \qquad (1.D.4)
\end{aligned}$$

Bibliography

[1] ALVAREZ, J., AND M. ARELLANO (2003), "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica* 71(4), 1121-1159.

[2] AMEMIYA, T. (1981), "Qualitative Response Models: A Survey," *Journal of Economic Literature* 19(4), 1483-1536.

[3] ANDERSEN, E. B. (1973), *Conditional Inference and Models for Measuring.* Mentalhygiejnisk Forlag. Copenhagen.

[4] ANGRIST, J. D. (2001), "Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics* 19(1), 2-16.

[5] ANGRIST, J. D., AND W. H. EVANS (1998), "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review* 88(3), 450-477.

[6] ARELLANO, M. (2003), "Discrete Choices with Panel Data," *Investigaciones Económicas* XXVII (3), 423-458.

[7] ARELLANO, M., AND R. CARRASCO (2003), "Discrete Choice Panel Data Models with Predetermined Variables," *Journal of Econometrics* 115 (1), 125-157.

[8] ARELLANO, M., AND B. HONORÉ (2001), "Panel Data Models: Some Recent Developments," in J. J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Vol. 5*, Amsterdam: North-Holland.

[9] BROWN, B. W., AND W. K. NEWEY (1998), "Efficient Semiparametric Estimation of Expectations," *Econometrica* 66, 453-464.

[10] CARD, D., AND D. SULLIVAN (1988), "Measuring the Effect of Subsidied Training Programs on Movements In and Out of Employment," *Econometrica* 56, 497-530.

[11] CARRO, J. M. (2003), "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects," CEMFI working paper 0304.

[12] CARRASCO, R. (2001), "Binary Choice With Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labor Participation," *Journal of Business and Economic Statistics* 19(4), 385-394.

[13] CHAMBERLAIN, G. (1980), "Analysis of Covariance with Qualitative Data," *Review of Economic Studies* XLVII, 225-238.

[14] CHAMBERLAIN, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics* 18, 5-46.

[15] CHAMBERLAIN, G. (1984), "Panel Data," in Z. GRILICHES AND M. INTRILIGATOR, eds., *Handbook of Econometrics, Vol. 2.* Amsterdam: North-Holland.

[16] CHAMBERLAIN, G. (1985), "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in J. J. HECKMAN AND B. SINGER, eds., *Longitudinal Analysis of Labor Market Data.* Cambridge University Press.

[17] CHAMBERLAIN, G., AND E. E. LEAMER (1976), "Matrix Weighted Averages and Posterior Bounds," *Journal of the Royal Statistical Society. Series B* 38(1), 73-84.

[18] CHAY, K. Y., AND D. R. HYSLOP (2000), "Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence using Alternative Approaches," unpublished manuscript, University of California at Berkeley.

[19] CHESHER, A., AND I. JEWITT (1987), "The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator," *Econometrica* 55, 1217-1222.

[20] COX, D. R. (1958), "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society. Series B* 20(2), 215-242.

[21] COX, D. R., AND N. REID (1987), "Parameter Orthogonality and Approximate Conditional Inference," *Journal of the Royal Statistical Society. Series B* 49, 1-39.

[22] FERNÁNDEZ-VAL, I. (2004), "Bias Correction in Panel Data Models with Individual Specific Parameters,"working paper, MIT Department of Economics.

[23] FERGUSON, H. (1992), "Asymptotic Properties of a Conditional Maximum-Lkelihood Estimator," *The Canadian Journal of Statistics* 20(1), 63-75.

[24] FERGUSON, H., N. REID, AND D. R. COX (1991), "Estimating Equations from Modified Profile Likelihood, " in V. P. GODAMBE, ed., *Estimating Functions*, 279-293.

[25] GREENE, W. H. (2000), *Econometric Analysis*, Fourth Edition, Prentice Hall, New Jersey.

[26] GREENE, W.H. (2002), "The Behavior of the Fixed Effects Estimator in Nonlinear Models,"unpublished manuscript, New York University.

[27] HAHN, J. (2001), "Comment: Binary Regressors in Nonlinear Panel-Data Models With Fixed Effects," *Journal of Business and Economic Statistics* 19(1), 16-17.

[28] HAHN, J., AND G. KUERSTEINER (2002), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," *Econometrica* 70, 1639-1657.

[29] HAHN, J., AND G. KUERSTEINER (2003), "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," unpublished manuscript. UCLA.

[30] HAHN, J., AND W. NEWEY (2004), "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica* 72, 1295-1319.

[31] HECKMAN, J. J. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process, " in C. F. MANSKI AND D. MCFADDEN, eds., *Structural Analysis of Discrete Panel Data with Econometric Applications*, 179-195.

[32] HECKMAN, J. J., AND T. E. MACURDY (1980), "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies* 47, 47-74.

[33] HECKMAN, J. J., AND T. E. MACURDY (1982), "Corrigendum on: A Life Cycle Model of Female Labor Supply," *Review of Economic Studies* 49, 659-660.

[34] HONORÉ, B. E., AND E. KYRIAZIDOU (2000), "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica* 68, 839-874.

[35] HYSLOP, D. R. (1999), "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica* 67(6), 1255-1294.

[36] KATZ, E. (2001), "Bias in Conditional and Unconditional Fixed Effects Logit Estimation," *Political Analysis* 9(4), 379-384.

[37] KIEFER, J., AND J. WOLFOWITZ (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *The Annals of Mathematical Statistics* 27(4), 887-906.

[38] LANCASTER, T. (2002), "Orthogonal Parameters and Panel Data," *Review of Economic Studies* 69, 647-666.

[39] LIANG, K. Y. (1987), "Estimating Functions and Approximate Conditional Likelihood," *Biometrika* 74, 695-702.

[40] MANSKI, C. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica* 55, 357-362.

[41] MCCULLAGH, P., AND R. TIBSHIRANI (1990), "A Simple Method for the Adjustment of Profile Likelihoods," *Journal of the Royal Statistical Society. Series B* 52, 325-344.

[42] MOON, C.-G., AND J. G. STOTSKY (1993), "The Effect of Rent Control on Housing Quality Change: A Longitudinal Analysis," *Journal of Political Economy* 101, 1114-1148.

[43] MUNDLAK, Y. (1978), "On the Pooling of Time Series and Cross Section Data," *Econometrica* 46, 69-85.

[44] NELSON, D. B. (1995), "Vector Attenuation Bias in the Classical Errors-in-Variables Model," *Economics Letters* 49, 345-349.

[45] NERLOVE, M. (1967), "Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections," *Economic Studies Quarterly* 18, 42-74.

[46] NERLOVE, M. (1971), "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections," *Econometrica* 39, 359-387.

[47] NEWEY, W. K., AND D. MCFADDEN (1994), "Large Sample Estimation and Hypothesis Testing," in R.F. ENGLE AND D.L. MCFADDEN, eds., *Handbook of Econometrics, Vol. 4*. Elsevier Science. Amsterdam: North-Holland.

[48] NEWEY, W. K., AND K. D. WEST (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55(3), 703-708.

[49] NEYMAN, J., AND E.L. SCOTT (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica* 16, 1-32.

[50] NICKELL, S. (1981), "Biases in Dynamic Models with Fixed Effects," *Econometrica* 49, 1417-1426.

[51] PHILLIPS, P. C. B., AND H. R. MOON (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica* 67, 1057-1111.

[52] PORTER, J. (2002), "Efficiency of Covariance Matrix Estimators for Maximum Likelihood Estimation," *Review of Business and Economic Statistics* 20(3), 431-440.

[53] RAO, C. R. (1973), *Linear Statistical Inference and its Applications*, 2nd Edition. New York: John Wiley.

[54] RILSTONE, P., V. K. SRIVASTAVA, AND A. ULLAH (1996), "The Second-Order Bias and Mean Squared Error of Nonlinear Estimators, " *Journal of Econometrics* 75, 369-395.

[55] ROBERTS, M. J., AND J. R. TYBOUT (1997), "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs, " *The American Economic Review* 87(4), 545-564.

[56] RUDD, P. A. (2000), *An Introduction to Classical Econometric Theory*, Oxford University Press, New York.

[57] STOKER, T. M. (1986), "Consistent Estimation of Scaled Coefficients," *Econometrica* 54, 1461-1481.

[58] WOOLDRIDGE, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.

[59] WOOLDRIDGE, J. M. (2003), "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity ," *Journal of Applied Econometrics*, forthcoming.

[60] WOUTERSEN, T. M. (2002), "Robustness Against Incidental Parameters," unpublished manuscript, University of Western Ontario.

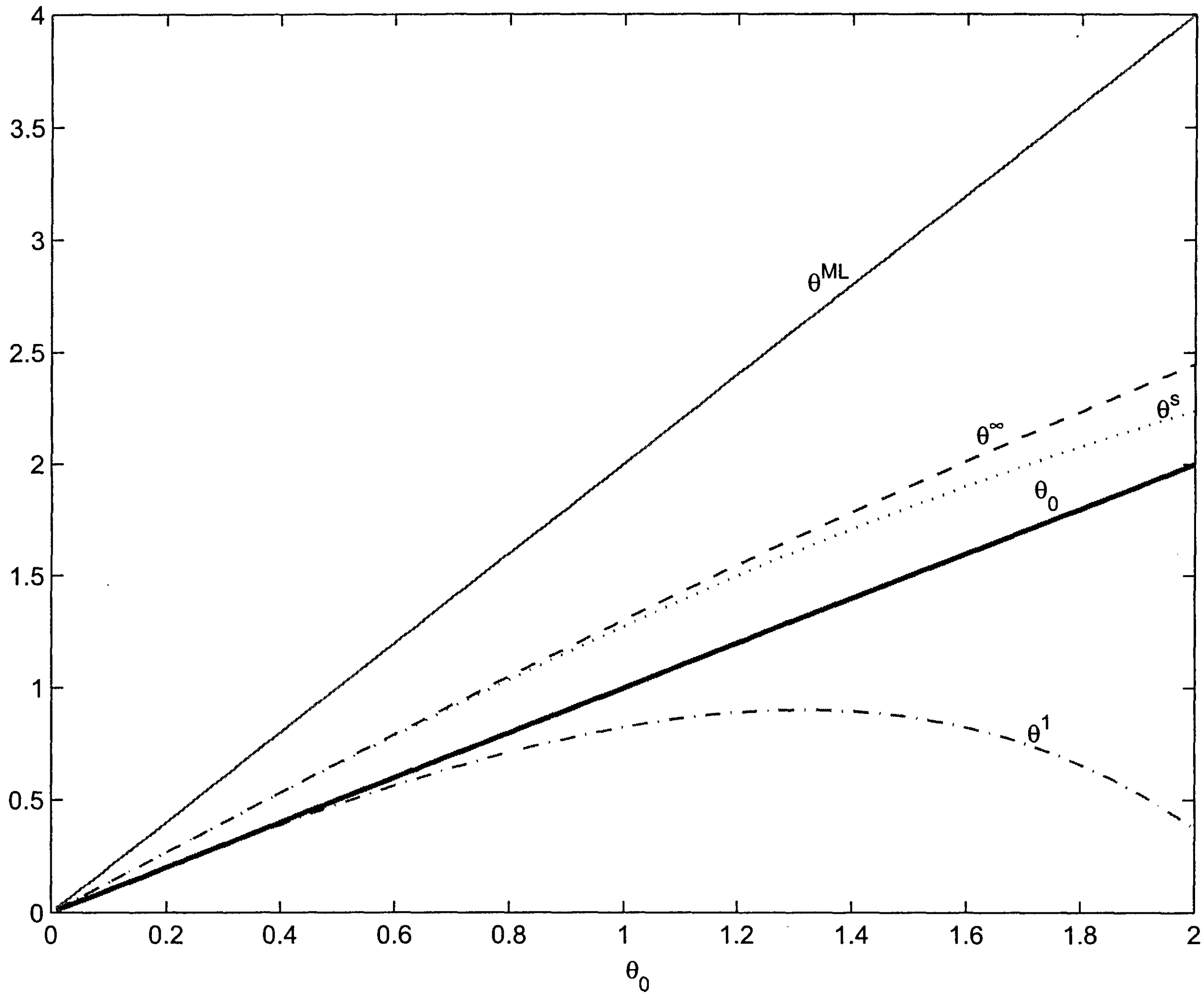

Figure 1-1: Asymptotic probability limit of estimators: Andersen (1973) two-period logit model. θ_0 is the limit of the conditional logit estimator (true parameter value); θ^{ML} is the limit of the fixed effects maximum likelihood logit estimator; θ^1 is the limit of the one-step bias-corrected estimator; θ^∞ is the limit of the iterated bias-corrected estimator; and θ^s is the limit of the score (estimating equation)-corrected estimator.

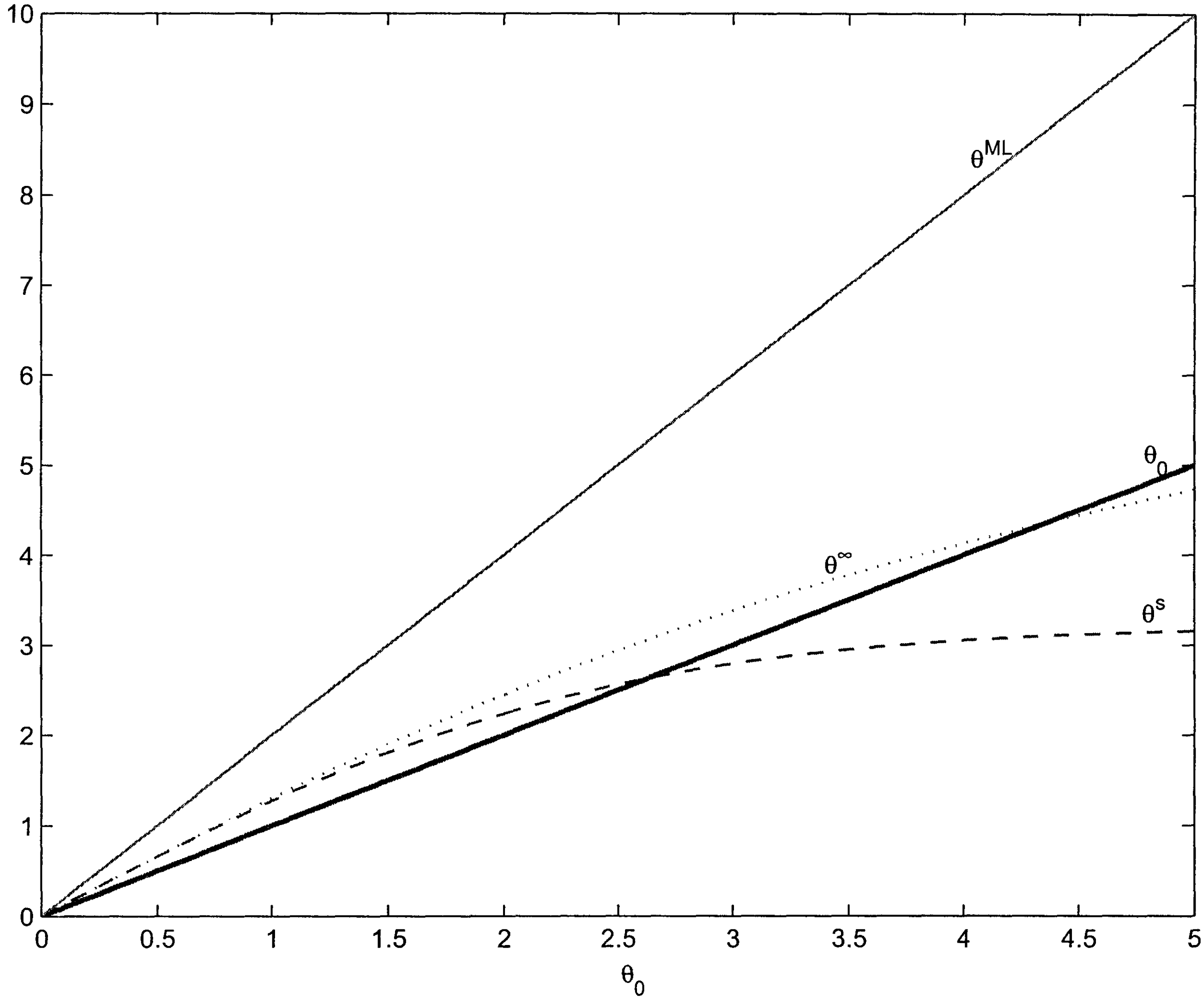

Figure 1-2: Asymptotic probability limit of estimators: Andersen (1973) two-period logit model. θ_0 is the limit of the conditional logit estimator (true parameter value); θ^{ML} is the limit of the fixed effects maximum likelihood estimator; θ^{∞} is the limit of the iterated bias-corrected estimator; and θ^{s} is the limit of the score (estimating equation)-corrected estimator.

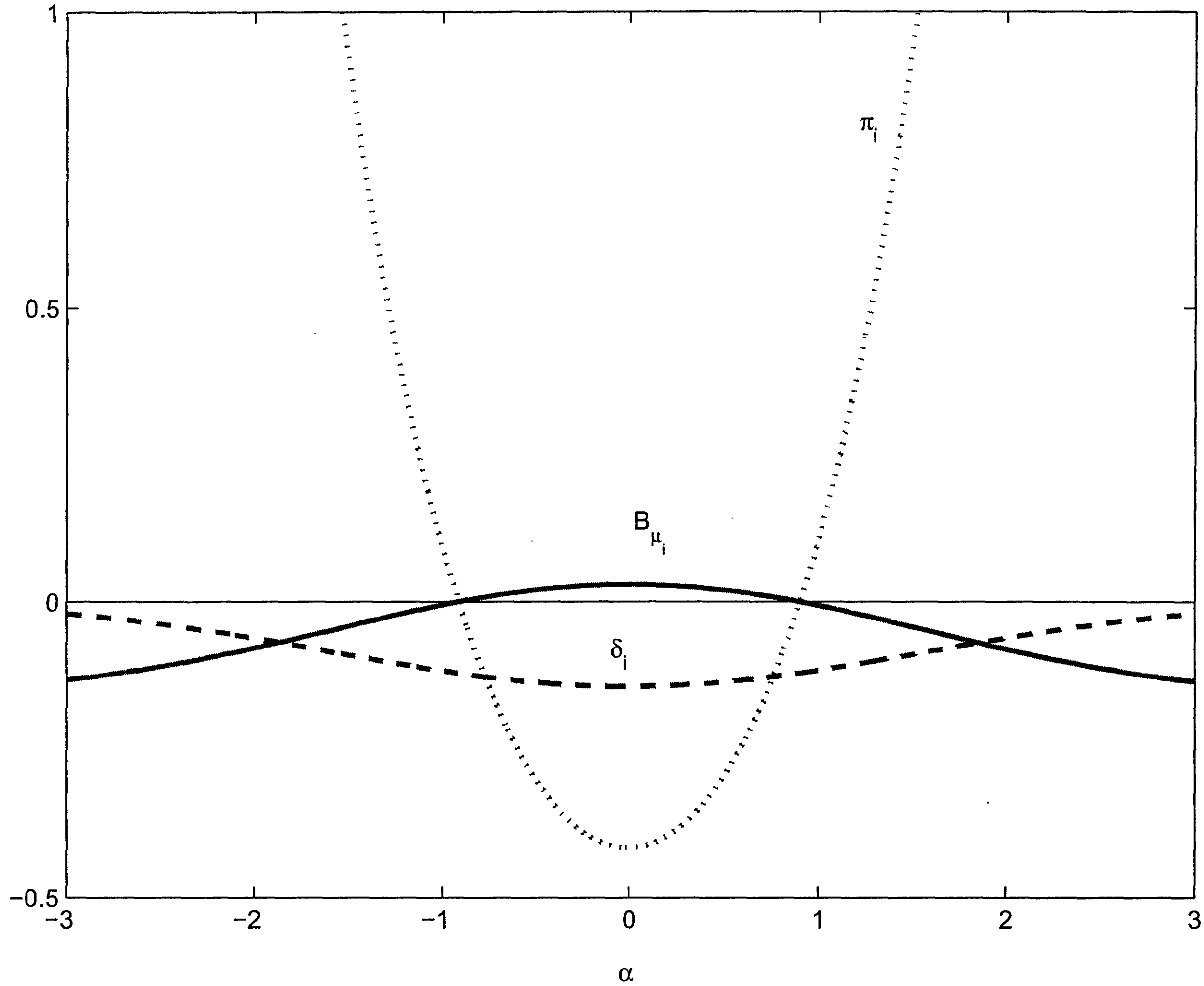

Figure 1-3: Components of the bias of the fixed effects estimator of the marginal effect: $\mathcal{B}_\mu = E_\alpha[\mathcal{B}_{\mu_i}] = E_\alpha[\delta_i \pi_i]$. Individual effects and regressor generated from independent standard normal distributions.

Table 1: First Order Bias, T = 8
(in percent of the true parameter value)

Individual Effects	Regressor				
	Nerlove	Normal	$\chi^2(1)$	$\chi^2(2)$	Bi(10,.9)
	A - Index Coefficients				
Normal	14.59	15.62	16.56	15.93	15.62
$\chi^2(1)$	12.57	14.77	13.82	14.00	14.85
$\chi^2(2)$	13.17	14.95	13.04	14.55	15.16
Bi(10,.9)	15.46	15.53	16.36	16.47	15.44
	B -Marginal Effects				
Normal	-0.27	-0.07	1.89	0.95	-0.03
$\chi^2(1)$	-0.21	-0.06	-0.18	-0.24	0.34
$\chi^2(2)$	-0.28	-0.05	-0.54	-0.10	0.31
Bi(10,.9)	-0.27	-0.08	1.56	1.59	-0.26

Notes: Bias formulae evaluated numerically using 10,000 replications.

Table 2: Estimators of θ ($\theta_0 = 1$), $\varepsilon \sim N(0,1)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 4				
PROBIT	1.41	1.40	0.393	0.25	0.36	0.82	0.410
JK-PROBIT	0.75	0.75	0.277	0.11	0.19	1.08	0.265
BC1-PROBIT	1.11	1.10	0.304	0.04	0.11	1.03	0.215
BC2-PROBIT	1.20	1.19	0.333	0.09	0.16	0.95	0.253
BC3-PROBIT	1.06	1.06	0.275	0.02	0.06	1.13	0.195
LOGIT	1.30	1.29	0.374	0.17	0.26	0.83	0.331
JK-LOGIT	0.71	0.70	0.239	0.18	0.28	1.15	0.307
BC1-LOGIT	0.97	0.96	0.266	0.04	0.08	1.08	0.180
BC2-LOGIT	0.95	0.94	0.263	0.03	0.09	1.09	0.178
BC3-LOGIT	0.94	0.94	0.253	0.04	0.07	1.13	0.173
CLOGIT	0.95	0.94	0.263	0.04	0.08	1.09	0.177
			T = 8				
PROBIT	1.18	1.18	0.151	0.28	0.37	0.90	0.180
JK-PROBIT	0.95	0.96	0.118	0.05	0.11	1.09	0.085
BC1-PROBIT	1.05	1.05	0.134	0.05	0.11	0.98	0.099
BC2-PROBIT	1.05	1.05	0.132	0.05	0.10	1.00	0.097
BC3-PROBIT	1.02	1.02	0.124	0.03	0.07	1.05	0.085
LOGIT	1.12	1.12	0.148	0.14	0.23	0.91	0.129
JK-LOGIT	0.91	0.91	0.114	0.12	0.20	1.09	0.105
BC1-LOGIT	0.97	0.97	0.127	0.06	0.13	1.00	0.087
BC2-LOGIT	0.96	0.95	0.122	0.07	0.13	1.03	0.087
BC3-LOGIT	0.95	0.95	0.121	0.07	0.13	1.04	0.089
CLOGIT	0.96	0.96	0.122	0.07	0.13	1.03	0.088
			T = 12				
PROBIT	1.13	1.13	0.096	0.30	0.41	0.94	0.129
JK-PROBIT	0.98	0.98	0.080	0.05	0.10	1.06	0.055
BC1-PROBIT	1.04	1.04	0.087	0.07	0.13	0.99	0.062
BC2-PROBIT	1.03	1.03	0.085	0.06	0.11	1.01	0.058
BC3-PROBIT	1.01	1.01	0.082	0.04	0.09	1.05	0.056
LOGIT	1.09	1.09	0.097	0.15	0.25	0.96	0.095
JK-LOGIT	0.95	0.95	0.080	0.08	0.15	1.07	0.068
BC1-LOGIT	0.99	0.99	0.086	0.06	0.10	1.01	0.059
BC2-LOGIT	0.98	0.97	0.083	0.06	0.11	1.04	0.060
BC3-LOGIT	0.98	0.97	0.082	0.06	0.10	1.05	0.059
CLOGIT	0.98	0.98	0.083	0.06	0.10	1.04	0.059

Notes: 1,000 replications. JK denotes Hahn and Newey (2004) Jackknife bias-corrected estimator; BC1 denotes Hahn and Newey (2004) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Newey (2004) bias-corrected estimator based on general estimating equations; BC3 denotes the bias-corrected estimator proposed in the paper; CLOGIT denotes conditional logit estimator.

Table 3: Estimators of μ (true value = 1), ε ~ N(0,1)

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 4				
PROBIT	0.99	0.99	0.242	0.10	0.16	0.82	0.163
JK-PROBIT	1.02	1.02	0.285	0.12	0.19	0.75	0.182
BC1-PROBIT	1.00	1.00	0.261	0.12	0.18	0.79	0.176
BC2-PROBIT	1.04	1.04	0.255	0.12	0.19	0.80	0.176
BC3-PROBIT	0.94	0.94	0.226	0.08	0.13	0.91	0.158
LOGIT	1.00	0.99	0.246	0.10	0.16	0.82	0.164
JK-LOGIT	1.01	1.01	0.279	0.13	0.19	0.76	0.189
BC1-LOGIT	0.98	0.97	0.257	0.11	0.17	0.81	0.178
BC2-LOGIT	0.94	0.94	0.236	0.09	0.15	0.88	0.164
BC3-LOGIT	0.93	0.93	0.230	0.08	0.14	0.90	0.166
BC-CLOGIT	0.94	0.94	0.237	0.09	0.15	0.87	0.168
LPM	0.98	0.98	0.233	0.09	0.15	0.84	0.156
LPM-FS	1.00	1.00	0.242	0.10	0.16	0.87	0.163
			T = 8				
PROBIT	0.99	0.99	0.104	0.08	0.14	0.82	0.070
JK-PROBIT	1.00	1.00	0.107	0.07	0.14	0.84	0.071
BC1-PROBIT	1.01	1.01	0.110	0.09	0.15	0.80	0.073
BC2-PROBIT	1.00	1.00	0.105	0.07	0.13	0.83	0.070
BC3-PROBIT	0.97	0.97	0.103	0.08	0.13	0.86	0.071
LOGIT	0.99	0.99	0.106	0.08	0.13	0.83	0.071
JK-LOGIT	0.99	1.00	0.108	0.07	0.13	0.84	0.072
BC1-LOGIT	1.00	1.00	0.112	0.08	0.15	0.81	0.073
BC2-LOGIT	0.98	0.98	0.106	0.07	0.13	0.85	0.071
BC3-LOGIT	0.98	0.98	0.106	0.08	0.13	0.85	0.071
BC-CLOGIT	0.98	0.98	0.106	0.08	0.13	0.85	0.071
LPM	0.98	0.98	0.104	0.07	0.14	0.84	0.071
LPM-FS	1.00	1.00	0.109	0.07	0.13	0.87	0.075
			T = 12				
PROBIT	0.99	0.99	0.062	0.05	0.11	0.75	0.043
JK-PROBIT	1.00	1.00	0.064	0.05	0.11	0.76	0.042
BC1-PROBIT	1.00	1.00	0.065	0.06	0.11	0.74	0.042
BC2-PROBIT	0.99	0.99	0.062	0.05	0.10	0.76	0.042
BC3-PROBIT	0.98	0.98	0.062	0.05	0.11	0.77	0.043
LOGIT	0.99	0.99	0.063	0.05	0.10	0.77	0.043
JK-LOGIT	1.00	1.00	0.065	0.05	0.11	0.77	0.042
BC1-LOGIT	1.01	1.00	0.067	0.06	0.12	0.74	0.044
BC2-LOGIT	0.99	0.99	0.064	0.06	0.10	0.77	0.043
BC3-LOGIT	0.99	0.99	0.064	0.05	0.10	0.77	0.044
BC-CLOGIT	0.99	0.99	0.064	0.05	0.10	0.77	0.044
LPM	0.99	0.99	0.065	0.06	0.11	0.76	0.041
LPM-FS	1.01	1.01	0.067	0.05	0.11	0.80	0.045

Notes: 1,000 replications. JK denotes Hahn and Newey (2004) Jackknife bias-corrected estimator; BC1 denotes Hahn and Newey (2004) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Newey (2004) bias-corrected estimator based on general estimating equations; BC3 denotes the bias-corrected estimator proposed in the paper; CLOGIT denotes conditional logit estimator; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model.

Table 4: Estimators of θ_Y ($\theta_{Y,0} = .5$), $\varepsilon \sim L(0,1)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
PROBIT	-0.26	-0.26	0.161	1.00	1.00	0.92	0.763
BC1-PROBIT	0.25	0.25	0.148	0.43	0.54	0.96	0.251
BC3-PROBIT	0.48	0.48	0.140	0.05	0.10	0.99	0.101
LOGIT	-0.24	-0.24	0.154	1.00	1.00	0.92	0.740
BC1-LOGIT	0.39	0.38	0.134	0.14	0.23	1.00	0.131
HK-LOGIT		0.45					0.131
MML-LOGIT		0.39		0.11			0.127
BC3-LOGIT	0.45	0.45	0.134	0.06	0.12	1.00	0.101
			T = 12				
PROBIT	0.06	0.06	0.111	0.97	0.99	0.99	0.435
BC1-PROBIT	0.39	0.39	0.103	0.16	0.26	1.04	0.115
BC3-PROBIT	0.50	0.50	0.101	0.04	0.10	1.05	0.066
LOGIT	0.06	0.06	0.105	0.98	0.99	1.00	0.436
BC1-LOGIT	0.44	0.44	0.094	0.07	0.13	1.07	0.079
BC3-LOGIT	0.47	0.48	0.095	0.05	0.10	1.06	0.064
			T = 16				
PROBIT	0.20	0.20	0.097	0.89	0.93	0.94	0.302
BC1-PROBIT	0.44	0.44	0.091	0.11	0.18	0.98	0.080
BC3-PROBIT	0.51	0.51	0.091	0.06	0.11	0.98	0.061
LOGIT	0.19	0.19	0.091	0.93	0.96	0.95	0.312
BC1-LOGIT	0.46	0.45	0.084	0.07	0.14	1.01	0.067
HK-LOGIT		0.45					0.074
MML-LOGIT		0.48					0.067
BC3-LOGIT	0.48	0.48	0.085	0.06	0.11	1.00	0.059

Notes: 1,000 replications. BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator; HK denotes Honoré and Kyriazidou (2000) bias-corrected estimator; MML denotes Carro (2003) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in the paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model. Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 5: Estimators of θ_X ($\theta_{X,0} = 1$), $\varepsilon \sim L(0,1)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
PROBIT	1.28	1.28	0.082	0.97	0.99	0.92	0.277
BC1-PROBIT	1.19	1.19	0.077	0.78	0.86	0.89	0.187
BC3-PROBIT	1.10	1.09	0.065	0.32	0.44	0.97	0.094
LOGIT	1.22	1.22	0.082	0.84	0.90	0.93	0.222
BC1-LOGIT	1.08	1.08	0.074	0.22	0.32	0.90	0.079
HK-LOGIT		1.01					0.050
MML-LOGIT		1.01		0.06			0.039
BC3-LOGIT	1.05	1.04	0.067	0.12	0.18	0.96	0.054
			T = 12				
PROBIT	1.19	1.19	0.057	0.94	0.97	0.93	0.185
BC1-PROBIT	1.11	1.11	0.053	0.61	0.73	0.92	0.110
BC3-PROBIT	1.08	1.08	0.050	0.35	0.48	0.95	0.076
LOGIT	1.13	1.13	0.057	0.67	0.77	0.94	0.126
BC1-LOGIT	1.03	1.03	0.051	0.10	0.16	0.94	0.039
BC3-LOGIT	1.02	1.02	0.050	0.08	0.13	0.96	0.035
			T = 16				
PROBIT	1.15	1.15	0.045	0.95	0.98	0.97	0.148
BC1-PROBIT	1.09	1.09	0.042	0.59	0.71	0.96	0.090
BC3-PROBIT	1.07	1.07	0.041	0.41	0.54	0.99	0.070
LOGIT	1.09	1.09	0.044	0.53	0.66	0.98	0.088
BC1-LOGIT	1.02	1.01	0.041	0.07	0.13	0.98	0.029
HK-LOGIT		1.01					0.023
MML-LOGIT		1.01					0.029
BC3-LOGIT	1.01	1.01	0.040	0.06	0.10	0.99	0.027

Notes: 1,000 replications. BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator; HK denotes Honoré and Kyriazidou (2000) bias-corrected estimator; MML denotes Carro (2003) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in the paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model. Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 6: Estimators of μ_Y (true value = 1), $\varepsilon \sim L(0,1)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
			T = 8				
PROBIT	-0.38	-0.39	0.235	1.00	1.00	0.92	1.386
BC1-PROBIT	0.43	0.42	0.254	0.68	0.77	0.87	0.577
BC3-PROBIT	0.86	0.85	0.255	0.14	0.22	0.88	0.211
LOGIT	-0.37	-0.37	0.236	1.00	1.00	0.93	1.374
BC1-LOGIT	0.72	0.70	0.257	0.28	0.38	0.88	0.304
BC3-LOGIT	0.85	0.85	0.260	0.14	0.24	0.87	0.212
LPM	-0.40	-0.40	0.238	1.00	1.00	0.92	1.399
BC-LPM	0.77	0.77	0.267	0.23	0.32	0.85	0.261
LPM-FS	-0.45	-0.46	0.257	1.00	1.00	0.92	1.459
BC-LPM-FS	0.84	0.83	0.289	0.16	0.24	0.85	0.236
			T = 12				
PROBIT	0.11	0.11	0.183	1.00	1.00	0.99	0.894
BC1-PROBIT	0.70	0.70	0.189	0.38	0.49	0.96	0.297
BC3-PROBIT	0.93	0.93	0.190	0.08	0.13	0.96	0.136
LOGIT	0.11	0.11	0.182	1.00	1.00	1.00	0.889
BC1-LOGIT	0.86	0.86	0.187	0.13	0.19	0.98	0.169
BC3-LOGIT	0.93	0.94	0.191	0.08	0.13	0.97	0.133
LPM	0.09	0.09	0.188	1.00	1.00	1.00	0.908
BC-LPM	0.91	0.91	0.202	0.09	0.15	0.94	0.151
LPM-FS	0.09	0.09	0.192	1.00	1.00	1.00	0.909
BC-LPM-FS	0.93	0.93	0.207	0.08	0.14	0.94	0.143
			T = 16				
PROBIT	0.34	0.34	0.167	0.98	0.99	0.94	0.660
BC1-PROBIT	0.81	0.80	0.170	0.25	0.35	0.93	0.201
BC3-PROBIT	0.95	0.94	0.171	0.09	0.15	0.92	0.125
LOGIT	0.35	0.34	0.166	0.97	0.99	0.95	0.658
BC1-LOGIT	0.91	0.90	0.168	0.10	0.18	0.95	0.139
BC3-LOGIT	0.95	0.94	0.171	0.08	0.14	0.93	0.128
LPM	0.34	0.33	0.170	0.97	0.99	0.96	0.666
BC-LPM	0.95	0.95	0.180	0.09	0.15	0.91	0.129
LPM-FS	0.34	0.34	0.172	0.97	0.99	0.96	0.665
BC-LPM-FS	0.96	0.95	0.181	0.08	0.15	0.92	0.128

Notes: 1,000 replications. BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator; HK denotes Honoré and Kyriazidou (2000) bias-corrected estimator; MML denotes Carro (2003) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in the paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model; BC-LPM denotes Nickell (1981) bias-corrected adjusted linear probability model; BC-LPM-FS denotes Nickell (1981) bias-corrected linear probability model. Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 7: Estimators of μ_X (true value = 1), $\varepsilon \sim L(0,1)$

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE
				T = 8			
PROBIT	0.97	0.97	0.041	0.11	0.19	0.93	0.034
BC1-PROBIT	1.02	1.02	0.044	0.11	0.17	0.85	0.032
BC3-PROBIT	0.98	0.97	0.041	0.11	0.19	0.88	0.033
LOGIT	0.98	0.98	0.042	0.07	0.14	0.93	0.031
BC1-LOGIT	1.01	1.01	0.046	0.09	0.15	0.83	0.031
BC3-LOGIT	0.99	0.99	0.043	0.08	0.15	0.86	0.030
LPM	0.96	0.96	0.040	0.19	0.29	0.91	0.043
BC-LPM	0.98	0.97	0.040	0.09	0.16	0.92	0.033
LPM-FS	0.97	0.97	0.040	0.10	0.17	0.97	0.034
BC-LPM-FS	0.99	0.99	0.040	0.05	0.10	0.98	0.028
				T = 12			
PROBIT	0.99	0.99	0.031	0.07	0.14	0.92	0.021
BC1-PROBIT	1.01	1.01	0.032	0.08	0.13	0.89	0.022
BC3-PROBIT	0.99	0.99	0.031	0.07	0.14	0.90	0.020
LOGIT	0.99	0.99	0.031	0.06	0.11	0.93	0.020
BC1-LOGIT	1.01	1.01	0.032	0.06	0.13	0.89	0.022
BC3-LOGIT	1.00	1.00	0.032	0.06	0.13	0.90	0.021
LPM	0.99	0.99	0.031	0.08	0.14	0.93	0.023
BC-LPM	0.99	0.99	0.031	0.06	0.11	0.93	0.021
LPM-FS	0.99	0.99	0.031	0.07	0.12	0.94	0.022
BC-LPM-FS	1.00	1.00	0.031	0.06	0.11	0.95	0.020
				T = 16			
PROBIT	0.99	0.99	0.025	0.05	0.10	0.98	0.017
BC1-PROBIT	1.01	1.01	0.025	0.06	0.11	0.95	0.017
BC3-PROBIT	1.00	1.00	0.025	0.05	0.10	0.96	0.017
LOGIT	1.00	1.00	0.025	0.04	0.09	0.98	0.017
BC1-LOGIT	1.00	1.00	0.026	0.05	0.11	0.95	0.017
BC3-LOGIT	1.00	1.00	0.025	0.05	0.10	0.96	0.017
LPM	0.99	0.99	0.026	0.06	0.12	0.97	0.019
BC-LPM	1.00	1.00	0.026	0.05	0.11	0.98	0.018
LPM-FS	0.99	0.99	0.026	0.05	0.11	0.98	0.019
BC-LPM-FS	1.00	1.00	0.026	0.05	0.10	0.98	0.018

Notes: 1,000 replications. BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator; HK denotes Honoré and Kyriazidou (2000) bias-corrected estimator; MML denotes Carro (2003) Modified Maximum Likelihood estimator; BC3 denotes the bias-corrected estimator proposed in the paper; LPM denotes adjusted linear probability model (see text); LPM-FS denotes linear probability model; BC-LPM denotes Nickell (1981) bias-corrected adjusted linear probability model; BC-LPM-FS denotes Nickell (1981) bias-corrected linear probability model. Honoré-Kyriazidou estimator is based on bandwidth parameter = 8.

Table 8: Descriptive Statistics, Married Women (n = 1461, T = 9)

	Full Sample	Always Participate		Never Participate		Movers		
	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Within (%)
Participation	0.45	1	0	0	0	0.57	0.49	72.51
Age (in 1985)	9.22	37.98	9.04	42.98	10.09	35.57	8.71	9.09
Black	0.40	0.24	0.43	0.25	0.43	0.16	0.37	0
Years of schooling	3.79	12.49	3.88	12.09	3.20	12.05	3.77	0
Kids 0-2	0.47	0.18	0.41	0.21	0.47	0.28	0.51	62.96
Kids 3-5	0.51	0.23	0.46	0.23	0.48	0.36	0.56	65.52
Kids 6-17	1.10	1.00	1.06	0.99	1.19	1.11	1.11	33.67
Husband income (1995 $1000)	40.01	38.33	25.15	53.27	74.62	44.32	42.69	20.99
No. Observations	13149	6084		1089		5976		

Source: PSID 1980-1988.

Table 9: Female Labor Force Participation (n = 1461, T = 9), Static Model

Estimator	PROBIT			LOGIT				LPM	
	FE [1]	JK [2]	BC3 [3]	FE [4]	JK [5]	BC3 [6]	C [7]	FE [8]	FE-FS [9]
	A - Index Coefficients								
Kids 0-2	-0.71	-0.61	-0.63	-0.68	-0.59	-0.60	-0.60		
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)		
Kids 3-5	-0.42	-0.37	-0.37	-0.40	-0.35	-0.35	-0.35		
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		
Kids 6-17	-0.13	-0.10	-0.11	-0.13	-0.11	-0.11	-0.11		
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		
Log(Husband income)	-0.25	-0.22	-0.22	-0.24	-0.21	-0.21	-0.21		
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		
	B - Marginal Effects (%)								
Kids 0-2	-9.22	-9.38	-9.07	-9.35	-9.35	-9.20	-9.20	-9.46	-11.19
	(0.70)	(0.71)	(0.70)	(0.72)	(0.72)	(0.72)	(0.72)	(0.75)	(0.88)
Kids 3-5	-5.45	-5.60	-5.36	-5.53	-5.59	-5.45	-5.45	-5.54	-6.09
	(0.66)	(0.66)	(0.66)	(0.67)	(0.67)	(0.67)	(0.67)	(0.70)	(0.81)
Kids 6-17	-1.68	-1.59	-1.66	-1.78	-1.72	-1.76	-1.76	-1.78	-1.25
	(0.53)	(0.53)	(0.53)	(0.54)	(0.54)	(0.54)	(0.54)	(0.58)	(0.56)
Log(Husband income)	-3.25	-3.31	-3.20	-3.26	-3.29	-3.22	-3.22	-3.17	-3.61
	(0.70)	(0.69)	(0.70)	(0.71)	(0.71)	(0.71)	(0.71)	(0.72)	(0.71)

Notes: All the specifications include time dummies and a quadratic function of age. FE denotes uncorrected fixed effects estimator; JK denotes Hahn and Newey (2004) Jackknife bias-corrected estimator; BC3 denotes the bias-corrected estimator proposed in the paper; C denotes conditional logit estimator; LPM-FE denotes adjusted linear probability model (see text); LPM-FE-FS denotes linear probability model. Logit estimates and standard errors of index coefficients are normalized to have the same scale as probit.

Source: PSID 1980-1988.

Table 10: Female Labor Force Participation (n = 1461, T = 10), Dynamic Model

	PROBIT			LOGIT		LPM			
Estimator	FE [1]	BC3 [2]	MML [3]	FE [4]	BC3 [5]	FE [6]	BC [7]	FE-FS [8]	BC-FS [9]
A - Index Coefficients									
$\text{Participation}_{t-1}$	0.76	1.04	1.08	0.69	0.96				
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
Kids 0-2	-0.55	-0.44	-0.40	-0.53	-0.43				
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)				
Kids 3-5	-0.29	-0.21	-0.18	-0.27	-0.20				
	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)				
Kids 6-17	-0.07	-0.05	-0.04	-0.07	-0.05				
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)				
Log(Husband income)	-0.25	-0.22	-0.21	-0.24	-0.21				
	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)				
B - Marginal Effects (%)									
$\text{Participation}_{t-1}$	10.69	16.86		10.47	16.80	11.42	17.80	25.58	39.06
	(0.62)	(0.65)		(0.61)	(0.64)	(0.63)	(0.65)	(1.30)	(1.36)
Kids 0-2	-6.76	-5.95		-6.81	-5.94	-6.87	-5.42	-8.02	-6.35
	(0.75)	(0.72)		(0.75)	(0.72)	(0.75)	(0.73)	(0.86)	(0.84)
Kids 3-5	-3.55	-2.79		-3.53	-2.72	-3.44	-2.27	-3.57	-2.24
	(0.69)	(0.67)		(0.69)	(0.67)	(0.69)	(0.66)	(0.79)	(0.77)
Kids 6-17	-0.91	-0.67		-0.95	-0.69	-0.93	-0.45	-0.49	-0.10
	(0.55)	(0.52)		(0.54)	(0.52)	(0.55)	(0.52)	(0.53)	(0.51)
Log(Husband income)	-3.08	-2.90		-3.07	-2.87	-2.98	-2.88	-3.29	-3.12
	(0.70)	(0.68)		(0.71)	(0.68)	(0.70)	(0.69)	(0.69)	(0.68)

Notes: All the specifications include time dummies and a quadratic function of age. FE denotes uncorrected fixed effects estimator; BC3 denotes the bias-corrected estimator proposed in the paper; MML denotes Carro (2003) Modified Maximum Likelihood estimator; LPM - FE denotes adjusted linear probability model (see text); LPM-BC denotes Nickell (1981) bias-corrected adjusted linear probability model; LPM-FE-FS denotes linear probability model; LPM-BC-FS denotes Nickell (1981) bias-corrected linear probability model. Column [3] taken from Carro (2003). The specification in Carro (2003) includes also a lag of Kids 0-2 with estimated coefficient -0.039 (0.054). Logit estimates and standard errors of index coefficients are normalized to have the same scale as probit. First period used as initial condition.
Source: PSID 1979-1988.

Chapter 2

Bias Correction in Panel Data Models with Individual-Specific Parameters

2.1 Introduction

Random coefficients panel models are attractive because they allow for heterogeneity in the individual response to the regressors. However, they pose important technical challenges in the estimation of average effects if the individual heterogeneity is left unrestricted. In particular, if some of the regressors are endogenous and different coefficients are estimated for each individual (fixed-effects approach), then averages of these individual IV estimates are biased in short panels due to the finite-sample bias of IV estimators. A way to overcome this problem is to neglect the individual heterogeneity and estimate the same coefficients for all the individuals. However, in the context of cross-section models, Imbens and Angrist (1994) and Angrist, Graddy and Imbens (1999) show that the estimands of these fixed coefficients IV estimators are weighted-averages of the underlying heterogenous individual effects. The implicit weights in these averages are typically correlated with the individual effects, and therefore these estimators do not converge to population average effects.[1]

In this chapter I introduce a new class of bias-corrected fixed effects estimators for panel models where the response to the regressors can be individual-specific in an unrestricted fashion. Thus, instead of imposing the same coefficients for all the individuals, I treat the sample real-

[1]Angrist (2004) finds homogeneity conditions for the estimands of fixed coefficients IV estimators to be the average effects for models with binary endogenous regressors.

ization of the individual-specific coefficients as parameters (fixed-effects) to be estimated. For linear models, the new estimators differ from the standard fixed effects estimators in which, not only the constant term, but also the slopes can be different for each individual. Moreover, unlike the classical random coefficients models, no restriction is imposed in the relationship between the regressors and the random coefficients. This allows me to incorporate, for instance, Roy (1951) type selection where the regressors are decision variables with levels determined at least in part by their returns. Treating the random coefficients as fixed effects also overcomes the identification problems for these models in the presence of endogenous regressors, see Kelejian (1974).

The models proposed here are semiparametric in the sense that they are based on moment conditions. These conditions can be nonlinear functions in parameters and variables, accommodating both linear and nonlinear models, and allowing for the presence of endogenous regressors. In addition to identifying the model parameters under mild assumptions about the nature of the stochastic component, these moment conditions can be used for estimation via GMM procedures. The resulting fixed effects GMM estimates, however, can be severely biased in short panels due to the incidental parameters problem, which in this case is a consequence of the finite-sample bias of GMM estimators (see, for e.g., Nagar, 1959; Buse, 1992; and Newey and Smith, 2004). Analytical corrections are then developed to reduce this bias.

These bias corrections are derived from large-T expansions of the finite-sample bias of GMM estimators. They reduce the order of this bias from $O(T^{-1})$ to $O(T^{-2})$ by removing an estimate of the leading term of the expansion from the fixed-effects estimator. As a result, the asymptotic distribution of the corrected estimators is normal and centered at the true parameter value under asymptotic sequences where $n = o\left(T^3\right)$. These corrections therefore aim to work in econometric applications that use long panels, for e.g., PSID or Penn World Table, where the most important part of the bias is captured by the first term of the expansion. Other recent studies that use a similar approach for the analysis of fixed effects estimators in panel data include Phillips and Moon (1999), Rathouz and Liang (1999), Alvarez and Arellano (2001), Hahn and Kuersteiner (2002), Lancaster (2002), Woutersen (2002), Hahn and Kuersteiner (2003), Li, Lindsay and Waterman (2003), and Hahn and Newey (2004)

A distinctive feature of the corrections proposed here is that they can be used in situations where the model is overidentified, that is when the number of moment restrictions is greater than the dimension of the parameter vector. This situation is very common in economic applications where the moment conditions come from rational expectation models, or more generally when there are multiple sources of exogenous variation to instrument for the endogenous variables. For example, Hansen and Singleton (1982), Holtz-Eakin, Newey, and Rosen (1988), Abowd and Card (1989), and Angrist and Krueger (1991), all use quite large numbers of moment conditions in their empirical work. This feature, however, complicates the analysis by introducing an initial stage for estimating optimal weighting matrices to combine these moment conditions. Thus, overidentification precludes the use of the existing bias correction methods.

To characterize the bias of fixed effects GMM estimators, I use higher-order asymptotics. The results obtained here extend the bias formulas of Newey and Smith (2004) for cross sectional GMM estimators to panel data GMM estimators with fixed effects. Analytical bias-correction methods are then described for model parameters and smooth functions of the individual-specific parameters, such as averages or other moments of the distribution of the random coefficients. These corrections are computationally more attractive than other alternatives, such as Bootstrap or Jackknife, especially when the moment conditions are nonlinear in parameters. Different methods are considered depending on whether the correction is made on the estimator or on the estimating equation, and on whether the correction is one-step or iterated.

The finite sample performance of these corrections is evaluated using a Monte Carlo example. Here, I consider a linear IV model with both common and individual-specific coefficients. I find that estimators that do not account for heterogeneity by imposing constant coefficients can have important biases for the common parameter and the mean of the individual effects. I also find that analytical bias corrections are effective in removing the bias of estimators of the standard deviation of the individual effects.

Finally, the estimators introduced in the chapter are illustrated in an empirical application. I use 14 waves of the National Longitudinal Survey (NLS) to estimate earnings equations for young men allowing the effect of the union status to be different for each individual. I consider both OLS and IV methods. The latter accounts for the possibility of endogeneity in the union membership decision. The results suggest that there is important heterogeneity across individuals in the effect of union status on earnings. Moreover, estimators that impose a constant coefficient for the union status overstate the average effect.

The rest of the chapter is organized as follows. Section 2.2 illustrates the type of models considered and discusses the nature of the bias with a simple example. Section 2.3 introduces the general model and GMM estimators. Section 2.4 derives the asymptotic properties of the estimators. The different bias correction methods and their asymptotic properties are given in Section 2.5. Section 2.6 compares the finite sample performance of these and other methods through a Monte Carlo experiment. Section 2.7 describes the empirical application and Section 2.8 concludes. Proofs and other technical details are given in the Appendices.

2.2 Example: A Linear IV Panel Model with Individual-Specific Coefficients

2.2.1 The Model

A simple example of the type of models considered is the following

$$y_{it} = \alpha_{0i} + \alpha_{1i} x_{it} + \epsilon_{it}, \tag{2.2.1}$$

where y_{it} is a response variable; x_{it} is a regressor; ϵ_{it} is an error term; and i and t usually index individual and time period, respectively.[2] This is a linear random coefficients model where the effect of the regressor is heterogenous across individuals, but no restriction is imposed on the distribution of the individual effects vector $\alpha_i \equiv (\alpha_{0i}, \alpha_{1i})'$.[3] The regressor can be correlated with the individual effects and the error term, and a valid instrument z_{it} is available for x_{it}, that is $E[z_{it}\epsilon_{it}|\alpha_i] = 0$ and $E[z_{it}x_{it}|\alpha_i] \neq 0$. All the random variables are i.i.d. across time periods and independent across individuals.

2.2.2 Example

An important case that is encompassed by this simple set-up is the panel version of the treatment-effect model, see, for e.g., Wooldridge (2002, Chapter 10.2.3) and Angrist and Hahn (2004). Here, the objective is to evaluate the effect of a treatment (D) on an outcome variable (Y) for some population of interest. The average causal effect for each level of treatment is defined as the difference between the potential outcome that the individual would get with and without the treatment, $Y_d - Y_0$, averaged across the individuals of the population of interest. If individuals can choose the level of treatment, then outcome and level of treatment are generally correlated, and an instrumental variable Z is needed to identify the causal effect. This instrument is typically a random offer of treatment that generates potential treatments (D_z) indexed by the possible values of the instrument. For the binary treatment-instrument case, the panel version

[2]More generally, i denotes a group index and t indexes the observations within the group. Examples of groups include individuals, states, households, schools, or twins.

[3]Random coefficients models usually assume that α_i is uncorrelated with the regressor x_{it}. See Hsiao and Pesaran (2004) for a survey of these models.

of this model is

$$Y_{it} = Y_{0it} + (Y_{1it} - Y_{0it})D_{it}, \tag{2.2.2}$$

$$D_{it} = D_{0it} + (D_{0it} - D_{1it})Z_{it}. \tag{2.2.3}$$

Potential outcomes and treatments can be represented as the sum of permanent individual components and transitory individual-time specific shocks, that is $Y_{jit} = Y_{ji} + \epsilon_{jit}$ and $D_{jit} = D_{ji} + v_{jit}$ for $j \in \{0, 1\}$, yielding

$$Y_{it} = \alpha_{0i} + \alpha_{1i} D_{it} + \epsilon_{it}, \tag{2.2.4}$$

$$D_{it} = \pi_{0i} + \pi_{1i} Z_{it} + v_{it}, \tag{2.2.5}$$

where $\alpha_{0i} = Y_{0i}$, $\alpha_{1i} = Y_{1i} - Y_{0i}$, $\epsilon_{it} = (1 - D_{it})\epsilon_{0it} + D_{it}\epsilon_{1it}$, $\pi_{0i} = D_{0i}$, $\pi_{1i} = D_{1i} - D_{0i}$, and $v_{it} = (1 - Z_{it})v_{0it} + Z_{it}v_{1it}$. In this model, under independence of the instrument with the transitory shocks to potential outcomes, and monotonicity on the effect of the instrument on the treatment, that is either $D_{1it} - D_{0it} \geq 0$ or $D_{1it} - D_{0it} \leq 0$ with probability one, local average treatment effects are identified for each individual as the number of time periods grow, see Imbens and Angrist (1994).

2.2.3 The Problem

Returning to the linear IV example, suppose that we are ultimately interested in the average response to the regressor, i.e. $E[\alpha_{1i}]$, and we run fixed effects OLS and IV regressions without accounting for heterogeneity on the slope of x_{it}, i.e. we estimate

$$y_{it} = \alpha_{i0} + \alpha_1 x_{it} + u_{it}, \tag{2.2.6}$$

where $u_{it} = x_{it}(\alpha_{1i} - \alpha_1) + \epsilon_{it}$ if the true model is (2.2.1). In this case, OLS and IV estimate weighted averages of the individual effects in the population, see for example Yitzhaki (1996) and Angrist and Krueger (1999) for OLS, and Angrist, Graddy and Imbens (2000) for IV. OLS puts more weight on individuals with higher variances of the regressor because they give more information about the slope; whereas IV weights individuals in proportion to the variance of the first stage fitted values because these variances reflect the amount of information that the individuals convey about the part of the slope affected by the instrument. These weighted averages, however, do not necessarily estimate the average effect because the weights can be correlated with the individual effects.

To see how these implicit OLS and IV weighting schemes affect the estimand of fixed-coefficients estimators, assume for simplicity that the reduced form of x_{it} is linear, that is

$$x_{it} = \pi_{0i} + \pi_{1i} z_{it} + v_{it}, \tag{2.2.7}$$

$\pi_i \equiv (\pi_{0i}, \pi_{1i})$, (ϵ_{it}, v_{it}) is normal conditional on $(z_{it}, \alpha_i, \pi_i)$, z_{it} is independent of (α_i, π_i), and (α_i, π_i) is normal. Then, it is straightforward to find that OLS and IV estimate the following population moments[4]

$$\alpha_1^{OLS} = E[\alpha_{1i}] + \frac{Cov[\epsilon_{it}, v_{it}] + 2E[\pi_{1i}]V[z_{it}]Cov[\alpha_{1i}, \pi_{1i}]}{Var[x_{it}]}, \tag{2.2.9}$$

$$\alpha_1^{IV} = E[\alpha_{1i}] + \frac{Cov[\alpha_{1i}, \pi_{1i}]}{E[\pi_{1i}]}. \tag{2.2.10}$$

These expressions show that the OLS estimand differs from the average effect in presence of endogeneity, i.e. correlation between the individual-time specific error terms, or whenever the individual coefficients are correlated; while the IV estimand differs from the average effect in the latter case.[5]

The extent of these correlations can be explained in the panel treatment-effects model. In this case, there exists correlation between the error terms in presence of endogeneity bias. Correlation between the individual effects arises when individuals who experience a higher permanent effect of the treatment are relatively more prone to accept the offer of treatment. In other words, constant effects estimators do not estimate average effect in the presence of Roy (1951) type selection, that is if individuals choose the level of the regressors knowing the effects of these regressors (up to random deviations). Angrist (2004) finds special circumstances where the estimand of the IV estimator coincides with the average effects in models with binary endogenous regressors. For models with continuous regressors, Angrist's conditional constant effects restriction corresponds to $Cov[\alpha_{1i}, \pi_{1i}] = 0$.

To overcome this problem, i.e. that estimands of constant coefficients estimators generally differ from average effects, I propose to estimate the population moments of interest based

[4]The limit of the IV estimator is obtained from a first stage equation that imposes also fixed coefficients, that is $x_{it} = \pi_{0i} + \pi_1 z_{it} + w_{it}$. When the first stage equation is different for each individual, the limit of the IV estimator is

$$\alpha_1^{IV} = E[\alpha_{1i}] + \frac{2E[\pi_{1i}]Cov[\alpha_{1i}, \pi_{1i}]}{E[\pi_{1i}]^2 + V[\pi_{1i}]}. \tag{2.2.8}$$

See Theorems 2 and 3 in Angrist and Imbens (1995) for a related discussion.

[5]This feature of the IV estimator is also pointed out in Angrist, Graddy and Imbens (1999), p. 507.

directly on estimators of the individual effects. This strategy consists of estimating each of the individual effects separately, and then obtaining the population moment of interest as the corresponding sample moment of the individual estimators. For example, the mean effect in the population is obtained using the sample average of the estimated individual effects. As a result, the estimands are the population moments of interest and consistency of the sample analogs follows from consistency of the individual estimators, as the number of time periods grows. However, since a different slope is estimated for each individual, the asymptotic distributions of these estimators can be biased due to the incidental parameters problem (Neyman and Scott, 1948).

2.2.4 Incidental Parameters Bias

To illustrate the nature of this bias, consider the estimator of the average slope constructed from individual IV fixed effects estimators.[6] In this case the incidental parameters problem is caused by the finite-sample bias of the individual IV estimators. This can be explained using some expansions. Thus, using independence across t, standard higher-order asymptotics gives (e.g. Rilstone et. al., 1996), for $T \to \infty$

$$\hat{\alpha}_{1i}^{IV} = \alpha_{1i} + \frac{1}{T}\beta_i + \frac{1}{T}\sum_{t=1}^{T}\psi_{it} + o_p(T^{-1}), \quad \psi_{it} = E[x_{it}z_{it}]^{-1}z_{it}\epsilon_{it}, \quad \beta_i = -\frac{E[z_{it}^2 x_{it}\epsilon_{it}]}{E[z_{it}x_{it}]^2}, \tag{2.2.11}$$

where β_i is the higher-order bias of $\hat{\alpha}_{1i}^{IV}$, see also Nagar (1959) and Buse (1992). In the previous expression the first order asymptotic distribution of the individual estimator is centered at the truth, as $T \to \infty$, because it is determined by the influence function, third term of the expansion, since

$$\sqrt{T}(\hat{\alpha}_{1i}^{IV} - \alpha_{1i}) = \underbrace{\frac{1}{\sqrt{T}}\beta_i}_{=o(1)} + \underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_{it}}_{=O_p(1)} + o_p(T^{-1/2}) \xrightarrow{d} N(0, \sigma_i^2 \equiv E[x_{it}z_{it}]^{-2}E[\epsilon_{it}^2 z_{it}^2]). \tag{2.2.12}$$

However, the asymptotic distribution of the sample mean of the $\hat{\alpha}_{1i}^{IV}$'s has bias in short panels, more precisely under asymptotic sequences where $T = o(n)$. Averaging reduces the order of the variance of the influence function without affecting the order of the bias. Thus, under regularity

[6] In the rest of the section I assume that the individual constant terms have been partialled out by taking differences of all the variables with respect of their individual means.

and uniform integrability conditions,[7] the expansion for the sample mean is

$$\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}_{1i}^{IV}-\alpha_{1i})=\underbrace{\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}\beta_i}_{=O(\sqrt{\frac{n}{T}})}+\underbrace{\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{t=1}^{T}\psi_{it}}_{=O_p(1)}+o_p(1). \tag{2.2.13}$$

This expression shows that the bias term becomes first order in the asymptotic distribution of the sample average of the individual estimators if $T=o(n)$.

2.2.5 Bias Corrections

One solution to center the asymptotic distribution of the average estimator is to use higher-order unbiased estimators for the α_i's (see, for e.g., Hahn and Hausman, 2002, and Newey and Smith, 2004). These estimators can be constructed by removing estimates of the higher-order biases from the $\hat{\alpha}_{1i}^{IV}$'s. For example, using the sample analog of the expression for β_i evaluated at the individual IV estimators, yields

$$\hat{\beta}_i\left(\hat{\alpha}_{1it}^{IV}\right)=-\frac{(1/T)\sum_{t=1}^{T}z_{it}^2x_{it}(y_{it}-x_{it}\hat{\alpha}_{1it}^{IV})}{\left((1/T)\sum_{t=1}^{T}z_{it}x_{it}\right)^2}, \tag{2.2.14}$$

and

$$\hat{\alpha}_{1i}^{(BC)}=\hat{\alpha}_{1i}^{IV}-\frac{1}{T}\hat{\beta}_i\left(\hat{\alpha}_{1it}^{IV}\right). \tag{2.2.15}$$

Then, the average of the bias-corrected estimators has an asymptotic distribution correctly centered at the true parameter in moderate length panels, that is if $n=o(T^3)$. To see this, note that $\sqrt{nT}\sum_{i=1}^{n}(\hat{\beta}_i-\beta_i)/nT\xrightarrow{p}0$ because β_i is a smooth function of α_{1i} and $\hat{\alpha}_{1i}^{IV}$ is consistent as $T\to\infty$. Then, for $T/n^{1/3}\to\infty$

$$\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}_{1i}^{(BC)}-\alpha_{1i})=\underbrace{\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}(\beta_i-\hat{\beta}_i)}_{=o(1)}+\underbrace{\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{t=1}^{T}\psi_{it}}_{=O_p(1)}+\underbrace{O_p\left(\sqrt{\frac{n}{T^3}}\right)}_{=o(1)}, \tag{2.2.16}$$

where this expansion is again correct under regularity and uniform integrability conditions.

Finally, note that the procedure for the bias correction of the individual estimators can be

[7] I will establish precisely these conditions in Section 2.4.

iterated by solving the equation

$$\hat{\alpha}_{1i}^{(IBC)} = \hat{\alpha}_{1i}^{IV} - \frac{1}{T}\hat{\beta}_i\left(\hat{\alpha}_{1i}^{(IBC)}\right), \tag{2.2.17}$$

which in this case has closed-form solution given by

$$\hat{\alpha}_{1i}^{(IBC)} = \frac{\sum_{t=1}^{T} z_{it}y_{it} + \frac{\sum_{t=1}^{T} z_{it}^2 x_{it} y_{it}}{\sum_{t=1}^{T} z_{it}x_{it}}}{\sum_{t=1}^{T} z_{it}x_{it} + \frac{\sum_{t=1}^{T} z_{it}^2 x_{it}^2}{\sum_{t=1}^{T} z_{it}x_{it}}}. \tag{2.2.18}$$

2.3 The Model and Estimators

The general panel model I consider is one with a finite number of moment conditions m. To describe it, let $\{z_{it}, t = 1, ..., T\}$, for $i = 1, ..., n$, be sequences of i.i.d. observations of a random variable z_i, where $\{z_i, i = 1, ..., n\}$ are independent across i (not necessarily identically distributed). Also, let θ be a $p \times 1$ vector of common parameters; α_i, $i = 1, ..., n$, be a sequence of $q \times 1$ vectors of individual effects; and $g(z; \theta, \alpha_i)$ be an $m \times 1$ vector of functions, where $m \geq p + q$. The model has true parameters θ_0 and α_{i0}, $i = 1, ..., n$, satisfying the moment conditions

$$E\left[g(z_{it}; \theta_0, \alpha_{i0})\right] = 0, \qquad i = 1, ..., n, \tag{2.3.1}$$

where $E[\cdot]$ denotes expectation taken with respect to the distribution of z_i. In this model, the ultimate quantities of interest are smooth functions of parameters and observations, which in some cases could be the parameters themselves,

$$\zeta = \lim_{n,T\to\infty} \zeta_{nT}, \quad \zeta_{nT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{T=1}^{T} f(z_{it}; \theta_0, \alpha_{i0}). \tag{2.3.2}$$

For example, in the linear IV model considered in the previous section $g(z_{it}; \theta_0, \alpha_{i0}) = z_{it}(y_{it} - \alpha_{0i0} - \alpha_{1i0})$, $\theta = \emptyset$, $p = 0$, $q = 2$ and $\zeta_{nT} = \frac{1}{n}\sum_{i=1}^{n} \alpha_{1i0}$.

2.3.1 Some Notation

Some more notation, which will be extensively used in the definition of the estimators and in the analysis of their asymptotic properties, is the following

$$
\begin{aligned}
E\left[g(z_{it};\theta,\alpha_i)g(z_{it};\theta,\alpha_i)'\right] &\equiv \Omega_i(\theta,\alpha_i), && (2.3.3)\\
E\left[\frac{\partial g(z_{it};\theta,\alpha_i)}{\partial \theta'}\right] &\equiv E[G_\theta(z_{it};\theta,\alpha_i)] = G_{\theta_i}(\theta,\alpha_i), && (2.3.4)\\
E\left[\frac{\partial g(z_{it};\theta,\alpha_i)}{\partial \alpha_i'}\right] &\equiv E[G_\alpha(z_{it};\theta,\alpha_i)] = G_{\alpha_i}(\theta,\alpha_i), && (2.3.5)
\end{aligned}
$$

where superscript $'$ denotes transpose and higher-order derivatives will be denoted by adding subscripts. Here Ω_i is the variance-covariance matrix of the moment conditions for individual i, and G_{θ_i} and G_{α_i} are individual average derivatives of these conditions. Analogously, for sample moments

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} g(z_{it};\theta,\alpha_i)g(z_{it};\theta,\alpha_i)' &\equiv \hat{\Omega}_i(\theta,\alpha_i), && (2.3.6)\\
\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(z_{it};\theta,\alpha_i)}{\partial \theta'} &\equiv \frac{1}{T}\sum_{t=1}^{T} G_\theta(z_{it};\theta,\alpha_i) = \hat{G}_{\theta_i}(\theta,\alpha_i), && (2.3.7)\\
\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(z_{it};\theta,\alpha_i)}{\partial \alpha_i'} &\equiv \frac{1}{T}\sum_{t=1}^{T} G_\alpha(z_{it};\theta,\alpha_i) = \hat{G}_{\alpha_i}(\theta,\alpha_i). && (2.3.8)
\end{aligned}
$$

In the sequel, the arguments of the expressions will be omitted when the functions are evaluated at the true parameter value $(\theta_0', \alpha_{i0}')'$, for example $g(z_{it})$ means $g(z_{it};\theta_0,\alpha_{i0})$.

2.3.2 Fixed Effects GMM Estimator (FE-GMM)

In cross-section models, parameters defined from moment conditions are usually estimated using the two-step GMM estimator of Hansen (1982). To describe how to adapt this method to models with fixed effects, let $\hat{g}_i(\theta,\alpha_i) \equiv T^{-1}\sum_{t=1}^{T} g(z_{it};\theta,\alpha_i)$, and let $(\tilde{\theta}', \{\tilde{\alpha}_i'\}_{i=1}^n)'$ be some preliminary one-step estimator, given by $(\tilde{\theta}', \{\tilde{\alpha}_i'\}_{i=1}^n)' = \arg\inf_{\{(\theta',\alpha_i')'\}_{i=1}^n \in \Upsilon} \sum_{i=1}^n \hat{g}_i(\theta,\alpha_i)'\ W_i^{-1}\ \hat{g}_i(\theta,\alpha_i)$, where Υ denotes the individual parameter space, and W_i, $i = 1,...,n$, is a sequence of positive definite symmetric $m \times m$ weighting matrices. The two-step fixed effects GMM estimator is the solution to the following program

$$
(\hat{\theta}', \{\hat{\alpha}_i'\}_{i=1}^n)' = \arg\inf_{\{(\theta',\alpha_i')'\}_{i=1}^n \in \Upsilon} \sum_{i=1}^{n} \hat{g}_i(\theta,\alpha_i)'\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)^{-1}\hat{g}_i(\theta,\alpha_i), \qquad (2.3.9)
$$

where $\hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha_i})$ is an estimate of the optimal weighting matrix Ω_i for individual i. Note that for exactly identified models without common parameters, i.e., $m = q$, the solution to this program is the same as to the preliminary estimator, and the choice of weighting matrices becomes irrelevant.[8]

2.3.3 First Order Conditions for FE-GMM

For the posterior analysis of the asymptotic properties of the estimator, it is convenient to consider the concentrated or profiled problem. This problem is a two-step procedure. In the first step the program is solved for the individual effects $\alpha = (\alpha_1', ..., \alpha_n')'$, given the value of the common parameter θ. The First Order Conditions (FOC) for this stage, reparametrized conveniently as in Newey and Smith (2004), are the following

$$\hat{t}_i(\hat{\gamma}_i, \theta) = -\begin{pmatrix} \hat{G}_{\alpha_i}(\theta, \hat{\alpha}_i(\theta))'\hat{\lambda}_i(\theta) \\ \hat{g}_i(\theta, \hat{\alpha}_i(\theta)) + \hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i)\hat{\lambda}_i(\theta) \end{pmatrix} = 0, \; i = 1, ..., n, \tag{2.3.10}$$

where λ_i is a $m \times 1$ vector of individual Lagrange multipliers for the moment conditions, and $\gamma_i = (\alpha_i', \lambda_i')'$ is the new $(q + m) \times 1$ vector of individual effects. Then, the solutions to the previous equations are plugged into the original problem, leading to the following first order conditions for θ

$$\hat{s}_n(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \hat{s}_i(\hat{\theta}, \hat{\gamma}_i(\hat{\theta})) = -\frac{1}{n}\sum_{i=1}^{n} \hat{G}_{\theta_i}(\hat{\theta}, \hat{\alpha}_i(\hat{\theta}))'\hat{\lambda}_i(\hat{\theta}) = 0. \tag{2.3.11}$$

Note that the problem in this form lies within Hahn and Newey's (2004) framework, with the difference that the individual effects are multidimensional.[9] Moreover, if the model is over-identified, then the initial estimation of the optimal weighting matrices complicates the analysis by introducing additional terms in the bias formulas.

[8] In this model the inclusion of common parameters θ gives rise to over-identification. Intuitively, this case corresponds to having different parameters θ_i for each individual with the additional restrictions $\theta_i = \theta_j \;\forall\; i \neq j$.

[9] In the original parametrization, the FOC can be written as

$$\frac{1}{n}\sum_{i=1}^{n} \hat{G}_{\theta_i}(\hat{\theta}, \hat{\alpha}_i(\hat{\theta}))'\hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i)^{-}\hat{g}_i(\theta, \hat{\alpha}_i(\theta)) = 0, \tag{2.3.12}$$

where the superscript $^{-}$ denotes a generalized inverse.

2.4 Asymptotic Properties of Panel Data GMM Estimators

In this section I analyze the asymptotic properties of one-step and two-step FE-GMM estimators. In both cases, I assume that the initial weighting matrices are deterministic (for example, the identity matrix for each individual).[10] Consistency and asymptotic distributions for estimators of individual effects and common parameter are derived, along with precise rates of convergence. Results and necessary conditions are established separately for one-step and two-step estimators. The section concludes with a simple example that illustrates the procedure followed to derive the results.

To simplify the notation I give results for scalar individual effects, i.e., $q = 1$, but all the formulas can be extended to the multidimensional case without special complications. The arguments for obtaining the properties of GMM estimating procedures are general, and can be used to analyze the properties of other estimators based on moment conditions. In particular, all the results can be extended to the class of generalized empirical likelihood (GEL) estimators, which includes empirical likelihood, continuous updating and exponential tilting estimators.

Condition 1 *(i)* $\{z_{it}, t = 1, 2, ...\}$ *are i.i.d. and independent across* i. *(ii)* $n, T \to \infty$ *such that* $O(T) \leq n = o(T^r)$, *for some* $r > 1$. *Alternatively,* $n = T^r / a_T$, *where* $a_T = o(T^\epsilon)$ *for any* $\epsilon > 0$ *and* $r \geq 1$. *(iii)* $\dim[g(\cdot; \theta, \alpha_i)] = m < \infty$.

The alternative definition of the relative rate of convergence in Condition 1 (ii) is just a representation and does not impose additional restrictions on this rate. It is adopted for notational convenience, because it allows me to jointly treat cases where $n = O(T^r)$ and more general relative rates. The parameter r can be interpreted as a minimum polynomial rate of T that is not dominated by n. For example, if $n = O(T)$, then $a_T = C$ for some constant $C > 0$, or if $n = O(T^r / \log T)$, then $a_T = C \log T$ for some constant $C > 0$.

Condition 2 *(i) The function* $g(\cdot; \theta, \alpha) = (g_1(\cdot; \theta, \alpha), ..., g_m(\cdot; \theta, \alpha))'$ *is continuous in* $(\theta, \alpha) \in \Upsilon$; *(ii) The parameter space* Υ *is compact. (iii)* $\dim(\theta, \alpha) = p + q \leq m$; *(iv) there exists a function* $M(z_{it})$ *such that* $|g_k(z_{it}; \theta, \alpha_i)| \leq M(z_{it})$, $\left|\frac{\partial g_k(z_{it}; \theta, \alpha_i)}{\partial(\theta, \alpha_i)}\right| \leq M(z_{it})$, *for* $k = 1, ..., m$, *and* $\sup_i E\left[M(z_{it})^{2s}\right] < \infty$; *(v)* $s \geq r$; *(vi) for each* $\eta > 0$ *and any sequence of symmetric*

[10] Alternatively, stochastic initial weighting matrices can be used by adding a condition on the stochastic expansions of these matrices. This condition is slightly more restrictive than in Newey and Smith (2004), because it requires uniform convergence of the remainder terms.

deterministic positive definite matrices $\{W_i, i = 1, 2, ...\}$,

$$\inf_i \left[Q_i^W(\theta_0, \alpha_{i0}) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}} Q_i^W(\theta, \alpha) \right] > 0, \tag{2.4.1}$$

where

$$Q_i^W(\theta, \alpha_i) \equiv -g_i(\theta, \alpha_i)' W_i^{-1} g_i(\theta, \alpha_i), \tag{2.4.2}$$

$$g_i(\theta, \alpha_i) \equiv E[\hat{g}_i(\theta, \alpha_i)]. \tag{2.4.3}$$

Conditions 2(i), 2(ii) and 2(iii) are standard in the GMM literature and guarantee the identification of the parameters based on time series variation, see, for e.g., Newey and McFadden (1994). Conditions 2(iv), 2(v) and 2(vi) are needed to establish uniform consistency of the estimators of the individual effects and are similar to Conditions 2 and 3 in Hahn and Newey (2004), but applied to moment conditions instead of log-likelihoods.

Theorem 1 (Consistency of First Stage Estimator for Common Parameters) *Suppose that Conditions 1 and 2 hold. Assume also that, for each i,* W_i *is a finite symmetric positive definite deterministic matrix. Then,*

$$\Pr\left\{ \left| \tilde{\theta} - \theta_0 \right| \geq \eta \right\} = o(T^{r-s}) = o(1) \quad \forall \eta > 0, \tag{2.4.4}$$

where $\tilde{\theta} \in \arg\max_{\{(\theta,\alpha_i)\}_{i=1}^n \in \Upsilon} \frac{1}{n} \sum_{i=1}^n \hat{Q}_i^W(\theta, \alpha_i)$, *and* $\hat{Q}_i^W(\theta, \alpha_i) \equiv -\hat{g}_i(\theta, \alpha_i)' W_i^{-1} \hat{g}_i(\theta, \alpha_i)$.

Proof. See Appendix 2.A. ∎

Theorem 2 (Consistency of First Stage Estimator for Individual Effects) *Suppose that Conditions 1 and 2 hold. Assume also that, for each i,* W_i *is a finite symmetric positive definite deterministic matrix. Then,*

$$\Pr\left\{ \max_{1 \leq i \leq n} |\tilde{\alpha}_i - \alpha_{i0}| \geq \eta \right\} = o\left(T^{r-s}\right) = o(1), \tag{2.4.5}$$

for any $\eta > 0$, *where* $\tilde{\alpha}_i \in \arg\max_\alpha \hat{Q}_i^W(\tilde{\theta}, \alpha)$. *Also,* $\Pr\left\{ \max_{1 \leq i \leq n} \left| \tilde{\lambda}_i \right| \geq \eta \right\} = o\left(T^{r-s}\right) = o(1)$, *for any* $\eta > 0$, *where* $\tilde{\lambda}_i = -W_i^{-1} \hat{g}_i(\tilde{\theta}, \tilde{\alpha}_i)$.

Proof. See Appendix 2.A. ∎

These theorems establish (uniform) consistency of first stage estimators of common parameters and individual effects. Let $J_{si}^W \equiv G_{\theta_i}' P_{\alpha_i}^W G_{\theta_i}$ and $J_s^W = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n J_{si}^W$.

Condition 3 *(i) For each i, $(\theta_0, \alpha_{i0}) \in int\,[\Upsilon]$. (ii) $r < 3$. (iii) For any sequence of symmetric positive definite deterministic matrices $\{W_i, i = 1, 2, ...\}$, J_s^W is finite positive definite, and $\{(G'_{\alpha_i} W_i^{-1} G_{\alpha_i})\}_{i=1}^n$ is a sequence of positive definite matrices $\forall n$.*

Condition 4 *There exists some $M(z_{it})$ such that, for $k = 1, ..., m$*

$$\left| \frac{\partial^{d_1+d_2} g_k(z_{it}; \theta, \alpha_i)}{\partial \theta^{d_1} \partial \alpha_i^{d_2}} \right| \leq M(z_{it}), \qquad 0 \leq d_1 + d_2 \leq 1, \ldots, 6 \tag{2.4.6}$$

and $\sup_i E\left[M(z_{it})^{2s}\right] < \infty$ *for some $s \geq 4r$.*

Conditions 3(i) and 3(iii) are analogous for panel data to the standard asymptotic normality conditions for GMM, see, for e.g., Newey and McFadden (1994). Condition 3(ii) establishes the minimum polynomial rate of T that is not dominated by n, i.e. it is a condition in the relative rate of convergence of the two dimensions of the panel, and it is the same as in other studies in nonlinear panel data (see Hahn and Newey, 2004; or Woutersen, 2002). Condition 4 extends Assumption 3 in Newey and Smith (2004) to the context of panel data, and guarantees the existence of higher order expansions for the GMM estimators and the uniform convergence of their remainder terms.

Let $\Sigma_{\alpha_i}^W = \left(G'_{\alpha_i} W_i^{-1} G_{\alpha_i}\right)^{-1}$, $H_{\alpha_i}^W = \Sigma_{\alpha_i}^W G'_{\alpha_i} W_i^{-1}$, and $P_{\alpha_i}^W = W_i^{-1} - W_i^{-1} G_{\alpha_i} H_{\alpha_i}^W$. Let e_j denote a $m \times 1$ unitary vector with 1 in column j, and $P_{\alpha_i,j}^W$ denote the j-th column of $P_{\alpha_i}^W$.

Theorem 3 (Asymptotic Limit of First Stage Estimators for Common Parameters) *Under Conditions 1, 2, 3 and 4, we have*

$$\begin{cases} \sqrt{nT}(\tilde{\theta} - \theta_0) \xrightarrow{d} N\left(-\sqrt{\rho}(J_s^W)^{-1} B_s^W, (J_s^W)^{-1} V_s^W (J_s^W)^{-1}\right), & \text{if} \quad n = \rho T; \\ T(\tilde{\theta} - \theta_0) \xrightarrow{p} -(J_s^W)^{-1} B_s^W, & \text{otherwise}. \end{cases} \tag{2.4.7}$$

where

$$J_s^W = \lim_{n\to\infty} \tfrac{1}{n} \textstyle\sum_{i=1}^n G'_{\theta_i} P_{\alpha_i}^W G_{\theta_i}, \qquad V_s^W = \lim_{n\to\infty} \tfrac{1}{n} \textstyle\sum_{i=1}^n G'_{\theta_i} P_{\alpha_i}^W \Omega_i P_{\alpha_i}^W G_{\theta_i},$$
$$B_s^W = \lim_{n\to\infty} \tfrac{1}{n} \textstyle\sum_{i=1}^n \left\{ B_{si}^{W,B} + B_{si}^{W,C} + B_{si}^{W,V} \right\}, \tag{2.4.8}$$

with

$$
\begin{aligned}
B_{si}^{W,C} &= E[G_{\theta_i}(z_{it})' P_{\alpha_i}^W g_i(z_{it})], & B_{si}^{W,V} &= -\tfrac{1}{2} G_{\theta\alpha_i}' \left(P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} + \textstyle\sum_{j=1}^m e_j H_{\alpha_i}^W \Omega_i P_{\alpha_i,j} \right), \\
B_{si}^{W,B} &= -G_{\theta_i}' \left(B_{\lambda_i}^{W,I} + B_{\lambda_i}^{W,G} + B_{\lambda_i}^{W,1S} \right), & B_{\lambda_i}^{W,I} &= P_{\alpha_i}^W E[G_{\alpha_i}(z_{it})' H_{\alpha_i}^W g_i(z_{it})] - \tfrac{1}{2} P_{\alpha_i}^W G_{\alpha\alpha_i} \Sigma_{\alpha_i}^W, \\
B_{\lambda_i}^{W,G} &= H_{\alpha_i}^{W\prime} E[G_{\alpha_i}(z_{it})' P_{\alpha_i}^W g_i(z_{it})], & B_{\lambda_i}^{W,1S} &= \tfrac{1}{2} H_{\alpha_i}^W G_{\alpha\alpha_i}' \left(P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} + \textstyle\sum_{j=1}^m e_j H_{\alpha_i}^W \Omega_i P_{\alpha_i,j}^W \right).
\end{aligned}
\tag{2.4.9}
$$

Proof. See Appendix 2.B. ■

The bias in the asymptotic distribution of the GMM estimator comes from the non-zero expectation of the concentrated score of the common parameter at the true parameter value. This bias in the estimating equation in turn is caused by the substitution of the unobserved individual effects for sample estimates. This estimates converge to their true parameter value slower than the sample size under asymptotic sequences where $T = o(n)$. Intuitively, only observations for each individual convey information about the corresponding individual effect. In nonlinear model, the bias in the estimation of the individual effects is transmitted to the estimates of the rest of parameters. This can be explained with an expansion of the first stage plug-in score around the true value of the individual effects[11]

$$
\begin{aligned}
E\left[\hat{s}_i^W(\theta_0, \tilde{\gamma}_i)\right] &= E\left[\hat{s}_i^W\right] + E\left[\hat{s}_{\gamma i}^W\right]' E\left[\tilde{\gamma}_i - \gamma_{i0}\right] + E\left[(\hat{s}_{\gamma i}^W - E\left[\hat{s}_{\gamma i}^W\right])'(\tilde{\gamma}_i - \gamma_{i0})\right] \\
&+ E\left[(\tilde{\gamma}_i - \gamma_{i0})' E\left[\hat{s}_{\gamma\gamma i}^W\right](\tilde{\gamma}_i - \gamma_{i0})\right]/2 + o(T^{-1}) \\
&= 0 + \left\{B_s^{W,B} + B_s^{W,C} + B_s^{W,V}\right\}/T + o(T^{-1}).
\end{aligned}
\tag{2.4.12}
$$

This expression shows that the bias has three components as in the MLE case, see Hahn and Newey (2004). The first component ($B_s^{W,B}$) comes from the higher-order bias of the estimator of the individual effects. The second component ($B_s^{W,C}$) is a correlation term and is present because individual effects and common parameters are estimated using the same observations. The third component ($B_s^{W,V}$) is a variance term. The bias of the individual effects ($B_s^{W,C}$) can be further decomposed in three terms corresponding to the asymptotic bias for a GMM estimator

[11] Using the notation introduced in Section 2.3, the plug-in score for the first stage is

$$
\hat{s}_n^W(\theta_0) = \frac{1}{n}\sum_{i=1}^n \hat{s}_i^W(\theta_0, \tilde{\gamma}_i) = -\frac{1}{n}\sum_{i=1}^n \hat{G}_{\theta_i}(\theta_0, \tilde{\alpha}_i)'\tilde{\lambda}_i, \tag{2.4.10}
$$

where $\tilde{\gamma}_i = (\tilde{\alpha}_i', \tilde{\lambda}_i'), i = 1, ..., n$, are the solutions to

$$
\hat{t}_i^W(\tilde{\gamma}_i, \theta_0) = -\begin{pmatrix} \hat{G}_{\alpha_i}(\theta_0, \tilde{\alpha}_i)'\tilde{\lambda}_i \\ \hat{g}_i(\theta_0, \tilde{\alpha}_i) + W_i \tilde{\lambda}_i \end{pmatrix} = 0, \; i = 1, ..., n. \tag{2.4.11}
$$

with the optimal score ($B_\lambda^{W,I}$), when W is used as the weighting function; the bias arising from estimation of G_{α_i} ($B_\lambda^{W,G}$); and the bias arising from not using an optimal weighting matrix ($B_\lambda^{W,1S}$).

Condition 5 *(i) There exists a function $M(z_{it})$ such that $|g_k(z_{it};\theta,\alpha_i)| \leq M(z_{it})$, $\left|\frac{\partial g_k(z_{it};\theta,\alpha_i)}{\partial(\theta,\alpha_i)}\right| \leq M(z_{it})$, for $k = 1,...,m$, and $\sup_i E\left[M(z_{it})^{4s}\right] < \infty$. (ii) $\{\Omega_i, i = 1, ...\}$ is a sequence of finite positive definite matrices. (iii) $s \geq 4r$.*

Condition 5 is needed to establish the uniform consistency of the estimators of the individual weighting matrices.

Theorem 4 (Consistency of Second Stage Estimator for Common Parameters) *Suppose that Conditions 1, 2, 3 and 5 hold. Then,*

$$\Pr\left\{\left|\hat{\theta}-\theta_0\right| \geq \eta\right\} = o\left(T^{r-s}\right) = o(1) \quad \forall \eta > 0, \tag{2.4.13}$$

where $\hat{\theta} \in \arg\max_{\{(\theta',\alpha_i')\}_{i=1}^n \in \Upsilon} \frac{1}{n}\sum_{i=1}^n \hat{Q}_i^\Omega(\theta,\alpha_i)$, and $\hat{Q}_i^\Omega(\theta,\alpha_i) \equiv -\hat{g}_i(\theta,\alpha_i)'\,\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)^{-1}\hat{g}_i(\theta,\alpha_i)$.

Proof. See Appendix 2.C. ∎

Theorem 5 (Consistency of Second Stage Estimators for Individual Effects) *Suppose that Conditions 1, 2, 3 and 5 hold. Then,*

$$\Pr\left\{\max_{1\leq i\leq n}|\hat{\alpha}_i-\alpha_0| \geq \eta\right\} = o\left(T^{r-s}\right) = o(1), \tag{2.4.14}$$

for any $\eta > 0$, where $\hat{\alpha}_i \in \arg\max_\alpha \hat{Q}_i^\Omega(\hat{\theta},\alpha)$. Also, $\Pr\left\{\max_{1\leq i\leq n}\left|\hat{\lambda}_i\right| \geq \eta\right\} = o\left(T^{r-s}\right) = o(1)$, for any $\eta > 0$, where $\hat{g}_i(\hat{\theta},\hat{\alpha}_i) + \hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)\hat{\lambda}_i = 0$.

Proof. See Appendix 2.C. ∎

Condition 6 *There exists some $M(z_{it})$ such that, for $k = 1,...,m$*

$$\left|\frac{\partial^{d_1+d_2} g_k(z_{it};\theta,\alpha_i)}{\partial\theta^{d_1}\partial\alpha_i^{d_2}}\right| \leq M(z_{it}) \qquad 0 \leq d_1+d_2 \leq 1,\ldots,6 \tag{2.4.15}$$

and $\sup_i E\left[M(z_{it})^{4s}\right] < \infty$ for some $s \geq 4r$.

Condition 6 guarantees the existence of higher order expansions for the estimators of the weighting matrices and uniform convergence of their remainder terms. Conditions 5, and 6 are stronger versions of conditions 2(iv) and 2(v), and 4, respectively. They are presented separately because they are only needed when there is a first stage where the weighting matrices are estimated.

Let $\Sigma_{\alpha_i} = \left(G'_{\alpha_i}\Omega_i^{-1}G_{\alpha_i}\right)^{-1}$, $H_{\alpha_i} = \Sigma_{\alpha_i}G'_{\alpha_i}\Omega_i^{-1}$, and $P_{\alpha_i} = \Omega_i^{-1} - \Omega_i^{-1}G_{\alpha_i}H_{\alpha_i}$.

Theorem 6 (Asymptotic limit of Second Stage Estimators for Common Parameters) *Under the Conditions 1, 2, 3, 4, 5 and 6, we have*

$$\begin{cases} \sqrt{nT}(\hat{\theta}-\theta_0) \xrightarrow{d} N\left(-\sqrt{\rho}(J_s)^{-1}B_s, J_s^{-1}\right), & \text{if} \quad n = \rho T; \\ T(\hat{\theta}-\theta_0) \xrightarrow{p} -J_s^{-1}B_s \equiv B\left(\hat{\theta}\right), & \text{otherwise.} \end{cases} \tag{2.4.16}$$

where

$$\begin{aligned} J_s &= \lim_{n\to\infty} \tfrac{1}{n}\textstyle\sum_{i=1}^n G'_{\theta_i}P_{\alpha_i}G_{\theta_i}, & B_s &= \lim_{n\to\infty} \tfrac{1}{n}\textstyle\sum_{i=1}^n \left[B_{si}^B + B_{si}^C + B_{si}^V\right], \\ B_{si}^C &= E\left[G_{\theta_i}(z_{it})'P_{\alpha_i}g(z_{it})\right], & B_{si}^V &= 0, \\ B_{si}^B &= -G'_{\theta_i}\left[B_{\lambda_i}^I + B_{\lambda_i}^G + B_{\lambda_i}^\Omega + B_{\lambda_i}^W\right], \end{aligned} \tag{2.4.17}$$

with

$$\begin{aligned} B_{\lambda_i}^I &= P_{\alpha_i}E\left[G_{\alpha_i}(z_{it})H_{\alpha_i}g(z_{it})\right] - \tfrac{1}{2}P_{\alpha_i}G_{\alpha\alpha_i}\Sigma_{\alpha_i}, & B_{\lambda_i}^G &= H'_{\alpha_i}E\left[G_{\alpha_i}(z_{it})'P_{\alpha_i}g(z_{it})\right], \\ B_{\lambda_i}^\Omega &= P_{\alpha_i}E\left[g(z_{it})g(z_{it})'P_{\alpha_i}g(z_{it})\right], & B_{\lambda_i}^W &= P_{\alpha_i}\Omega_{\alpha_i}\left(H_{\alpha_i}^W - H_{\alpha_i}\right). \end{aligned} \tag{2.4.18}$$

Proof. See Appendix 2.D. ■

Theorem 6 establishes that one iteration of the GMM procedure not only improves asymptotic efficiency by reducing the variance of the influence function, but also removes the variance and non-optimal weighting matrices components from the bias. The higher-order bias of the estimator of the individual effects (B_λ^B) now has four components, as in Newey and Smith (2004). These components correspond to the asymptotic bias for a GMM estimator with the optimal score (B_λ^I); the bias arising from estimation of G_{α_i} (B_λ^G); the bias arising from estimation of Ω_i (B_λ^Ω); and the bias arising from the choice of the preliminary first step estimator (B_λ^W). Moreover, an additional iteration of the GMM estimator removes the B_λ^W term, which is the most difficult to estimate in practice because it involves Ω_{α_i}.

The general procedure for deriving the asymptotic properties of the estimators consists of several expansions. First, higher-order asymptotic expansions for the estimators of the individual effects are derived, with the common parameter fixed at the true value θ_0. Next, the asymptotic expansion for the plug-in score of the common parameter is obtained from the individual expansions. Finally, the properties of estimator for the common parameter follow from the properties of its plug-in score. This procedure can be illustrated in the classical Neyman-

Scott example of estimation of the common variance in a panel data with different means for each individual. This model has been extensively studied in the literature. It is included for analytical convenience, because it allows me to obtain exact stochastic expansions and closed-form bias-corrected estimators.

Example 2 (Neyman-Scott) *Consider the model*

$$x_{it} = \alpha_{i0} + \epsilon_{it}, \quad \epsilon_{it}|\alpha_{i0} \sim i.i.d.\,(0,\theta_0), \qquad t = 1, ..., T, \tag{2.4.19}$$

where the observations are independent across i, and ϵ_{it} has finite fourth moment. The parameter of interest is the common variance of the disturbance θ_0. I analyze the properties of the following estimators for the individual effects and common parameter

$$\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}, \ i = 1, ..., n, \qquad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}(x_{it} - \hat{\alpha}_i)^2. \tag{2.4.20}$$

First, the stochastic expansion for the estimator of the individual effects can be obtained by substituting (2.4.19) into the expression of $\hat{\alpha}_i$. This yields the following exact expansion

$$\sqrt{T}(\hat{\alpha}_i - \alpha_{i0}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{it} = \tilde{\psi}_{\alpha_i} \quad \xrightarrow{d} \quad N(0,\theta_0), \tag{2.4.21}$$

which does not depend on θ. Note that, unlike in the linear IV model, the estimators of the individual effects do not have higher-order bias. Second, plugging this expansion into (2.4.20), the exact stochastic expansion for the estimator of the common parameter is[12]

$$\sqrt{nT}(\hat{\theta} - \theta_0) = -\sqrt{\frac{n}{T}}\theta_0 + \tilde{\psi}_\theta - \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\tilde{\psi}^2_{\alpha_i} - \theta_0\right), \tag{2.4.22}$$

with

$$\tilde{\psi}_\theta = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}(\epsilon^2_{it} - \theta_0) \xrightarrow{d} N(0, E[\epsilon^4_{it}] - \theta_0^2), \quad \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\tilde{\psi}^2_{\alpha_i} - \theta_0\right) \xrightarrow{p} 0, \tag{2.4.23}$$

where the second result follows from the properties of the V-statistics.[13] *Therefore, the asymp-*

[12]In this case the plug-in score is linear in θ, and therefore the expansion of the common parameter coincides with the expansion of the plug-in score.

[13]See Lemma 59 in Appendix 2.J.

totic bias for $\hat{\theta}$ is

$$E[\hat{\theta}] - \theta_0 = -\frac{1}{T}\theta_0, \tag{2.4.24}$$

which is the well-known bias formula for the estimator of the variance that does not correct for degrees of freedom. Note that in this case the sources of the bias are the correlation between the estimators of the individual effects and common parameter, and the variance of the estimators of the individual effects. Moreover, in this simple example the first term of the expansion captures all the incidental parameters bias because there are no remainder terms in the previous expansions.

2.5 Bias Corrections

When T grows large, the FE-GMM estimators, while consistent, have asymptotic distributions not centered around the true parameter value under asymptotic sequences where $T = o(n)$. These sequences seem appropriate to capture the behavior of the estimators in empirical applications, since the time dimension is usually moderate but much smaller than the number of individuals in commonly used micro-panels, for e.g., the PSID. The presence of this bias has important practical implications because invalidates any inference that uses the standard asymptotic distribution. In this section I describe different methods to adjust the asymptotic distribution of the FE-GMM estimators of the common parameter and smooth functions of the data and model parameters, and therefore to provide valid inference for these quantities. All the methods considered are analytical. These corrections have computational advantages for nonlinear models over other alternative methods, such as Bootstrap and Jackknife. The main disadvantage is that they require closed-form expressions for the bias, but the results derived in previous section can be used. Moreover, Hahn, Kuersteiner and Newey (2004) show that analytical, Bootstrap, and Jackknife bias corrections methods are asymptotically equivalent up to third order for MLE. This result seems likely to apply also to GMM estimators, but the formal proof is beyond the scope of this chapter.

The methods proposed differ in whether the bias is corrected from the estimator directly, or from the estimating equation or plug-in score. For methods that correct the bias of the estimator, one-step and fully iterated procedures are considered. All these methods are first order asymptotically equivalent, reduce the incidental parameters bias from $O(T^{-1})$ to $O(T^{-2})$, and yield asymptotic normal distributions centered at the true parameter value under asymptotic sequences where $n = o(T^3)$. However, they are numerically different in finite samples, and vary in their computational complexity. Thus, score methods are the most computationally intensive because they require solving highly nonlinear equations. Fully iterated methods can also be computationally cumbersome when they do not have closed-form solution. In both cases, bias-corrected estimators can always be obtained by iterative procedures, using as initial value the FE-GMM estimates. Here, it is important to note that, despite the faster rate of convergence of the estimator of the common parameter, it is necessary to recompute the estimators of the individual effects at least in the first iteration. This is because the bias of the estimator of the common parameter affects the higher order bias of the estimator of the individual effects, which, in turn, is part of the bias of the estimator of the common parameter. This can be seen in the following expansion

$$\hat{\alpha}_i(\hat{\theta}) - \hat{\alpha}_i(\theta_0) = \frac{1}{T}\frac{\partial \hat{\alpha}_i}{\partial \theta'} B(\hat{\theta}) + o(T^{-1}). \tag{2.5.1}$$

However, no re-computation is necessary in subsequent iterations because the one-step estimator of the common parameter is already consistent, up to order $O(T^{-2})$.

2.5.1 Bias Correction of the Estimator

This bias correction consists of removing an estimate of the expression of the bias given in Theorem 6 from the estimator of the common parameter. Hence, it only requires computing

$$\hat{B}_n(\hat{\theta}) = -\hat{J}_{sn}^{-1} \hat{B}_{sn}, \tag{2.5.2}$$

where $\hat{J}_{sn} = \frac{1}{n}\sum_{i=1}^{n} \hat{J}_{si}$ and $\hat{B}_{sn} = \frac{1}{n}\sum_{i=1}^{n} \hat{B}_{si}$. The components of $\hat{J}_{si}$ and $\hat{B}_{si}$ can be calculated using individual averages evaluated at the initial estimates of the parameters. The bias-corrected estimator is then $\hat{\theta}^{(BC)} = \hat{\theta} - \frac{1}{T}\hat{B}_n(\hat{\theta})$.

In practice, this bias correction is straightforward to implement because it only requires one (possibly nonlinear) optimization. The estimates of θ and α_i's are obtained from a GMM procedure, including dummy variables for the individual effects, and then the expressions of the bias are calculated using these estimates. The computational complexity of the initial estimation of the FE-GMM can be reduced by constructing an appropriate algorithm, e.g., adapting the methods of Greene (2002) to GMM. This procedure can be fully iterated in order to improve the finite sample properties of the bias-corrected estimators.[14] This is equivalent to solving the

[14]See MacKinnon and Smith (1998) for a comparison of one-step and iterated bias correction methods.

following nonlinear equation

$$\hat{\theta}^{(IBC)} = \hat{\theta} - \frac{1}{T}\hat{B}_n(\hat{\theta}^{(IBC)}). \tag{2.5.3}$$

When the expression of the bias is a linear function in θ, or more generally when $\theta + \hat{B}_n(\theta)$ is invertible, it is possible to obtain a closed-form solution to the previous equation. Otherwise, an iterative procedure is needed.

2.5.2 Bias Correction of the Plug-in Score

An alternative to the previous method consists of removing the bias directly from the first order conditions, instead of from the estimator. This procedure, while computationally more intensive, has the attractive feature that both estimator and bias are obtained simultaneously. This feature contrasts with the previous (one-step) method, where the estimator of the bias is calculated using uncorrected estimates of the parameters. The score bias-corrected estimator is the solution to the following estimating equation

$$\hat{s}_n(\hat{\theta}^{(SBC)}) - \frac{1}{T}\hat{B}_{sn}(\hat{\theta}^{(SBC)}) = 0, \tag{2.5.4}$$

where $\hat{B}_{sn}$ is a sample analog of the expression for the bias of the plug-in score. This adjusted estimating equation can be highly nonlinear increasing the computational complexity of the estimator.

This method is related to the fully iterated bias correction of the common parameter. Thus, using a Taylor expansion of $\hat{s}_n(\hat{\theta}^{(SBC)})$ around θ_0, (2.5.4) can be written as

$$\hat{s}_n(\theta_0) + J_{sn}(\overline{\theta})\left(\hat{\theta}^{(SBC)} - \theta_0\right) - \frac{1}{T}\hat{B}_{sn}(\hat{\theta}^{(SBC)}) = 0, \tag{2.5.5}$$

where $\overline{\theta}$ lies between $\hat{\theta}^{(SBC)}$ and θ_0, and can be different across rows of $J_{sn}(\cdot)$. For the iterated method, noting that $\hat{s}_n(\hat{\theta}) = 0$ and $\hat{B}_n(\theta) = -\hat{J}_{sn}(\theta)^{-1}\hat{B}_{sn}(\theta)$, (2.5.3) can be written

$$\hat{s}_n(\hat{\theta}) + J_{sn}(\hat{\theta}^{(IBC)})\left(\hat{\theta}^{(IBC)} - \hat{\theta}\right) - \frac{1}{T}\hat{B}_{sn}(\hat{\theta}^{(IBC)}) = 0. \tag{2.5.6}$$

Comparing (2.5.5) with (2.5.6), we have that both methods solve the same estimating equation but centered at different points: θ_0 for the score correction and $\hat{\theta}$ for the iterated correction. Moreover, the two estimating equations are the same when the score is linear in θ. This can be seen here, by noting that

$$\hat{s}_n(\hat{\theta}) = \hat{s}_n(\theta_0) + J_{sn}(\tilde{\theta})(\hat{\theta} - \theta_0), \tag{2.5.7}$$

where $\tilde{\theta}$ lies between $\hat{\theta}$ and θ_0. Then, when $J_{sn}(\theta) = J_{sn}$, replacing the previous expression in (2.5.6) yields

$$\hat{s}_n(\theta_0) + J_{sn}\left(\hat{\theta}^{(IBC)} - \theta_0\right) - \frac{1}{T}\hat{B}_{sn}(\hat{\theta}^{(IBC)}) = 0, \tag{2.5.8}$$

which is precisely the estimating equation for the score method (see also Hahn and Newey, 2004).

Example 3 (Neyman-Scott model) *The previous methods can be illustrated in the Neyman-Scott example. Here, the one-step bias-corrected estimator of θ_0 is given by*

$$\hat{\theta}^{(1)} = \hat{\theta} + \hat{\theta}/T. \tag{2.5.9}$$

This procedure can be iterated using the recursion formula $\hat{\theta}^{(k)} = \hat{\theta} + \hat{\theta}^{(k-1)}/T$. In this case, it is straightforward to show that

$$\begin{aligned} \hat{\theta}^{(k)} &= \left[1 - \frac{1}{T^{k+1}}\right]\theta_0 + \left[1 + \frac{1}{T} + \ldots + \frac{1}{T^k}\right]\frac{\tilde{\psi}_\theta}{\sqrt{nT}} \\ &- \left[1 + \frac{1}{T} + \ldots + \frac{1}{T^k}\right]\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\psi}^2_{\alpha_i} - \theta_0\right). \end{aligned} \tag{2.5.10}$$

Hence, each iteration reduces the order of the bias by T^{-1}. The fully iterated bias-corrected estimator is obtained by taking the limit as $k \to \infty$

$$\hat{\theta}^{IBC} = \hat{\theta}^{(\infty)} = \theta_0 + \frac{T}{T-1}\frac{\tilde{\psi}_\theta}{\sqrt{nT}} - \frac{T}{T-1}\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{n}\left(\tilde{\psi}^2_{\alpha_i} - \theta_0\right) = \frac{T}{T-1}\hat{\theta}, \tag{2.5.11}$$

which corresponds to the standard unbiased estimator of the variance.

Alternatively, the bias can be removed directly from the first order conditions by centering the score of the common parameter. In this case the plug-in score and its bias are given by

$$\hat{s}_i(\theta, \hat{\alpha}_i) = \frac{1}{T}\sum_{t=1}^{T}(x_{it} - \hat{\alpha}_i)^2 - \theta, \quad E\left[\hat{s}_i(\theta_0, \hat{\alpha}_i)\right] = -\frac{1}{T}\theta_0. \tag{2.5.12}$$

Then, the bias-corrected score can be constructed as

$$\hat{s}_i^{SBC}(\theta, \hat{\alpha}_i) = \hat{s}_i(\theta, \hat{\alpha}_i) + \frac{1}{T}\theta = \frac{1}{T}\sum_{t=1}^{T}(x_{it} - \hat{\alpha}_i)^2 - \frac{T-1}{T}\theta. \tag{2.5.13}$$

Finally, the score bias-corrected estimator for θ can be obtained by solving the first order condi-

tion in the the modified score, that is

$$\frac{1}{n}\sum_{i=1}^{n}\hat{s}_i^{SBC}(\hat{\theta}^{SBC},\hat{\alpha}_i) \quad = \quad 0 \Rightarrow \hat{\theta}^{SBC} = \hat{\theta}^{IBC} = \frac{T}{T-1}\hat{\theta}. \tag{2.5.14}$$

In this simple example bias correction of the estimator (when iterated) and bias correction of the score are equivalent and remove completely the bias, even for fixed T.

2.5.3 Asymptotic Properties of bias-corrected FE-GMM Estimators

The bias reduction methods described before remove the incidental parameters bias of the common parameter estimator up to order $O(T^{-2})$, and yield normal asymptotic distributions centered at the true parameter value for panels where $n = o(T^3)$. This result is formally stated in Theorem 7, which establishes that all the methods are asymptotically equivalent, up to first order.

Theorem 7 (Asymptotic Distribution of bias-corrected FE-GMM) *Under Conditions 1, 2, 3, 4, 5 and 6, for $C \in \{BC, SBC, IBC\}$ we have*

$$\sqrt{nT}(\hat{\theta}^{(C)} - \theta_0) \xrightarrow{d} N\left(0, J_s^{-1}\right). \tag{2.5.15}$$

where

$$\hat{\theta}^{(BC)} \quad = \quad \hat{\theta} - \frac{1}{T}\hat{B}_n\left(\hat{\theta}\right), \tag{2.5.16}$$

$$\hat{\theta}^{(SBC)} \quad : \quad \hat{s}_n\left(\hat{\theta}^{(SBC)}\right) - \frac{1}{T}\hat{B}_{sn}\left(\hat{\theta}^{(SBC)}\right) = 0, \tag{2.5.17}$$

$$\hat{\theta}^{(IBC)} \quad : \quad \hat{\theta}^{(IBC)} = \hat{\theta} - \frac{1}{T}\hat{B}_n\left(\hat{\theta}^{(IBC)}\right), \tag{2.5.18}$$

$$J_s \quad = \quad \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} G_{\theta_i}' P_{\alpha_i} G_{\theta_i}, \tag{2.5.19}$$

Proof. See Appendix 2.E. ■

Theorem 7 also shows that all the bias-corrected estimators considered are first order asymptotically efficient, since their variances achieve the semiparametric efficiency bound for this model, see Chamberlain (1992).

In nonlinear models the ultimate quantities of interest are usually functions of the data, model parameters and individual effects. For example, average marginal effects of the regressors on the conditional probability of the response are often reported in probit and logit models.

The following corollary establishes bias corrected estimators for this quantities, along with the asymptotic distributions for these estimators.

Corollary 1 (Bias correction of smooth functions of the parameters) *Let $f(z;\theta,\alpha_i)$ be a twice differentiable function in its third argument. For $C \in \{BC, SBC, IBC\}$, let*

$$\hat{\zeta}_{nT}^{(C)} = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} f(z_{it};\hat{\theta}^{(C)},\hat{\alpha}_i(\hat{\theta}^{(C)})) - \frac{1}{T}\hat{\Delta}_{nT}(\hat{\theta}^{(C)}), \tag{2.5.20}$$

$$\hat{\Delta}_{nT}(\theta) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{f_{\alpha}(z_{it};\theta,\hat{\alpha}_i(\theta))\left[\hat{B}_{\alpha_i}(\theta)+\hat{\tilde{\psi}}_{\alpha_{it}}(\theta)\right] + f_{\alpha\alpha}(z_{it};\theta,\hat{\alpha}_i(\theta))\hat{\Sigma}_{\alpha_i}(\theta)\right\}, \tag{2.5.21}$$

where the subscripts on f denote partial derivatives, $\hat{B}_{\alpha_i}(.)$ is a consistent estimator of the first component of B_{γ_i}, $\hat{\tilde{\psi}}_{\alpha_{it}}(.)$ is a consistent estimator of $-H_{\alpha_i}g(z_{it})$, and $\hat{\Sigma}_{\alpha_i}(.)$ is an estimator of Σ_{α_i}. Then, under the conditions of Theorem 7

$$\sqrt{nT}(\hat{\zeta}_{nT}^{(C)} - \zeta) \xrightarrow{d} N(0, V_{\zeta}), \tag{2.5.22}$$

where

$$\zeta = \lim_{n,T\to\infty}\zeta_{nT} = \lim_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} f(z_{it};\theta_0,\alpha_{i0}), \tag{2.5.23}$$

$$\begin{aligned} V_{\zeta} &= \lim_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[f_{\alpha}^2(z_{it};\theta_0,\alpha_{i0})\Sigma_{\alpha_i} + f_{\theta}(z_{it};\theta_0,\alpha_{i0})'J_s^{-1}f_{\theta}(z_{it};\theta_0,\alpha_{i0})\right] \\ &+ \lim_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[(f(z_{it};\theta_0,\alpha_{i0})-\zeta)^2\right]. \end{aligned} \tag{2.5.24}$$

A consistent estimator for V_{ζ} is given by

$$\begin{aligned} \hat{V}_{\zeta} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[f_{\alpha}^2(z_{it};\hat{\theta}^{(C)},\hat{\alpha}_i(\hat{\theta}^{(C)}))\hat{\Sigma}_{\alpha_i} + f_{\theta}(z_{it};\hat{\theta}^{(C)},\hat{\alpha}_i(\hat{\theta}^{(C)}))'\hat{J}_s^{-1}f_{\theta}(z_{it};\hat{\theta}^{(C)},\hat{\alpha}_i(\hat{\theta}^{(C)}))\right] \\ &+ \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[(f(z_{it};\theta_0,\alpha_{i0})-\hat{\zeta}_{nT}^{(C)})^2\right]. \end{aligned} \tag{2.5.25}$$

2.6 Monte Carlo Experiment

For a Monte Carlo example I consider a linear IV model with both common and individual-specific parameters. This model is a simplified version of the specification that I use for the empirical application in next section. The model design is

$$\begin{cases} y_{it} = x_{1it}\alpha_{i0} + x_{2it}\theta_0 + \epsilon_{y,it}, \\ x_{1it} = z_{it}\pi_{i0} + x_{2it}\pi_{2,0} + \epsilon_{x,it}, \end{cases} \tag{2.6.1}$$

where

$$\begin{pmatrix} \epsilon_{y,it} \\ \epsilon_{x,it} \end{pmatrix} \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \quad \begin{pmatrix} \alpha_{i0} \\ \pi_{i0} \end{pmatrix} \sim i.i.d.N\left(\begin{pmatrix} \alpha_0 \\ \pi_0 \end{pmatrix}, \sigma_0^2 \begin{pmatrix} 1 & \eta \\ \eta & 1 \end{pmatrix}\right),$$
$$\begin{pmatrix} x_{2it} \\ z_{it} \end{pmatrix} \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right), \tag{2.6.2}$$

and

$$\begin{aligned} n &= 100; \; T = 8; \; \alpha_0 = \pi_0 = \theta_0 = \pi_{2,0} = \sigma_0^2 = 1; \\ \rho &= 0, \; 0.1, \; 0.5, \; 0.9; \; \eta = 0, \; 0.1, \; 0.5, \; 0.9. \end{aligned} \tag{2.6.3}$$

I report results for estimators of θ_0, and for the mean and standard deviation of the α_{i0}'s, that is for α_0 and σ_0. OLS and IV estimators that impose a constant coefficient on x_{1it} are compared to OLS and IV estimators that allow for the coefficient to be different for each individual. In the estimators that allow for individual-specific coefficients I impose a constant coefficient for x_{2it} in the reduced form of x_{1it} to avoid identification problems due to weak instruments at the individual level. In particular, these estimators are the result of two GMM procedures. In the first GMM an instrument, $\hat{x}_{1it}$, is obtained for x_{1it} using the moment conditions $E\left[Z_{it}(x_{1it} - z_{it}\pi_i - x_{2it}\pi_2)\right] = 0$ with $Z_{it} = (z_{it}, x_{2it})'$. Then, estimates of θ_0 and α_{i0}'s are obtained from the moment conditions $E\left[\hat{X}_{it}(y_{it} - \hat{x}_{1it}\alpha_i - x_{2it}\theta)\right] = 0$ with $\hat{X} = (\hat{x}_{1it}, x_{2it})'$.

Throughout the tables SE/SD denotes the ratio of the average standard error to standard deviation; p; .05 is the rejection frequency with nominal value .05; FC refers to fixed coefficient for x_{1it}; RC refers to random (individual-specific) coefficient for x_{1it}; BC refers to one-step bias correction; and IBC refers to fully iterated bias correction. Standard errors are clustered at the individual level for fixed coefficients estimators (see Arellano, 1987), and are robust to heteroskedasticity for random coefficients estimators.[15]

Table 1a reports the Monte Carlo results for the estimators of θ_0 when $T = 8$ and $\theta_0 = 1$. I find important bias in the estimators of the common parameter that do not account for heterogeneity, with the bias of OLS depending on the degree of endogeneity, ρ, and the correlation between the fixed effects, η; whereas the bias of IV depends only on η. For random coefficients estimators, OLS is only consistent in the absence of endogeneity and the bias corrections have little impact. This is because the true conditional expectation is linear and therefore the first order term of the bias is zero. Random effects IV estimators work fairly well in terms of having small biases for all the configurations of the parameters, but their performances deteriorate slightly with the degree of endogeneity (ρ). Random coefficients estimators also show large improvements in rejection frequencies, which are smaller, due to overestimation of the dispersion, but much closer to their nominal values than for fixed coefficient estimators. Bias corrections have little impact here.

Table 1b shows similar results for estimators of the (mean) coefficient of x_{1it}. Here, as in Section 2.2, fixed coefficient OLS estimates the average effect in the absence of both endogeneity ($\rho = 0$), and correlation between the individual coefficients of the structural and reduced form ($\eta = 0$). IV estimates the average effect in the latter case. On the other hand, random coefficients IV estimators have small biases and give accurate rejection probabilities in all the configurations of the simulation parameters.

Table 1c gives properties for the estimators of the standard deviation (σ) of the individual coefficients of x_{1it}. I find that the bias corrections are relevant for IV and remove most of the bias. Surprisingly, OLS estimates have small bias for most parameter configurations, and only deteriorate when both ρ and η are large. This is probably due to cancellation of bias terms in this particular design.

2.7 Empirical Application

The effect of unions on the structure of wages is a longstanding question in labor economics, see Freeman (1984), Lewis (1986), Robinson (1989), Green (1991), and Card, Lemieux, and Riddell (2004) for surveys and additional references. Most empirical studies that look at the causal effect of union membership on earnings or union premium (increase in worker wage for being

[15]The standard errors for IV random coefficients models should account for estimated regressors. However, this bias downward is more than compensated by the heteroskedasticity adjustment. In particular, the standard errors reported without accounting for estimated regressors are biased upward.

member of a union) recognize the presence of unobserved differences between union and nonunion workers. This heterogeneity has been accounted for by estimating different equations for union and nonunion sectors, e.g., Lee (1978); by using longitudinal estimators, e.g., Chamberlain (1983), Chowdhury and Nickell (1985), Jakubson (1991), and Angrist and Newey (1991); by including sample selection corrections, e.g., Lee (1978), Vella and Verbeek (1998), and Lemieux (1998); or by using instrumental variables techniques, e.g., Robinson (1989), and Card (1996). The validity of all these approaches depends on the underlying selection mechanism of the workers into union and nonunion workers, see Green (1991).

Individual heterogeneity in the union premium arises naturally as a consequence, for instance, of differences in the union presence at the firm level. Thus, bigger gains of union membership are expected for individuals in firms with a higher proportion of unionized workers, even comparing to other firms in the same industry. Here, I consider a different estimation strategy to model this heterogeneity in the union effect. I impose no structure in the selection mechanism by estimating wage equations allowing the union effect to be different for each individual. A similar idea is explored in Lemieux (1998), but he uses a more restrictive random effects specification.

Both OLS and IV methods are considered. For the IV, I follow Vella and Verbeek (1998) and use lagged union status as an instrument for current union status. The identification assumption is that past union statuses have no effects on wages, except through current union status. This is likely to be satisfied when there is queuing for union employment, or when union employment produces long-term advantages. To simplify the analysis, I also assume that union status is measured without error. See Chowdhury and Nickell (1985), Card (1996), and Kane, Rouse and Staiger (1999) for examples of how to deal with measurement error problems when the dependent variable is binary.

The empirical specification is also similar to Vella and Verbeek (1998). In particular, I estimate the following equation

$$w_{it} = \delta_t + U_{it}\alpha_{1i} + X_{it}'\theta + \alpha_{0i} + \epsilon_{it}, \tag{2.7.1}$$

where w_{it} is the log of the real hourly rate of pay on the most recent job; δ_t is a period-specific intercept; U_{it} is the union status; and X_{it} is a vector of covariates that includes completed years of schooling, log of potential experience (age - schooling - 6), and married, health disability, region and industry dummies.

The sample, selected from the National Longitudinal Survey (Youth Sample), consists of full-time young working males who had completed their schooling by 1980, and are then followed over the period 1980 to 1993. I exclude individuals who fail to provide sufficient information for each year, are in the active forces any year, have negative potential experience in at least one year, their schooling decrease in any year or increase by more than two years between two interviews, report too high (more than \$500 per hour) or too low (less than \$1 per hour) wages, or do not change union status during the sample.[16] The final sample includes 294 men. The first period is used as initial condition for the lagged union variable.

Table 2 reports descriptive statistics for the sample used. Union membership is based on a question reflecting whether or not the individual had his wage set in collective bargaining agreement. Roughly 40 % of the sample are union members. Union and nonunion workers have similar observed characteristics, though union workers are slightly more educated, are more frequently married, live more often in the northern central region, and live less often in the South. By industries, there are relatively more union workers in transportation, manufacturing and public administration, and fewer in trade and business. Union membership reduces wage dispersion and has high persistence. Note that all variables, except for the Black and Hispanic dummies, display time variation over the period considered. The unconditional union premium is around 20 %.

Table 3 reports pooled, fixed effects with constant union coefficient, and uncorrected and bias-corrected fixed effects estimates with individual-specific union coefficients of the wage equation. Pooled estimators include time-invariant covariates (dummies for Black and Hispanic). Pooled and standard fixed effects estimates display the same pattern observed in previous studies, with the effect of union increasing when instrumental variables are used, and decreasing when omitted individual time-invariant characteristics are taken into account, see for example Vella and Verbeek (1998). Estimates based on individual-specific coefficients show a slightly different picture, with the average union effect around 9 % and large heterogeneity across workers, although this dispersion is not precisely estimated with the IV individual estimators. Here, as in the Monte Carlo example, the bias corrections have significant impact on the estimator of the dispersion of the union effect. The results also show that estimators that impose a constant coefficient for the union status overstate the average effect. Overall, the pattern of the results is consistent with strong Roy type selection and negative endogeneity bias.

[16]Note that for individuals that do not change union status the union effect is not identified.

2.8 Conclusion

This chapter introduces a new class of fixed effects GMM estimators for panel data models where the effect of some variables can be heterogenous across agents in an unrestricted manner. Bias correction methods are developed because these estimators suffer from the incidental parameters problem. An empirical example that analyzes the union premium shows that not accounting for this heterogeneity on the effect of the regressors can lead to biased estimates of the average effects. In particular, fixed-coefficients estimates overstate the average effect of union membership on earnings.

Other estimators based on moment conditions, like the class of GEL estimators, can be analyzed using a similar methodology. I leave this extension to future research. Over-identified dynamic models, extending the results of Anatolyev (2005) for cross sectional GMM estimators to panel data, is another promising avenue of future research.

Appendix

Throughout the appendices O_{up} and o_{up} will denote uniform orders in probability. For example, for a sequence of random variables $\{\xi_i, i = 1, ..., n\}$, $\xi_i = O_{up}(1)$ means $\max_{1\leq i\leq n} \xi_i = O_p(1)$, and $\xi_i = o_{up}(1)$ means $\max_{1\leq i\leq n} \xi_i = o_p(1)$. e_j denotes a unitary vector with a one in position j. For a matrix $A = (a_{ij}), i = 1, ..., m, j = 1, ..., n$, $|A|$ denotes Euclidean norm, that is $|A|^2 = trace[AA']$. The rest of notation is standard.

2.A Consistency of the First Stage GMM Estimator

2.A.1 Some Lemmas

Lemma 1 *(HN) Assume that ξ_t are i.i.d. with $E[\xi_t] = 0$ and $E[\xi_t^{2s}] < \infty$. Then,*

$$E\left[\left(\sum_{t=1}^{T} \xi_t\right)^{2s}\right] = C(s)T^s + o(T^s) \tag{2.A.1}$$

for some constant $C(s)$.

Proof. See HN. ■

Lemma 2 *(Modification of HN) Suppose that $\{\xi_t, t = 1, 2, \ldots\}$ is a sequence of zero mean i.i.d. random variables. We also assume that $E\left[|\xi_t|^{2s}\right] < \infty$. We then have*

$$\Pr\left[\left|\frac{1}{T}\sum_{t=1}^{T} \xi_t\right| > \eta\right] = O\left(T^{-s}\right) \tag{2.A.2}$$

for every $\eta > 0$.

Proof. Using Lemma 1, we obtain

$$E\left[\left|\sum_{t=1}^{T} \xi_t\right|^{2s}\right] \leq CT^s \cdot E\left[|\xi_t|^{2s}\right], \tag{2.A.3}$$

where $C > 0$ is a constant. Therefore, by Markov's inequality

$$\Pr\left[\left|\frac{1}{T}\sum_{t=1}^{T} \xi_t\right| > \eta\right] \leq \frac{CT^s}{T^{2s}\eta^{2s}} E\left[|\xi_t|^{2s}\right] = O(T^{-s}). \tag{2.A.4}$$

■

Lemma 3 *Suppose that, for each i, $\{z_{it}, t = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables. We assume that $\{z_{it}, t = 1, 2, \ldots\}$ are independent across i. Let $h(z_{it}; \theta, \alpha_i)$ be a function such that (i)*

$h(z_{it};\theta,\alpha_i)$ *is continuous in* $(\theta,\alpha_i)\in\Upsilon$*; (ii)* Υ *is compact; (iii) there exists a function* $M(z_{it})$ *such that* $|h(z_{it};\theta,\alpha_i)|\leq M(z_{it})$ *and* $\left|\frac{\partial h(z_{it};\theta,\alpha_i)}{\partial(\theta,\alpha_i)}\right|\leq M(z_{it})$*, with* $E\left[M(z_{it})^{2s}\right]<\infty$*. Then, we have*

$$\Pr\left\{\left|\hat{h}_i(\theta,\alpha_i)-h_i(\theta,\alpha_i)\right|\geq\eta\right\}=O\left(T^{-s}\right), \tag{2.A.5}$$

for every $\eta>0$*, where*

$$\hat{h}_i(\theta,\alpha_i)=\frac{1}{T}\sum_{t=1}^{T}h(z_{it};\theta,\alpha_i), \tag{2.A.6}$$

$$h_i(\theta,\alpha_i)=E[\hat{h}_i(\theta,\alpha_i)]. \tag{2.A.7}$$

Proof. For any $\eta>0$

$$\Pr\left\{\left|\hat{h}_i(\theta,\alpha_i)-h_i(\theta,\alpha_i)\right|\geq\eta\right\}\leq\frac{E\left[\left|\hat{h}_i(\theta,\alpha_i)-h_i(\theta,\alpha_i)\right|^{2s}\right]}{\eta^{2s}}=\frac{E\left[\left|\sum_{t=1}^{T}\left(h(z_{it};\theta,\alpha_i)-h_i(\theta,\alpha_i)\right)\right|^{2s}\right]}{\eta^{2s}T^{2s}}$$

$$\leq\frac{CT^sE\left[|h(z_{it};\theta,\alpha_i)-h_i(\theta,\alpha_i)|^{2s}\right]}{\eta^{2s}T^{2s}}=O\left(T^{-s}\right), \tag{2.A.8}$$

where C is some positive constant. The first inequality follows by Markov Inequality, the second inequality is immediate and the third inequality follows by Lemma 4 and Lemma 1. ■

Lemma 4 *Under the conditions of Lemma 3*

$$E\left[|h(z_{it};\theta,\alpha_i)-h_i(\theta,\alpha_i)|^{2s}\right]<\infty. \tag{2.A.9}$$

Proof. By the triangle inequality, we have

$$E\left[|h(z_{it};\theta,\alpha_i)-h_i(\theta,\alpha_i)|^{2s}\right]\leq E\left[\left(|h(z_{it};\theta,\alpha_i)|+|h_i(\theta,\alpha_i)|\right)^{2s}\right]\leq E\left[\left(M(z_{it})+E[M(z_{it})]\right)^{2s}\right]$$

$$\leq C(s)\cdot E\left[M(z_{it})^{2s}\right]<\infty, \tag{2.A.10}$$

for some constant $C(s)$. ■

Lemma 5 *(Modification of HN) Under the conditions of Lemma 3*

$$\Pr\left\{\sup_{(\theta,\alpha)}\left|\hat{h}_i(\theta,\alpha)-h_i(\theta,\alpha)\right|>\eta\right\}=O(T^{-s}). \tag{2.A.11}$$

Proof. Let $\varepsilon>0$ be chosen such that $2\varepsilon\max_i\ E\left[M\left(x_{it}\right)\right]<\frac{\eta}{3}$. Divide Υ into subsets Υ_1, Υ_2, ..., $\Upsilon_{M(\varepsilon)}$, such that $|(\theta,\alpha)-(\theta',\alpha')|<\varepsilon$ whenever (θ,α) and (θ',α') are in the same subset. Note that $M\left(\varepsilon\right)$ is finite by compactness of Υ. Let (θ_j,α_j) denote *some* point in Υ_j for each j. Then,

$$\sup_{(\theta,\alpha)}\left|\hat{h}_i\left(\theta,\alpha\right)-h_i\left(\theta,\alpha\right)\right|=\max_j\sup_{\Upsilon_j}\left|\hat{h}_i\left(\theta,\alpha\right)-h_i\left(\theta,\alpha\right)\right|, \tag{2.A.12}$$

and therefore

$$\Pr\left[\sup_{(\theta,\alpha)}\left|\hat{h}_i(\theta,\alpha)-h_i(\theta,\alpha)\right|\geq\eta\right]\leq\sum_{j=1}^{M(\varepsilon)}\Pr\left[\sup_{\Upsilon_j}\left|\hat{h}_i(\theta,\alpha)-h_i(\theta,\alpha)\right|\geq\eta\right]. \qquad (2.A.13)$$

For $(\theta,\alpha)\in\Upsilon_j$, we have

$$\left|\hat{h}_i(\theta,\alpha)-h_i(\theta,\alpha)\right|\leq\left|\hat{h}_i(\theta_j,\alpha_j)-h_i(\theta_j,\alpha_j)\right|+\frac{\varepsilon}{T}\left|\sum_{t=1}^{T}\left(M(z_{it})-E\left[M(z_{it})\right]\right)\right|+2\varepsilon E\left[M(z_{it})\right]. \qquad (2.A.14)$$

Then,

$$\begin{aligned}\Pr\left[\sup_{\Upsilon_j}\left|\hat{h}_i(\theta,\alpha)-h_i(\theta,\alpha)\right|\geq\eta\right] &\leq \Pr\left[\left|\hat{h}_i(\theta_j,\alpha_j)-h_i(\theta_j,\alpha_j)\right|\geq\frac{\eta}{3}\right]\\ &\quad+\Pr\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(M(z_{it})-E\left[M(z_{it})\right]\right)\right|\geq\frac{\eta}{3\varepsilon}\right]\\ &= O\left(T^{-s}\right) \qquad (2.A.15)\end{aligned}$$

by Lemmas 2 and 3. ■

Lemma 6 *Under the conditions of Lemma 3, if $n=o(T^r)$, then*

$$\Pr\left\{\max_{1\leq i\leq n}\sup_{(\theta,\alpha)\in\Upsilon}|\hat{h}_i(\theta,\alpha)-h_i(\theta,\alpha)|\geq\eta\right\}=o(T^{r-s}). \qquad (2.A.16)$$

for every $\eta>0$.

Proof. Given $\eta>0$, note that

$$\begin{aligned}\Pr\left\{\max_{1\leq i\leq n}\sup_{(\theta,\alpha)\in\Upsilon}|\hat{h}_i(\theta,\alpha)-h_i(\theta,\alpha)|\geq\eta\right\}&\leq\sum_{i=1}^{n}\Pr\left\{\sup_{(\theta,\alpha)\in\Upsilon}|\hat{h}_i(\theta,\alpha)-h_i(\theta,\alpha)|\geq\eta\right\}\\ &=no(T^{-s})=o(T^{r-s}). \qquad (2.A.17)\end{aligned}$$

■

Lemma 7 *Suppose Conditions 1 and 2 hold and for each i, W_i is a finite positive definite symmetric deterministic matrix. Then, for every $\eta>0$*

$$\Pr\left\{\max_{1\leq i\leq n}\sup_{(\theta,\alpha)\in\Upsilon}\left|\hat{Q}_i^W(\theta,\alpha)-Q_i^W(\theta,\alpha)\right|\geq\eta\right\}=o(T^{r-s}), \qquad (2.A.18)$$

where

$$\hat{Q}_i^W(\theta,\alpha) = -\hat{g}_i(\theta,\alpha)'W_i^{-1}\hat{g}_i(\theta,\alpha), \tag{2.A.19}$$

$$Q_i^W(\theta,\alpha) = -g_i(\theta,\alpha)'W_i^{-1}g_i(\theta,\alpha), \tag{2.A.20}$$

$$g_i(\theta,\alpha) = E[\hat{g}_i(\theta,\alpha)]. \tag{2.A.21}$$

Proof. First, note that

$$\begin{aligned}\left|\hat{Q}_i^W(\theta,\alpha)-Q_i^W(\theta,\alpha)\right| &\le \left|(\hat{g}_i(\theta,\alpha)-g_i(\theta,\alpha))'W_i^{-1}(\hat{g}_i(\theta,\alpha)-g_i(\theta,\alpha))\right| \\ &\quad +2\cdot\left|g_i(\theta,\alpha)'W_i^{-1}(\hat{g}_i(\theta,\alpha)-g_i(\theta,\alpha))\right| \\ &\le m^2\max_{1\le k\le m}\left|(\hat{g}_{k,i}(\theta,\alpha)-g_{k,i}(\theta,\alpha))\right|^2|W_i|^{-1} \\ +2m^2\max_{1\le i\le n}E[M(z_{it})]\,|W_i|^{-1}&\max_{1\le k\le m}\left|(\hat{g}_{k,i}(\theta,\alpha)-g_{k,i}(\theta,\alpha))\right|.\end{aligned} \tag{2.A.22}$$

Then, we have

$$\begin{aligned} &\Pr\left\{\max_{1\le i\le n}\sup_{(\theta,\alpha)\in\Upsilon}\left|\hat{Q}_i^W(\theta,\alpha)-Q_i^W(\theta,\alpha)\right|\ge\eta\right\} \\ \le\;&\Pr\left\{\max_{1\le i\le n}\sup_{(\theta,\alpha)\in\Upsilon}m^2\max_{1\le k\le m}\left|(\hat{g}_{k,i}(\theta,\alpha)-g_{k,i}(\theta,\alpha))\right|^2|W_i|^{-1}\ge\frac{\eta}{2}\right\} \\ +\;&\Pr\left\{\max_{1\le i\le n}|W_i|^{-1}\sup_{(\theta,\alpha)\in\Upsilon}\max_{1\le k\le m}\left|(\hat{g}_{k,i}(\theta,\alpha)-g_{k,i}(\theta,\alpha))\right|\ge\frac{\eta}{4\cdot m^2\max_{1\le i\le n}E[M(z_{it})]}\right\}.\end{aligned} \tag{2.A.23}$$

The conclusion follows by Condition 2 and Lemma 6. ■

Lemma 8 *Suppose that Conditions 1 and 2 hold. Assume also that, for each i, W_i is a finite symmetric positive definite deterministic matrix. Then,*

$$\sup_{\alpha}\left|Q_i^W(\theta,\alpha)-Q_i^W(\theta',\alpha)\right|\le C\cdot E[M(z_{it})]^2\,|\theta-\theta'| \tag{2.A.24}$$

for some constant $C>0$.

Proof. Note that

$$\begin{aligned}\left|Q_i^W(\theta,\alpha)-Q_i^W(\theta',\alpha)\right| &\le \left|g_i(\theta,\alpha)'W_i^{-1}(g_i(\theta,\alpha)-g_i(\theta',\alpha))\right| \\ &+ \left|(g_i(\theta,\alpha)-g_i(\theta',\alpha))'W_i^{-1}g_i(\theta',\alpha)\right| \\ &\le 2\cdot m^2E[M(z_{it})]^2\,|W_i|^{-1}\,|\theta-\theta'|.\end{aligned} \tag{2.A.25}$$

■

2.A.2 Proof of Theorem 1

Proof. Let η be given, and let $\varepsilon \equiv \inf_i \left[Q_i^W(\theta_0, \alpha_{i0}) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}} Q_i^W(\theta,\alpha)\right] > 0$ by Condition 2. With probability $1 - o(T^{r-s})$, we have

$$\begin{aligned}
\max_{|\theta-\theta_0|>\eta,\alpha_1,\ldots,\alpha_n} n^{-1}\sum_{i=1}^{n} \hat{Q}_i^W(\theta,\alpha_i) &\leq \max_{|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta} n^{-1}\sum_{i=1}^{n} \hat{Q}_i^W(\theta,\alpha_i) \\
&< \max_{|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta} n^{-1}\sum_{i=1}^{n} Q_i^W(\theta,\alpha_i) + \frac{1}{3}\varepsilon \\
&< n^{-1}\sum_{i=1}^{n} Q_i^W(\theta_0,\alpha_{i0}) - \frac{2}{3}\varepsilon \\
&< n^{-1}\sum_{i=1}^{n} \hat{Q}_i^W(\theta_0,\alpha_{i0}) - \frac{1}{3}\varepsilon, \qquad (2.A.26)
\end{aligned}$$

where the second and fourth inequalities are based on Lemma 7. Because

$$\max_{\theta,\alpha_1,\ldots,\alpha_n} n^{-1}\sum_{i=1}^{n} \hat{Q}_i^W(\theta,\alpha_i) \geq n^{-1}\sum_{i=1}^{n} \hat{Q}_i^W(\theta_0,\alpha_{i0}) \qquad (2.A.27)$$

by definition, we can conclude that $\Pr\left[\left|\tilde{\theta} - \theta_0\right| \geq \eta\right] = o(T^{r-s}) = o(1)$. ■

2.A.3 Proof of Theorem 2

Proof. Part I: Consistency of $\tilde{\alpha}_i$.

We first prove that

$$\Pr\left[\max_{1\leq i\leq n}\sup_{\alpha}\left|\hat{Q}_i^W\left(\tilde{\theta},\alpha\right) - Q_i^W(\theta_0,\alpha)\right| \geq \eta\right] = o(1) \qquad (2.A.28)$$

for every $\eta > 0$. Note that

$$\begin{aligned}
&\max_{1\leq i\leq n}\sup_{\alpha}\left|\hat{Q}_i^W\left(\tilde{\theta},\alpha\right) - Q_i^W(\theta_0,\alpha)\right| \\
\leq\ &\max_{1\leq i\leq n}\sup_{\alpha}\left|\hat{Q}_i^W\left(\tilde{\theta},\alpha\right) - Q_i^W\left(\tilde{\theta},\alpha\right)\right| + \max_{1\leq i\leq n}\sup_{\alpha}\left|Q_i^W\left(\tilde{\theta},\alpha\right) - Q_i^W(\theta_0,\alpha)\right| \\
\leq\ &\max_{1\leq i\leq n}\sup_{(\theta,\alpha)}\left|\hat{Q}_i^W(\theta,\alpha) - Q_i^W(\theta,\alpha)\right| + \max_{1\leq i\leq n} C\cdot E\left[M(z_{it})\right]^2\cdot\left|\tilde{\theta}-\theta_0\right|, \qquad (2.A.29)
\end{aligned}$$

where the last inequality hold by Lemma 8. Therefore,

$$\begin{aligned}
\Pr\left[\max_{1\leq i\leq n}\sup_{\alpha}\left|\hat{Q}_i^W\left(\tilde{\theta},\alpha\right) - Q_i^W(\theta_0,\alpha)\right| \geq \eta\right] &\leq \Pr\left[\max_{1\leq i\leq n}\sup_{(\theta,\alpha)}\left|\hat{Q}_i^W(\theta,\alpha) - Q_i^W(\theta,\alpha)\right| \geq \frac{\eta}{2}\right] \\
&\quad + \Pr\left[\left|\tilde{\theta}-\theta_0\right| \geq \frac{\eta}{2\cdot C\left(\max_{1\leq i\leq n} E\left[M(z_{it})\right]^2\right)}\right] \\
&= o\left(T^{r-s}\right), \qquad (2.A.30)
\end{aligned}$$

by Lemma 7 and Theorem 1.

We now get back to the proof of Theorem 2. Let η be given, and let

$$\varepsilon \equiv \inf_i \left[Q_i^W(\theta_0, \alpha_{i0}) - \sup_{\{\alpha_i : |\alpha_i - \alpha_{i0}| > \eta\}} Q_i^W(\theta_0, \alpha_i) \right] > 0. \tag{2.A.31}$$

Condition on the event

$$\left\{ \max_{1 \le i \le n} \sup_\alpha \left| \hat{Q}_i^W\left(\tilde{\theta}, \alpha\right) - Q_i^W(\theta_0, \alpha) \right| \le \frac{1}{3}\varepsilon \right\}, \tag{2.A.32}$$

which has a probability equal to $1 - o(T^{r-s})$ by (2.A.28). We then have

$$\max_{|\alpha_i - \alpha_{i0}| > \eta} \hat{Q}_i^W\left(\tilde{\theta}, \alpha_i\right) < \max_{|\alpha_i - \alpha_{i0}| > \eta} Q_i^W(\theta_0, \alpha_i) + \frac{1}{3}\varepsilon < Q_i^W(\theta_0, \alpha_{i0}) - \frac{2}{3}\varepsilon < \hat{Q}_i^W\left(\tilde{\theta}, \alpha_{i0}\right) - \frac{1}{3}\varepsilon \tag{2.A.33}$$

This is inconsistent with $\hat{Q}_i^W\left(\tilde{\theta}, \tilde{\alpha}_i\right) \ge \hat{Q}_i^W\left(\tilde{\theta}, \alpha_{i0}\right)$, and therefore, $|\tilde{\alpha}_i - \alpha_{i0}| \le \eta$ for every i.

Part II: Consistency of $\tilde{\lambda}_i$.

First, note that

$$\begin{aligned}
\left|\tilde{\lambda}_i\right| &= \left|W_i^{-1}\hat{g}_i(\tilde{\theta}, \tilde{\alpha}_i)\right| \le m\left|W_i\right|^{-1} \max_{1 \le k \le m} \left(\left|\hat{g}_{k,i}(\tilde{\theta}, \tilde{\alpha}_i) - g_{k,i}(\tilde{\theta}, \tilde{\alpha}_i)\right| + \left|g_{k,i}(\tilde{\theta}, \tilde{\alpha}_i)\right| \right) \\
&\le m\left|W_i\right|^{-1} \max_{1 \le k \le m} \left[\sup_{(\theta, \alpha_i) \in \Upsilon} \left|\hat{g}_{k,i}(\theta, \alpha_i) - g_{k,i}(\theta, \alpha_i)\right| \right] \\
&+ m\left|W_i\right|^{-1} M(z_{it}) \left|\tilde{\theta} - \theta_0\right| + m\left|W_i\right|^{-1} M(z_{it}) \left|\tilde{\alpha}_i - \alpha_{i0}\right|.
\end{aligned} \tag{2.A.34}$$

Then, the result follows by Lemma 6, Theorem 1 and the first statement of Theorem 2. ■

Corollary 2 *Suppose that Conditions 1 and 2 hold. Assume also that, for each i, W_i is a finite symmetric positive definite deterministic matrix. Then,*

$$\Pr\left\{ \max_{1 \le i \le n} |\tilde{\alpha}_{i0} - \alpha_{i0}| \ge \eta \right\} = o(T^{r-s}) = o(1), \tag{2.A.35}$$

for any $\eta > 0$, where

$$\tilde{\alpha}_{i0} \in \arg\max_\alpha \hat{Q}_i^W(\theta_0, \alpha). \tag{2.A.36}$$

Also,

$$\Pr\left\{ \max_{1 \le i \le n} \left|\tilde{\lambda}_{i0}\right| \ge \eta \right\} = o(T^{r-s}) = o(1), \tag{2.A.37}$$

for any $\eta > 0$, where

$$\tilde{\lambda}_{i0} = -W_i^{-1}\hat{g}_i(\theta_0, \tilde{\alpha}_{i0}). \tag{2.A.38}$$

Proof. The arguments in the proof of Theorem 1 go through replacing $\tilde{\theta}$ for θ_0. ■

2.B Asymptotic Distribution of the First Stage GMM Estimator

2.B.1 Some Lemmas

Lemma 9 *(Modification of HN) Suppose that, for each i, $\{\xi_{it}(\phi), t = 1, 2, \ldots\}$ is a sequence of zero mean i.i.d. random variables indexed by some parameter $\phi \in \Phi$. We assume that $\{\xi_{it}(\phi), t = 1, 2, \ldots\}$ are independent across i. We also assume that $\sup_{\phi\in\Phi} |\xi_{it}(\phi)| \leq B_{it}$ for some sequence of random variables B_{it} that is i.i.d. across t and independent across i. Finally, we assume that $\max_i E\left[B_{it}^{2s}\right] < \infty$, and $n = o(T^r)$, with $s \geq r$. We then have*

$$\Pr\left[\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}(\phi_i)\right| > \eta T^{\frac{r}{2s}}\right] = o(1) \tag{2.B.1}$$

for every $\eta > 0$. Here, $\{\phi_i\}$ is an arbitrary sequence in Φ.

Proof. For each i, we have

$$\begin{aligned}
\Pr\left[\sup_{\phi\in\Phi}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}(\phi_i)\right| > \eta T^{\frac{r}{2s}}\right] &= \Pr\left[\sup_{\phi\in\Phi}\left|\sum_{t=1}^{T}\xi_{it}(\phi_i)\right| > \eta T^{\frac{1}{2}+\frac{r}{2s}}\right] \\
&\leq \frac{E\left[\sup_{\phi\in\Phi}\left|\sum_{t=1}^{T}\xi_{it}(\phi_i)\right|^{2s}\right]}{\eta^{2s}\, T^{2s\left(\frac{1}{2}+\frac{r}{2s}\right)}} \\
&\leq \frac{C(s)\sup_{\phi\in\Phi} T^s E\left[|\xi_{it}(\phi_i)|^{2s}\right]}{\eta^{2s}T^{s+r}} \\
&\leq \frac{C(s)\max_i E\left[B_{it}^{2s}\right]}{\eta^{2s}T^r} = \frac{C}{T^r},
\end{aligned} \tag{2.B.2}$$

for some $C > 0$, where the second inequality is based on Lemma 2. Therefore, we have

$$\Pr\left[\max_i\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}(\phi_i)\right| > \eta T^{\frac{r}{2s}}\right] \leq \sum_{i=1}^{n}\Pr\left\{\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}(\phi_i)\right| > \eta T^{\frac{r}{2s}}\right\} \leq n\frac{C}{T^r} = o(1). \tag{2.B.3}$$

■

Lemma 10 *Suppose that Conditions 1 and 2 hold. Then, for every $\eta > 0$, we have*

$$\Pr\left\{\left|\overline{\theta} - \theta_0\right| \geq \eta\right\} = o(1) \tag{2.B.4}$$

for any $\overline{\theta}$ between θ_0 and $\tilde{\theta}$,

$$\Pr\left\{\max_{1\leq i\leq n}|\overline{\alpha}_{i0} - \alpha_{i0}| \geq \eta\right\} = o(1) \tag{2.B.5}$$

for any $\overline{\alpha}_{i0}$ *between* α_{i0} *and* $\tilde{\alpha}_{i0}$*,*

$$\Pr\left\{\max_{1\leq i\leq n}\left|\overline{\alpha}_i - \alpha_{i0}\right| \geq \eta\right\} = o(1) \tag{2.B.6}$$

for any $\overline{\alpha}_i$ *between* α_{i0} *and* $\tilde{\alpha}_i$*,*

$$\Pr\left\{\max_{1\leq i\leq n}\left|\overline{\lambda}_{i0}\right| \geq \eta\right\} = o(1) \tag{2.B.7}$$

for any $\overline{\lambda}_i$ *between* $\lambda_{i0} = 0$ *and* $\tilde{\lambda}_{i0}$*, and*

$$\Pr\left\{\max_{1\leq i\leq n}\left|\overline{\lambda}_i\right| \geq \eta\right\} = o(1) \tag{2.B.8}$$

for any $\overline{\lambda}_i$ *between* $\lambda_{i0} = 0$ *and* $\tilde{\lambda}_i$*.*

Proof. The four statements hold by Theorems 1 and 2, and Corollary 2, noting that

$$\left|\overline{\zeta} - \zeta_0\right| \leq \left|\tilde{\zeta} - \zeta_0\right|, \tag{2.B.9}$$

for $\zeta \in \{\theta, \alpha_i, \lambda_i\}$. ∎

Lemma 11 *Suppose that, for each* i*,* $\{z_{it}, t = 1, 2, \ldots\}$ *is a sequence of i.i.d. random variables. We assume that* $\{z_{it}, t = 1, 2, \ldots\}$ *are independent across* i*. Let* $h(z_{it}; \theta, \alpha_i)$ *be a function such that (i)* $h(z_{it}; \theta, \alpha_i)$ *is five times continuously differentiable in* $(\theta, \alpha_i) \in \Upsilon$*; (ii)* Υ *is compact; (iii) there exists a function* $M(z_{it})$ *such that* $|h(z_{it}; \theta, \alpha_i)| \leq M(z_{it})$ *and* $\left|\frac{\partial^{d_1+d_2} h(z_{it};\theta,\alpha_i)}{\partial\theta^{d_1}\alpha_i^{d_2}}\right| \leq M(z_{it})$ *for* $1 \leq d_1 + d_2 \leq 5$*, with* $E\left[M(z_{it})^{2s}\right] < \infty$*. Let* $H_i(z_{it}; \theta, \alpha_i)$ *denote*

$$\frac{\partial^{d_1+d_2} h(z_{it}; \theta, \alpha_i)}{\partial\theta^{d_1}\alpha_i^{d_2}}, \tag{2.B.10}$$

$\hat{H}_i(\theta, \alpha_i)$ *denote*

$$\frac{1}{T}\sum_{t=1}^{T} H_i(z_{it}; \theta, \alpha_i), \tag{2.B.11}$$

and $H_i(\theta, \alpha_i)$ *denote*

$$E\left[\hat{H}_i(\theta, \alpha_i)\right], \tag{2.B.12}$$

for some $0 \leq d_1 + d_2 \leq 5$*. Let* α_i^* *be defined as*

$$\alpha_i^* = \arg\sup_{\alpha} \hat{Q}_i^W(\theta^*, \alpha), \tag{2.B.13}$$

such that $\alpha_i^* - \alpha_{i0} = o_{up}(T^{a_\alpha})$ *and* $\theta^* - \theta_0 = o_p(T^{a_\theta})$*, with* $-\frac{1}{2}(1 - r/s) \leq a = \max(a_\alpha, a_\theta) \leq 0$*.*

Assume also that $n = o(T^r)$*, with* $s \geq r$*. Then, for any* $\overline{\theta}$ *between* θ^* *and* θ_0*,* $\overline{\alpha}_i$ *between* α_i^* *and* α_{i0}*,* *and* $\eta > 0$, *we have*

$$\Pr\left[\max_{1\le i\le n}\left|\hat{H}_i(\overline{\theta},\overline{\alpha}_i) - H_i(\theta_0,\alpha_{i0})\right| > \eta T^a\right] = o\left(1\right), \tag{2.B.14}$$

and

$$\Pr\left[\max_{1\le i\le n}\sqrt{T}\left|\hat{H}_i(\overline{\theta},\overline{\alpha}_i) - H_i(\overline{\theta},\overline{\alpha}_i)\right| > \eta T^{\frac{r}{2s}}\right] = o\left(1\right). \tag{2.B.15}$$

Proof. Note that we can write

$$\begin{aligned}\hat{H}_i(\overline{\theta},\overline{\alpha}_i) - E\left[\hat{H}_i(\theta_0,\alpha_{i0})\right] &= \left(\hat{H}_i(\overline{\theta},\overline{\alpha}_i) - \hat{H}_i(\theta_0,\alpha_{i0})\right) + \left(\hat{H}_i(\theta_0,\alpha_{i0}) - E\left[\hat{H}_i(\theta_0,\alpha_{i0})\right]\right)\\ &= \frac{\partial \hat{H}_i(\theta^{**},\alpha_i^{**})}{\partial\theta'}(\overline{\theta}-\theta_0) + \frac{\partial \hat{H}_i(\theta^{**},\alpha_i^{**})}{\partial\alpha_i}(\overline{\alpha}_i-\alpha_{i0}) + \left(\hat{H}_i(\theta_0,\alpha_{i0}) - E\left[\hat{H}_i(\theta_0,\alpha_{i0})\right]\right),\end{aligned} \tag{2.B.16}$$

where $(\theta^{**},\alpha_i^{**})$ lies between (θ_0,α_{i0}) and $(\overline{\theta},\overline{\alpha}_i)$. Next, observe that

$$\begin{aligned}\left|\hat{H}_i(\overline{\theta},\overline{\alpha}_i) - E\left[\hat{H}_i(\theta_0,\alpha_{i0})\right]\right| \le\;& \underbrace{\left|\overline{\theta}-\theta_0\right|}_{=o_{up}(T^a)}\underbrace{\left|\frac{1}{T}\sum_{t=1}^{T}M(z_{it})\right|}_{=O_u(1)+o_{up}(1)} + \underbrace{\left|\overline{\alpha}_i-\alpha_{i0}\right|}_{=o_{up}(T^a)}\underbrace{\left|\frac{1}{T}\sum_{t=1}^{T}M(z_{it})\right|}_{=O_u(1)+o_{up}(1)}\\ &+ \underbrace{\left|\hat{H}_i\left(\theta_0,\alpha_{i0}\right) - E\left[\hat{H}_i\left(\theta_0,\alpha_{i0}\right)\right]\right|}_{=o_{up}\left(T^{-\frac{1}{2}(1-r/s)}\right)} = o_{up}(T^a),\end{aligned} \tag{2.B.17}$$

by the conditions of this Lemma and Lemma 9.

Finally, the second statement holds by Lemma 9 for $\xi_{it}(\phi_i) = H_i(z_{it};\overline{\theta},\overline{\alpha}_i) - E\left[H_i(z_{it};\overline{\theta},\overline{\alpha}_i)\right]$, with $\phi_i = (\overline{\theta}',\overline{\alpha}_i)'$ and $B_{it} = M(z_{it})$. ∎

Lemma 12 *Let* $\{\hat{\xi}_{j,i}, i = 1,2,\ldots\}$, $j = 1,2$, *be two sequences of independent random variables such that* $\hat{\xi}_{1,i} = o_{up}(T^{a_1})$, $\hat{\xi}_{2,i} = o_{up}(T^{a_2})$ *and* $a_1 \ge a_2$, *then*

$$\hat{\xi}_{1,i} + \hat{\xi}_{2,i} = o_{up}(T^{a_1}), \tag{2.B.18}$$

$$\hat{\xi}_{1,i}\cdot\hat{\xi}_{2,i} = o_{up}(T^{a_1+a_2}). \tag{2.B.19}$$

Proof. For the first statement, note that

$$\left|\max_{1\le i\le n}\left(\hat{\xi}_{1,i} + \hat{\xi}_{2,i}\right)\right| \le \max_{1\le i\le n}\left|\hat{\xi}_{1,i}\right| + \max_{1\le i\le n}\left|\hat{\xi}_{1,i}\right| = o_p(T^{a_1}). \tag{2.B.20}$$

The second statement follows similarly by

$$\left|\max_{1\le i\le n}\left(\hat{\xi}_{1,i}\cdot\hat{\xi}_{2,i}\right)\right| \le \max_{1\le i\le n}\left|\hat{\xi}_{1,i}\right|\cdot\max_{1\le i\le n}\left|\hat{\xi}_{1,i}\right| = o_p(T^{a_1})\cdot o_p(T^{a_2}) = o_p(T^{a_1+a_2}). \tag{2.B.21}$$

∎

Lemma 13 *Let* $\{\hat{\xi}_{j,i}, i = 1, 2, \ldots\}$, $j = 1, 2$, *be two sequences of independent random variables such that* $\hat{\xi}_{1,i} = \xi_{1,i} + o_{up}(T^{a_1})$, $\hat{\xi}_{2,i} = \xi_{2,i} + (T^{a_2})$, $|\xi_{j,i}| < \infty$, $j = 1, 2$, *and* $0 \geq a_1 \geq a_2$; *then*

$$\hat{\xi}_{1,i} \cdot \hat{\xi}_{2,i} - \xi_{1,i} \cdot \xi_{2,i} = o_{up}(T^{a_1}). \tag{2.B.22}$$

Proof. Note that

$$\begin{aligned} \left|\hat{\xi}_{1,i} \cdot \hat{\xi}_{2,i} - \xi_{1,i} \cdot \xi_{2,i}\right| &\leq \left|\left(\hat{\xi}_{1,i} - \xi_{1,i}\right) \cdot \left(\hat{\xi}_{2,i} - \xi_{2,i}\right)\right| + \left|(\hat{\xi}_{1,i} - \xi_{1,i}) \cdot \xi_{2,i}\right| + \left|(\hat{\xi}_{2,i} - \xi_{2,i}) \cdot \xi_{1,i}\right| \\ &\leq \left|\hat{\xi}_{1,i} - \xi_{1,i}\right| \cdot \left|\hat{\xi}_{2,i} - \xi_{2,i}\right| + \left|\hat{\xi}_{1,i} - \xi_{1,i}\right| \cdot |\xi_{2,i}| + \left|\hat{\xi}_{2,i} - \xi_{2,i}\right| \cdot |\xi_{1,i}| \\ &= o_{up}(T^{a_1+a_2}) + o_{up}(T^{a_1}) + o_{up}(T^{a_2}) = o_{up}(T^{a_1}), \end{aligned} \tag{2.B.23}$$

by Lemma 12. ■

Lemma 14 *Assume that Conditions 1, 2, 3 and 4 hold. Let* $\hat{t}_i^W(\gamma_i; \theta)$ *denote the first stage GMM score of the fixed effects, that is*

$$\hat{t}_i^W(\gamma_i; \theta) = -\begin{pmatrix} \hat{G}_{\alpha_i}(\theta, \alpha_i)'\lambda_i \\ \hat{g}_i(\theta, \alpha_i) + W_i\lambda_i \end{pmatrix} \tag{2.B.24}$$

where $\gamma_i = (\alpha_i, \lambda_i')'$. *Let* $\hat{T}_i^W(\gamma_i; \theta)$ *denote* $\frac{\partial \hat{t}_i^W(\gamma_i;\theta)}{\partial \gamma_i'}$. *Define* $\tilde{\gamma}_{i0}$ *as the solution to* $\hat{t}_i^W(\gamma_i; \theta_0) = 0$.

Then, for any $\overline{\gamma}_{i0}$ *between* $\tilde{\gamma}_{i0}$ *and* γ_{i0}, *we have*

$$\begin{aligned} \sqrt{T}\hat{t}_i^W(\gamma_{i0}) &= o_{up}\left(T^{\frac{r}{2s}}\right), \end{aligned} \tag{2.B.25}$$

$$\begin{aligned} \hat{T}_i^W(\overline{\gamma}_{i0}) - T_i^W &= o_{up}(1). \end{aligned} \tag{2.B.26}$$

Proof. The results follow by inspection of the score and its derivative (see Appendix 2.K), Lemma 10, Lemma 11 applied to $\theta^* = \theta_0$ and $\alpha_i^* = \alpha_{i0}$, Lemma 11 applied to $\theta^* = \theta_0$ and $\alpha_i^* = \tilde{\alpha}_{i0}$, and Lemmas 12 and 13. ■

Lemma 15 *Suppose that Conditions 1, 2, 3, and 4 hold. We then have*

$$\sqrt{T}(\tilde{\gamma}_{i0} - \gamma_{i0}) = o_{up}(T^{r/2s}). \tag{2.B.27}$$

where $\tilde{\gamma}_{i0}$ *is the solution to* $\hat{t}_i^W(\gamma_i; \theta_0) = 0$.

Proof. By a first order Taylor Expansion of the FOC for $\tilde{\gamma}_{i0}$, we have

$$\begin{aligned} 0 &= \hat{t}_i^W(\tilde{\gamma}_{i0}) = \hat{t}_i^W(\gamma_{i0}) + \hat{T}_i^W(\overline{\gamma}_i)(\tilde{\gamma}_{i0} - \gamma_{i0}) \\ &= \hat{t}_i^W(\gamma_{i0}) + T_i^W(\tilde{\gamma}_{i0} - \gamma_{i0}) + \left(\hat{T}_i^W(\overline{\gamma}_i) - T_i^W\right)(\tilde{\gamma}_{i0} - \gamma_{i0}), \end{aligned} \tag{2.B.28}$$

where $\overline{\gamma}_i$ lies between $\tilde{\gamma}_{i0}$ and γ_{i0}. Next

$$\begin{aligned}
\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) &= -\underbrace{\left(T_i^W\right)^{-1}}_{=O_u(1)}\underbrace{\sqrt{T}\hat{t}_i^W(\gamma_{i0})}_{=o_{up}(T^{r/2s})} - \underbrace{\left(T_i^W\right)^{-1}}_{=O_u(1)}\underbrace{\left(\hat{T}_i^W(\overline{\gamma}_i)-T_i^W\right)}_{=o_{up}(1)}\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) \\
&= o_{up}(T^{r/2s}) + o_{up}\left(\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0})\right), \qquad (2.B.29)
\end{aligned}$$

by Condition 3 and Lemma 14. Therefore

$$(1+o_{up}(1))\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) = o_{up}(T^{r/2s}) \Rightarrow \sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) = o_{up}(T^{r/2s}). \qquad (2.B.30)$$

■

Lemma 16 *Assume that Conditions 1, 2, 3 and 4 hold. Let $\hat{t}_i^W(\gamma_i;\theta)$ denote the first stage GMM score for the fixed effects, that is*

$$\hat{t}_i^W(\gamma_i;\theta) = -\begin{pmatrix} \hat{G}_{\alpha_i}(\theta,\alpha_i)'\lambda_i \\ \hat{g}_i(\theta,\alpha_i)+W_i\lambda_i \end{pmatrix} \qquad (2.B.31)$$

where $\gamma_i = (\alpha_i,\lambda_i')'$. Let $\hat{s}_i^W(\theta,\tilde{\gamma}_i(\theta))$ denote the first stage GMM plug-in score for the common parameter, that is

$$\hat{s}_i^W(\theta,\tilde{\gamma}_i(\theta)) = -\hat{G}_{\theta i}(\theta,\tilde{\alpha}_i(\theta))'\tilde{\lambda}_i(\theta), \qquad (2.B.32)$$

where $\tilde{\gamma}_i(\theta)$ is such that $\hat{t}_i^W(\tilde{\gamma}_i(\theta);\theta)=0$. Let $\hat{T}_{i,d}^W(\gamma_i;\theta)$ denote $\frac{\partial^d \hat{t}_i^W(\gamma_i;\theta)}{\partial\gamma_i^{d'}}$, for some $1\le d\le 4$. Let $\hat{N}_i^W(\gamma_i;\theta)$ denote $\frac{\partial \hat{t}_i^W(\gamma_i;\theta)}{\partial\theta'}$. Let $\hat{M}_{i,d}^W(\theta,\tilde{\gamma}_i)$ denote $\frac{\partial \hat{s}_i^W(\tilde{\gamma}_i,\theta)}{\partial\tilde{\gamma}_i^{d'}}$, for some $1\le d\le 4$. Let $\hat{S}_i^W(\theta,\tilde{\gamma}_i)$ denote $\frac{\partial \hat{s}_i^W(\tilde{\gamma}_i,\theta)}{\partial\theta'}$. Let $(\tilde{\theta},\{\tilde{\gamma}_i\}_{i=1}^n)$ be the first stage GMM estimators.

Then, for any $\overline{\theta}$ between $\tilde{\theta}$ and θ_0, and $\overline{\gamma}_i$ between $\tilde{\gamma}_i$ and γ_{i0}, we have

$$\begin{aligned}
\hat{T}_{i,d}^W(\overline{\theta},\overline{\gamma}_i) - T_{i,d}^W &= o_{up}(1), \qquad (2.B.33)\\
\hat{M}_{i,d}^W(\overline{\theta},\overline{\gamma}_i) - M_{i,d}^W &= o_{up}(1), \qquad (2.B.34)\\
\hat{N}_i^W(\overline{\theta},\overline{\gamma}_i) - N_i^W &= o_{up}(1), \qquad (2.B.35)\\
\hat{S}_i^W(\overline{\theta},\overline{\gamma}_i) - S_i^W &= o_{up}(1). \qquad (2.B.36)
\end{aligned}$$

Proof. The results follow by inspection of the scores and their derivatives (see Appendices 2.K and 2.L), Theorem 1, Theorem 2, Lemma 10, Lemma 11 applied to $\theta^* = \tilde{\theta}$ and $\alpha_i^* = \tilde{\alpha}_i$ with $a=0$, and Lemmas 12 and 13. ■

Lemma 17 *Assume that Conditions 1, 2, 3 and 4 hold. Let $\hat{t}_i^W(\gamma_i;\theta)$ denote the first stage GMM score for the fixed effects, that is*

$$\hat{t}_i^W(\gamma_i;\theta) = -\begin{pmatrix} \hat{G}_{\alpha_i}(\theta,\alpha_i)'\lambda_i \\ \hat{g}_i(\theta,\alpha_i)+W_i\lambda_i \end{pmatrix} \qquad (2.B.37)$$

where $\gamma_i = (\alpha_i, \lambda_i')'$. Let $\hat{s}_i^W(\theta, \tilde{\gamma}_i(\theta))$ denote the first stage GMM plug-in score for the common parameter, that is

$$\hat{s}_i^W(\theta, \tilde{\gamma}_i(\theta)) = -\hat{G}_{\theta i}(\theta, \tilde{\alpha}_i(\theta))'\tilde{\lambda}_i(\theta), \tag{2.B.38}$$

where $\tilde{\gamma}_i(\theta)$ is such that $\hat{t}_i^W(\tilde{\gamma}_i(\theta); \theta) = 0$. Let $\hat{T}_{i,d}^W(\gamma_i; \theta)$ denote $\frac{\partial^d \hat{t}_i^W(\gamma_i;\theta)}{\partial \gamma_i^{d'}}$, for some $1 \leq d \leq 4$. Let $\hat{N}_i^W(\gamma_i; \theta)$ denote $\frac{\partial \hat{t}_i^W(\gamma_i;\theta)}{\partial \theta'}$. Let $\hat{M}_{i,d}^W(\theta, \tilde{\gamma}_i)$ denote $\frac{\partial \hat{s}_i^W(\tilde{\gamma}_i,\theta)}{\partial \tilde{\gamma}_i^{d'}}$, for some $1 \leq d \leq 4$. Let $\hat{S}_i^W(\theta, \tilde{\gamma}_i)$ denote $\frac{\partial \hat{s}_i^W(\tilde{\gamma}_i,\theta)}{\partial \theta'}$. Let $\tilde{\gamma}_{i0}$ denote $\tilde{\gamma}_i(\theta_0)$.

Then, for any $\overline{\gamma}_i$ between $\tilde{\gamma}_{i0}$ and γ_{i0}, we have

$$\sqrt{T}\hat{t}_i^W(\overline{\gamma}_i) = o_{up}\left(T^{\frac{r}{2s}}\right), \tag{2.B.39}$$

$$\sqrt{T}\left(\hat{T}_{i,d}^W(\overline{\gamma}_i) - T_{i,d}^W\right) = o_{up}\left(T^{\frac{r}{2s}}\right), \tag{2.B.40}$$

$$\sqrt{T}\left(\hat{M}_{i,d}^W(\overline{\gamma}_i) - M_{i,d}^W\right) = o_{up}\left(T^{\frac{r}{2s}}\right), \tag{2.B.41}$$

Proof. The results follow by inspection of the scores and their derivatives (see Appendices 2.K and 2.L), Theorem 1, Theorem 2, Lemma 15, Lemma 11 applied to $\theta^* = \theta_0$ and $\alpha_i^* = \tilde{\alpha}_{i0}$ with $a = T^{\frac{1}{2}(1-r/s)}$, and Lemmas 12 and 13. ■

2.B.2 Proof of Theorem 3

Proof. From a Taylor Expansion of the FOC for $\tilde{\theta}$, we have

$$0 = s_n^W(\tilde{\theta}) = s_n^W(\theta_0) + \frac{ds_n^W(\overline{\theta})}{d\theta'}(\tilde{\theta} - \theta_0), \tag{2.B.42}$$

where $\overline{\theta}$ lies between $\tilde{\theta}$ and θ_0.

Part I: Asymptotic limit of $\frac{ds_n^W(\overline{\theta})}{d\theta'}$. Note that

$$\frac{ds_n^W(\overline{\theta},)}{d\theta'} = \frac{1}{n}\sum_{i=1}^{n}\frac{ds_i^W(\overline{\theta}, \tilde{\gamma}_i(\overline{\theta}))}{d\theta'}, \tag{2.B.43}$$

$$\frac{ds_i^W(\overline{\theta}, \tilde{\gamma}_i(\overline{\theta}))}{d\theta'} = \frac{\partial s_i^W(\overline{\theta}, \tilde{\gamma}_i(\overline{\theta}))}{\partial \theta'} + \frac{\partial s_i^W(\overline{\theta}, \tilde{\gamma}_i(\overline{\theta}))}{\partial \tilde{\gamma}_i'}\frac{\partial \tilde{\gamma}_i(\overline{\theta})}{\theta'}. \tag{2.B.44}$$

From Lemma 16, we have

$$\frac{\partial s_i^W(\overline{\theta}, \tilde{\gamma}_i(\overline{\theta}))}{\partial \theta'} = S_i^W + o_{up}(1) = o_{up}(1), \tag{2.B.45}$$

$$\frac{\partial s_i^W(\overline{\theta}, \tilde{\gamma}_i(\overline{\theta}))}{\partial \tilde{\gamma}_i'} = M_i^W + o_{up}(1). \tag{2.B.46}$$

Then, differentiation of the FOC for $\tilde{\gamma}_i$, $\hat{t}_i^W(\tilde{\gamma}_i(\overline{\theta}); \overline{\theta}) = 0$, with respect to θ and $\tilde{\gamma}_i$ gives

$$\hat{T}_i^W(\tilde{\gamma}_i(\overline{\theta}); \overline{\theta})\frac{\partial \tilde{\gamma}_i(\overline{\theta})}{\theta'} + \hat{N}_i^W(\tilde{\gamma}_i(\overline{\theta}); \overline{\theta}) = 0, \tag{2.B.47}$$

By repeated application of Lemma 16 and Condition 3, we can write

$$\frac{\partial \tilde{\gamma}_i(\overline{\theta})}{\theta'} = -\left(T_i^W\right)^{-1} N_i^W + o_{up}(1). \tag{2.B.48}$$

Finally, replacing the expressions for the components in (2.B.44) and using the formulae for the derivatives from Appendices 2.K and 2.L, we have

$$\begin{aligned} \frac{ds_n^W(\overline{\theta})}{d\theta'} &= J_{sn}^W + o_p(1) = \frac{1}{n}\sum_{i=1}^{n} G'_{\theta_i} P_{\alpha_i}^W G_{\theta_i} + o_p(1) = J_s^W + o_p(1), && \text{(2.B.49)} \\ J_s^W &= \lim_{n\to\infty} J_{sn}^W = O(1). && \text{(2.B.50)} \end{aligned}$$

Part II: Asymptotic Expansion for $\tilde{\theta} - \theta_0$. For the case $n = O(T)$, from (2.B.50) and Lemma 46 we have

$$0 = \underbrace{\sqrt{nT} s_n^W(\theta_0)}_{O_p(1)} + \underbrace{\frac{ds_n^W(\overline{\theta})}{d\theta'}}_{O(1)} \sqrt{nT}(\tilde{\theta} - \theta_0). \tag{2.B.51}$$

Therefore, $\sqrt{nT}(\tilde{\theta} - \theta_0) = O_p(1)$. Then the result follows by using again (2.B.50) and Lemma 46.

Similarly for the general case $T = o(n)$, from (2.B.50) and Lemma 46 we have

$$0 = \underbrace{T s_n^W(\theta_0)}_{O(1)} + \underbrace{\frac{ds_n^W(\overline{\theta})}{d\theta'}}_{O(1)} T(\tilde{\theta} - \theta_0). \tag{2.B.52}$$

Therefore, $T(\tilde{\theta} - \theta_0) = O(1)$. Then the result also follows by (2.B.50) and Lemma 46. ■

Corollary 3 *Under Conditions 1, 2, 3 and 4, we have*

$$T(\tilde{\theta} - \theta_0) = Q_{1\theta}^W + \frac{a_T}{T^{(r-1)/2}} \psi_\theta^W + \frac{1}{T} R_{2\theta}^W, \tag{2.B.53}$$

where

$$\begin{aligned} a_T &= \begin{cases} C, & \text{if } n = O(T^a) \text{ for some } a \in \Re; \\ o(T^\epsilon) \text{ for any } \epsilon > 0, & \text{otherwise,} \end{cases} && \text{(2.B.54)} \\ Q_{1\theta}^W &= -(J_s^W)^{-1} \frac{1}{n} \sum_{i=1}^{n} Q_{1si}^W, && \text{(2.B.55)} \\ \psi_\theta^W &= -(J_s^W)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\psi}_{si}^W, && \text{(2.B.56)} \\ R_{2\theta}^W &= o_{up}(\sqrt{T}). && \text{(2.B.57)} \end{aligned}$$

Proof. The result follows by using the expansion of $T\hat{s}_n^W(\theta_0)$ in the proof of Lemma 46. ■

2.C Consistency of Second Stage GMM Estimator

2.C.1 Some Lemmas

Lemma 18 *Assume that Conditions 1, 2, 3 and 5 hold. Then, for any $\eta > 0$, we have*

$$\Pr\left\{\max_{1\leq i\leq n}\left|\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i) - \Omega_i(\theta_0,\alpha_{i0})\right| \geq \eta\right\} = o(T^{r-s}), \tag{2.C.1}$$

where

$$\hat{\Omega}_i(\theta,\alpha_i) = \frac{1}{T}\sum_{t=1}^{T} g(z_{it};\theta,\alpha)g(z_{it};\theta,\alpha)', \tag{2.C.2}$$

$$\Omega_i(\theta,\alpha_i) = E\left[\hat{\Omega}_i(\theta,\alpha_i)\right]. \tag{2.C.3}$$

Proof. By Triangle inequality and Conditions 2 and 5, we can write

$$\begin{aligned}\left|\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i) - \Omega_i(\theta_0,\alpha_{i0})\right| &\leq \left|\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i) - \Omega_i(\tilde{\theta},\tilde{\alpha}_i)\right| + \left|\Omega_i(\tilde{\theta},\tilde{\alpha}_i) - \Omega_i(\theta_0,\alpha_{i0})\right| \\ &\leq \left|\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i) - \Omega_i(\tilde{\theta},\tilde{\alpha}_i)\right| + m^2 E\left[M(z_{it})^2\right]\left|(\tilde{\theta},\tilde{\alpha}_i) - (\theta_0,\alpha_{i0})\right|. \end{aligned}\tag{2.C.4}$$

Therefore, we have

$$\begin{aligned}\Pr\left\{\max_{1\leq i\leq n}\left|\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i) - \Omega_i(\theta_0,\alpha_{i0})\right| \geq \eta\right\} &\leq \Pr\left\{\max_{1\leq i\leq n}\left|\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i) - \Omega_i(\tilde{\theta},\tilde{\alpha}_i)\right| \geq \eta/2\right\} \\ &+ \Pr\left\{\left|(\tilde{\theta},\tilde{\alpha}_i) - (\theta_0,\alpha_{i0})\right| \geq \frac{\eta}{2m^2\max_i E\left[M(z_{it})^2\right]}\right\}. \end{aligned}\tag{2.C.5}$$

Then, the result follows by Lemma 6 (applied to $\max_{1\leq k,l\leq m}|g_k g_l - E[g_k g_l]|$) and Theorems 1 and 2. ■

Lemma 19 *Let $\{\hat{\xi}_i, i = 1,2,\ldots\}$, be a sequence of independent random variables such that $\hat{\xi}_i = \xi_i + o_{up}(T^a)$, with $a \leq 0$ and $\min_i \xi_i > 0$, then*

$$\hat{\xi}_i^{-1} = \xi_i^{-1} + o_{up}(T^a). \tag{2.C.6}$$

Proof. By Mean Value Theorem

$$\hat{\xi}_i^{-1} = \xi_i^{-1} - \left(\overline{\xi}_i\right)^{-1}\left(\hat{\xi}_i - \xi_i\right)\left(\overline{\xi}_i\right)^{-1} = \xi_i^{-1} + o_{up}(T^a), \tag{2.C.7}$$

where $\overline{\xi}_i$ lies between $\hat{\xi}_i$ and ξ_i, and is non-singular with probability $1 - o_{up}(T^a)$. ■

Lemma 20 *Assume that Conditions 1, 2, 3 and 5 hold. Then, for any $\eta > 0$, we have*

$$\Pr\left\{\max_{1\leq i\leq n}\left|\sup_{(\theta,\alpha)\in\Upsilon}\hat{Q}_i^{\Omega}(\theta,\alpha) - Q_i^{\Omega}(\theta,\alpha)\right| \geq \eta\right\} = o(T^{r-s}), \tag{2.C.8}$$

where

$$
\begin{aligned}
\hat{Q}_i^{\Omega}(\theta,\alpha_i) &= -\hat{g}_i(\theta,\alpha)'\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)^{-1}\hat{g}_i(\theta,\alpha), & (2.C.9)\\
Q_i^{\Omega}(\theta,\alpha_i) &= -g_i(\theta,\alpha)'\Omega_i^{-1}g_i(\theta,\alpha), & (2.C.10)\\
g_i(\theta,\alpha) &= E\left[\hat{g}_i(\theta,\alpha)\right]. & (2.C.11)
\end{aligned}
$$

Proof. First, note that with probability $1-o(T^{r-s})$

$$
\begin{aligned}
\left|\hat{Q}_i^{\Omega}(\theta,\alpha)-Q_i^{\Omega}(\theta,\alpha)\right| &\leq \left|(\hat{g}_i(\theta,\alpha)-g_i(\theta,\alpha))'\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)^{-1}(\hat{g}_i(\theta,\alpha)-g_i(\theta,\alpha))\right|\\
&+ 2\cdot\left|g_i(\theta,\alpha)'\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)^{-1}(\hat{g}_i(\theta,\alpha)-g_i(\theta,\alpha))\right|\\
&+ \left|g_i(\theta,\alpha_i)'\left(\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)^{-1}-\Omega_i^{-1}\right)g_i(\theta,\alpha_i)\right|\\
&\leq m^2\max_{1\leq k\leq m}\left(\sup_{(\theta,\alpha)\in\Upsilon}\left|(\hat{g}_{k,i}(\theta,\alpha)-g_{k,i}(\theta,\alpha))\right|\right)^2\left|\Omega_i\right|^{-1}\\
&+ 2m^2\max_{1\leq i\leq n}E[M(z_{it})]\left|\Omega_i\right|^{-1}\max_{1\leq k\leq m}\left(\sup_{(\theta,\alpha)\in\Upsilon}\left|(\hat{g}_{k,i}(\theta,\alpha)-g_{k,i}(\theta,\alpha))\right|\right)\\
&+ m^2\cdot\max_{1\leq i\leq n}E\left[M(z_{it})\right]^2\left|\hat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)^{-1}-\Omega_i^{-1}\right|. \qquad (2.C.12)
\end{aligned}
$$

Then, the conclusion follows Condition 5, and Lemmas 6 and 19. ■

Lemma 21 *Suppose that Conditions 1, 2, 3 and 5 hold. Then, with probability* $1-o(1)$

$$
\sup_{\alpha}\left|Q_i^{\Omega}(\theta,\alpha)-Q_i^{\Omega}(\theta',\alpha)\right|\leq C\cdot E[M(z_{it})]^2\left|\theta-\theta'\right|, \qquad (2.C.13)
$$

for some constant $C>0$, *where*

$$
\begin{aligned}
Q_i^{\Omega}(\theta,\alpha_i) &= -g_i(\theta,\alpha)'\Omega_i^{-1}g_i(\theta,\alpha), & (2.C.14)\\
g_i(\theta,\alpha) &= E\left[\hat{g}_i(\theta,\alpha)\right]. & (2.C.15)
\end{aligned}
$$

Proof. Note that

$$
\begin{aligned}
\left|Q_i^{\Omega}(\theta,\alpha)-Q_i^{\Omega}(\theta',\alpha)\right| &\leq \left|g_i(\theta,\alpha)'\Omega_i^{-1}(g_i(\theta,\alpha)-g_i(\theta',\alpha))\right|\\
&+ \left|(g_i(\theta,\alpha)-g_i(\theta',\alpha))'\Omega_i^{-1}g_i(\theta',\alpha)\right|\\
&\leq 2\cdot m^2E[M(z_{it})]^2\left|\Omega_i\right|^{-1}\left|\theta-\theta'\right|. \qquad (2.C.16)
\end{aligned}
$$

■

2.C.2 Proof of Theorem 4

Proof. Let η be given, and let $\varepsilon \equiv \inf_i \left[Q_i^{\Omega}(\theta_0, \alpha_{i0}) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}} Q_i^{\Omega}(\theta, \alpha)\right] > 0$ with probability $1 - o(T^{r-s})$ by Conditions 2 and 5. With probability equal to $1 - o(T^{r-s})$, we have

$$\begin{aligned} \max_{|\theta-\theta_0|>\eta,\alpha_1,\ldots,\alpha_n} n^{-1}\sum_{i=1}^{n} \hat{Q}_i^{\Omega}(\theta, \alpha_i) &\le \max_{|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta} n^{-1}\sum_{i=1}^{n} \hat{Q}_i^{\Omega}(\theta, \alpha_i) \\ &< \max_{|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta} n^{-1}\sum_{i=1}^{n} Q_i^{\Omega}(\theta, \alpha_i) + \frac{1}{3}\varepsilon \\ &< n^{-1}\sum_{i=1}^{n} Q_i^{\Omega}(\theta_0, \alpha_{i0}) - \frac{2}{3}\varepsilon \\ &< n^{-1}\sum_{i=1}^{n} \hat{Q}_i^{\Omega}(\theta_0, \alpha_{i0}) - \frac{1}{3}\varepsilon, \end{aligned} \tag{2.C.17}$$

where the second and fourth inequalities follow from Lemma 20. Because

$$\max_{\theta,\alpha_1,\ldots,\alpha_n} n^{-1}\sum_{i=1}^{n} \hat{Q}_i^{\Omega}(\theta, \alpha_i) \ge n^{-1}\sum_{i=1}^{n} \hat{Q}_i^{\Omega}(\theta_0, \alpha_{i0}) \tag{2.C.18}$$

by definition, we can conclude that $\Pr\left[\left|\hat{\theta} - \theta_0\right| \ge \eta\right] = o(T^{r-s})$. ■

2.C.3 Proof of Theorem 5

Proof. Part I: Consistency of $\hat{\alpha}_i$. We first prove that

$$\Pr\left[\max_{1\le i\le n} \sup_{\alpha} \left|\hat{Q}_i^{\Omega}\left(\hat{\theta}, \alpha\right) - Q_i^{\Omega}(\theta_0, \alpha)\right| \ge \eta\right] = o\left(T^{r-s}\right) \tag{2.C.19}$$

for every $\eta > 0$. Note that

$$\begin{aligned} &\max_{1\le i\le n} \sup_{\alpha} \left|\hat{Q}_i^{\Omega}\left(\hat{\theta}, \alpha\right) - Q_i^{\Omega}(\theta_0, \alpha)\right| \\ \le\ &\max_{1\le i\le n} \sup_{\alpha} \left|\hat{Q}_i^{\Omega}\left(\hat{\theta}, \alpha\right) - Q_i^{\Omega}\left(\hat{\theta}, \alpha\right)\right| + \max_{1\le i\le n} \sup_{\alpha} \left|Q_i^{\Omega}\left(\hat{\theta}, \alpha\right) - Q_i^{\Omega}(\theta_0, \alpha)\right| \\ \le\ &\max_{1\le i\le n} \sup_{(\theta,\alpha)} \left|\hat{Q}_i^{\Omega}(\theta, \alpha) - Q_i^{\Omega}(\theta, \alpha)\right| + \max_{1\le i\le n} C \cdot E\left[M(z_{it})\right]^2 \cdot \left|\hat{\theta} - \theta_0\right|, \end{aligned} \tag{2.C.20}$$

by Lemma 21. Therefore,

$$\begin{aligned} \Pr\left[\max_{1\le i\le n} \sup_{\alpha} \left|\hat{Q}_i^{\Omega}\left(\hat{\theta}, \alpha\right) - Q_i^{\Omega}(\theta_0, \alpha)\right| \ge \eta\right] &\le \Pr\left[\max_{1\le i\le n} \sup_{(\theta,\alpha)} \left|\hat{Q}_i^{\Omega}(\theta, \alpha) - Q_i^{\Omega}(\theta, \alpha)\right| \ge \frac{\eta}{2}\right] \\ &\quad + \Pr\left[\left|\hat{\theta} - \theta_0\right| \ge \frac{\eta}{2 \cdot C\left(\max_{1\le i\le n} E\left[M(z_{it})\right]^2\right)}\right] \\ &= o\left(T^{r-s}\right), \end{aligned} \tag{2.C.21}$$

by Lemma 20 and Theorem 4.

We now get back to the proof of Theorem 5. Let η be given, and let

$$\varepsilon \equiv \inf_i \left[Q_i^{\Omega}\left(\theta_0, \alpha_{i0}\right) - \sup_{\{\alpha_i : |\alpha_i - \alpha_{i0}| > \eta\}} Q_i^{\Omega}\left(\theta_0, \alpha_i\right) \right] > 0 \tag{2.C.22}$$

by Conditions 2 and 5. Condition on the event

$$\left\{ \max_{1 \leq i \leq n} \sup_{\alpha} \left| \hat{Q}_i^{\Omega}\left(\hat{\theta}, \alpha\right) - Q_i^{\Omega}\left(\theta_0, \alpha\right) \right| \leq \frac{1}{3}\varepsilon \right\}, \tag{2.C.23}$$

which has a probability equal to $1 - o\left(T^{r-s}\right)$ by (2.C.19). We then have

$$\max_{|\alpha_i - \alpha_{i0}| > \eta} \hat{Q}_i^{\Omega}\left(\hat{\theta}, \alpha_i\right) < \max_{|\alpha_i - \alpha_{i0}| > \eta} Q_i^{\Omega}\left(\theta_0, \alpha_i\right) + \frac{1}{3}\varepsilon < Q_i^{\Omega}\left(\theta_0, \alpha_{i0}\right) - \frac{2}{3}\varepsilon < \hat{Q}_i^{\Omega}\left(\hat{\theta}, \alpha_{i0}\right) - \frac{1}{3}\varepsilon, \tag{2.C.24}$$

This is inconsistent with $\hat{Q}_i^{\Omega}\left(\hat{\theta}, \hat{\alpha}_i\right) \geq \hat{Q}_i^{\Omega}\left(\hat{\theta}, \alpha_{i0}\right)$, and therefore, $|\hat{\alpha}_i - \alpha_{i0}| \leq \eta$ for every i.

Part II: Consistency of $\hat{\lambda}_i$. First, note that with probability $1 - o\left(T^{r-s}\right)$

$$\begin{aligned}
\left|\hat{\lambda}_i\right| &\leq |\Omega_i|^{-1} m \max_k \left(\left|\hat{g}_{k,i}(\hat{\theta}, \hat{\alpha}_i) - g_{k,i}(\hat{\theta}, \hat{\alpha}_i)\right| + \left|g_{k,i}(\hat{\theta}, \hat{\alpha}_i) - g_{k,i}(\theta_0, \alpha_{i0})\right| \right) \\
&\leq |\Omega_i|^{-1} m \max_k \sup_{(\theta, \alpha_i) \in \Upsilon} \left|\hat{g}_{k,i}(\theta, \alpha_i) - g_{k,i}(\theta, \alpha_i)\right| \\
&+ |\Omega_i|^{-1} M(z_{it}) m \left|\overline{\theta} - \theta_0\right| + |\Omega_i|^{-1} M(z_{it}) m \left|\overline{\alpha}_i - \alpha_{i0}\right|,
\end{aligned} \tag{2.C.25}$$

where $(\overline{\theta}, \overline{\alpha}_i)$ are between (θ_0, α_{i0}) and $(\hat{\theta}, \hat{\alpha}_i)$. Then, the results follow from Lemma 6, Theorem 4 and the first statement of Theorem 5. ∎

Corollary 4 *Suppose that Conditions 1, 2, 3 and 5 hold. Then,*

$$\Pr\left\{ \max_{1 \leq i \leq n} |\hat{\alpha}_{i0} - \alpha_{i0}| \geq \eta \right\} = o(T^{r-s}) = o(1), \tag{2.C.26}$$

for any $\eta > 0$, where

$$\hat{\alpha}_{i0} \in \arg\max_{\alpha} \hat{Q}_i^{\Omega}(\theta_0, \alpha). \tag{2.C.27}$$

Also,

$$\Pr\left\{ \max_{1 \leq i \leq n} \left|\hat{\lambda}_{i0}\right| \geq \eta \right\} = o(T^{r-s}) = o(1), \tag{2.C.28}$$

for any $\eta > 0$, where

$$\hat{\lambda}_{i0} = -W_i^{-1} \hat{g}_i(\theta_0, \hat{\alpha}_{i0}). \tag{2.C.29}$$

Proof. The arguments in the proof of Theorem 4 go through replacing $\hat{\theta}$ for θ_0. ∎

2.D Asymptotic Distribution of the Second Stage GMM Estimator

2.D.1 Some Lemmas

Lemma 22 *Assume that Conditions 1, 2, 3, 4 and 5 hold. We then have*

$$\sqrt{T}(\tilde{\gamma}_i - \gamma_{i0}) = o_{up}(T^{r/2s}). \tag{2.D.1}$$

where $\tilde{\gamma}_i$ is the solution to $\hat{t}_i^W(\gamma_i;\tilde{\theta}) = 0$, i.e. the first stage estimator for the fixed effects.

Proof. We show that

$$\sqrt{T}(\tilde{\gamma}_i - \tilde{\gamma}_{i0}) = o_{up}(1), \tag{2.D.2}$$

and then the result follows by Lemma 15.

Note that

$$\sqrt{T}(\tilde{\gamma}_i - \tilde{\gamma}_{i0}) = \frac{\partial \gamma_i(\overline{\theta})}{\partial \theta'} \sqrt{T}(\tilde{\theta} - \theta_0), \tag{2.D.3}$$

where $\overline{\theta}$ lies between $\tilde{\theta}$ and θ_0. Following an analogous argument as in the proof of Theorem 3, we have

$$\sqrt{T}(\tilde{\gamma}_i - \tilde{\gamma}_{i0}) = -\underbrace{\left(T_i^W\right)^{-1} N_i^W}_{=O_u(1)} \underbrace{\sqrt{T}(\tilde{\theta} - \theta_0)}_{=o_{up}(1)} + o_{up}\left(\sqrt{T}(\tilde{\theta} - \theta_0)\right) = o_{up}(1). \tag{2.D.4}$$

■

Lemma 23 *Suppose that, for each i, $\{z_{it}, t = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables. We assume that $\{z_{it}, t = 1, 2, \ldots\}$ are independent across i. Let $h_j(z_{it};\theta,\alpha_i)$, $j = 1, 2$ be two functions such that (i) $h_j(z_{it};\theta,\alpha_i)$ is five times continuously differentiable in $(\theta,\alpha_i) \in \Upsilon$; (ii) Υ is compact; (iii) there exists a function $M(z_{it})$ such that $|h_j(z_{it};\theta,\alpha_i)| \leq M(z_{it})$ and $\left|\frac{\partial^{d_1+d_2} h_j(z_{it};\theta,\alpha_i)}{\partial \theta^{d_1} \alpha_i^{d_2}}\right| \leq M(z_{it})$ for $0 \leq d_1 + d_2 \leq 5$, with $E\left[M(z_{it})^{4s}\right] < \infty$. Let $f(z_{it};\theta,\alpha_i)$ denote $h_1(z_{it};\theta,\alpha_i) \cdot h_2(z_{it};\theta,\alpha_i)$. Let $F_i(z_{it};\theta,\alpha_i)$ denote*

$$\frac{\partial^{d_1+d_2} f(z_{it};\theta,\alpha_i)}{\partial \theta^{d_1} \alpha_i^{d_2}}, \tag{2.D.5}$$

$\hat{F}_i(\theta,\alpha_i)$ denote

$$\frac{1}{T}\sum_{t=1}^{T} F_i(z_{it};\theta,\alpha_i), \tag{2.D.6}$$

and $F_i(\theta,\alpha_i)$ denote

$$E\left[\hat{F}_i(\theta,\alpha_i)\right], \tag{2.D.7}$$

for some $0 \leq d_1 + d_2 \leq 5$. *Let* α_i^* *be defined as*

$$\alpha_i^* = \arg\sup_{\alpha} \hat{Q}_i^W(\theta^*, \alpha), \tag{2.D.8}$$

such that $\alpha_i^* - \alpha_{i0} = o_{up}(T^{a_\alpha})$ *and* $\theta^* - \theta_0 = o_p(T^{a_\theta})$, *with* $0 \geq a = \min(a_\alpha, a_\theta)$.

Assume also that $n = o(T^r)$, *with* $s \geq r$. *Then, for any* $\overline{\theta}$ *between* θ^* *and* θ_0, $\overline{\alpha}_i$ *between* α_i^* *and* α_{i0}, *and* $\eta > 0$, *we have*

$$\Pr\left[\max_{1\leq i\leq n}\left|\hat{F}_i(\overline{\theta}, \overline{\alpha}_i) - F_i(\theta_0, \alpha_{i0})\right| > \eta T^a\right] = o(1), \tag{2.D.9}$$

and

$$\Pr\left[\max_{1\leq i\leq n}\sqrt{T}\left|\hat{F}_i(\overline{\theta}, \overline{\alpha}_i) - F_i(\overline{\theta}, \overline{\alpha}_i)\right| > \eta T^{\frac{r}{2s}}\right] = o(1). \tag{2.D.10}$$

Proof. Same as for Lemma 11, replacing H_i for F_i, and $M(z_{it})$ for $M(z_{it})^2$.

■

Lemma 24 *Assume that Conditions 1, 2, 3, 4 and 5 hold. Let* $\hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i)$ *denote the estimator of the weighting functions*

$$\frac{1}{T}\sum_{t=1}^{T} g(z_{it}; \tilde{\theta}, \tilde{\alpha}_i) g(z_{it}; \tilde{\theta}, \tilde{\alpha}_i)' \quad i = 1, ..., n, \tag{2.D.11}$$

where $(\tilde{\theta}', \tilde{\alpha}_i)'$ *are the first stage GMM estimators. Let* $\hat{\Omega}_{\alpha^{d_1}\theta^{d_2}i}(\tilde{\theta}, \tilde{\alpha}_i)$ *denote its derivatives*

$$\frac{\partial^{d_1+d_2}\hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i)}{\partial^{d_1}\alpha_i \partial^{d_2}\theta}, \tag{2.D.12}$$

for $0 \leq d_1 + d_2 \leq 3$ *Then, we have*

$$\sqrt{T}\left(\hat{\Omega}_{\alpha^{d_1}\theta^{d_2}\ i}(\tilde{\theta}, \tilde{\alpha}_i) - \Omega_{\alpha^{d_1}\theta^{d_2}\ i}\right) = o_{up}\left(T^{\frac{r}{2s}}\right). \tag{2.D.13}$$

Proof. Note that

$$\begin{aligned}&\left|g(z_{it}; \tilde{\theta}, \tilde{\alpha}_i) g(z_{it}; \tilde{\theta}, \tilde{\alpha}_i)' - E\left[g(z_{it}; \tilde{\theta}, \tilde{\alpha}_i) g(z_{it}; \tilde{\theta}, \tilde{\alpha}_i)'\right]\right| \\ \leq\ & m^2 \max_{1\leq k,l\leq m}\left|g_k(z_{it}; \tilde{\theta}, \tilde{\alpha}_i) g_l(z_{it}; \tilde{\theta}, \tilde{\alpha}_i)' - E\left[g_k(z_{it}; \tilde{\theta}, \tilde{\alpha}_i) g_l(z_{it}; \tilde{\theta}, \tilde{\alpha}_i)'\right]\right|\end{aligned} \tag{2.D.14}$$

Then, by Theorem 3 and Lemma 22, we can use Lemma 23 for $f = g_k g_l - E[g_k g_l']$ with $a = o\left(T^{\frac{1}{2}(1-r/s)}\right)$. A similar argument applies for the derivatives, since they are sums of products of elements that satisfy the assumptions of the Lemma 23. ■

Lemma 25 *Assume that Conditions 1, 2, 3, 4 and 5 hold. We then have*

$$\hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i) = \Omega_i + \frac{1}{\sqrt{T}}\tilde{\psi}_{\Omega i}^W + \frac{1}{T}Q_{1\Omega i}^W + \frac{a_T}{T^{(r+1)/2}}Q_{1r\Omega i}^W + \frac{1}{T^{3/2}}Q_{2\Omega i}^W + \frac{a_T}{T^{(r+2)/2}}Q_{2r\Omega i}^W + \frac{1}{T^2}R_{3\Omega i}^W, \tag{2.D.15}$$

where

$$
\begin{aligned}
a_T &= \begin{cases} C, & \textit{if } n = O(T^a) \textit{ for some } a \in \Re; \\ o(T^\epsilon) \textit{ for any } \epsilon > 0, & \textit{otherwise,} \end{cases} && \text{(2.D.16)} \\
\hat{\Omega}_i(\theta, \alpha_i) &= \frac{1}{T}\sum_{t=1}^{T} g(z_{it}; \theta, \alpha) g(z_{it}; \theta, \alpha)', && \text{(2.D.17)} \\
\Omega_i(\theta, \alpha_i) &= E\left[\hat{\Omega}_i(\theta, \alpha_i)\right] && \text{(2.D.18)} \\
\tilde{\psi}^W_{\Omega i} &= \sqrt{T}\left(\hat{\Omega}_i - \Omega_i\right) + \Omega_{\alpha_i}\tilde{\psi}^{W'}_i e_1 = o_{up}(T^{r/2s}), && \text{(2.D.19)} \\
Q^W_{1\Omega i} &= \Omega_{\alpha_i} Q^{W'}_{1i} e_1 + \sqrt{T}(\hat{\Omega}_{\alpha_i} - \Omega_{\alpha_i})\tilde{\psi}^{W'}_i e_1 + \sum_{j=1}^{p} \Omega_{\theta_j i} Q^{W'}_{1\theta} e_j + \frac{1}{2}\Omega_{\alpha\alpha_i}\left(\tilde{\psi}^{W'}_i e_1\right)^2 \\
&= o_{up}(T^{r/s}), && \text{(2.D.20)} \\
Q^W_{1r\Omega i} &= \sum_{j=1}^{p} \Omega_{\theta_j i}\psi^{W'}_\theta e_j = O_{up}(1), && \text{(2.D.21)} \\
Q^W_{2\Omega i} &= \Omega_{\alpha_i} Q^{W'}_{2i} e_1 + \sqrt{T}(\hat{\Omega}_{\alpha_i} - \Omega_{\alpha_i}) Q^{W'}_{1i} e_1 + \sum_{j=1}^{p} \sqrt{T}\left(\hat{\Omega}_{\theta_j i} - \Omega_{\theta_j i}\right) Q^{W'}_{1\theta} e_j \\
&+ \frac{1}{2}\Omega_{\alpha\alpha_i}\left(\tilde{\psi}^{W'}_i e_1\right)^2 + \Omega_{\alpha\alpha_i}\tilde{\psi}^{W'}_i e_1 Q^{W'}_{1i} e_1 + \frac{1}{2}\sqrt{T}\left(\hat{\Omega}_{\alpha\alpha_i} - \Omega_{\alpha\alpha_i}\right)\left(\tilde{\psi}^{W'}_i e_1\right)^2 \\
&+ \sum_{j=1}^{p} Q^{W'}_{1\theta} e_j \Omega_{\alpha\theta_j i}\tilde{\psi}^{W'}_i e_1 + \frac{1}{6}\Omega_{\alpha\alpha\alpha_i}\left(\tilde{\psi}^{W'}_i e_1\right)^3 = O_{up}(T^{3r/2s}), && \text{(2.D.22)} \\
Q^W_{2r\Omega i} &= \sum_{j=1}^{p} \sqrt{T}\left(\hat{\Omega}_{\theta_j i} - \Omega_{\theta_j i}\right)\psi^{W'}_\theta e_j + \sum_{j=1}^{p} \psi^{W'}_\theta e_j \tilde{\psi}^{W'}_i e_1 = o_{up}(T^{r/2s}), && \text{(2.D.23)} \\
R^W_{2\Omega i} &= o_{up}(T^{2r/s}) = o_{up}(\sqrt{T}). && \text{(2.D.24)}
\end{aligned}
$$

Proof. By a standard Taylor expansion around (θ_0, α_{i0}), we have

$$
\begin{aligned}
\hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i) &= \hat{\Omega}_i + \hat{\Omega}_{\alpha_i}(\tilde{\alpha}_i - \alpha_{i0}) + \sum_{j=1}^{p}\hat{\Omega}_{\theta_j}(\tilde{\theta}_j - \theta_{0,j}) + \frac{1}{2}\hat{\Omega}_{\alpha\alpha_i}(\tilde{\alpha}_i - \alpha_{i0})^2 \\
&+ \sum_{j=1}^{p}\hat{\Omega}_{\alpha\theta_j i}(\overline{\theta}, \overline{\alpha}_i)(\tilde{\theta}_j - \theta_{0,j})(\tilde{\alpha}_i - \alpha_{i0}) + \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p}\hat{\Omega}_{\theta_k\theta_j i}(\overline{\theta}, \overline{\alpha}_i)(\tilde{\theta}_j - \theta_{0,j})(\tilde{\theta}_k - \theta_{0,k}) \\
&+ \frac{1}{6}\Omega_{\alpha\alpha\alpha_i}(\overline{\theta}, \overline{\alpha}_i)(\tilde{\alpha}_i - \alpha_{i0})^3 + \frac{1}{6}\sum_{j=1}^{p}\hat{\Omega}_{\alpha\alpha\theta_j i}(\overline{\theta}, \overline{\alpha}_i)(\tilde{\theta}_j - \theta_{0,j})(\tilde{\alpha}_i - \alpha_{i0})^2, && \text{(2.D.25)}
\end{aligned}
$$

where $(\overline{\theta}, \overline{\alpha}_i)$ lies between $(\tilde{\theta}, \tilde{\alpha}_i)$ and (θ_0, α_{0i}). The expressions for $\tilde{\psi}^W_{\Omega i}$, $Q^W_{1\Omega i}$, $Q^W_{1r\Omega i}$, $Q^W_{2\Omega i}$, and $Q^W_{2r\Omega i}$ can be obtained using the expansions for $\tilde{\gamma}_{i0}$ in Lemma 42 and for $\tilde{\theta}$ in Corollary 3, after some algebra. The rest of the properties for these terms follow by the properties of their components and Lemmas 12, 58, 59, 60, and 61. For the remainder term, we have

$$
\begin{aligned}
R^W_{3\Omega i} &= \Omega_{\alpha_i} R^{W\prime}_{3i} e_1 + \sqrt{T}(\hat{\Omega}_{\alpha_i} - \Omega_{\alpha_i}) R^{W\prime}_{2i} e_1 + \sum_{j=1}^{p} \Omega_{\theta_j} R^{W\prime}_{2\theta} e_j + \sum_{j=1}^{p} \sqrt{T}\left(\hat{\Omega}_{\theta_j i} - \Omega_{\theta_j i}\right) R^{W\prime}_{2\theta} e_j \\
&+ \frac{1}{2}\Omega_{\alpha\alpha_i}\left[\tilde{\psi}^{W\prime}_i e_1 R^{W\prime}_{2i} e_1 + Q^{W\prime}_{1i} e_1 R^{W\prime}_{1i} e_1 + R^{W\prime}_{2i} e_1 \sqrt{T}\left(\tilde{\alpha}_i - \alpha_{i0}\right)\right] \\
&+ \frac{1}{2}\sqrt{T}(\hat{\Omega}_{\alpha\alpha_i} - \Omega_{\alpha\alpha_i})\left[\tilde{\psi}^{W\prime}_i e_1 R^{W\prime}_{1i} e_1 + R^{W\prime}_{1i} e_1 \sqrt{T}\left(\tilde{\alpha}_i - \alpha_{i0}\right)\right] \\
&+ \sum_{j=1}^{p} \Omega_{\alpha\theta_j i}\left[\tilde{\psi}^{W\prime}_i e_1 R^{W\prime}_{2\theta} e_j + R^{W\prime}_{1i} e_1 T\left(\tilde{\theta}_j - \theta_{0,j}\right)\right] \\
&+ \sum_{j=1}^{p} \sqrt{T}\left(\hat{\Omega}_{\alpha\theta_j i}(\overline{\theta}, \overline{\alpha}_i) - \Omega_{\alpha\theta_j i}\right) T\left(\tilde{\theta}_j - \theta_{0,j}\right)\sqrt{T}\left(\tilde{\alpha}_i - \alpha_{i0}\right) \\
&+ \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p} \sqrt{T}\hat{\Omega}_{\theta_k\theta_j i}(\overline{\theta}, \overline{\alpha}_i) T\left(\tilde{\theta}_j - \theta_{0,j}\right) T\left(\tilde{\theta}_k - \theta_{0,k}\right) \\
&+ \frac{1}{6}\Omega_{\alpha\alpha\alpha_i}\left[\left(\tilde{\psi}^{W\prime}_i e_1\right)^2 R^{W\prime}_{1i} e_1 + \tilde{\psi}^{W\prime}_i e_1 R^{W\prime}_{1i} e_1 \sqrt{T}\left(\tilde{\alpha}_i - \alpha_{i0}\right) + R^{W\prime}_{1i} e_1 \sqrt{T}\left[\sqrt{T}\left(\tilde{\alpha}_i - \alpha_{i0}\right)\right]^2\right] \\
&+ \frac{1}{6}\sqrt{T}\left(\hat{\Omega}_{\alpha\alpha\alpha_i}(\overline{\theta}, \overline{\alpha}_i) - \Omega_{\alpha\alpha\alpha_i}\right)\left[\sqrt{T}\left(\tilde{\alpha}_i - \alpha_{i0}\right)\right]^3 + \frac{1}{6}\sum_{j=1}^{p} \hat{\Omega}_{\alpha\alpha\theta_j i}(\overline{\theta}, \overline{\alpha}_i)(\tilde{\theta}_j - \theta_{0,j})\left[\sqrt{T}(\tilde{\alpha}_i - \alpha_{i0})\right]^2 \\
&= O_{up}(T^{2r/s}) = o_{up}(\sqrt{T}). \qquad (2.D.26)
\end{aligned}
$$

The uniform rate of convergence then follows by Lemmas 12, 15, 40, 41 and 24, and Condition 3.

■

Lemma 26 *Assume that Conditions 1, 2, 3, 4 and 5 hold. Let $\hat{t}_i(\gamma_i; \theta)$ denote the first stage GMM score of the fixed effects, that is*

$$\hat{t}_i(\gamma_i; \theta) = \hat{t}^W_i(\gamma_i; \theta) + \hat{t}^R_i(\gamma_i; \theta), \qquad (2.D.27)$$

where $\gamma_i = (\alpha_i, \lambda_i')'$. Let $\hat{T}_i(\gamma_i; \theta)$ denote $\frac{\partial \hat{t}_i(\gamma_i; \theta)}{\partial \gamma_i'}$. Define $\hat{\gamma}_{i0}$ as the solution to $\hat{t}_i(\gamma_i; \theta_0) = 0$.

Then, for any $\overline{\gamma}_{i0}$ between $\hat{\gamma}_{i0}$ and γ_{i0}, we have

$$
\begin{aligned}
\sqrt{T}\hat{t}_i(\gamma_{i0}) &= o_{up}\left(T^{\frac{r}{2s}}\right), \qquad (2.D.28) \\
\hat{T}_i(\overline{\gamma}_{i0}) - T_i &= o_{up}(1). \qquad (2.D.29)
\end{aligned}
$$

Proof. The results follow by inspection of the score and its derivative (see Appendix 2.M), Corollary 4, Lemma 11 applied to $\theta^* = \theta_0$ and $\alpha_i^* = \alpha_{i0}$, Lemma 11 applied to $\theta^* = \theta_0$ and $\alpha_i^* = \hat{\alpha}_{i0}$, and Lemmas 12 and 13. ■

Lemma 27 *Suppose that Conditions 1, 2, 3, 4, and 5 hold. We then have*

$$\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) = o_{up}\left(T^{r/2s}\right). \qquad (2.D.30)$$

Proof. By a first order Taylor Expansion of the FOC for $\hat{\gamma}_{i0}$, we have

$$0 = \hat{t}_i(\hat{\gamma}_{i0}) = \hat{t}_i(\hat{\gamma}_{i0}) + \hat{t}_i^R(\hat{\gamma}_{i0}) = \hat{t}_i^{\Omega}(\gamma_{i0}) + \hat{T}_i(\overline{\gamma}_i)(\hat{\gamma}_{i0} - \gamma_{i0}), \tag{2.D.31}$$

where $\overline{\gamma}_i$ is between $\hat{\gamma}_{i0}$ and γ_{i0}. Next

$$\begin{aligned}
\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) &= -\underbrace{(T_i)^{-1}}_{=O_u(1)}\underbrace{\sqrt{T}\hat{t}_i(\gamma_{i0})}_{=o_{up}(T^{r/2s})} - \underbrace{(T_i)^{-1}}_{=O_u(1)}\underbrace{\left(\hat{T}_i(\overline{\gamma}_i) - T_i\right)}_{=o_{up}(1)}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) \\
&= o_{up}(T^{r/2s}) + o_{up}\left(\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})\right),
\end{aligned} \tag{2.D.32}$$

by Conditions 3 and 4, and Lemma 26. Therefore

$$(1 + o_{up}(1))\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) = o_{up}(T^{r/2s}) \Rightarrow \sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) = o_{up}(T^{r/2s}). \tag{2.D.33}$$

■

Lemma 28 *Assume that Conditions 1, 2, 3, 4 and 5 hold. Let $\hat{t}_i(\gamma_i;\theta)$ denote the second stage GMM score for the fixed effects, that is*

$$\hat{t}_i(\gamma_i;\theta) = \hat{t}_i^{\Omega}(\gamma_i;\theta) + \hat{t}_i^R(\gamma_i;\theta), \tag{2.D.34}$$

where $\gamma_i = (\alpha_i, \lambda_i')'$. Let $\hat{s}_i(\theta, \hat{\gamma}_i(\theta))$ denote the second stage GMM plug-in score for the common parameter, that is

$$\hat{s}_i(\theta, \hat{\gamma}_i(\theta)) = -\hat{G}_{\theta i}(\theta, \hat{\alpha}_i(\theta))'\hat{\lambda}_i(\theta), \tag{2.D.35}$$

where $\hat{\gamma}_i(\theta)$ is such that $\hat{t}_i(\hat{\gamma}_i(\theta);\theta) = 0$. Let $\hat{T}_{i,d}(\gamma_i;\theta)$ denote $\frac{\partial^d \hat{t}_i(\gamma_i;\theta)}{\partial \gamma_i^{d'}}$, for some $1 \leq d \leq 4$. Let $\hat{N}_i(\gamma_i;\theta)$ denote $\frac{\partial \hat{t}_i(\gamma_i;\theta)}{\partial \theta'}$. Let $\hat{M}_{i,d}(\theta, \hat{\gamma}_i)$ denote $\frac{\partial \hat{s}_i(\hat{\gamma}_i,\theta)}{\partial \hat{\gamma}_i^{d'}}$, for some $1 \leq d \leq 4$. Let $\hat{S}_i(\theta, \hat{\gamma}_i)$ denote $\frac{\partial \hat{s}_i(\hat{\gamma}_i,\theta)}{\partial \theta'}$. Let $(\hat{\theta}, \{\hat{\gamma}_i\}_{i=1}^n)$ be the GMM second stage estimators.

Then, for any $\overline{\theta}$ between $\hat{\theta}$ and θ_0, and $\overline{\gamma}_i$ between $\hat{\gamma}_i$ and γ_{i0}, we have

$$\hat{T}_{i,d}(\overline{\theta}, \overline{\gamma}_i) - T_{i,d} = o_{up}(1), \tag{2.D.36}$$

$$\hat{M}_{i,d}(\overline{\theta}, \overline{\gamma}_i) - M_{i,d} = o_{up}(1), \tag{2.D.37}$$

$$\hat{N}_i(\overline{\theta}, \overline{\gamma}_i) - N_i = o_{up}(1), \tag{2.D.38}$$

$$\hat{S}_i(\overline{\theta}, \overline{\gamma}_i) - S_i = o_{up}(1). \tag{2.D.39}$$

Proof. The results follow by inspection of the scores and their derivatives (see Appendices 2.M and 2.N), Theorem 3, Theorem 4, Lemma 11 applied to $\theta^* = \hat{\theta}$ and $\alpha_i^* = \hat{\alpha}_i$ with $a = 0$, and Lemmas 12 and 13. ■

Lemma 29 *Assume that Conditions 1, 2, 3 and 4 hold. Let $\hat{t}_i(\gamma_i;\theta)$ denote the second stage GMM score for the fixed effects, that is*

$$\hat{t}_i(\gamma_i;\theta) = \hat{t}_i^{\Omega}(\gamma_i;\theta) + \hat{t}_i^{R}(\gamma_i;\theta), \tag{2.D.40}$$

where $\gamma_i = (\alpha_i, \lambda_i')'$. Let $\hat{s}_i(\theta, \tilde{\gamma}_i(\theta))$ denote the second stage GMM plug-in score for the common parameter, that is

$$\hat{s}_i(\theta, \hat{\gamma}_i(\theta)) = -\hat{G}_{\theta i}(\theta, \hat{\alpha}_i(\theta))'\hat{\lambda}_i(\theta), \tag{2.D.41}$$

where $\hat{\gamma}_i(\theta)$ is such that $\hat{t}_i(\hat{\gamma}_i(\theta);\theta) = 0$. Let $\hat{T}_{i,d}(\gamma_i;\theta)$ denote $\frac{\partial^d \hat{t}_i(\gamma_i;\theta)}{\partial \gamma_i^{d'}}$, for some $1 \leq d \leq 4$. Let $\hat{N}_i(\gamma_i;\theta)$ denote $\frac{\partial \hat{t}_i(\gamma_i;\theta)}{\partial \theta'}$. Let $\hat{M}_{i,d}(\theta, \hat{\gamma}_i)$ denote $\frac{\partial \hat{s}_i(\hat{\gamma}_i,\theta)}{\partial \hat{\gamma}_i^{d'}}$, for some $1 \leq d \leq 4$. Let $\hat{S}_i(\theta, \hat{\gamma}_i)$ denote $\frac{\partial \hat{s}_i(\hat{\gamma}_i,\theta)}{\partial \theta'}$. Let $\hat{\gamma}_{i0}$ denote $\hat{\gamma}_i(\theta_0)$.

Then, for any $\overline{\gamma}_i$ between $\hat{\gamma}_{i0}$ and γ_{i0}, we have

$$\begin{aligned} \sqrt{T}\hat{t}_i(\overline{\gamma}_i) &= o_{up}\left(T^{\frac{r}{2s}}\right), && (2.D.42) \\ \sqrt{T}\left(\hat{T}_{i,d}(\overline{\gamma}_i) - T_{i,d}\right) &= o_{up}\left(T^{\frac{r}{2s}}\right), && (2.D.43) \\ \sqrt{T}\left(\hat{M}_{i,d}(\overline{\gamma}_i) - M_{i,d}\right) &= o_{up}\left(T^{\frac{r}{2s}}\right), && (2.D.44) \end{aligned}$$

Proof. The results follow by inspection of the scores and their derivatives (see Appendices 2.M and 2.N), Lemma 27, Lemma 11 applied to $\theta^* = \theta_0$ and $\alpha_i^* = \hat{\alpha}_{i0}$ with $a = T^{\frac{1}{2}(1-r/s)}$, and Lemmas 12 and 13. ■

2.D.2 Proof of Theorem 6

Proof. From a Taylor Expansion of the FOC for $\hat{\theta}$, we have

$$0 = s_n(\hat{\theta}) = s_n(\theta_0) + \frac{ds_n(\overline{\theta})}{d\theta'}(\hat{\theta} - \theta_0), \tag{2.D.45}$$

where $\overline{\theta}$ lies between $\hat{\theta}$ and θ_0.

Part I: Asymptotic limit of $\frac{ds_n(\overline{\theta})}{d\theta'}$. Note that

$$\begin{aligned} \frac{ds_n(\overline{\theta})}{d\theta'} &= \frac{1}{n}\sum_{i=1}^{n} \frac{ds_i(\overline{\theta}, \hat{\gamma}_i(\overline{\theta}))}{d\theta'}, && (2.D.46) \\ \frac{ds_i(\overline{\theta}, \hat{\gamma}_i(\overline{\theta}))}{d\theta'} &= \frac{\partial s_i(\overline{\theta}, \hat{\gamma}_i(\overline{\theta}))}{\partial \theta'} + \frac{\partial s_i(\overline{\theta}, \hat{\gamma}_i(\overline{\theta}))}{\partial \hat{\gamma}_i'}\frac{\partial \hat{\gamma}_i(\overline{\theta})}{\theta'}. && (2.D.47) \end{aligned}$$

From Lemma 28 and Appendix 2.N, we have

$$\begin{aligned} \frac{\partial s_i(\overline{\theta}, \hat{\gamma}_i(\overline{\theta}))}{\partial \theta'} &= S_i + o_{up}(1) = o_{up}(1), && (2.D.48) \\ \frac{\partial s_i(\overline{\theta}, \hat{\gamma}_i(\overline{\theta}))}{\partial \hat{\gamma}_i'} &= T_i + o_{up}(1). && (2.D.49) \end{aligned}$$

Then, differentiation of the FOC of $\hat{\gamma}_i$, $\hat{t}_i(\hat{\gamma}_i(\overline{\theta});\overline{\theta}) = 0$, with respect to θ and $\hat{\gamma}_i$ gives

$$\hat{T}_i(\hat{\gamma}_i(\overline{\theta});\overline{\theta})\frac{\partial\hat{\gamma}_i(\overline{\theta})}{\theta'} + \hat{N}_i(\hat{\gamma}_i(\overline{\theta});\overline{\theta}) = 0, \tag{2.D.50}$$

By repeated application of Lemma 28, we can write

$$\frac{\partial\hat{\gamma}_i(\overline{\theta})}{\theta'} = -\left(T_i\right)^{-1} N_i + o_{up}(1). \tag{2.D.51}$$

Finally, replacing the expressions for the components in (2.D.47) and using the formulae for the derivatives from Appendix 2.M, we have

$$\frac{ds_n(\overline{\theta},\hat{\gamma}(\overline{\theta}))}{d\theta'} = J_{sn} + o_p(1) = \frac{1}{n}\sum_{i=1}^{n} G'_{\theta_i} P_{\alpha_i} G_{\theta_i} + o_p(1) = J_s + o_p(1), \tag{2.D.52}$$

$$J_s = \lim_{n\to\infty} J_{sn} = O(1). \tag{2.D.53}$$

Part II: Asymptotic Expansion for $\hat{\theta} - \theta_0$. For the case $n = O(T)$, from (2.D.53) and Lemma 54 we have

$$0 = \underbrace{\sqrt{nT}s_n(\theta_0)}_{O_p(1)} + \underbrace{\frac{ds_n(\overline{\theta})}{d\theta'}}_{O(1)}\sqrt{nT}(\hat{\theta} - \theta_0). \tag{2.D.54}$$

Therefore, $\sqrt{nT}(\hat{\theta} - \theta_0) = O_p(1)$. Then the result follows by using again (2.D.53) and Lemma 54.

Similarly for the case $T = o(n)$, from (2.D.53) and Lemma 54 we have

$$0 = \underbrace{Ts_n(\theta_0)}_{O(1)} + \underbrace{\frac{ds_n(\overline{\theta})}{d\theta'}}_{O(1)}T(\hat{\theta} - \theta_0). \tag{2.D.55}$$

Therefore, $T(\hat{\theta} - \theta_0) = O(1)$. Then the result also follows by (2.D.53) and Lemma 54. ∎

Corollary 5 *Under Conditions 1, 2, 3, 4 and 5, we have*

$$T(\hat{\theta} - \theta_0) = Q_{1\theta} + \frac{a_T}{T^{(r-1)/2}}\psi_\theta + \frac{1}{T}R_{2\theta}, \tag{2.D.56}$$

where

$$
\begin{aligned}
a_T &= \begin{cases} C, & \text{if } n = O(T^a) \text{ for some } a \in \Re; \\ o(T^\epsilon) \text{ for any } \epsilon > 0, & \text{otherwise,} \end{cases} && \text{(2.D.57)} \\
Q_{1\theta} &= -(J_s)^{-1} \frac{1}{n} \sum_{i=1}^{n} Q_{1si} = O_{up}(1), && \text{(2.D.58)} \\
\psi_\theta &= -(J_s)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\psi}_{si} = O_{up}(1), && \text{(2.D.59)} \\
R_{2\theta} &= o_{up}(\sqrt{T}). && \text{(2.D.60)}
\end{aligned}
$$

Proof. The result follows by using the expansion of $T\hat{s}_n(\theta_0)$ in the proof of Lemma 54. ■

2.E Asymptotic Distribution of the bias-corrected Second Stage GMM Estimator

2.E.1 Some Lemmas

Lemma 30 *Assume that Conditions 1, 2, 3, 4, and 5 hold. We then have*

$$\sqrt{T}(\hat{\gamma}_i - \gamma_{i0}) = o_{up}(T^{r/2s}). \tag{2.E.1}$$

where $\hat{\gamma}_i$ is the solution to $\hat{t}_i(\gamma_i; \hat{\theta}) = 0$, i.e. the second stage estimator of the fixed effects.

Proof. We show that

$$\sqrt{T}(\hat{\gamma}_i - \hat{\gamma}_{i0}) = o_{up}(1), \tag{2.E.2}$$

and then the result follows by Lemma 27.

Note that

$$\sqrt{T}(\hat{\gamma}_i - \hat{\gamma}_{i0}) = \frac{\partial \gamma_i(\overline{\theta})}{\partial \theta'} \sqrt{T}(\hat{\theta} - \theta_0), \tag{2.E.3}$$

where $\overline{\theta}$ lies between $\hat{\theta}$ and θ_0. Following an analogous argument as in the proof of Theorem 6, we have

$$\sqrt{T}(\hat{\gamma}_i - \hat{\gamma}_{i0}) = -\underbrace{(T_i)^{-1}}_{=O_u(1)} N_i \underbrace{\sqrt{T}(\hat{\theta} - \theta_0)}_{=O_{up}(T^{-1/2})} + o_{up}\left(\sqrt{T}(\hat{\theta} - \theta_0)\right) = O_{up}(T^{-1/2}) = o_{up}(1). \tag{2.E.4}$$

■

Lemma 31 *Assume that Conditions 1, 2, 3, 4 and 5 hold. Let $\hat{t}_i(\gamma_i; \theta)$ denote the second stage GMM* *score for the fixed effects, that is*

$$\hat{t}_i(\gamma_i;\theta) = \hat{t}_i^{\Omega}(\gamma_i;\theta) + \hat{t}_i^{R}(\gamma_i;\theta), \tag{2.E.5}$$

where $\gamma_i = (\alpha_i, \lambda_i')'$. *Let* $\hat{s}_i(\theta, \tilde{\gamma}_i(\theta))$ *denote the second stage GMM plug-in score for the common parameter, that is*

$$\hat{s}_i(\theta, \hat{\gamma}_i(\theta)) = -\hat{G}_{\theta i}(\theta, \hat{\alpha}_i(\theta))'\hat{\lambda}_i(\theta), \tag{2.E.6}$$

where $\hat{\gamma}_i(\theta)$ *is such that* $\hat{t}_i(\hat{\gamma}_i(\theta);\theta) = 0$. *Let* $\hat{T}_{i,d}(\gamma_i;\theta)$ *denote* $\frac{\partial^d \hat{t}_i(\gamma_i;\theta)}{\partial \gamma_i^{d'}}$, *for some* $1 \le d \le 4$. *Let* $\hat{N}_i(\gamma_i;\theta)$ *denote* $\frac{\partial \hat{t}_i(\gamma_i;\theta)}{\partial \theta'}$. *Let* $\hat{M}_{i,d}(\theta, \hat{\gamma}_i)$ *denote* $\frac{\partial \hat{s}_i(\hat{\gamma}_i,\theta)}{\partial \hat{\gamma}_i^{d'}}$ *, for some* $1 \le d \le 4$. *Let* $\hat{S}_i(\theta, \hat{\gamma}_i)$ *denote* $\frac{\partial \hat{s}_i(\hat{\gamma}_i,\theta)}{\partial \theta'}$. *Let* $(\hat{\theta}, \{\hat{\gamma}_i\}_{i=1}^n)$ *be the second stage GMM estimators.*

Then, for any $\overline{\theta}$ *between* $\hat{\theta}$ *and* θ_0, *and* $\overline{\gamma}_i$ *between* $\hat{\gamma}_i$ *and* γ_{i0}, *we have*

$$\begin{aligned}
\sqrt{T}\left(\hat{T}_{i,d}(\overline{\theta}, \overline{\gamma}_i) - T_{i,d}\right) &= o_{up}\left(T^{r/2s}\right), && (2.E.7)\\
\sqrt{T}\left(\hat{M}_{i,d}(\overline{\theta}, \overline{\gamma}_i) - M_{i,d}\right) &= o_{up}\left(T^{r/2s}\right), && (2.E.8)\\
\sqrt{T}\left(\hat{N}_i(\overline{\theta}, \overline{\gamma}_i) - N_i\right) &= o_{up}\left(T^{r/2s}\right), && (2.E.9)\\
\sqrt{T}\left(\hat{S}_i(\overline{\theta}, \overline{\gamma}_i) - S_i\right) &= o_{up}\left(T^{r/2s}\right). && (2.E.10)
\end{aligned}$$

Proof. The results follow by inspection of the scores and their derivatives (see Appendices 2.M and 2.N), Theorem 6, Lemma 30, Lemma 11 applied to $\theta^* = \hat{\theta}$ and $\alpha_i^* = \hat{\alpha}_i$ with $a = T^{\frac{1}{2}(1-r/s)}$, and Lemmas 12 and 13. ■

Lemma 32 *Assume that Conditions 1, 2, 3, 4, 5, and 6 hold. Let* $\hat{\Omega}_i(\overline{\theta}, \overline{\alpha}_i)$ *denote the estimator of the weighting functions*

$$\frac{1}{T}\sum_{t=1}^{T} g(z_{it};\overline{\theta}, \overline{\alpha}_i) g(z_{it};\overline{\theta}, \overline{\alpha}_i)' \quad i = 1, ..., n, \tag{2.E.11}$$

where $(\overline{\theta}', \overline{\alpha}_i)'$ *lies between* $(\hat{\theta}', \hat{\alpha}_i)'$ *and* $(\theta_0', \alpha_{i0})'$. *Let* $\hat{\Omega}_{\alpha^{d_1}\theta^{d_2}i}(\overline{\theta}, \overline{\alpha}_i)$ *denote its derivatives*

$$\frac{\partial^{d_1+d_2}\hat{\Omega}_i(\overline{\theta}, \overline{\alpha}_i)}{\partial^{d_1}\alpha_i \partial^{d_2}\theta}, \tag{2.E.12}$$

for $0 \le d_1 + d_2 \le 3$. *Then, we have*

$$\sqrt{T}\left(\hat{\Omega}_{\alpha^{d_1}\theta^{d_2}\ i}(\overline{\theta}, \overline{\alpha}_i) - \Omega_{\alpha^{d_1}\theta^{d_2}\ i}\right) = o_{up}\left(T^{\frac{r}{2s}}\right). \tag{2.E.13}$$

Proof. Note that

$$\begin{aligned}
&\left|g(z_{it};\overline{\theta}, \overline{\alpha}_i) g(z_{it};\overline{\theta}, \overline{\alpha}_i)' - E\left[g(z_{it};\overline{\theta}, \overline{\alpha}_i) g(z_{it};\overline{\theta}, \overline{\alpha}_i)'\right]\right| \\
\le\ & m^2 \max_{1\le k\le l\le m} \left|g_k(z_{it};\overline{\theta}, \overline{\alpha}_i) g_l(z_{it};\overline{\theta}, \overline{\alpha}_i)' - E\left[g_k(z_{it};\overline{\theta}, \overline{\alpha}_i) g_l(z_{it};\overline{\theta}, \overline{\alpha}_i)'\right]\right| && (2.E.14)
\end{aligned}$$

Then we can use Lemma 23 for $f = g_k g_l - E[g_k g_l']$ with $a = o\left(T^{\frac{1}{2}(1-r/s)}\right)$. A similar argument applies for the derivatives, since they are sums of products of elements that satisfy the assumption of the Lemma 23. ■

Lemma 33 *Assume that Conditions 1, 2, 3, 4, 5, and 6 hold. Let*

$$\hat{\Sigma}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) = \left(\hat{G}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i)' \hat{\Omega}_i(\overline{\theta}, \overline{\alpha}_i)^{-1} \hat{G}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i)\right)^{-1}, \tag{2.E.15}$$

$$\hat{H}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) = \hat{\Sigma}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) \hat{G}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i)' \hat{\Omega}_i(\overline{\theta}, \overline{\alpha}_i)^{-1}, \tag{2.E.16}$$

$$\hat{P}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) = \hat{\Omega}_i(\overline{\theta}, \overline{\alpha}_i)^{-1} - \hat{\Omega}_i(\overline{\theta}, \overline{\alpha}_i)^{-1} \hat{G}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) \hat{H}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i), \tag{2.E.17}$$

$$\hat{\Sigma}^W_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) = \left(\hat{G}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i)' W_i^{-1} \hat{G}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i)\right)^{-1}, \tag{2.E.18}$$

$$\hat{H}^W_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) = \hat{\Sigma}^W_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) \hat{G}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i)' W_i^{-1}, \tag{2.E.19}$$

be estimators of

$$\Sigma_{\alpha_i} = \left(G'_{\alpha_i} \Omega_i^{-1} G_{\alpha_i}\right)^{-1}, \tag{2.E.20}$$

$$\Sigma^W_{\alpha_i} = \left(G'_{\alpha_i} W_i^{-1} G_{\alpha_i}\right)^{-1}, \tag{2.E.21}$$

$$H^W_{\alpha_i} = \Sigma^W_{\alpha_i} G'_{\alpha_i} W_i^{-1}, \tag{2.E.22}$$

$$H_{\alpha_i} = \Sigma_{\alpha_i} G'_{\alpha_i} \Omega_i^{-1}, \tag{2.E.23}$$

$$P_{\alpha_i} = \Omega_i^{-1} - \Omega_i^{-1} G_{\alpha_i} H_{\alpha_i}, \tag{2.E.24}$$

where $(\overline{\theta}', \overline{\alpha}_i)'$ lies between $(\hat{\theta}', \hat{\alpha}_i)'$ and $(\theta_0', \alpha_{i0})'$. Let $\hat{F}_{\alpha^{d_1}\theta^{d_2}i}(\overline{\theta}, \overline{\alpha}_i)$ and $F_{\alpha^{d_1}\theta^{d_2}i}$, with $F \in \{\Sigma, H, P, \Sigma^W, H^W\}$ denote their derivatives for $0 \leq d_1 + d_2 \leq 3$. Then, we have

$$\sqrt{T}\left(\hat{\Sigma}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) - \Sigma_{\alpha_i}\right) = o_{up}\left(T^{r/2s}\right), \tag{2.E.25}$$

$$\sqrt{T}\left(\hat{H}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) - H_{\alpha_i}\right) = o_{up}\left(T^{r/2s}\right), \tag{2.E.26}$$

$$\sqrt{T}\left(\hat{P}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) - P_{\alpha_i}\right) = o_{up}\left(T^{r/2s}\right), \tag{2.E.27}$$

$$\sqrt{T}\left(\hat{\Sigma}^W_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) - \Sigma^W_{\alpha_i}\right) = o_{up}\left(T^{r/2s}\right), \tag{2.E.28}$$

$$\sqrt{T}\left(\hat{H}^W_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) - H^W_{\alpha_i}\right) = o_{up}\left(T^{r/2s}\right), \tag{2.E.29}$$

$$\sqrt{T}\left(\hat{F}_{\alpha^{d_1}\theta^{d_2}i}(\overline{\theta}, \overline{\alpha}_i) - F_{\alpha^{d_1}\theta^{d_2}i}\right) = o_{up}\left(T^{r/2s}\right). \tag{2.E.30}$$

Proof. The results follow by Lemmas 12, 13, 19, 31 and 32. ■

Lemma 34 *Assume that Conditions 1, 2, 3, 4, 5, and 6 hold. Let*

$$\hat{J}_{si}(\overline{\theta}, \overline{\alpha}_i) = \hat{G}_{\theta_i}(\overline{\theta}, \overline{\alpha}_i)' \hat{P}_{\alpha_i}(\overline{\theta}, \overline{\alpha}_i) \hat{G}_{\theta_i}(\overline{\theta}, \overline{\alpha}_i), \tag{2.E.31}$$

be an estimator of

$$J_{si} = G'_{\theta_i} P_{\alpha_i} G_{\theta_i}, \tag{2.E.32}$$

where $(\overline{\theta}', \overline{\alpha}_i)'$ *lies between* $(\hat{\theta}', \hat{\alpha}_i)'$ *and* $(\theta'_0, \alpha_{i0})'$. *Let* $\hat{J}_{s\alpha^{d_1}\theta^{d_2}i}(\overline{\theta}, \overline{\alpha}_i)$ *and* $J_{s\alpha^{d_1}\theta^{d_2}i}$ *denote their derivatives for* $0 \leq d_1 + d_2 \leq 3$. *Then, we have*

$$\sqrt{T}\left(\hat{J}_{s\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) - J_{s\alpha_i}\right) = o_{up}\left(T^{r/2s}\right), \tag{2.E.33}$$

$$\sqrt{T}\left(\hat{J}_{s\alpha^{d_1}\theta^{d_2}i}\left(\overline{\theta}, \overline{\alpha}_i\right) - J_{s\alpha^{d_1}\theta^{d_2}i}\right) = o_{up}\left(T^{r/2s}\right). \tag{2.E.34}$$

Proof. The results follow by Lemmas 12, 31 and 33. ■

Lemma 35 *Assume that Conditions 1, 2, 3, 5, and 4 hold. Let*

$$\begin{aligned}
\hat{B}_{si}\left(\overline{\theta}, \overline{\alpha}_i\right) &= \frac{1}{2}\hat{G}_{\theta_i}\left(\overline{\theta}, \overline{\alpha}_i\right)' \hat{P}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) \hat{G}_{\alpha\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) \hat{\Sigma}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) \\
&- \hat{G}_{\theta_i}\left(\overline{\theta}, \overline{\alpha}_i\right)' \hat{P}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) \frac{1}{T}\sum_{t=1}^{T}\left[G_{\alpha_i}(z_{it}; \overline{\theta}, \overline{\alpha}_i)\hat{H}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) g(z_{it}; \overline{\theta}, \overline{\alpha}_i)\right] \\
&- \hat{G}_{\theta_i}\left(\overline{\theta}, \overline{\alpha}_i\right)' \hat{H}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right)' \frac{1}{T}\sum_{t=1}^{T}\left[G_{\alpha_i}(z_{it}; \overline{\theta}, \overline{\alpha}_i)'\hat{P}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) g(z_{it}; \overline{\theta}, \overline{\alpha}_i)\right] \\
&- \hat{G}_{\theta_i}\left(\overline{\theta}, \overline{\alpha}_i\right)' \hat{P}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) \frac{1}{T}\sum_{t=1}^{T}[g(z_{it}; \overline{\theta}, \overline{\alpha}_i)g(z_{it}; \overline{\theta}, \overline{\alpha}_i)'\hat{P}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) g(z_{it}; \overline{\theta}, \overline{\alpha}_i)] \\
&- \hat{G}_{\theta_i}\left(\overline{\theta}, \overline{\alpha}_i\right)' \hat{P}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) \hat{\Omega}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right)\left(\hat{H}^W_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) - \hat{H}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right)\right) \\
&+ \frac{1}{T}\sum_{t=1}^{T}\left[G_{\theta_i}(z_{it}; \overline{\theta}, \overline{\alpha}_i)'\hat{P}_{\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) g(z_{it}; \overline{\theta}, \overline{\alpha}_i)\right],
\end{aligned} \tag{2.E.35}$$

be an estimator of

$$\begin{aligned}
B_{si} &= \frac{1}{2}G'_{\theta_i} P_{\alpha_i} G_{\alpha\alpha_i} \Sigma_{\alpha_i} - G'_{\theta_i} P_{\alpha_i} E\left[G_{\alpha_i}(z_{it}) H_{\alpha_i} g(z_{it})\right] \\
&- G'_{\theta_i} H'_{\alpha_i} E\left[G_{\alpha_i}(z_{it})' P_{\alpha_i} g(z_{it})\right] - G'_{\theta_i} P_{\alpha_i} E[g(z_{it}) g(z_{it})' P_{\alpha_i} g(z_{it})] \\
&- G'_{\theta_i} P_{\alpha_i} \Omega_{\alpha_i}\left(H^W_{\alpha_i} - H_{\alpha_i}\right) + E\left[G_{\theta_i}(z_{it})' P_{\alpha_i} g(z_{it})\right],
\end{aligned} \tag{2.E.36}$$

where $(\overline{\theta}', \overline{\alpha}_i)'$ *lies between* $(\hat{\theta}', \hat{\alpha}_i)'$ *and* $(\theta'_0, \alpha_{i0})'$. *Let* $\hat{B}_{s\alpha^{d_1}\theta^{d_2}i}(\overline{\theta}, \overline{\alpha}_i)$ *and* $B_{s\alpha^{d_1}\theta^{d_2}i}$ *denote their derivatives for* $0 \leq d_1 + d_2 \leq 3$. *Then, we have*

$$\sqrt{T}\left(\hat{B}_{s\alpha_i}\left(\overline{\theta}, \overline{\alpha}_i\right) - B_{s\alpha_i}\right) = o_{up}\left(T^{r/2s}\right), \tag{2.E.37}$$

$$\sqrt{T}\left(\hat{B}_{s\alpha^{d_1}\theta^{d_2}i}\left(\overline{\theta}, \overline{\alpha}_i\right) - B_{s\alpha^{d_1}\theta^{d_2}i}\right) = o_{up}\left(T^{r/2s}\right). \tag{2.E.38}$$

Proof. The results follow by Lemmas 12, 31 and 33. ■

Lemma 36 *Assume that Conditions 1, 2, 3, 4, 5, and 6 hold. We then have*

$$\hat{J}_{si}(\hat{\theta},\hat{\alpha}_i) = J_{si} + \frac{1}{\sqrt{T}}\tilde{\psi}_{Jsi} + \frac{1}{T}Q_{1Jsi} + \frac{a_T}{T^{(r+1)/2}}Q_{1rJsi} + \frac{1}{T^{3/2}}R_{2Jsi}, \tag{2.E.39}$$

where

$$\begin{aligned} \tilde{\psi}_{Jsi} &= \sqrt{T}\left(\hat{J}_{si} - J_{si}\right) + J_{\alpha_i}\tilde{\psi}_i' e_1 = o_{up}\left(T^{r/2s}\right), && (2.E.40) \\ Q_{1Jsi} &= J_{s\alpha_i}Q_{1i}'e_1 + \sqrt{T}(\hat{J}_{s\alpha_i} - J_{s\alpha_i})\tilde{\psi}_i' e_1 + \sum_{j=1}^{p} J_{s\theta_j i}Q_{1\theta}' e_j + \frac{1}{2}J_{s\alpha\alpha_i}\left(\tilde{\psi}_i' e_1\right) = o_{up}\left(T^{r/s}\right), && (2.E.41) \\ Q_{1rJsi} &= \sum_{j=1}^{p} J_{s\theta_j i}\psi_\theta' e_j = O_{up}(1), && (2.E.42) \\ R_{2Jsi} &= o_{up}\left(T^{3r/2s}\right) = o_{up}\left(\sqrt{T}\right). && (2.E.43) \end{aligned}$$

Also

$$\hat{J}_{sn}\left(\hat{\theta}\right) = \frac{1}{n}\sum_{i=1}^{n}\hat{J}_{si}(\hat{\theta},\hat{\alpha}_i) = J_{sn} + \frac{1}{T}Q_{1Js} + \frac{1}{T^{3/2}}R_{1Js}, \tag{2.E.44}$$

where

$$\begin{aligned} Q_{1Js} &= \frac{1}{n}\sum_{i=1}^{n} E\left[Q_{1rJsi} + \frac{a_T}{T^{(r-1)/2}}Q_{1rJsi}\right] = O(1), && (2.E.45) \\ R_{1Js} &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\psi}_{Jsi} + \frac{1}{\sqrt{T}}\frac{1}{n}\sum_{i=1}^{n}(Q_{1Jsi} - E\left[Q_{1Jsi}\right]) + \frac{a_T}{T^{r/2}}\frac{1}{n}\sum_{i=1}^{n}(Q_{1rJsi} - E\left[Q_{1rJsi}\right]) \\ &+ \frac{1}{T}\frac{1}{n}\sum_{i=1}^{n} R_{2Jsi} = o_p(1). && (2.E.46) \end{aligned}$$

Proof. By a standard Taylor expansion around (θ_0, α_{i0}), we have

$$\begin{aligned} \hat{J}_{si}(\hat{\theta},\hat{\alpha}_i) &= \hat{J}_{si} + \hat{J}_{s\alpha_i}(\hat{\alpha}_i - \alpha_{i0}) + \sum_{j=1}^{p}\hat{J}_{s\theta_j}(\overline{\theta},\overline{\alpha}_i)(\hat{\theta}_j - \theta_{0,j}) + \frac{1}{2}\hat{J}_{s\alpha\alpha_i}(\overline{\theta},\overline{\alpha}_i)(\hat{\alpha}_i - \alpha_{i0})^2 \\ &+ \frac{1}{2}\sum_{j=1}^{p}\hat{J}_{s\alpha\theta_j i}(\overline{\theta},\overline{\alpha}_i)(\hat{\theta}_j - \theta_{0,j})(\hat{\alpha}_i - \alpha_{i0}), && (2.E.47) \end{aligned}$$

where $(\overline{\theta},\overline{\alpha}_i)$ lies between $(\tilde{\theta},\tilde{\alpha}_i)$ and (θ_0, α_{0i}). The expressions for $\tilde{\psi}_{Jsi}^W$, Q_{1sJi}^W, and Q_{1rJsi}^W can be obtained using the expansions for $\hat{\gamma}_{i0}$ in Lemma 49 and for $\hat{\theta}$ in Corollary 5, after some algebra. The rest of the properties for these terms follow by the properties of their components and Lemmas 12, 58, 59, 60, and 61. For the remainder term, we have

$$
\begin{aligned}
R^W_{2Jsi} &= J_{s\alpha_i} R^{W'}_{2i} e_1 + \sqrt{T}(\hat{J}_{s\alpha_i} - J_{s\alpha_i}) R^{W'}_{1i} e_1 + \sum_{j=1}^{p} J_{s\theta_j} R^{W'}_{2\theta} e_j + \sum_{j=1}^{p} \sqrt{T}\left(\hat{J}_{s\theta_j i}\left(\overline{\theta},\overline{\alpha}_i\right) - J_{\theta_j i}\right)\left(\hat{\theta}_j - \theta_{0,j}\right) \\
&+ \frac{1}{2} J_{s\alpha\alpha_i}\left[\tilde{\psi}^{W'}_i e_1 R^{W'}_{1i} e_1 + R^{W'}_{1i} e_1 \sqrt{T}\left(\hat{\alpha}_i - \alpha_{i0}\right)\right] + \frac{1}{2}\sqrt{T}(\hat{J}_{s\alpha\alpha_i}\left(\overline{\theta},\overline{\alpha}_i\right) - J_{\alpha\alpha_i})\left[\sqrt{T}\left(\hat{\alpha}_i - \alpha_{i0}\right)\right]^2 \\
&+ \sum_{j=1}^{p} \hat{J}_{\alpha\theta_j i}(\overline{\theta},\overline{\alpha}_i) T\left(\hat{\theta}_j - \theta_{0,j}\right)\sqrt{T}\left(\hat{\alpha}_i - \alpha_{i0}\right). \qquad (2.E.48)
\end{aligned}
$$

The uniform rate of convergence then follows by Lemmas 12, 15, 40, 41 and 34. ■

Lemma 37 *Assume that Conditions 1, 2, 3, 4, 5, and 6 hold. We then have*

$$
\hat{B}_{si}(\hat{\theta},\hat{\alpha}_i) = B_{si} + \frac{1}{\sqrt{T}}\tilde{\psi}_{Bsi} + \frac{1}{T} Q_{1Bsi} + \frac{a_T}{T^{(r+1)/2}} Q_{1rBsi} + \frac{1}{T^{3/2}} R_{2Bsi}, \qquad (2.E.49)
$$

where

$$
\begin{aligned}
\tilde{\psi}_{Bsi} &= \sqrt{T}\left(\hat{B}_{si} - B_{si}\right) + B_{\alpha_i}\tilde{\psi}_i' e_1 = o_{up}\left(T^{r/2s}\right), \qquad (2.E.50) \\
Q_{1Bsi} &= B_{s\alpha_i} Q_{1i}' e_1 + \sqrt{T}(\hat{B}_{s\alpha_i} - B_{s\alpha_i})\tilde{\psi}_i' e_1 + \sum_{j=1}^{p} B_{s\theta_j i} Q_{1\theta}' e_j + \frac{1}{2} B_{\alpha\alpha_i}\left(\tilde{\psi}_i' e_1\right) = o_{up}\left(T^{r/s}\right), \qquad (2.E.51) \\
Q_{1rBsi} &= \sum_{j=1}^{p} B_{s\theta_j i}\psi_\theta' e_j = O_{up}(1), \qquad (2.E.52) \\
R_{2Bsi} &= o_{up}\left(T^{3r/2s}\right) = o_{up}\left(\sqrt{T}\right). \qquad (2.E.53)
\end{aligned}
$$

Also

$$
\hat{B}_{sn}\left(\hat{\theta}\right) = \frac{1}{n}\sum_{i=1}^{n} \hat{B}_{si}(\hat{\theta},\hat{\alpha}_i) = B_{sn} + \frac{1}{T} Q_{1Bs} + \frac{1}{T^{3/2}} R_{1Bs}, \qquad (2.E.54)
$$

where

$$
\begin{aligned}
Q_{1Bs} &= \frac{1}{n}\sum_{i=1}^{n} E\left[Q_{1rBsi} + \frac{a_T}{T^{(r-1)/2}} Q_{1rBsi}\right] = O(1), \qquad (2.E.55) \\
\tilde{R}_{1Bs} &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\psi}_{Jsi} + \frac{1}{\sqrt{T}}\frac{1}{n}\sum_{i=1}^{n}\left(Q_{1Bsi} - E\left[Q_{1Bsi}\right]\right) + \frac{a_T}{T^{r/2}}\frac{1}{n}\sum_{i=1}^{n}\left(Q_{1rBsi} - E\left[Q_{1rBsi}\right]\right) \\
&+ \frac{1}{T}\frac{1}{n}\sum_{i=1}^{n} R_{2Bsi} = o_p(1). \qquad (2.E.56)
\end{aligned}
$$

Proof. Analogous to the proof of Lemma 36 replacing J_s by B_s, and Lemma 34 by Lemma 35. ■

Lemma 38 *Suppose that* $\hat{\xi}_j = \xi_j + a_{nT}\tilde{\psi}_{ji} + b_{nT}R_{ji}$, *for* $j = 1, 2$, $a_{nT} = o(1)$, $b_{nT} = o(a_{nT})$, *and* $\hat{\xi}_{2i}$

and ξ_{2i} are non singular. Then, we have

$$\hat{\xi}_{2i}^{-1}\hat{\xi}_{1i} = \xi_{2i}^{-1}\xi_{1i} + a_{nT}\left[\xi_{2i}^{-1}\tilde{\psi}_{1i} - \xi_{2i}^{-1}\tilde{\psi}_{2i}\xi_{2i}^{-1}\xi_{1i}\right] + b_{nT}R_i. \tag{2.E.57}$$

Proof. Note that

$$\begin{aligned}
\hat{\xi}_{2i}^{-1}\hat{\xi}_{1i} &= \left(\xi_{2i} + a_{nT}\tilde{\psi}_{2i} + b_{nT}R_{2i}\right)^{-1}\left(\xi_{1i} + a_{nT}\tilde{\psi}_{1i} + b_{nT}R_{1i}\right) \\
&= \xi_{2i}^{-1}\left(\xi_{1i} + a_{nT}\tilde{\psi}_{1i} + b_{nT}R_{1i}\right) + \underbrace{\left[\left(\xi_{2i} + a_{nT}\tilde{\psi}_{2i} + b_{nT}R_{2i}\right)^{-1} - \xi_{2i}^{-1}\right]}_{\equiv A}\left(\xi_{1i} + a_{nT}\tilde{\psi}_{1i} + b_{nT}R_{1i}\right)
\end{aligned}$$

$$. \tag{2.E.58}$$

We can rewrite A as

$$\begin{aligned}
A &= \left(\xi_{2i} + a_{nT}\tilde{\psi}_{2i} + b_{nT}R_{2i}\right)^{-1}\left[\xi_{2i} - \left(\xi_{2i} + a_{nT}\tilde{\psi}_{2i} + b_{nT}R_{2i}\right)\right]\xi_{2i}^{-1} \\
&= -a_{nT}\underbrace{\left(\xi_{2i} + a_{nT}\tilde{\psi}_{2i} + b_{nT}R_{2i}\right)^{-1}\tilde{\psi}_{2i}\xi_{2i}^{-1}}_{\equiv B} - b_{nT}\left(\xi_{2i} + a_{nT}\tilde{\psi}_{2i} + b_{nT}R_{2i}\right)^{-1}R_{2i}\xi_{2i}^{-1}.
\end{aligned} \tag{2.E.59}$$

Similarly, we can write B

$$B = \xi_{2i}^{-1}\tilde{\psi}_{2i}\xi_{2i}^{-1} - a_{nT}\left(\xi_{2i} + a_{nT}\tilde{\psi}_{2i} + b_{nT}R_{2i}\right)^{-1}\left(\tilde{\psi}_{2i} + a_{nT}R_{2i}\right)\xi_{2i}^{-1}\tilde{\psi}_{2i}\xi_{2i}^{-1}. \tag{2.E.60}$$

Finally, replacing back in (2.E.58) we get the result with

$$R_i = \left(\xi_{2i} + a_{nT}\tilde{\psi}_{2i} + b_{nT}R_{2i}\right)^{-1}\left\{R_{1i} + \left[\left(\tilde{\psi}_{2i} + a_{nT}R_{2i}\right)\xi_{2i}^{-1}\tilde{\psi}_{2i} - R_{2i}\right]\xi_{2i}^{-1}\tilde{\psi}_{1i}\right\}. \tag{2.E.61}$$

∎

Lemma 39 *Assume that Conditions 1, 2, 3, 4, 5, and 6 hold. We then have*

$$\hat{B}_n(\hat{\theta}) = -\hat{J}_{sn}(\hat{\theta})^{-1}\hat{B}_{sn}(\hat{\theta}) = -J_{sn}^{-1}B_{sn} - \frac{1}{T}J_{sn}^{-1}\left[Q_{1Bs} - Q_{1Js}J_{sn}^{-1}B_{sn}\right] - \frac{1}{T^{3/2}}R_{1B}. \tag{2.E.62}$$

Also

$$\sqrt{\frac{n}{T}}\hat{B}_n(\hat{\theta}) = -\sqrt{\frac{n}{T}}J_{sn}^{-1}B_{sn} + o_p(1). \tag{2.E.63}$$

Proof. The results follow from Lemmas 36, 37 and 38 ∎

2.E.2 Proof of Theorem 7

Proof. Case I: C = BC. First, note that for any $\overline{\theta}$ such that $\overline{\theta} - \theta_0 = O_p(T^{-1})$, by a first order Taylor expansion

$$J_{sn}\left(\overline{\theta}\right) = J_{sn} + O_p\left(T^{-1}\right). \tag{2.E.64}$$

Next, by Theorem 6 and Lemma 56

$$\sqrt{nT}\left(\hat{\theta} - \theta_0\right) = -J_{sn}\left(\overline{\theta}\right)^{-1}\hat{s}_n(\theta_0) = -J_{sn}^{-1}\hat{s}_n(\theta_0) + O_p(T^{-1})O_p\left(\sqrt{\frac{n}{T}}\right) = -J_{sn}^{-1}\hat{s}_n(\theta_0) + o_p(1). \tag{2.E.65}$$

Finally, by Lemmas 38 and 56

$$\begin{aligned}\sqrt{nT}\left(\hat{\theta}^{(BC)} - \theta_0\right) &= \sqrt{nT}\left(\hat{\theta} - \theta_0\right) - \sqrt{nT}\frac{1}{T}\hat{B}_n\left(\hat{\theta}\right) = -J_{sn}^{-1}\hat{s}_n(\theta_0) + \sqrt{\frac{n}{T}}J_{sn}^{-1}B_{sn} + o_p(1)\\ &= -J_{sn}^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}_{si} + \sqrt{\frac{n}{T}}B_{sn} - \sqrt{\frac{n}{T}}B_{sn}\right] + o_p(1) \xrightarrow{d} N(0, J_s^{-1}).\end{aligned} \tag{2.E.66}$$

Case II: C = SBC. First, note that since the correction of the score is of order $O_p(T^{-1})$, we have that $\hat{\theta}^{(SBC)} - \theta_0 = O(T^{-1})$. Then, by a Taylor expansion of the FOC

$$0 = \hat{s}_n\left(\hat{\theta}^{(SBC)}\right) - \frac{1}{T}\hat{B}_{sn}\left(\hat{\theta}^{(SBC)}\right) = \hat{s}_n(\theta_0) + J_{sn}\left(\overline{\theta}\right)\left(\hat{\theta}^{(SBC)} - \theta_0\right) - \frac{1}{T}B_{sn} + o_p(T^{-2}), \tag{2.E.67}$$

where $\overline{\theta}$ lies between $\hat{\theta}^{(SBC)}$ and θ_0. Then by Lemma 56

$$\begin{aligned}\sqrt{nT}\left(\hat{\theta}^{(SBC)} - \theta_0\right) &= -J_{sn}\left(\overline{\theta}\right)^{-1}\left[\sqrt{nT}\hat{s}_n(\theta_0) - \sqrt{\frac{n}{T}}B_{sn}\right] + o_p(1)\\ &= -J_{sn}\left(\overline{\theta}\right)^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}_{si} + \sqrt{\frac{n}{T}}B_{sn} - \sqrt{\frac{n}{T}}B_{sn}\right] + o_p(1) \xrightarrow{d} N(0, J_s^{-1}).\end{aligned} \tag{2.E.68}$$

Case II: C = IBC. A similar argument applies to the estimating equation (2.5.6), since $\hat{\theta}$ is in a $O(T^{-1})$ neighborhood of θ_0. ∎

2.F Stochastic Expansion for $\tilde{\gamma}_{i0} = \tilde{\gamma}_i(\theta_0)$

Lemma 40 *Suppose that Conditions 1, 2, 3, and 4 hold. We then have*

$$\sqrt{T}(\tilde{\gamma}_{i0} - \gamma_{i0}) = \tilde{\psi}_i^W + \frac{1}{\sqrt{T}}R_{1i}^W, \tag{2.F.1}$$

where

$$\tilde{\psi}_i^W = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_{it}^W = -\left(T_i^W\right)^{-1}\sqrt{T}\hat{t}_i^W(\gamma_{i0}) = o_{up}(T^{r/2s}), \tag{2.F.2}$$

$$R_{1i}^W = o_{up}(T^{r/s}) = o_{up}(\sqrt{T}). \tag{2.F.3}$$

Also

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i^W = O_p(1). \tag{2.F.4}$$

Proof. The statements about $\tilde{\psi}_i^W$ follow by Lemmas 14 and 58. From the proof of Lemma 15, we have

$$R_{1i}^W = -\underbrace{\left(T_i^W\right)^{-1}}_{=O_u(1)}\underbrace{\sqrt{T}(\hat{T}_i^W(\overline{\gamma}_i) - T_i^W)}_{=o_u(T^{r/2s})}\underbrace{\sqrt{T}(\tilde{\gamma}_{i0} - \gamma_{i0})}_{=o_u(T^{r/2s})} = o_{up}(T^{r/s}) = o_{up}(\sqrt{T}), \tag{2.F.5}$$

by Lemmas 12, 17 and 15, and Conditions 3 and 4. ∎

Lemma 41 *Suppose that Conditions 1, 2, 3, and 4 hold. We then have*

$$\sqrt{T}(\tilde{\gamma}_{i0} - \gamma_{i0}) = \tilde{\psi}_i^W + \frac{1}{\sqrt{T}}Q_{1i}^W + \frac{1}{T}R_{2i}^W, \tag{2.F.6}$$

where

$$Q_{1i}^W = -\left(T_i^W\right)^{-1}\left[\tilde{A}_i^W\tilde{\psi}_i^W + \frac{1}{2}\sum_{j=1}^{m+1}\tilde{\psi}_{i,j}^W T_{i,j}^W\tilde{\psi}_i^W\right] = o_{up}(T^{r/s}), \tag{2.F.7}$$

$$\tilde{A}_i^W = \sqrt{T}(\hat{T}_i^W - T_i^W) = o_{up}(T^{r/2s}) \tag{2.F.8}$$

$$R_{2i}^W = o_{up}(T^{3r/2s}) = o_{up}(\sqrt{T}). \tag{2.F.9}$$

Also,

$$\frac{1}{n}\sum_{i=1}^{n}\left(Q_{1i}^W - E\left[Q_{1i}^W\right]\right) = o_p(1), \tag{2.F.10}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Q_{1i}^W - E\left[Q_{1i}^W\right]\right) = O_p(1). \tag{2.F.11}$$

Proof. By a second order Taylor expansion of the FOC for $\tilde{\gamma}_{i0}$, we have

$$0 = \hat{t}_i^W(\tilde{\gamma}_{i0}) = \hat{t}_i^W(\gamma_{i0}) + \hat{T}_i(\tilde{\gamma}_{i0})(\tilde{\gamma}_{i0} - \gamma_{i0}) + \frac{1}{2}\sum_{j=1}^{m+1}(\tilde{\gamma}_{i0,j} - \gamma_{i0,j})\hat{T}_{i,j}(\overline{\gamma}_i)(\tilde{\gamma}_{i0} - \gamma_{i0}), \tag{2.F.12}$$

where $\overline{\gamma}_i$ is between $\tilde{\gamma}_{i0}$ and γ_{i0}. The expression for Q_{1i}^W can be obtained in a similar fashion as in Lemma A4 in Newey and Smith (2004). The rest of the properties for Q_{1i}^W follow by Lemmas 12, 17, 40 and 59.

For the remainder term, we have

$$\begin{aligned}
R_{2i}^{W} &= -\left(T_i^W\right)^{-1}\left[\tilde{A}_i^W R_{1i}^W + \frac{1}{2}\sum_{j=1}^{m+1}\left[R_{1i,j}^W T_{i,j}^W \sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) + \tilde{\psi}_{i,j}^W T_{i,j}^W R_{1i}^W\right]\right] \\
&+ -\left(T_i^W\right)^{-1}\left[\frac{1}{2}\sum_{j=1}^{m+1}\sqrt{T}(\tilde{\gamma}_{i0,j}-\gamma_{i0,j})\sqrt{T}(\hat{T}_{i,j}^W(\overline{\gamma}_i) - T_{i,j}^W)\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0})\right] \\
&= o_{up}(T^{3r/2s}) = o_{up}(\sqrt{T}). \qquad (2.F.13)
\end{aligned}$$

The uniform rate of convergence follows by Lemmas 12, 17, 15 and 40, and Conditions 3 and 4. ■

Lemma 42 *Suppose that Conditions 1, 2, 3, and 4 hold. We then have*

$$\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) = \tilde{\psi}_i^W + \frac{1}{\sqrt{T}}Q_{1i}^W + \frac{1}{T}Q_{2i}^W + \frac{1}{T^{3/2}}R_{3i}^W, \qquad (2.F.14)$$

where

$$\begin{aligned}
Q_{2i}^{W} &= -\left(T_i^W\right)^{-1}\left[\tilde{A}_i^W Q_{1i}^W + \frac{1}{2}\sum_{j=1}^{m+1}\left[\tilde{\psi}_{i,j}^W T_{i,j}^W Q_{1i}^W + Q_{1i,j}^W T_{i,j}^W \tilde{\psi}_i^W + \tilde{\psi}_{i,j}^W \tilde{B}_{i,j}^W \tilde{\psi}_i^W\right]\right] \\
&+ -\left(T_i^W\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\tilde{\psi}_{i,j}^W\tilde{\psi}_{i,k}^W T_{i,jk}^W \tilde{\psi}_i^W\right] = o_{up}(T^{3r/2s}), \qquad (2.F.15) \\
\tilde{B}_{i,j}^W &= \sqrt{T}(\hat{T}_{i,j}^W - T_{i,j}^W) = o_{up}(T^{r/2s}), \qquad (2.F.16) \\
R_{3i}^W &= o_{up}(T^{2r/s}) = o_{up}(\sqrt{T}). \qquad (2.F.17)
\end{aligned}$$

Also,

$$\begin{aligned}
\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}Q_{2i}^W &= o_p(1), \qquad (2.F.18) \\
\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}R_{3i}^W &= o_p(1). \qquad (2.F.19)
\end{aligned}$$

Proof. By a third order Taylor expansion of the FOC for $\tilde{\gamma}_{i0}$, we have

$$\begin{aligned}
0 = \hat{t}_i^W(\tilde{\gamma}_{i0}) &= \hat{t}_i^W + \hat{T}_i^W(\tilde{\gamma}_{i0})(\tilde{\gamma}_{i0}-\gamma_{i0}) + \frac{1}{2}\sum_{j=1}^{m+1}(\tilde{\gamma}_{i0,j}-\gamma_{i0,j})\hat{T}_{i,j}^W(\tilde{\gamma}_{i0}-\gamma_{i0}) \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}(\tilde{\gamma}_{i0,j}-\gamma_{i0,j})(\tilde{\gamma}_{i0,k}-\gamma_{i0,k})\hat{T}_{i,jk}^W(\overline{\gamma}_i)(\tilde{\gamma}_{i0}-\gamma_{i0}), \qquad (2.F.20)
\end{aligned}$$

where $\overline{\gamma}_i$ lies between $\tilde{\gamma}_{i0}$ and γ_{i0}. The expression for Q_{2i}^W can be obtained in a similar fashion as in Lemma A4 in Newey and Smith (2004). The rest of the properties for Q_{2i}^W follow by Lemmas 12, 17, 40, 41 and 60. For the remainder term, we have

$$\begin{aligned}
R_{3i}^W &= -\left(T_i^W\right)^{-1}\left[\tilde{A}_i^W R_{2i}^W + \frac{1}{2}\sum_{j=1}^{m+1}\left[R_{2i,j}^W T_{i,j}^W \sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) + \tilde{\psi}_{i,j}^W T_{i,j}^W R_{2i}^W\right]\right] \\
&+ -\left(T_i^W\right)^{-1}\left[\frac{1}{2}\sum_{j=1}^{m+1}\left[Q_{1i,j}^W T_{i,j}^W R_{1i}^W + R_{1i,j}^W \tilde{B}_{i,j}^W \sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) + \tilde{\psi}_{i,j}^W \tilde{B}_{i,j}^W R_{1i}^W\right]\right] \\
&+ -\left(T_i^W\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[R_{1i,j}^W \sqrt{T}(\tilde{\gamma}_{i0,k}-\gamma_{i0,k}) T_{i,jk}^W \sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) + \tilde{\psi}_{i,j}^W R_{1i,k}^W T_{i,jk}^W \sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0})\right]\right] \\
&+ -\left(T_i^W\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\tilde{\psi}_{i,j}^W \tilde{\psi}_{i,k}^W T_{i,jk}^W R_{1i}^W\right] \\
&+ -\left(T_i^W\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sqrt{T}(\tilde{\gamma}_{i0,j}-\gamma_{i0,j})\sqrt{T}(\tilde{\gamma}_{i0,k}-\gamma_{i0,k})\sqrt{T}(\hat{T}_{i,jk}^W(\overline{\gamma}_i) - T_{i,jk}^W)\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0})\right] \\
&= o_{up}(T^{2r/s}) = o_{up}(\sqrt{T}). \qquad (2.F.21)
\end{aligned}$$

The uniform rate of convergence then follows by Lemmas 12, 17, 15, 40 and 41, and Conditions 3 and 4. ∎

Lemma 43 *Suppose that Conditions 1, 2, 3, and 4 hold. We then have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i^W \xrightarrow{d} N(0, V^W), \qquad (2.F.22)$$

$$V^W = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} H_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} & H_{\alpha_i}^W \Omega_i P_{\alpha_i}^W \\ P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} & P_{\alpha_i}^W \Omega_i P_{\alpha_i}^W \end{pmatrix}, \qquad (2.F.23)$$

$$\frac{1}{n}\sum_{i=1}^{n} Q_{1i}^W \xrightarrow{p} B_\gamma^W = \lim_{n\to\infty}\frac{1}{n}E\left[Q_{1i}^W\right] = \lim_{n\to\infty}\frac{1}{n}B_{\gamma_i}^W, \qquad (2.F.24)$$

$$B_{\gamma_i}^W = B_{\gamma_i}^{W,I} + B_{\gamma_i}^{W,G} + B_{\gamma_i}^{W,1S}, \qquad (2.F.25)$$

$$B_{\gamma_i}^{W,I} = \begin{pmatrix} B_{\alpha_i}^{W,I} \\ B_{\lambda_i}^{W,I} \end{pmatrix} = \begin{pmatrix} H_{\alpha_i}^W E\left[G_{\alpha_i}(z_{it}) H_{\alpha_i}^W g(z_{it})\right] - \frac{1}{2} H_{\alpha_i}^W G_{\alpha\alpha_i} H_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} \\ P_{\alpha_i}^W E\left[G_{\alpha_i}(z_{it}) H_{\alpha_i}^W g(z_{it})\right] - \frac{1}{2} P_{\alpha_i}^W G_{\alpha\alpha_i} H_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} \end{pmatrix}, \qquad (2.F.26)$$

$$B_{\gamma_i}^{W,G} = \begin{pmatrix} B_{\alpha_i}^{W,G} \\ B_{\lambda_i}^{W,G} \end{pmatrix} = \begin{pmatrix} -\Sigma_{\alpha_i}^W E\left[G_{\alpha_i}(z_{it})' P_{\alpha_i}^W g(z_{it})\right] \\ H_{\alpha_i}^{W\prime} E\left[G_{\alpha_i}(z_{it})' P_{\alpha_i}^W g(z_{it})\right] \end{pmatrix}, \qquad (2.F.27)$$

$$B_{\gamma_i}^{W,1S} = \begin{pmatrix} B_{\alpha_i}^{W,1S} \\ B_{\lambda_i}^{W,1S} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\Sigma_{\alpha_i}^W G_{\alpha\alpha_i}' P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} + \frac{1}{2}\Sigma_{\alpha_i}^W \sum_{j=1}^{m} G_{\alpha\alpha_i}' e_j H_{\alpha_i}^W \Omega_i P_{\alpha_i,j}^W \\ -\frac{1}{2} H_{\alpha_i}^{W\prime} G_{\alpha\alpha_i}' P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} - \frac{1}{2} H_{\alpha_i}^{W\prime} \sum_{j=1}^{m} G_{\alpha\alpha_i}' e_j H_{\alpha_i}^W \Omega_i P_{\alpha_i,j}^W \end{pmatrix}, \qquad (2.F.28)$$

where

$$
\begin{aligned}
\Sigma_{\alpha_i}^W &= \left(G'_{\alpha_i} W_i^{-1} G_{\alpha_i}\right)^{-1}, && (2.F.29)\\
H_{\alpha_i}^W &= \Sigma_{\alpha_i}^W G'_{\alpha_i} W_i^{-1}, && (2.F.30)\\
P_{\alpha_i}^W &= W_i^{-1} - W_i^{-1} G_{\alpha_i} H_{\alpha_i}^W. && (2.F.31)
\end{aligned}
$$

Proof. The results follow by Lemmas 40 and 41, noting that

$$
\begin{aligned}
\left(T_i^W\right)^{-1} &= -\begin{pmatrix} -\Sigma_{\alpha_i}^W & H_{\alpha_i}^W \\ H_{\alpha_i}^{W\prime} & P_{\alpha_i}^W \end{pmatrix}, && (2.F.32)\\
\psi_{it}^W &= -\begin{pmatrix} H_{\alpha_i}^W \\ P_{\alpha_i}^W \end{pmatrix} g(z_{it}), && (2.F.33)\\
E\left[\psi_i^W \psi_i^{W\prime}\right] &= \begin{pmatrix} H_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} & H_{\alpha_i}^W \Omega_i P_{\alpha_i}^W \\ P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} & P_{\alpha_i}^W \Omega_i P_{\alpha_i}^W \end{pmatrix}, && (2.F.34)\\
E\left[\tilde{A}_i^W \tilde{\psi}_i^W\right] &= \begin{pmatrix} E\left[G_{\alpha_i}(z_{it})' P_{\alpha_i}^W g(z_{it})\right] \\ E\left[G_{\alpha_i}(z_{it}) H_{\alpha_i}^W g(z_{it})\right] \end{pmatrix}, && (2.F.35)\\
E\left[\tilde{\psi}_{i,j}^W T_{i,j}^W \tilde{\psi}_i^W\right] &= \begin{cases} -\begin{pmatrix} G'_{\alpha\alpha_i} P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} \\ G'_{\alpha\alpha_i} H_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} \end{pmatrix}, & \text{if } j = 1; \\ G'_{\alpha\alpha_i} e_{j-1} H_{\alpha_i}^W \Omega_i P_{\alpha_i,j}^W, & \text{if } j > 1. \end{cases} && (2.F.36)
\end{aligned}
$$

■

2.G Stochastic Expansion for $\hat{s}_i^W(\theta_0, \tilde{\gamma}_{i0})$

Lemma 44 *Suppose that Conditions 1, 2, 3, and 4 hold. We then have*

$$
\hat{s}_i^W(\theta_0, \tilde{\gamma}_{i0}) = \frac{1}{\sqrt{T}} \tilde{\psi}_{si}^W + \frac{1}{T} Q_{1si}^W + \frac{1}{T^{3/2}} Q_{2si}^W + \frac{1}{T^2} R_{3si}^W, \tag{2.G.1}
$$

where

$$
\begin{aligned}
\tilde{\psi}_{si}^{W} &= M_i^W \tilde{\psi}_i^W = o_{up}(T^{r/2s}), && (2.G.2)\\
Q_{1si}^{W} &= M_i^W Q_i^W + \tilde{C}_i^W \tilde{\psi}_i^W + \frac{1}{2}\sum_{j=1}^{m+1} \tilde{\psi}_{i,j}^W M_{i,j}^W \tilde{\psi}_i^W = o_{up}(T^{r/s}), && (2.G.3)\\
\tilde{C}_i^W &= \sqrt{T}(\hat{M}_i^W - M_i^W) = o_{up}(T^{r/2s}), && (2.G.4)\\
Q_{2si}^{W} &= M_i^W Q_{2i}^W + \tilde{C}_i^W Q_{1i}^W + \frac{1}{2}\sum_{j=1}^{m+1} \left[\tilde{\psi}_{i,j}^W M_{i,j}^W Q_{1i}^W + Q_{1i,j}^W M_{i,j}^W \tilde{\psi}_i^W + \tilde{\psi}_{i,j}^W \tilde{D}_{i,j}^W \tilde{\psi}_i^W\right] \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1} \tilde{\psi}_{i,j}^W \tilde{\psi}_{i,k}^W M_{i,jk}^W \tilde{\psi}_i^W = o_{up}(T^{3r/2s}), && (2.G.5)\\
\tilde{D}_{i,j}^{W} &= \sqrt{T}(\hat{M}_{i,j}^W - M_{i,j}^W) = o_{up}(T^{r/2s}), && (2.G.6)\\
R_{3si}^{W} &= o_{up}(T^{2r/s}) = o_{up}(\sqrt{T}). && (2.G.7)
\end{aligned}
$$

Also,

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \tilde{\psi}_{si}^W &= O_p(1), && (2.G.8)\\
\frac{1}{n}\sum_{i=1}^{n} \left(Q_{1si}^W - E\left[Q_{1si}^W\right]\right) &= o_p(1), && (2.G.9)\\
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(Q_{1si}^W - E\left[Q_{1si}^W\right]\right) &= O_p(1), && (2.G.10)\\
\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n} Q_{2si}^W &= o_p(1), && (2.G.11)\\
\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sqrt{T}} R_{3si}^W &= o_p(1). && (2.G.12)
\end{aligned}
$$

Proof. By a third order Taylor expansion of $\hat{s}_i^W(\theta_0, \tilde{\gamma}_{i0})$, we have

$$
\begin{aligned}
\hat{s}_i^W(\theta_0, \tilde{\gamma}_{i0}) &= \hat{s}_i^W(\theta_0, \gamma_{i0}) + \hat{M}_i(\tilde{\gamma}_{i0})(\tilde{\gamma}_{i0} - \gamma_{i0}) + \frac{1}{2}\sum_{j=1}^{m+1}(\tilde{\gamma}_{i0,j} - \gamma_{i0,j})\hat{M}_{i,j}(\tilde{\gamma}_{i0} - \gamma_{i0}) \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}(\tilde{\gamma}_{i0,j} - \gamma_{i0,j})(\tilde{\gamma}_{i0,k} - \gamma_{i0,k})\hat{M}_{i,jk}^W(\overline{\gamma}_i)(\tilde{\gamma}_{i0} - \gamma_{i0}), && (2.G.13)
\end{aligned}
$$

where $\overline{\gamma}_i$ is between $\tilde{\gamma}_{i0}$ and γ_{i0}. Noting that $\hat{s}_i^W(\theta_0, \gamma_{i0}) = 0$ and using the expansion for $\tilde{\gamma}_{i0}$ in Lemma 42, we can obtain the expressions for $\tilde{\psi}_{si}^W$, Q_{1si}^W, and Q_{2si}^W, after some algebra. The rest of the properties for these terms follow by the properties of $\tilde{\psi}_i^W$, Q_{1i}^W, and Q_{2i}^W, and Lemmas 58, 59 and 60. For the remainder term, we have

$$\begin{aligned}
R_{3si}^{W} &= M_i^{W} R_{3i}^{W} + \tilde{C}_i^{W} R_{2i}^{W} + \frac{1}{2}\sum_{j=1}^{m+1}\left[R_{2i,j}^{W} M_{i,j}^{W}\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) + \tilde{\psi}_{i,j}^{W} M_{i,j}^{W} R_{2i}^{W}\right] \\
&+ \frac{1}{2}\sum_{j=1}^{m+1}\left[Q_{1i,j}^{W} M_{i,j}^{W} R_{1i}^{W} + R_{1i,j}^{W}\tilde{D}_{i,j}^{W}\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) + \tilde{\psi}_{i,j}^{W}\tilde{D}_{i,j}^{W} R_{1i}^{W}\right] \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[R_{1i,j}^{W}\sqrt{T}(\tilde{\gamma}_{i0,k}-\gamma_{i0,k}) M_{i,jk}^{W}\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}) + \tilde{\psi}_{i,j}^{W} R_{1i,k}^{W} M_{i,jk}^{W}\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0})\right] \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\tilde{\psi}_{i,j}^{W}\tilde{\psi}_{i,k}^{W} M_{i,jk}^{W} R_{1i}^{W} \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sqrt{T}(\tilde{\gamma}_{i0,j}-\gamma_{i0,j})\sqrt{T}(\tilde{\gamma}_{i0,k}-\gamma_{i0,k})\sqrt{T}(\hat{M}_{i,jk}^{W}(\overline{\gamma}_i) - M_{i,jk}^{W})\sqrt{T}(\tilde{\gamma}_{i0}-\gamma_{i0}).
\end{aligned} \tag{2.G.14}$$

Then, the results for R_{3si}^{W} follow by the properties of the components in the expansion of $\tilde{\gamma}_{i0}$, Lemmas 12, 17 and 59, and Conditions 3 and 4. ■

Lemma 45 *Suppose that Conditions 1, 2, 3, and 4 hold. We then have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{si}^{W} \xrightarrow{d} N(0, V_s^{W}), \tag{2.G.15}$$

$$V_s^{W} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} G_{\theta_i}' P_{\alpha_i}^{W}\Omega_i P_{\alpha_i}^{W} G_{\theta_i} \tag{2.G.16}$$

$$\frac{1}{n}\sum_{i=1}^{n} Q_{1i}^{W} \xrightarrow{p} B_s^{W} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E\left[Q_{1i}^{W}\right] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} B_{si}^{W}, \tag{2.G.17}$$

$$B_{si}^{W} = B_{si}^{W,B} + B_{si}^{W,C} + B_{si}^{W,V}, \tag{2.G.18}$$

$$B_{si}^{W,B} = -G_{\theta_i}' B_{\lambda_i}^{W} = -G_{\theta_i}'\left(B_{\lambda_i}^{W,I} + B_{\lambda_i}^{W,G} + B_{\lambda_i}^{W,1S}\right), \tag{2.G.19}$$

$$B_{si}^{W,C} = E\left[G_{\theta_i}(z_{it})' P_{\alpha_i}^{W} g(z_{it})\right], \tag{2.G.20}$$

$$B_{si}^{W,V} = -\frac{1}{2} G_{\alpha\theta_i}' P_{\alpha_i}^{W}\Omega_i H_{\alpha_i}^{W\prime} - \frac{1}{2}\sum_{j=1}^{m} G_{\alpha\theta_i}' e_j H_{\alpha_i}^{W}\Omega_i P_{\alpha_i,j}^{W}, \tag{2.G.21}$$

where

$$\Sigma_{\alpha_i}^{W} = \left(G_{\alpha_i}' W_i^{-1} G_{\alpha_i}\right)^{-1}, \tag{2.G.22}$$

$$H_{\alpha_i}^{W} = \Sigma_{\alpha_i}^{W} G_{\alpha_i}' W_i^{-1}, \tag{2.G.23}$$

$$P_{\alpha_i}^{W} = W_i^{-1} - W_i^{-1} G_{\alpha_i} H_{\alpha_i}^{W}. \tag{2.G.24}$$

Proof. The results follow by Lemmas 44 and 43, noting that

$$
\begin{aligned}
E\left[\psi_{si}^{W}\psi_{si}^{W\prime}\right] &= M_i^W \begin{pmatrix} H_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} & H_{\alpha_i}^W \Omega_i P_{\alpha_i}^W \\ P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime} & P_{\alpha_i}^W \Omega_i P_{\alpha_i}^W \end{pmatrix} M_i^{W\prime}, && (2.G.25)\\
E\left[\tilde{C}_i^W \tilde{\psi}_i^W\right] &= E\left[G_{\theta_i}(z_{it})' P_{\alpha_i}^W g(z_{it})\right], && (2.G.26)\\
E\left[\tilde{\psi}_{i,j}^W M_{i,j}^W \tilde{\psi}_i^W\right] &= \begin{cases} -G'_{\alpha\theta_i} P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime}, & \text{if } j=1; \\ -G'_{\alpha\theta_i} e_{j-1} H_{\alpha_i}^W \Omega_i P_{\alpha_i,j}^W, & \text{if } j>1. \end{cases} && (2.G.27)
\end{aligned}
$$

■

Lemma 46 *Suppose that Conditions 1, 2, 3, and 4 hold. We then have*

$$
\begin{cases} \sqrt{nT}\hat{s}_n^W(\theta_0) \xrightarrow{d} N\left(\sqrt{\rho}B_s^W, V_s^W\right), & \textit{if } n=\rho T; \\ T\hat{s}_n^W(\theta_0) \xrightarrow{p} B_s^W, & \textit{otherwise}; \end{cases} \tag{2.G.28}
$$

where

$$
\hat{s}_n^W(\theta_0) = \frac{1}{n}\sum_{i=1}^n \hat{s}_i^W(\theta_0, \tilde{\gamma}_{i0}), \tag{2.G.29}
$$

and B_s^W and V_s^W are defined in Lemma 45.

Proof. Case I: $n = O(T)$ From Lemma 44, we have

$$
\begin{aligned}
\sqrt{nT}\hat{s}_n^W(\theta_0) &= \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\psi}_{si}^W}_{=O_p(1)} + \underbrace{\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^n Q_{1si}^W}_{=O_p(1)} + \underbrace{\sqrt{\frac{n}{T^2}}\frac{1}{n}\sum_{i=1}^n Q_{2si}^W}_{=o_p(1)} + \underbrace{\sqrt{\frac{n}{T^3}}\frac{1}{n}\sum_{i=1}^n R_{3si}^W}_{=o_p(1)} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\psi}_{si}^W + \sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^n Q_{1si}^W + o_p(1). && (2.G.30)
\end{aligned}
$$

Then, the result follows by Lemma 45.

Case II: $T = o(n)$ Similarly to the previous case, we have from Lemma 44

$$
\begin{aligned}
T\hat{s}_n^W(\theta_0) &= \underbrace{\sqrt{\frac{T}{n}}\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\psi}_{si}^W}_{=o_p(1)} + \underbrace{\frac{1}{n}\sum_{i=1}^n Q_{1si}^W}_{=O_p(1)} + \underbrace{\sqrt{\frac{a_T}{T^2}}\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^n Q_{2si}^W}_{=o_p(1)} + \underbrace{\frac{1}{\sqrt{T}}\frac{1}{n}\sum_{i=1}^n \frac{1}{\sqrt{T}}R_{3si}^W}_{=o_p(1)} \\
&= \frac{1}{n}\sum_{i=1}^n Q_{1si}^W + o_p(1). && (2.G.31)
\end{aligned}
$$

Then, the result follows by Lemma 45. ■

2.H Stochastic Expansion for $\hat{\gamma}_{i0} = \hat{\gamma}_i(\theta_0)$

Lemma 47 *Suppose that Conditions 1, 2, 3, 4, 5, and 6 hold. We then have*

$$\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) = \tilde{\psi}_i + \frac{1}{\sqrt{T}} R_{1i}, \tag{2.H.1}$$

where

$$\begin{aligned}
\tilde{\psi}_i &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_{it} = -\left(T_i^{\Omega}\right)^{-1} \sqrt{T} \hat{t}_i^{\Omega}(\gamma_{i0}) = o_{up}\left(T^{r/2s}\right), &(2.H.2)\\
R_{1i} &= o_{up}\left(\sqrt{T}\right). &(2.H.3)
\end{aligned}$$

Also

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_i = O_p(1) \tag{2.H.4}$$

Proof. The statements about $\tilde{\psi}_i$ follow by Lemmas 26 and 58. From the proof of Lemma 27, we have

$$\begin{aligned}
R_{1i} &= -\underbrace{\left(T_i^{\Omega}\right)^{-1}}_{=O_u(1)} \underbrace{\sqrt{T}(\hat{T}_i^{\Omega}(\overline{\gamma}_i) - T_i^{\Omega})}_{=o_u(T^{r/2s})} \underbrace{\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})}_{=o_u(T^{r/2s})} - \underbrace{\left(T_i^{\Omega}\right)^{-1}}_{=O_u(1)} \underbrace{\sqrt{T}(\hat{T}_i^{R}(\overline{\gamma}_i) - T_i^{R})}_{=o_u(T^{r/2s})} \underbrace{\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})}_{=o_u(T^{r/2s})}\\
&= o_{up}(T^{r/s}) = o_{up}(\sqrt{T}), &(2.H.5)
\end{aligned}$$

by Lemmas 12, 29 and 27, and Conditions 3 and 4. ∎

Lemma 48 *Suppose that Conditions 1, 2, 3, 5, and 4 hold. We then have*

$$\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) = \tilde{\psi}_i + \frac{1}{\sqrt{T}} Q_{1i} + \frac{1}{T} R_{2i}, \tag{2.H.6}$$

where

$$\begin{aligned}
Q_{1i}(\tilde{\psi}_i, \tilde{a}_i) &= -\left(T_i^{\Omega}\right)^{-1}\left[\tilde{A}_i^{\Omega}\tilde{\psi}_i + \frac{1}{2}\sum_{j=1}^{m+1} \tilde{\psi}_{i,j} T_{i,j}^{\Omega} \tilde{\psi}_i\right] - \left(T_i^{\Omega}\right)^{-1} diag[0, \tilde{\psi}_{\Omega_i}^{W}]\tilde{\psi}_i = o_{up}\left(T^{r/s}\right) &(2.H.7)\\
\tilde{A}_i^{\Omega} &= \sqrt{T}(\hat{T}_i^{\Omega} - T_i^{\Omega}) = o_{up}\left(T^{r/2s}\right) &(2.H.8)\\
R_{2i} &= o_{up}\left(T^{3r/2s}\right) = o_{up}\left(\sqrt{T}\right). &(2.H.9)
\end{aligned}$$

Also,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(Q_{1i} - E[Q_{1i}]) &= o_p(1), &(2.H.10)\\
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Q_{1i} - E[Q_{1i}]) &= O_p(1). &(2.H.11)
\end{aligned}$$

Proof. By a second order Taylor expansion of the FOC for $\hat{\gamma}_{i0}$, we have

$$0 = \hat{t}_i(\hat{\gamma}_{i0}) = \hat{t}_i^{\Omega}(\gamma_{i0}) + \hat{T}_i(\hat{\gamma}_{i0} - \gamma_{i0}) + \frac{1}{2}\sum_{j=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})\hat{T}_{i,j}(\overline{\gamma}_i)(\hat{\gamma}_{i0} - \gamma_{i0}), \tag{2.H.12}$$

where $\overline{\gamma}_i$ is between $\hat{\gamma}_{i0}$ and γ_{i0}. The expression for Q_{1i} can be obtained in a similar fashion as in Lemma A4 in Newey-Smith (2003). The rest of the properties for Q_{1i} follow by Lemmas 29,47 and 59. For the remainder term, we have

$$\begin{aligned} R_{2i} &= -\left(T_i^{\Omega}\right)^{-1}\left[\tilde{A}_i^{\Omega} R_{1i} + \frac{1}{2}\sum_{j=1}^{m+1}\left[R_{1i,j}T_{i,j}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}T_{i,j}^{\Omega}R_{1i}\right]\right] \\ &+ -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{2}\sum_{j=1}^{m+1}\sqrt{T}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})\sqrt{T}(\hat{T}_{i,j}^{\Omega}(\overline{\gamma}_i) - T_{i,j}^{\Omega})\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})\right] \\ &+ -\left(T_i^{\Omega}\right)^{-1}\left[diag[0, R_{1\Omega_i}^{W}]\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + diag[0, \tilde{\psi}_{\Omega_i}^{W}]R_{1i}\right]. \end{aligned} \tag{2.H.13}$$

The uniform rate of convergence then follows by Lemmas 12, 27, 29 and 47, and Conditions 3 and 4. ∎

Lemma 49 *Suppose that Conditions 1, 2, 3, 4, 5, and 6 hold. We then have*

$$\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) = \tilde{\psi}_i + \frac{1}{\sqrt{T}}Q_{1i} + \frac{1}{T}Q_{2i} + \frac{a_T}{T^{(r+1)/2}}Q_{2ri} + \frac{1}{T^{3/2}}R_{3i}, \tag{2.H.14}$$

where

$$a_T = \begin{cases} C, & \text{if } n = O(T^a) \text{ for some } a \in \Re; \\ o(T^{\epsilon}) \text{ for any } \epsilon > 0, & \text{otherwise,} \end{cases} \tag{2.H.15}$$

$$\begin{aligned} Q_{2i} &= -\left(T_i^{\Omega}\right)^{-1}\left[\tilde{A}_i^{\Omega}Q_{1i} + \frac{1}{2}\sum_{j=1}^{m+1}\left[\tilde{\psi}_{i,j}T_{i,j}^{\Omega}Q_{1i} + Q_{1i,j}T_{i,j}^{\Omega}\tilde{\psi}_i + \tilde{\psi}_{i,j}\tilde{B}_{i,j}^{\Omega}\tilde{\psi}_i\right]\right] \\ &+ -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}T_{i,jk}^{\Omega}\tilde{\psi}_i\right] \\ &+ -\left(T_i^{\Omega}\right)^{-1}\left[diag[0, \tilde{\psi}_{\Omega_i}^{W}]Q_{1i} + diag[0, Q_{1\Omega_i}^{W}]\tilde{\psi}_i\right] = o_{up}\left(T^{3r/2s}\right), \end{aligned} \tag{2.H.16}$$

$$\tilde{B}_{i,j}^{\Omega} = \sqrt{T}(\hat{T}_{i,j}^{\Omega} - T_{i,j}^{\Omega}) = O_{up}\left(T^{r/2s}\right) \tag{2.H.17}$$

$$Q_{2ri} = -\left(T_i^{\Omega}\right)^{-1} diag[0, Q_{1r\Omega_i}^{W}]\tilde{\psi}_i = o_{up}\left(T^{r/2s}\right), \tag{2.H.18}$$

$$R_{3i} = o_{up}\left(T^{2r/s}\right) = o_{up}\left(\sqrt{T}\right). \tag{2.H.19}$$

Also,

$$\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n} Q_{2i} = O_p(1), \tag{2.H.20}$$

$$\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n} Q_{2ri} = O_p(1). \tag{2.H.21}$$

Proof. By a third order Taylor expansion of the FOC for $\hat{\gamma}_{i0}$, we have

$$\begin{aligned} 0 = \hat{t}_i(\hat{\gamma}_{i0}) &= \hat{t}_i^{\Omega}(\gamma_{i0}) + \hat{T}_i(\hat{\gamma}_{i0} - \gamma_{i0}) + \frac{1}{2}\sum_{j=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})\hat{T}_{i,j}^{\Omega}(\hat{\gamma}_{i0} - \gamma_{i0}) \\ &+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})(\hat{\gamma}_{i0,k} - \gamma_{i0,k})\hat{T}_{i,jk}^{\Omega}(\overline{\gamma}_i)(\hat{\gamma}_{i0} - \gamma_{i0}), \end{aligned} \tag{2.H.22}$$

where $\overline{\gamma}_i$ is between $\hat{\gamma}_{i0}$ and γ_{i0}. The expressions for Q_{2i}^W and Q_{2ri}^W can be obtained in a similar fashion as in Lemma A4 in Newey and Smith (2004). The rest of the properties for Q_{2i}^W and Q_{2i}^W follow by Lemmas 29, 47, 48 and 60. For the remainder term, we have

$$\begin{aligned} R_{3i} &= -\left(T_i^{\Omega}\right)^{-1}\left[\tilde{A}_i^{\Omega}R_{2i} + \frac{1}{2}\sum_{j=1}^{m+1}\left[R_{2i,j}T_{i,j}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}T_{i,j}^{\Omega}R_{2i}\right]\right] \\ &+ -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{2}\sum_{j=1}^{m+1}\left[Q_{1i,j}T_{i,j}^{\Omega}R_{1i} + R_{1i,j}\tilde{B}_{i,j}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}\tilde{B}_{i,j}^{\Omega}R_{1i}\right]\right] \\ &+ -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[R_{1i,j}\sqrt{T}(\hat{\gamma}_{i0,k} - \gamma_{i0,k})T_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}R_{1i,k}T_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})\right]\right] \\ &+ -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}T_{i,jk}^{\Omega}R_{1i}\right] \\ &+ -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sqrt{T}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})\sqrt{T}(\hat{\gamma}_{i0,k} - \gamma_{i0,k})\sqrt{T}(\hat{T}_{i,jk}^{\Omega}(\overline{\gamma}_i) - T_{i,jk}^{\Omega})\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})\right] \\ &+ -\left(T_i^{\Omega}\right)^{-1}\left[diag[0, \tilde{\psi}_{\Omega_i}^W]R_{2i} + diag[0, Q_{1\Omega_i}^W]R_{1i} + diag[0, R_{2\Omega_i}^W]\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})\right]. \end{aligned} \tag{2.H.23}$$

The uniform rate of convergence then follows by Lemmas 12, 29, 27, 47 and 48, and Conditions 3 and 4. ■

Lemma 50 *Suppose that Conditions 1, 2, 3, 4, 5, and 6 hold. We then have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{it} \xrightarrow{d} N(0,V), \tag{2.H.24}$$

$$V = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\Sigma_{\alpha_i} & 0\\ 0 & P_{\alpha_i}\end{pmatrix}, \tag{2.H.25}$$

$$\frac{1}{n}\sum_{i=1}^{n}Q_{1i} \xrightarrow{p} B_\gamma = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left[Q_{1i}\right] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}B_{\gamma_i}, \tag{2.H.26}$$

$$B_{\gamma_i} = B^I_{\gamma_i} + B^G_{\gamma_i} + B^\Omega_{\gamma_i} + B^W_{\gamma_i}, \tag{2.H.27}$$

$$B^I_{\gamma_i} = \begin{pmatrix} B^I_{\alpha_i}\\ B^I_{\lambda_i}\end{pmatrix} = \begin{pmatrix} -\frac{1}{2}H_{\alpha_i}G_{\alpha\alpha_i}\Sigma_{\alpha_i} + H_{\alpha_i}E\left[G_{\alpha_i}(z_{it})H_{\alpha_i}g(z_{it})\right]\\ -\frac{1}{2}P_{\alpha_i}G_{\alpha\alpha_i}\Sigma_{\alpha_i} + P_{\alpha_i}E\left[G_{\alpha_i}(z_{it})H_{\alpha_i}g(z_{it})\right]\end{pmatrix}, \tag{2.H.28}$$

$$B^G_{\gamma_i} = \begin{pmatrix} B^G_{\alpha_i}\\ B^G_{\lambda_i}\end{pmatrix} = \begin{pmatrix} -\Sigma_{\alpha_i}E\left[G_{\alpha_i}(z_{it})'P_{\alpha_i}g(z_{it})\right]\\ H'_{\alpha_i}E\left[G_{\alpha_i}(z_{it})'P_{\alpha_i}g(z_{it})\right]\end{pmatrix}, \tag{2.H.29}$$

$$B^\Omega_{\gamma_i} = \begin{pmatrix} B^\Omega_{\alpha_i}\\ B^\Omega_{\lambda_i}\end{pmatrix} = \begin{pmatrix} H_{\alpha_i}E[g(z_{it})g(z_{it})'P_{\alpha_i}g(z_{it})]\\ P_{\alpha_i}E[g(z_{it})g(z_{it})'P_{\alpha_i}g(z_{it})]\end{pmatrix}, \tag{2.H.30}$$

$$B^W_{\gamma_i} = \begin{pmatrix} B^W_{\alpha_i}\\ B^W_{\lambda_i}\end{pmatrix} = \begin{pmatrix} H_{\alpha_i}\Omega_{\alpha_i}\left(H^W_{\alpha_i} - H_{\alpha_i}\right)\\ P_{\alpha_i}\Omega_{\alpha_i}\left(H^W_{\alpha_i} - H_{\alpha_i}\right)\end{pmatrix}, \tag{2.H.31}$$

where

$$\Sigma_{\alpha_i} = \left(G'_{\alpha_i}\Omega_i^{-1}G_{\alpha_i}\right)^{-1}, \tag{2.H.32}$$

$$H_{\alpha_i} = \Sigma_{\alpha_i}G'_{\alpha_i}\Omega_i^{-1}, \tag{2.H.33}$$

$$P_{\alpha_i} = \Omega_i^{-1} - \Omega_i^{-1}G_{\alpha_i}H_{\alpha_i}. \tag{2.H.34}$$

Proof. The results follow by Lemmas 47 and 48, noting that

$$\left(T_i^\Omega\right)^{-1} = -\begin{pmatrix} -\Sigma_{\alpha_i} & H_{\alpha_i}\\ H'_{\alpha_i} & P_{\alpha_i}\end{pmatrix}, \tag{2.H.35}$$

$$\psi_{it} = -\begin{pmatrix} H_{\alpha_i}\\ P_{\alpha_i}\end{pmatrix} g(z_{it}), \tag{2.H.36}$$

$$E\left[\psi_{it}\psi'_{it}\right] = \begin{pmatrix}\Sigma_{\alpha_i} & 0\\ 0 & P_{\alpha_i}\end{pmatrix}, \tag{2.H.37}$$

$$E\left[\tilde{A}_i^{\Omega}\tilde{\psi}_i\right] = \begin{pmatrix} E\left[G_{\alpha_i}(z_{it})'P_{\alpha_i}g(z_{it})\right] \\ E\left[G_{\alpha_i}(z_{it})H_{\alpha_i}g(z_{it})\right] \end{pmatrix}, \tag{2.H.38}$$

$$E\left[\tilde{\psi}_{i,j}T_{i,j}^{\Omega}\tilde{\psi}_i\right] = \begin{cases} -\begin{pmatrix} 0 \\ G'_{\alpha\alpha_i}\Sigma_{\alpha_i} \end{pmatrix}, & \text{if } j=1; \\ 0, & \text{if } j>1. \end{cases} \tag{2.H.39}$$

$$E\left[diag[0,\tilde{\psi}_{\Omega_i}^W]\tilde{\psi}_i\right] = \begin{pmatrix} 0 \\ E\left[g(z_{it})g(z_{it})'P_{\alpha_i}g(z_{it})\right] + \Omega_{\alpha_i}\left(H_{\alpha_i}^W - H_{\alpha_i}\right) \end{pmatrix}. \tag{2.H.40}$$

■

Lemma 51 *Suppose that Conditions 1, 2, 3, 4, 5 and 6 hold. We then have*

$$\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0}) = \tilde{\psi}_i + \frac{1}{\sqrt{T}}Q_{1i} + \frac{1}{T}Q_{2i} + \frac{a_T}{T^{(r+1)/2}}Q_{2ri} + \frac{1}{T^{3/2}}Q_{3i} + \frac{a_T}{T^{(r+2)/2}}Q_{3ri} + \frac{1}{T^2}R_{4i}, \tag{2.H.41}$$

where

$$a_T = \begin{cases} C, & \textit{if } n = O(T^a) \textit{ for some } a\in\Re; \\ o(T^{\epsilon}) \textit{ for any } \epsilon>0, & \textit{otherwise,} \end{cases} \tag{2.H.42}$$

$$\begin{aligned} Q_{3i} = {} & -\left(T_i^{\Omega}\right)^{-1}\left[\tilde{A}_i^{\Omega}Q_{2i} + \frac{1}{2}\sum_{j=1}^{m+1}\left[\tilde{\psi}_{i,j}T_{i,j}^{\Omega}Q_{2i} + Q_{1i,j}T_{i,j}^{\Omega}Q_{1i} + Q_{2i,j}T_{i,j}^{\Omega}\tilde{\psi}_i + \tilde{\psi}_{i,j}\tilde{B}_{i,j}^{\Omega}Q_{1i} + Q_{1i,j}\tilde{B}_{i,j}^{\Omega}\tilde{\psi}_i\right]\right] \\ & + -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}T_{i,jk}^{\Omega}Q_{1i} + \tilde{\psi}_{i,j}Q_{1i,k}T_{i,jk}^{\Omega}\tilde{\psi}_i + Q_{1i,j}\tilde{\psi}_{i,k}T_{i,jk}^{\Omega}\tilde{\psi}_i + \tilde{\psi}_{i,j}\tilde{\psi}_{i,k}\tilde{E}_{i,j}^{\Omega}\tilde{\psi}_i\right]\right] \\ & + -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{l=1}^{m+1}\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}\tilde{\psi}_{i,l}\hat{T}_{i,jkl}^{\Omega}\tilde{\psi}_i\right] \\ & + -\left(T_i^{\Omega}\right)^{-1}\left[diag[0,\tilde{\psi}_{\Omega_i}^W]Q_{2i} + diag[0,Q_{1\Omega_i}^W]Q_{1i} + diag[0,Q_{2\Omega_i}^W]\tilde{\psi}_i\right] = o_{up}\left(T^{2r/s}\right), \end{aligned} \tag{2.H.43}$$

$$\tilde{E}_{i,jk}^{\Omega} = \sqrt{T}(\hat{T}_{i,jk}^{\Omega} - T_{i,jk}^{\Omega}) = o_{up}\left(T^{r/2s}\right) \tag{2.H.44}$$

$$Q_{3ri} = -\left(T_i^{\Omega}\right)^{-1}\left[diag[0,Q_{1r\Omega_i}^W]Q_{1i} + diag[0,Q_{2r\Omega_i}^W]\tilde{\psi}_i\right] = o_{up}\left(T^{r/s}\right), \tag{2.H.45}$$

$$R_{4i} = o_{up}\left(T^{5r/2s}\right) = o_{up}\left(\sqrt{T}\right). \tag{2.H.46}$$

Also,

$$\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}Q_{3i} = O_p(1), \tag{2.H.47}$$

$$\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}Q_{3ri} = O_p(1), \tag{2.H.48}$$

$$\sqrt{\frac{n}{T^r}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}R_{4i} = o_p(1). \tag{2.H.49}$$

Proof. By a fourth order Taylor expansion of the FOC for $\hat{\gamma}_{i0}$, we have

$$\begin{aligned}
0 = \hat{t}_i(\hat{\gamma}_{i0}) &= \hat{t}_i^{\Omega}(\gamma_{i0}) + \hat{T}_i(\hat{\gamma}_{i0} - \gamma_{i0}) + \frac{1}{2}\sum_{j=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})\hat{T}_{i,j}^{\Omega}(\hat{\gamma}_{i0} - \gamma_{i0}) \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})(\hat{\gamma}_{i0,k} - \gamma_{i0,k})\hat{T}_{i,jk}^{\Omega}(\hat{\gamma}_{i0} - \gamma_{i0}) \\
&+ \frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{l=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})(\hat{\gamma}_{i0,k} - \gamma_{i0,k})(\hat{\gamma}_{i0,l} - \gamma_{i0,l})\hat{T}_{i,jkl}^{\Omega}(\overline{\gamma}_i)(\hat{\gamma}_{i0} - \gamma_{i0}),
\end{aligned} \tag{2.H.50}$$

where $\overline{\gamma}_i$ is between $\hat{\gamma}_{i0}$ and γ_{i0}. The expressions for Q_{3i} and Q_{3ri} can be obtained in a similar fashion as in Lemma A4 in Newey and Smith (2004), with an additional iteration of the procedure. The rest of the properties for Q_{3i} and Q_{3ri} follow by Lemmas 29, 47, 48, 49, 60 and 61. For the remainder term, we have

$$\begin{aligned}
R_{4i} &= -\left(T_i^{\Omega}\right)^{-1}\left[\tilde{A}_i^{\Omega}R_{3i} + \frac{1}{2}\sum_{j=1}^{m+1}\left[R_{3i,j}T_{i,j}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}T_{i,j}^{\Omega}R_{3i}\right]\right] \\
&+ -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{2}\sum_{j=1}^{m+1}\left[Q_{1i,j}T_{i,j}^{\Omega}R_{2i} + Q_{2i,j}T_{i,j}^{\Omega}R_{1i}\right]\right] \\
&+ -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{2}\sum_{j=1}^{m+1}\left[R_{2i,j}\tilde{B}_{i,j}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}\tilde{B}_{i,j}^{\Omega}R_{2i} + Q_{1i,j}\tilde{B}_{i,j}^{\Omega}R_{1i}\right]\right] \\
&+ -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[R_{2i,j}\sqrt{T}(\hat{\gamma}_{i0,k} - \gamma_{i0,k})T_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}R_{2i,k}T_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})\right]\right]
\end{aligned}$$

$$
\begin{aligned}
&+ \quad -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}T_{i,jk}^{\Omega}R_{2i}+\tilde{\psi}_{i,j}Q_{1i,k}T_{i,jk}^{\Omega}R_{1i}\right]\right]\\
&+ \quad -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[Q_{1i,j}\tilde{\psi}_{i,k}T_{i,jk}^{\Omega}R_{1i}+Q_{1i,j}R_{1i,k}T_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right]\right]\\
&+ \quad -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}\tilde{E}_{i,jk}^{\Omega}R_{1i}+\tilde{\psi}_{i,j}R_{1i,k}\tilde{E}_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right]\right]\\
&+ \quad -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[R_{1i,j}\sqrt{T}(\hat{\gamma}_{i0,k}-\gamma_{i0,k})\tilde{E}_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right]\right]\\
&+ \quad -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{k=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}\tilde{\psi}_{i,l}T_{i,jkl}^{\Omega}R_{1i}+\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}R_{1i,l}T_{i,jkl}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right]\right]\\
&+ \quad -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{k=1}^{m+1}\left[\tilde{\psi}_{i,j}R_{1i,k}\sqrt{T}(\hat{\gamma}_{i0,l}-\gamma_{i0,l})T_{i,jkl}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right]\right]\\
&+ \quad -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{k=1}^{m+1}\left[R_{1i,j}\sqrt{T}(\hat{\gamma}_{i0,k}-\gamma_{i0,k})\sqrt{T}(\hat{\gamma}_{i0,l}-\gamma_{i0,l})T_{i,jkl}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right]\right]\\
&+ \quad -\left(T_i^{\Omega}\right)^{-1}\left[\frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{l=1}^{m+1}\sqrt{T}(\hat{\gamma}_{i0,j}-\gamma_{i0,j})\sqrt{T}(\hat{\gamma}_{i0,k}-\gamma_{i0,k})\sqrt{T}(\hat{\gamma}_{i0,l}-\gamma_{i0,l})\sqrt{T}(\hat{T}_{i,jkl}^{\Omega}(\overline{\gamma}_i)-T_{i,jkl}^{\Omega})\right]\\
&\times \quad \sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\\
&+ \quad -\left(T_i^{\Omega}\right)^{-1}\left[diag[0,\tilde{\psi}_{\Omega_i}^{W}]R_{3i}+diag[0,Q_{1\Omega_i}^{W}]R_{2i}+diag[0,Q_{2\Omega_i}^{W}]R_{1i}+diag[0,R_{3\Omega_i}^{W}]\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right]\\
&+ \quad -\frac{1}{T^{(r-1)/2}}\left(T_i^{\Omega}\right)^{-1}\left[diag[0,Q_{1r\Omega_i}^{W}]R_{2i}+diag[0,Q_{2r\Omega_i}^{W}]R_{1i}\right]. \qquad (2.H.51)
\end{aligned}
$$

The uniform rate of convergence then follows by Lemmas 12, 27, 29, 47, 48 and 49, and Conditions 3 and 4. The result for the average follows by the observation that the sample mean is dominated in probability by the maximum. ■

2.I Stochastic Expansion for $\hat{s}_i(\theta_0, \hat{\gamma}_{i0})$

Lemma 52 *Suppose that Conditions 1, 2, 3, 4, 5, and 6 hold. We then have*

$$
\hat{s}_i(\theta_0,\hat{\gamma}_{i0})=\frac{1}{\sqrt{T}}\tilde{\psi}_{si}+\frac{1}{T}Q_{1si}+\frac{1}{T^{3/2}}Q_{2si}+\frac{a_T}{T^{(r+2)/2}}Q_{2rsi}+\frac{1}{T^2}R_{3si}, \qquad (2.I.1)
$$

where

$$a_T = \begin{cases} C, & \text{if } n = O(T^a) \text{ for some } a \in \Re; \\ o(T^\epsilon) \text{ for any } \epsilon > 0, & \text{otherwise,} \end{cases} \tag{2.I.2}$$

$$\tilde{\psi}_{si} = M_i^\Omega \tilde{\psi}_i = o_{up}\left(T^{r/2s}\right), \tag{2.I.3}$$

$$Q_{1si} = M_i^\Omega Q_{1i} + \tilde{C}_i^\Omega \tilde{\psi}_i + \frac{1}{2}\sum_{j=1}^{m+1} \tilde{\psi}_{i,j} M_{i,j}^\Omega \tilde{\psi}_i = o_{up}\left(T^{r/s}\right), \tag{2.I.4}$$

$$\tilde{C}_i = \sqrt{T}(\hat{M}_i^\Omega - M_i^\Omega) = o_{up}\left(T^{r/2s}\right), \tag{2.I.5}$$

$$\begin{aligned} Q_{2si} &= M_i^\Omega Q_{2i} + \tilde{C}_i^\Omega Q_{1i} + \frac{1}{2}\sum_{j=1}^{m+1}\left[\tilde{\psi}_{i,j} M_{i,j}^\Omega Q_{1i} + Q_{1i,j} M_{i,j}^\Omega \tilde{\psi}_i + \tilde{\psi}_{i,j}\tilde{D}_{i,j}^\Omega \tilde{\psi}_i\right] \\ &+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1} \tilde{\psi}_{i,j}\tilde{\psi}_{i,k} M_{i,jk}^\Omega \tilde{\psi}_i = o_{up}\left(T^{3r/2s}\right), \end{aligned} \tag{2.I.6}$$

$$\tilde{D}_{i,j}^\Omega = \sqrt{T}(\hat{M}_{i,j}^\Omega - M_{i,j}^\Omega) = o_{up}\left(T^{r/2s}\right), \tag{2.I.7}$$

$$Q_{2rsi} = M_i^\Omega Q_{2ri}, \tag{2.I.8}$$

$$R_{3si}^W = o_{up}(\sqrt{T}). \tag{2.I.9}$$

Also,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\psi}_{si} = O_p(1), \tag{2.I.10}$$

$$\frac{1}{n}\sum_{i=1}^n (Q_{1si} - E\left[Q_{1si}\right]) = o_p(1), \tag{2.I.11}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (Q_{1si} - E\left[Q_{1si}\right]) = O_p(1), \tag{2.I.12}$$

$$\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^n Q_{2si} = O_p(1), \tag{2.I.13}$$

$$\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^n Q_{2rsi} = O_p(1), \tag{2.I.14}$$

$$\frac{1}{n}\sum_{i=1}^n R_{3si} = o_p(\sqrt{T}). \tag{2.I.15}$$

Proof. By a third order Taylor expansion of $\hat{s}_i(\theta_0, \hat{\gamma}_{i0})$, we have

$$\begin{aligned}\hat{s}_i(\theta_0, \hat{\gamma}_{i0}) &= \hat{s}_i(\theta_0, \gamma_{i0}) + \hat{M}_i^{\Omega}(\hat{\gamma}_{i0} - \gamma_{i0}) + \frac{1}{2}\sum_{j=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})\hat{M}_{i,j}^{\Omega}(\hat{\gamma}_{i0} - \gamma_{i0}) \\ &+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})(\hat{\gamma}_{i0,k} - \gamma_{i0,k})\hat{M}_{i,jk}^{\Omega}(\overline{\gamma}_i)(\hat{\gamma}_{i0} - \gamma_{i0}), \qquad (2.\mathrm{I}.16)\end{aligned}$$

where $\overline{\gamma}_i$ is between $\hat{\gamma}_{i0}$ and γ_{i0}. Noting that $\hat{s}_i(\theta_0, \gamma_{i0}) = 0$ and using the expansion for $\hat{\gamma}_{i0}$ in Lemma 49, we can obtain the expressions for $\tilde{\psi}_{si}$, Q_{1si}, and Q_{2si}, after some algebra. The rest of the properties for these terms follow by the properties of $\tilde{\psi}_i$, Q_{1i}, and Q_{2i}, and Lemmas 12, 58, 59 and 60. For the remainder term, we have

$$\begin{aligned}R_{3si} &= M_i^{\Omega}R_{3i} + \tilde{C}_i^{\Omega}R_{2i} + \frac{1}{2}\sum_{j=1}^{m+1}\left[R_{2i,j}M_{i,j}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}M_{i,j}^{\Omega}R_{2i}\right] \\ &+ \frac{1}{2}\sum_{j=1}^{m+1}\left[Q_{1i,j}M_{i,j}^{\Omega}R_{1i} + R_{1i,j}\tilde{D}_{i,j}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}\tilde{D}_{i,j}^{\Omega}R_{1i}\right] \\ &+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[R_{1i,j}\sqrt{T}(\hat{\gamma}_{i0,k} - \gamma_{i0,k})M_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}R_{1i,k}M_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})\right] \\ &+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}M_{i,jk}^{\Omega}R_{1i} \\ &+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sqrt{T}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})\sqrt{T}(\hat{\gamma}_{i0,k} - \gamma_{i0,k})\sqrt{T}(\hat{M}_{i,jk}^{\Omega}(\overline{\gamma}_i) - M_{i,jk}^{\Omega})\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}). \quad (2.\mathrm{I}.17)\end{aligned}$$

Then, the results for R_{3si} follow by the properties of the components in the expansion of $\hat{\gamma}_{i0}$, Lemmas 12, 27, 29, 47, 48, 49, and Conditions 3 and 4. ∎

Lemma 53 *Suppose that Conditions 1, 2, 3, 4, 5, and 6 hold. We then have*

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{sit} &\xrightarrow{d} N(0, J_s), && (2.\mathrm{I}.18)\\ J_s &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}G_{\theta_i}'P_{\alpha_i}G_{\theta_i} && (2.\mathrm{I}.19)\\ \frac{1}{n}\sum_{i=1}^{n}Q_{1si} &\xrightarrow{p} B_s = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left[Q_{1si}\right] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}B_{si}, && (2.\mathrm{I}.20)\\ B_{si} &= B_{si}^B + B_{si}^C + B_{si}^V, && (2.\mathrm{I}.21)\\ B_{si}^B &= -G_{\theta_i}'B_{\lambda_i} = -G_{\theta_i}'\left(B_{\lambda_i}^I + B_{\lambda_i}^G + B_{\lambda_i}^{\Omega} + B_{\lambda_i}^W\right), && (2.\mathrm{I}.22)\\ B_{si}^C &= E\left[G_{\theta_i}(z_{it})'P_{\alpha_i}g(z_{it})\right], && (2.\mathrm{I}.23)\\ B_{si}^V &= 0, && (2.\mathrm{I}.24)\end{aligned}$$

where

$$\Sigma_{\alpha_i} = \left(G'_{\alpha_i}\Omega_i^{-1}G_{\alpha_i}\right)^{-1}, \tag{2.I.25}$$

$$\Sigma^W_{\alpha_i} = \left(G'_{\alpha_i}W_i^{-1}G_{\alpha_i}\right)^{-1}, \tag{2.I.26}$$

$$H^W_{\alpha_i} = \Sigma^W_{\alpha_i}G'_{\alpha_i}W_i^{-1}, \tag{2.I.27}$$

$$H_{\alpha_i} = \Sigma_{\alpha_i}G'_{\alpha_i}\Omega_i^{-1}, \tag{2.I.28}$$

$$P_{\alpha_i} = \Omega_i^{-1} - \Omega_i^{-1}G_{\alpha_i}H_{\alpha_i}. \tag{2.I.29}$$

Proof. The results follow by Lemmas 47, 48, 50 and 52, noting that

$$E\left[\psi_{sit}\psi'_{sit}\right] = M_i^{\Omega}\begin{pmatrix} \Sigma_{\alpha_i} & 0 \\ 0 & P_{\alpha_i}\end{pmatrix}M_i^{\Omega'}, \tag{2.I.30}$$

$$E\left[\tilde{C}_i^{\Omega}\tilde{\psi}_i\right] = E\left[G_{\theta_i}(z_{it})'P_{\alpha_i}g(z_{it})\right], \tag{2.I.31}$$

$$E\left[\tilde{\psi}_{i,j}M_{i,j}^{\Omega}\tilde{\psi}_i\right] = \begin{cases} 0, & \text{if } j = 1; \\ 0, & \text{if } j > 1. \end{cases} \tag{2.I.32}$$

■

Lemma 54 *Suppose that Conditions 1, 2, 3, 4, 5, and 6 hold. We then have*

$$\begin{cases} \sqrt{nT}\hat{s}_n(\theta_0) \xrightarrow{d} N\left(\sqrt{\rho}B_s, J_s\right), & \textit{if } n = \rho T; \\ T\hat{s}_i(\theta_0) \xrightarrow{p} B_s, & \textit{otherwise}; \end{cases} \tag{2.I.33}$$

where

$$\hat{s}_n(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\hat{s}_i(\theta_0, \hat{\gamma}_{i0}), \tag{2.I.34}$$

and B_s and J_s are defined in Lemma 53.

Proof. Case I: $n = O(T)$ From Lemma 52, we have

$$\begin{aligned} \sqrt{nT}\hat{s}_n(\theta_0) &= \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}_{si}}_{=O_p(1)} + \underbrace{\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}Q_{1si}}_{=O_p(1)} + \underbrace{\frac{1}{T}\sqrt{\frac{n}{T^{1-1}}}\frac{1}{n}\sum_{i=1}^{n}\left[Q_{2si} + a_T Q_{2rsi}\right]}_{=o_p(1)} \\ &+ \underbrace{\frac{1}{\sqrt{T}}\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}R_{3si}}_{=o_p(1)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}_{si} + \sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}Q_{1si} + o_p(1). \end{aligned} \tag{2.I.35}$$

Then, the result follows by Lemma 53.

Case II: $T = o(n)$ Similarly to the previous case, we have from Lemma 52

$$\begin{aligned}
T\hat{s}_n(\theta_0) &= \underbrace{\frac{a_T}{\sqrt{T^{r-1}}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}_{si}}_{=o_p(1)} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}Q_{1si}}_{=O_p(1)} + \underbrace{\frac{a_T}{T}\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}Q_{2si}}_{=o_p(1)} \\
&+ \underbrace{\frac{a_T}{\sqrt{T^{r+1}}}\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}Q_{2rsi}}_{=o_p(1)} + \underbrace{\frac{1}{\sqrt{T}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}R_{3si}}_{=o_p(1)} \\
&= \frac{1}{n}\sum_{i=1}^{n}Q_{1si} + o_p(1). \qquad (2.I.36)
\end{aligned}$$

Then, the result follows by Lemma 53. ■

Lemma 55 *Suppose that Conditions 1, 2, 3, 4, 5 and 6 hold. We then have*

$$\hat{s}_i(\theta_0, \hat{\gamma}_{i0}) = \frac{1}{\sqrt{T}}\tilde{\psi}_{si} + \frac{1}{T}Q_{1si} + \frac{1}{T^{3/2}}Q_{2si} + \frac{a_T}{T^{(r+2)/2}}Q_{2rsi} + \frac{1}{T^2}Q_{3si} + \frac{a_T}{T^{(r+3)/2}}Q_{2rsi} + \frac{1}{T^{5/2}}R_{4si}, \qquad (2.I.37)$$

where

$$\begin{aligned}
a_T &= \begin{cases} C, & \text{if } n = O(T^a) \text{ for some } a \in \Re; \\ o(T^\epsilon) \text{ for any } \epsilon > 0, & \text{otherwise,} \end{cases} \qquad (2.I.38) \\
Q_{3si} &= M_i^\Omega Q_{3i} + \tilde{C}_i^\Omega Q_{2i} + \frac{1}{2}\sum_{j=1}^{m+1}\left[\tilde{\psi}_{i,j}M_{i,j}^\Omega Q_{2i} + Q_{1i,j}M_{i,j}^\Omega Q_{1i} + Q_{2i,j}M_{i,j}^\Omega\tilde{\psi}_i\right] \\
&+ \frac{1}{2}\sum_{j=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{D}_{i,j}^\Omega Q_{1i} + Q_{1i,j}\tilde{D}_{i,j}^\Omega\tilde{\psi}_i\right] \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}M_{i,jk}^\Omega Q_{1i} + \tilde{\psi}_{i,j}Q_{1i,k}M_{i,jk}^\Omega\tilde{\psi}_i + Q_{1i,j}\tilde{\psi}_{i,k}M_{i,jk}^\Omega\tilde{\psi}_i\right] \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}\tilde{F}_{i,jk}^\Omega\tilde{\psi}_i\right] + \frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{l=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}\tilde{\psi}_{i,k}M_{i,jkl}^\Omega\tilde{\psi}_i\right] \\
&= O_{up}\left(T^{2r/s}\right), \qquad (2.I.39) \\
\tilde{F}_{i,jk}^\Omega &= \sqrt{T}(\hat{M}_{i,jk}^\Omega - M_{i,jk}^\Omega) = O_{up}\left(T^{r/2s}\right), \qquad (2.I.40) \\
Q_{3rsi} &= M_i^\Omega Q_{3ri} + \tilde{C}_i^\Omega Q_{2ri} + \frac{1}{2}\sum_{j=1}^{m+1}\left[\tilde{\psi}_{i,j}M_{i,j}^\Omega Q_{2ri} + Q_{2ri,j}M_{i,j}^\Omega\tilde{\psi}_i\right], \qquad (2.I.41) \\
R_{4si}^W &= o_{up}(T^{3r/2s}) = o_{up}(\sqrt{T}). \qquad (2.I.42)
\end{aligned}$$

Also,

$$\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}Q_{3si} = O_p(1), \tag{2.I.43}$$

$$\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}Q_{3rsi} = O_p(1), \tag{2.I.44}$$

$$\frac{1}{n}\sum_{i=1}^{n}R_{4si} = o_p(\sqrt{T}). \tag{2.I.45}$$

Proof. By a fourth order Taylor expansion of $\hat{s}_i(\theta_0, \hat{\gamma}_{i0})$, we have

$$\begin{aligned}
\hat{s}_i(\theta_0, \hat{\gamma}_{i0}) &= \hat{s}_i(\theta_0, \gamma_{i0}) + \hat{M}_i^{\Omega}(\hat{\gamma}_{i0} - \gamma_{i0}) + \frac{1}{2}\sum_{j=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})\hat{M}_{i,j}^{\Omega}(\hat{\gamma}_{i0} - \gamma_{i0}) \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})(\hat{\gamma}_{i0,k} - \gamma_{i0,k})\hat{M}_{i,jk}^{\Omega}(\hat{\gamma}_{i0} - \gamma_{i0}) \\
&+ \frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{k=1}^{m+1}(\hat{\gamma}_{i0,j} - \gamma_{i0,j})(\hat{\gamma}_{i0,k} - \gamma_{i0,k})(\hat{\gamma}_{i0,l} - \gamma_{i0,l})\hat{M}_{i,jk}^{\Omega}(\overline{\gamma}_i)(\hat{\gamma}_{i0} - \gamma_{i0}),
\end{aligned} \tag{2.I.46}$$

where $\overline{\gamma}_i$ is between $\hat{\gamma}_{i0}$ and γ_{i0}. Noting that $\hat{s}_i(\theta_0, \gamma_{i0}) = 0$ and using the expansion for $\hat{\gamma}_{i0}$ in Lemma 51, we can obtain the expressions for $\tilde{\psi}_{si}$, Q_{1si}, Q_{2si}, Q_{2rsi}, Q_{3si}, and Q_{3rsi}, after some algebra. The rest of the properties for these terms follow by the properties of their components and Lemmas 12, 58, 59 and 60. For the remainder term, we have

$$\begin{aligned}
R_{4si} &= M_i^{\Omega}R_{4i} + \tilde{C}_i^{\Omega}R_{3i} + \frac{1}{2}\sum_{j=1}^{m+1}\left[R_{3i,j}M_{i,j}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}M_{i,j}^{\Omega}R_{3i}\right] \\
&+ \frac{1}{2}\sum_{j=1}^{m+1}\left[Q_{1i,j}M_{i,j}^{\Omega}R_{2i} + Q_{2i,j}M_{i,j}^{\Omega}R_{1i} + \frac{1}{T^{(r-1)/2}}Q_{2ri,j}M_{i,j}^{\Omega}R_{1i}\right] \\
&+ \frac{1}{2}\sum_{j=1}^{m+1}\left[R_{2i,j}\tilde{D}_{i,j}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}\tilde{D}_{i,j}^{\Omega}R_{2i} + Q_{1i,j}\tilde{D}_{i,j}^{\Omega}R_{1i}\right] \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[R_{2i,j}\sqrt{T}(\hat{\gamma}_{i0,k} - \gamma_{i0,k})M_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}R_{2i,k}M_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})\right] \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}M_{i,jk}^{\Omega}R_{2i} + Q_{1i,j}R_{1i,k}M_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0}) + \tilde{\psi}_{i,j}Q_{1i,k}M_{i,jk}^{\Omega}R_{1i}\right] \\
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[Q_{1i,j}\tilde{\psi}_{i,k}M_{i,jk}^{\Omega}R_{1i} + \tilde{\psi}_{i,j}\tilde{\psi}_{i,k}\tilde{F}_{i,jk}^{\Omega}R_{1i} + \tilde{\psi}_{i,j}R_{1i,k}\tilde{F}_{i,jk}^{\Omega}\sqrt{T}(\hat{\gamma}_{i0} - \gamma_{i0})\right]
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{6}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\left[R_{1i,j}\sqrt{T}(\hat{\gamma}_{i0,k}-\gamma_{i0,k})\tilde{F}^{\Omega}_{i,jk}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right] \\
&+ \frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{l=1}^{m+1}\left[R_{1i,j}\sqrt{T}(\hat{\gamma}_{i0,k}-\gamma_{i0,k})\sqrt{T}(\hat{\gamma}_{i0,l}-\gamma_{i0,l})M^{\Omega}_{i,jkl}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right] \\
&+ \frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{l=1}^{m+1}\left[\tilde{\psi}_{i,j}R_{1i,k}\sqrt{T}(\hat{\gamma}_{i0,l}-\gamma_{i0,l})M^{\Omega}_{i,jkl}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right] \\
&+ \frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{l=1}^{m+1}\left[\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}\tilde{\psi}_{i,l}M^{\Omega}_{i,jkl}R_{1i}+\tilde{\psi}_{i,j}\tilde{\psi}_{i,k}R_{1i,l}M^{\Omega}_{i,jkl}\sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0})\right] \\
&+ \frac{1}{24}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{l=1}^{m+1}\left[\sqrt{T}(\hat{\gamma}_{i0,j}-\gamma_{i0,j})\sqrt{T}(\hat{\gamma}_{i0,k}-\gamma_{i0,k})\sqrt{T}(\hat{\gamma}_{i0,l}-\gamma_{i0,l})\sqrt{T}(\hat{M}^{\Omega}_{i,jkl}(\overline{\gamma}_i)-M^{\Omega}_{i,jk})\right] \\
&\times \sqrt{T}(\hat{\gamma}_{i0}-\gamma_{i0}). \qquad (2.I.47)
\end{aligned}
$$

Then, the results for R_{4si} follow by the properties of the components in the expansion of $\hat{\gamma}_{i0}$, Lemmas 12, 27, 29, 47, 48, 49 and 59, and Conditions 3 and 4. ■

Lemma 56 *Suppose that Conditions 1, 2, 3, 5, and 4 hold. We then have*

$$
\sqrt{nT}\hat{s}_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}_{si} + \sqrt{\frac{n}{T}}B_{sn} + o_p(1). \qquad (2.I.48)
$$

where

$$
\begin{aligned}
\hat{s}_n(\theta_0) &= \frac{1}{n}\sum_{i=1}^{n}\hat{s}_i(\theta_0,\hat{\gamma}_{i0}), \qquad (2.I.49) \\
B_{sn} &= \frac{1}{n}\sum_{i=1}^{n}E\left[Q_{1si}\right]. \qquad (2.I.50)
\end{aligned}
$$

Proof. We have from Lemma 55

$$
\begin{aligned}
\sqrt{nT}\hat{s}_n(\theta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}_{si} + \sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}E\left[Q_{1si}\right] + \underbrace{\frac{1}{\sqrt{T}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[Q_{1si}-E\left[Q_{1si}\right]\right]}_{=o_p(1)} \\
&+ \underbrace{\frac{1}{\sqrt{T^{3-r}}}\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}Q_{2si}}_{=o_p(1)} + \underbrace{\frac{1}{T}\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}Q_{2rsi}}_{=o_p(1)} + \underbrace{\frac{1}{\sqrt{T^{4-r}}}\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}Q_{3si}}_{=o_p(1)} \\
&+ \underbrace{\frac{1}{T^{3/2}}\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}Q_{3rsi}}_{=o_p(1)} + \underbrace{\frac{1}{\sqrt{T^{3-r}}}\sqrt{\frac{n}{T^{r}}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}R_{4si}}_{=o_p(1)} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}_{si} + \sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}E\left[Q_{1si}\right] + o_p(1). \qquad (2.I.51)
\end{aligned}
$$

Then, the result follows by Lemma 53. ■

2.J V - Statistics

2.J.1 Properties of Normalized V-Statistics

Definition 1 *(HN) Consider a statistic of the form*

$$W_{i,T}^{(v)} \equiv \frac{1}{T^{v/2}} \sum_{t=1}^{T} k_1(z_{it}) \sum_{t=1}^{T} k_2(z_{it}) \cdots \sum_{t=1}^{T} k_v(z_{it}). \tag{2.J.1}$$

We will call the average

$$\frac{1}{n} \sum_{i=1}^{n} W_{i,T}^{(v)} \tag{2.J.2}$$

of such $W_{i,T}^{(v)}$ the normalized V-statistic of order v. We will focus on the normalized V-statistic up to order 4.

Condition 7 *(i) $E[k_j(z_{it})] = 0$, $j = 1, ..., v$; (ii) $|k_j(z_{it})| \leq CM(z_{it})$ such that $\sup_i E\left[M(z_{it})^8\right] < \infty$, where C denotes a generic constant; (iii) $n = o(T^r)$ with $r \leq 3$.*

Lemma 57 *Suppose that Conditions 1 and 7 hold. Then for any $\{r_j, j = 1, ..., J \leq 4\}$ such that $0 < r_j \leq 2$*

$$\begin{aligned}
E\left[k_1(z_{it})^{r_1} k_2(z_{it})^{r_2}\right] &< \infty, &(2.J.3)\\
E\left[k_1(z_{it})^{r_1} k_2(z_{it})^{r_2} k_3(z_{it})^{r_3}\right] &< \infty, &(2.J.4)\\
E\left[k_1(z_{it})^{r_1} k_2(z_{it})^{r_2} k_3(z_{it})^{r_3} k_4(z_{it})^{r_4}\right] &< \infty. &(2.J.5)
\end{aligned}$$

Proof. For the first statement, note that by Hölder's Inequality and Condition 7

$$\begin{aligned}
E\left[k_1(z_{it})^{r_1} k_2(z_{it})^{r_2}\right] &\leq E\left[k_1(z_{it})^{\sum_{j=1}^2 r_j}\right]^{\frac{r_1}{\sum_{j=1}^2 r_j}} E\left[k_2(z_{it})^{\sum_{j=1}^2 r_j}\right]^{\frac{r_2}{\sum_{j=1}^2 r_j}} \\
&\leq C^{\sum_{j=1}^2 r_j} E\left[M(z_{it})^{\sum_{j=1}^2 r_j}\right] \leq CE\left[M(z_{it})^4\right] < \infty. &(2.J.6)
\end{aligned}$$

The second statement follows similarly by repeated application of Hölder's Inequality and Condition 7

$$\begin{aligned}
E\left[k_1(z_{it})^{r_1} k_2(z_{it})^{r_2} k_3(z_{it})^{r_3}\right] &\leq E\left[k_1(z_{it})^{\sum_{j=1}^3 r_j}\right]^{\frac{r_1}{\sum_{j=1}^3 r_j}} E\left[k_2(z_{it})^{\frac{r_2 \sum_{j=1}^3 r_j}{r_2+r_3}} k_3(z_{it})^{\frac{r_3 \sum_{j=1}^3 r_j}{r_2+r_3}}\right]^{\frac{r_2+r_3}{\sum_{j=1}^3 r_j}} \\
&\leq E\left[k_1(z_{it})^{\sum_{j=1}^3 r_j}\right]^{\frac{r_1}{\sum_{j=1}^3 r_j}} E\left[k_2(z_{it})^{\sum_{j=1}^3 r_j}\right]^{\frac{r_2}{\sum_{j=1}^3 r_j}} E\left[k_3(z_{it})^{\sum_{j=1}^3 r_j}\right]^{\frac{r_3}{\sum_{j=1}^3 r_j}} \\
&\leq C^{\sum_{j=1}^3 r_j} E\left[M(z_{it})^{\sum_{j=1}^3 r_j}\right] \leq CE\left[M(z_{it})^6\right] < \infty. &(2.J.7)
\end{aligned}$$

The third statement can be proven with an additional iteration of the previous argument. In particular, we have

$$E\left[k_1(z_{it})^{r_1} k_2(z_{it})^{r_2} k_3(z_{it})^{r_3} k_4(z_{it})^{r_4}\right] \leq CE\left[M(z_{it})^8\right] < \infty. \tag{2.J.8}$$

■

Lemma 58 *(HN) Suppose that Condition 7 holds. We then have*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{i,T}^{(1)} = O_p(1). \tag{2.J.9}$$

Proof. By Chebyshev's inequality, for any $\eta > 0$ we have

$$\begin{aligned}
\Pr\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{i,T}^{(1)}\right| \geq \eta\right] &= \Pr\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} k_1(z_{it})\right)\right| \geq \eta\right] \\
&\leq \frac{E\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} k_1(z_{it})\right)\right|^2\right]}{\eta^2} \\
&= \frac{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} E\left[k_1(z_{it})^2\right]}{\eta^2} \\
&\leq \frac{C^2 \sup_i E\left[M(z_{it})^2\right]}{\eta^2}
\end{aligned} \tag{2.J.10}$$

It therefore follows that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{i,T}^{(1)} = O_p(1)$, from which we obtain the desired conclusion. ■

Lemma 59 *Suppose that Conditions 1 and 7 hold. We then have*

$$\frac{1}{n}\sum_{i=1}^{n} W_{i,T}^{(2)} \xrightarrow{p} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E\left[W_{i,T}^{(2)}\right] = O(1), \tag{2.J.11}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(W_{i,T}^{(2)} - E\left[W_{i,T}^{(2)}\right]\right) = O_p(1). \tag{2.J.12}$$

Proof. Note that for $v = 2$ Condition 1 implies

$$
\begin{aligned}
E\left[W_{i,T}^{(2)}\right] &= E\left[k_1\left(z_{it}\right)k_2\left(z_{it}\right)\right] = O(1), && \text{(2.J.13)} \\
V\left[W_{i,T}^{(2)}\right] &= E\left[\left(W_{i,T}^{(2)}\right)^2\right] - E\left[W_{i,T}^{(2)}\right]^2 \\
&= \frac{1}{T^2}\underbrace{E\left[\sum_{t=1}^{T} k_1\left(z_{it}\right)^2 \sum_{t=1}^{T} k_2\left(z_{it}\right)^2\right]}_{O(T^2)} \\
&+ 2\frac{1}{T^2}\underbrace{E\left[\sum_{t=1}^{T} k_1\left(z_{it}\right)^2 \sum_{t=1}^{T}\sum_{s>t}^{T} k_2\left(z_{it}\right)k_2\left(z_{is}\right)\right]}_{o(1)} \\
&+ 2\frac{1}{T^2}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{n} k_1\left(z_{it}\right)k_1\left(z_{is}\right) \sum_{t=1}^{T} k_2\left(z_{it}\right)^2\right]}_{o(1)} \\
&+ 4\frac{1}{T^2}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{T} k_1\left(z_{it}\right)k_1\left(z_{is}\right) \sum_{t=1}^{T}\sum_{s>t}^{T} k_2\left(z_{it}\right)k_2\left(z_{is}\right)\right] + O(1)}_{O(T^2)} \\
&= O(1). && \text{(2.J.14)}
\end{aligned}
$$

where the rates of convergence of the sums are computed by counting the number of terms with non-zero expectation. Note also that all the expectations are finite by Lemma 57. Then, the results follow by WLLN and CLT. ■

Lemma 60 *Suppose that Conditions 1 and 7 hold. Then, for $v = 3, 4$, we have*

$$
\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n} W_{i,T}^{(v)} = O_p(1). \quad \text{(2.J.15)}
$$

Proof. For $v = 3$ Condition 1 implies

$$
\begin{aligned}
E\left[W_{i,T}^{(3)}\right] &= T^{-1/2}E\left[k_1\left(z_{it}\right)k_2\left(z_{it}\right)k_3\left(z_{it}\right)\right] = O(T^{-1/2}), && \text{(2.J.16)} \\
V\left[W_{i,T}^{(3)}\right] &= E\left[\left(W_{i,T}^{(3)}\right)^2\right] - E\left[W_{i,T}^{(3)}\right]^2 \\
&= \frac{1}{T^3}\underbrace{E\left[\sum_{t=1}^{T} k_1\left(z_{it}\right)^2 \sum_{t=1}^{T} k_2\left(z_{it}\right)^2 \sum_{t=1}^{T} k_3\left(z_{it}\right)^2\right]}_{O(T^3)}
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad \underbrace{2\frac{1}{T^3}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}k_3(z_{it})^2\right]}_{O(T^2)} \\
&+ \quad \underbrace{2\frac{1}{T^3}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\right]}_{O(T^2)} \\
&+ \quad \underbrace{2\frac{1}{T^3}E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}k_3(z_{it})^2\right]}_{O(T^2)} \\
&+ \quad \underbrace{4\frac{1}{T^3}E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}k_3(z_{it})^2\right]}_{O(T^3)} \\
&+ \quad \underbrace{4\frac{1}{T^3}E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\right]}_{O(T^3)} \\
&+ \quad \underbrace{4\frac{1}{T^3}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\right]}_{O(T^3)} \\
&+ \quad \underbrace{8\frac{1}{T^3}E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\right]}_{O(T^2)} \\
&+ \quad O(T^{-1}) = O(1), \qquad (2.J.17)
\end{aligned}
$$

where the rates of convergence of the sums are computed by counting the number of terms with non-zero expectation. Note also that all the expectations are finite by Lemma 57. Therefore by Condition 7 (iii)

$$
E\left[\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}W_{i,T}^{(3)}\right] = O(\sqrt{\frac{n}{T^r}}) = o(1), \qquad (2.J.18)
$$

$$
V\left[\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}W_{i,T}^{(3)}\right] = \frac{n}{T^{r-1}}\frac{1}{n}O(1) = O(T^{1-r}). \qquad (2.J.19)
$$

Then, the results follows by Chebyshev LLN.

Similarly, for $v = 4$ Condition 1 implies

$$
E\left[W_{i,T}^{(4)}\right] = T^{-1}E\left[k_1(z_{it})k_2(z_{it})k_3(z_{it})k_4(z_{it})\right] = O(T^{-1}), \qquad (2.J.20)
$$

$$
V\left[W_{i,T}^{(4)}\right] = E\left[\left(W_{i,T}^{(4)}\right)^2\right] - E\left[W_{i,T}^{(4)}\right]^2
$$

$$
\begin{aligned}
= & \underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}k_3(z_{it})^2\sum_{t=1}^{T}k_4(z_{it})^2\right]}_{O(T^4)} \\
+ & 2\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}k_3(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_4(z_{it})k_4(z_{is})\right]}_{O(T^3)} \\
+ & 2\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\sum_{t=1}^{T}k_4(z_{it})^2\right]}_{O(T^3)} \\
+ & 2\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}k_3(z_{it})^2\sum_{t=1}^{T}k_4(z_{it})^2\right]}_{O(T^3)} \\
+ & 2\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}k_3(z_{it})^2\sum_{t=1}^{T}k_4(z_{it})^2\right]}_{O(T^3)} \\
+ & 4\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}k_3(z_{it})^2\sum_{t=1}^{T}k_4(z_{it})^2\right]}_{O(T^4)} \\
+ & 4\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\sum_{t=1}^{T}k_4(z_{it})^2\right]}_{O(T^4)} \\
+ & 4\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}k_3(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_4(z_{it})k_4(z_{is})\right]}_{O(T^4)} \\
+ & 4\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\sum_{t=1}^{T}k_4(z_{it})^2\right]}_{O(T^4)} \\
+ & 4\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}k_3(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_4(z_{it})k_4(z_{is})\right]}_{O(T^4)} \\
+ & 4\underbrace{\frac{1}{T^4}E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_4(z_{it})k_4(z_{is})\right]}_{O(T^4)}
\end{aligned}
$$

$$+ \quad 8\frac{1}{T^4}E\underbrace{\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})\,k_1(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})\,k_2(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})\,k_3(z_{is})\sum_{t=1}^{T}k_4(z_{it})^2\right]}_{O(T^3)}$$

$$+ \quad 8\frac{1}{T^4}E\underbrace{\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})\,k_1(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})\,k_2(z_{is})\sum_{t=1}^{T}k_3(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_4(z_{it})\,k_4(z_{is})\right]}_{O(T^3)}$$

$$+ \quad 8\frac{1}{T^4}E\underbrace{\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})\,k_1(z_{is})\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})\,k_3(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_4(z_{it})\,k_4(z_{is})\right]}_{O(T^3)}$$

$$+ \quad 8\frac{1}{T^4}E\underbrace{\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})\,k_2(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})\,k_3(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_4(z_{it})\,k_4(z_{is})\right]}_{O(T^3)}$$

$$+ \quad 16\frac{1}{T^4}E\underbrace{\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})\,k_1(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})\,k_2(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})\,k_3(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_4(z_{it})\,k_4(z_{is})\right]}_{O(T^4)}$$

$$+ \quad O(T^{-2}) = O(1), \tag{2.J.21}$$

where the rates of convergence of the sums are computed by counting the number of terms with non-zero expectation. Note also that all the expectations are finite by Lemma 57. Therefore by Condition 7 (iii)

$$E\left[\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}W_{i,T}^{(4)}\right] = O(\sqrt{\frac{n}{T^{r+1}}}) = o(1), \tag{2.J.22}$$

$$V\left[\sqrt{\frac{n}{T^{r-1}}}\frac{1}{n}\sum_{i=1}^{n}W_{i,T}^{(3)}\right] = \frac{n}{T^{r-1}}\frac{1}{n}O(1) = O(T^{-2}) = O(T^{1-r}). \tag{2.J.23}$$

Then, the results follows by Chebyshev LLN. ■

2.J.2 Properties of Modified Normalized V-Statistics

Definition 2 *Consider now the following modified version of the V-statistics*

$$\tilde{W}_{i,T}^{(v)} \equiv \frac{1}{T^{v/2}}\frac{1}{n^{(v-1)/2}}\sum_{t=1}^{T}k_1(z_{it})\sum_{i=1}^{n}\left[\sum_{t=1}^{T}k_2(z_{it})\cdots\sum_{t=1}^{T}k_v(z_{it})\right]. \tag{2.J.24}$$

We will call the average

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{W}_{i,T}^{(v)} \tag{2.J.25}$$

of such $\tilde{W}_{i,T}^{(v)}$ the normalized V-statistic of order v. We will focus on the normalized V-statistic up to order 3.

Lemma 61 *Suppose that Conditions 1 and 7 hold. We then have*

$$\sqrt{\frac{n}{T^2}}\frac{1}{n}\sum_{i=1}^{n}\tilde{W}_{i,T}^{(v)} = o_p(1), \tag{2.J.26}$$

for $v = 2, 3$.

Proof. First, for $v = 2$ note that Condition 1 implies

$$E\left[\tilde{W}_{i,T}^{(2)}\right] = \frac{1}{\sqrt{n}}E\left[k_1(z_{it})\,k_2(z_{it})\right] = O(n^{-1/2}), \tag{2.J.27}$$

$$\begin{aligned}
V\left[\tilde{W}_{i,T}^{(2)}\right] &= E\left[\left(\tilde{W}_{i,T}^{(2)}\right)^2\right] - E\left[\tilde{W}_{i,T}^{(2)}\right]^2 \\
&= \frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{i=1}^{n}\sum_{t=1}^{T}k_2(z_{it})^2\right]}_{O(nT^2)} \\
&+ 2\frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})\,k_2(z_{is})\right]}_{o(1)} \\
&+ 2\frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})\,k_1(z_{is})\sum_{i=1}^{n}\sum_{t=1}^{T}k_2(z_{it})^2\right]}_{o(1)} \\
&+ 4\frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})\,k_1(z_{is})\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})\,k_2(z_{is})\right]}_{O(T^2)} \\
&+ 2\frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}\sum_{t=1}^{T}k_1(z_{it})^2\sum_{j>i}^{n}\sum_{t=1}^{T}k_2(z_{it})\sum_{t=1}^{T}k_2(z_{j,t})\right]}_{o(1)} \\
&+ 2\frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})\,k_1(z_{is})\sum_{j>i}^{n}\sum_{t=1}^{T}k_2(z_{it})\sum_{t=1}^{T}k_2(z_{j,t})\right]}_{o(1)} + O(1) \\
&= O(1),
\end{aligned} \tag{2.J.28}$$

$$
\begin{aligned}
C\left[\tilde{W}_{i,T}^{(2)},\tilde{W}_{j,T}^{(2)}\right] &= E\left[\tilde{W}_{i,T}^{(2)}\tilde{W}_{j,T}^{(2)}\right]-E\left[\tilde{W}_{i,T}^{(2)}\right]E\left[\tilde{W}_{j,T}^{(2)}\right] \\
&= \frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}k_1\left(z_{it}\right)\sum_{t=1}^{T}k_1\left(z_{j,t}\right)\sum_{i=1}^{n}\left(\sum_{t=1}^{T}k_2\left(z_{it}\right)\right)^2\right]}_{o(1)} \\
&+ 2\frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}k_1\left(z_{it}\right)\sum_{t=1}^{T}k_1\left(z_{j,t}\right)\sum_{j>i}^{n}\sum_{t=1}^{T}\sum_{s>t}^{T}k_2\left(z_{it}\right)k_2\left(z_{is}\right)\right]}_{O(T^2)}+O(n^{-1}) \\
&= O(n^{-1}). \qquad (2.J.29)
\end{aligned}
$$

where the rates of convergence of the sums are computed by counting the number of terms with non-zero expectation. Note also that all the expectations are finite by Lemma 57. Then, for the average we have the following

$$
E\left[\sqrt{\frac{n}{T^2}}\frac{1}{n}\sum_{i=1}^{n}\tilde{W}_{i,T}^{(2)}\right] = O(T^{-1})=o(1), \qquad (2.J.30)
$$

$$
V\left[\sqrt{\frac{n}{T^2}}\frac{1}{n}\sum_{i=1}^{n}\tilde{W}_{i,T}^{(2)}\right] = \frac{n}{T^2}\frac{1}{n^2}\underbrace{\sum_{i=1}^{n}V\left[\tilde{W}_{i,T}^{(2)}\right]}_{O(n)}+\frac{n}{T^2}\frac{2}{n^2}\underbrace{\sum_{j>i}^{n}C\left[\tilde{W}_{i,T}^{(2)},\tilde{W}_{j,T}^{(2)}\right]}_{O(n)}=O(T^{-2})=o(1). \qquad (2.J.31)
$$

Then, the result follows by Chebyshev's LLN.

For $v=3$ Condition 1 implies

$$
E\left[\tilde{W}_{i,T}^{(3)}\right] = \frac{1}{\sqrt{T}n}E\left[k_1\left(z_{it}\right)k_2\left(z_{it}\right)k_3\left(z_{it}\right)\right]=O(n^{-1}T^{-1/2}), \qquad (2.J.32)
$$

$$
\begin{aligned}
V\left[\tilde{W}_{i,T}^{(3)}\right] &= E\left[\left(\tilde{W}_{i,T}^{(3)}\right)^2\right]-E\left[\tilde{W}_{i,T}^{(3)}\right]^2 \\
&= \frac{1}{n^2T^3}\underbrace{E\left[\sum_{t=1}^{T}k_1\left(z_{it}\right)^2\sum_{i=1}^{n}\sum_{t=1}^{T}k_2\left(z_{it}\right)^2\sum_{t=1}^{T}k_3\left(z_{it}\right)^2\right]}_{O(nT^3)}
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad 2\frac{1}{n^2T^3}\underbrace{E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{i=1}^{n}\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\right]}_{O(T^2)} \\
&+ \quad 2\frac{1}{n^2T^3}\underbrace{E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}k_3(z_{it})^2\right]}_{O(T^2)} \\
&+ \quad 4\frac{1}{n^2T^3}\underbrace{E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\right]}_{O(T^2)} \\
&+ \quad 2\frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{i=1}^{n}\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}k_3(z_{it})^2\right]}_{O(T^2)} \\
&+ \quad 4\frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{i=1}^{n}\sum_{t=1}^{T}k_2(z_{it})^2\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\right]}_{O(T^2)} \\
&+ \quad 4\frac{1}{nT^2}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}k_3(z_{it})^2\right]}_{O(T^2)} \\
&+ \quad 8\frac{1}{n^2T^3}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s>t}^{T}k_2(z_{it})k_2(z_{is})\sum_{t=1}^{T}\sum_{s>t}^{T}k_3(z_{it})k_3(z_{is})\right]}_{O(T^2)} \\
&+ \quad 2\frac{1}{n^2T^3}\underbrace{E\left[\sum_{t=1}^{T}k_1(z_{it})^2\sum_{i=1}^{n}\sum_{j>i}^{n}\sum_{t=1}^{T}k_2(z_{it})\sum_{t=1}^{T}k_3(z_{it})\sum_{t=1}^{T}k_2(z_{j,t})\sum_{t=1}^{T}k_3(z_{j,t})\right]}_{O(n^2T^3)} \\
&+ \quad 4\frac{1}{n^2T^3}\underbrace{E\left[\sum_{t=1}^{T}\sum_{s>t}^{T}k_1(z_{it})k_1(z_{is})\sum_{i=1}^{n}\sum_{j>i}^{n}\sum_{t=1}^{T}k_2(z_{it})\sum_{t=1}^{T}k_3(z_{it})\sum_{t=1}^{T}k_2(z_{j,t})\sum_{t=1}^{T}k_3(z_{j,t})\right]}_{O(n^2T^2)} \\
&+ \quad O(n^{-2}T^{-1}) = O(1), \qquad (2.J.33)
\end{aligned}
$$

$$
\begin{aligned}
C\left[\tilde{W}_{i,T}^{(3)}, \tilde{W}_{j,T}^{(3)}\right] &= E\left[\tilde{W}_{i,T}^{(3)}\tilde{W}_{j,T}^{(3)}\right] - E\left[\tilde{W}_{i,T}^{(3)}\right] E\left[\tilde{W}_{j,T}^{(3)}\right] \\
&= \frac{1}{n^2T^3} \underbrace{E\left[\sum_{t=1}^{T} k_1(z_{it}) \sum_{t=1}^{T} k_1(z_{j,t}) \sum_{i=1}^{n} \left(\sum_{t=1}^{T} k_2(z_{it})\right)^2 \left(\sum_{t=1}^{T} k_3(z_{it})\right)^2\right]}_{o(1)} \\
&+ 2\frac{1}{n^{-2}T^3} \underbrace{E\left[\sum_{t=1}^{T} k_1(z_{it}) \sum_{t=1}^{T} k_1(z_{j,t}) \sum_{j>i}^{n} \sum_{t=1}^{T} k_2(z_{it}) \sum_{t=1}^{T} k_3(z_{it}) \sum_{t=1}^{T} k_2(z_{j,t}) \sum_{t=1}^{T} k_3(z_{j,t})\right]}_{O(T^2)} \\
&+ O(n^{-2}T^{-1}) = O(n^{-2}T^{-1}). \qquad (2.J.34)
\end{aligned}
$$

where the rates of convergence of the sums are computed by counting the number of terms with non-zero expectation. Note also that all the expectations are finite by Lemma 57. Then, for the average we have the following

$$
E\left[\sqrt{\frac{n}{T^2}} \frac{1}{n} \sum_{i=1}^{n} \tilde{W}_{i,T}^{(3)}\right] = O(n^{-1/2}T^{-3}) = o(1), \qquad (2.J.35)
$$

$$
V\left[\sqrt{\frac{n}{T^2}} \frac{1}{n} \sum_{i=1}^{n} \tilde{W}_{i,T}^{(3)}\right] = \frac{n}{T^2} \frac{1}{n^2} \underbrace{\sum_{i=1}^{n} V\left[\tilde{W}_{i,T}^{(3)}\right]}_{O(n)} + \frac{n}{T^2} \frac{2}{n^2} \underbrace{\sum_{j>i}^{n} C\left[\tilde{W}_{i,T}^{(3)}, \tilde{W}_{j,T}^{(3)}\right]}_{O(T^{-1})} = O(T^{-2}) = o(1). \qquad (2.J.36)
$$

Then, the result follows by Chebyshev's LLN.

∎

2.K First Stage Score and Derivatives: Fixed Effects

2.K.1 Score

$$
\hat{t}_i^W(\gamma_i; \theta) = -\frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} G_{\alpha_i}(z_{it}; \theta, \alpha_i)'\lambda_i \\ g(z_{it}; \theta, \alpha_i) + W_i\lambda_i \end{pmatrix} = -\begin{pmatrix} \hat{G}_{\alpha_i}(\theta, \alpha_i)'\lambda_i \\ \hat{g}_i(\theta, \alpha_i) + W_i\lambda_i \end{pmatrix} \qquad (2.K.1)
$$

2.K.2 Derivatives with respect to the fixed effects

First Derivatives

$$\hat{T}_i^W(\gamma_i;\theta) = \frac{\partial \hat{t}_i^W(\gamma_i;\theta)}{\partial \gamma_i'} = -\begin{pmatrix} \hat{G}_{\alpha\alpha_i}(\theta,\alpha_i)'\lambda_i & \hat{G}_{\alpha_i}(\theta,\alpha_i)' \\ \hat{G}_{\alpha_i}(\theta,\alpha_i) & W_i \end{pmatrix} \tag{2.K.2}$$

$$T_i^W = E\left[\hat{T}_i^W(\gamma_{i0};\theta_0)\right] = -\begin{pmatrix} 0 & G_{\alpha_i}' \\ G_{\alpha_i} & W_i \end{pmatrix} \tag{2.K.3}$$

$$\left(T_i^W\right)^{-1} = -\begin{pmatrix} -\Sigma_{\alpha_i}^W & H_{\alpha_i}^W \\ H_{\alpha_i}^{W\prime} & P_{\alpha_i}^W \end{pmatrix} \tag{2.K.4}$$

Second Derivatives

$$\hat{T}_{i,j}^W(\gamma_i;\theta) = \frac{\partial^2 \hat{t}_i^W(\gamma_i;\theta)}{\partial \gamma_{i,j}\partial \gamma_i'} = \begin{cases} -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i)'\lambda_i & \hat{G}_{\alpha\alpha_i}(\theta,\alpha_i)' \\ \hat{G}_{\alpha\alpha_i}(\theta,\alpha_i) & 0 \end{pmatrix}, & \text{if } j=1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i)'e_{j-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j>1. \end{cases} \tag{2.K.5}$$

$$T_{i,j}^W = E\left[\hat{T}_{i,j}^W(\gamma_{i0};\theta_0)\right] = \begin{cases} -\begin{pmatrix} 0 & G_{\alpha\alpha_i}' \\ G_{\alpha\alpha_i} & 0 \end{pmatrix}, & \text{if } j=1; \\ -\begin{pmatrix} G_{\alpha\alpha\alpha_i}'e_{j-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j>1. \end{cases} \tag{2.K.6}$$

Third Derivatives

$$\hat{T}_{i,jk}^W(\gamma_i;\theta) = \frac{\partial^3 \hat{t}_i^W(\gamma_i;\theta)}{\partial \gamma_{i,k}\partial \gamma_{i,j}\partial \gamma_i'} = \begin{cases} -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha\alpha_i}(\theta,\alpha_i)'\lambda_i & \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i)' \\ \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i) & 0 \end{pmatrix}, & \text{if } j=1,k=1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i)'e_{k-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k>1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i)'e_{j-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j>1,k=1; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k=1. \end{cases} \tag{2.K.7}$$

$$T_{i,jk}^{W} \quad = \quad E\left[\hat{T}_{i,jk}^{W}(\gamma_{i0};\theta_0)\right] = \begin{cases} -\begin{pmatrix} 0 & G'_{\alpha\alpha\alpha_i} \\ G_{\alpha\alpha\alpha_i} & 0 \end{pmatrix}, & \text{if } j=1, k=1; \\ -\begin{pmatrix} G'_{\alpha\alpha\alpha_i} e_{k-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1, k>1; \\ -\begin{pmatrix} G'_{\alpha\alpha\alpha_i} e_{j-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j>1, k=1; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1, k=1. \end{cases} \tag{2.K.8}$$

Fourth Derivatives

$$\hat{T}^W_{i,jkl}(\gamma_i;\theta) = \frac{\partial^4 \hat{t}^W_i(\gamma_i;\theta)}{\partial\gamma_{i,l}\partial\gamma_{i,k}\partial\gamma_{i,j}\partial\gamma_i'} = \begin{cases} -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha\alpha_i}(\theta,\alpha_i)'\lambda_i & \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i)' \\ \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i) & 0 \end{pmatrix}, & \text{if } j=1,k=1,l=1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i)'e_{l-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k=1,l>1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i)'e_{k-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k>1,l=1; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k>1,l>1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha_i}(\theta,\alpha_i)'e_{j-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j>1,k=1,l=1; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k=1,l>1. \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k=1,\forall l. \end{cases} \tag{2.K.9}$$

$$T^W_{i,jkl} = E\left[\hat{T}^W_{i,jkl}(\gamma_{i0};\theta_0)\right] = \begin{cases} -\begin{pmatrix} 0 & G'_{\alpha\alpha\alpha_i} \\ G_{\alpha\alpha\alpha_i} & 0 \end{pmatrix}, & \text{if } j=1,k=1,l=1; \\ -\begin{pmatrix} G'_{\alpha\alpha\alpha_i}e_{l-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k=1,l>1; \\ -\begin{pmatrix} G'_{\alpha\alpha\alpha_i}e_{k-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k>1,l=1; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k>1,l>1; \\ -\begin{pmatrix} G'_{\alpha\alpha\alpha_i}e_{j-1} & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j>1,k=1,l=1; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k=1,l>1. \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & \text{if } j=1,k=1,\forall l. \end{cases} \tag{2.K.10}$$

2.K.3 Derivatives with respect to the common parameter

First Derivatives

$$\hat{N}_{i,j}^{W}(\theta,\gamma_i) = \frac{\partial \hat{t}_i^W(\gamma_i;\theta)}{\partial \theta_j} = -\begin{pmatrix} \hat{G}_{\theta_j \alpha_i}(\theta,\alpha_i)'\lambda_i \\ \hat{G}_{\theta_j}(\theta,\alpha_i) \end{pmatrix} \tag{2.K.11}$$

$$N_{i,j}^{W} = E\left[\hat{N}_{i,j}^{W}(\theta_0,\gamma_{i0})\right] = -\begin{pmatrix} 0 \\ G_{\theta_j i} \end{pmatrix} \tag{2.K.12}$$

2.L First Stage Score and Derivatives: Common Parameters

2.L.1 Score

$$\hat{s}_i^W(\theta,\tilde{\gamma}_i(\theta)) = -\frac{1}{T}\sum_{t=1}^{T} G_\theta(z_{it};\theta,\tilde{\alpha}_i(\theta))'\tilde{\lambda}_i(\theta) = -\hat{G}_{\theta i}(\theta,\tilde{\alpha}_i(\theta))'\tilde{\lambda}_i(\theta), \tag{2.L.1}$$

2.L.2 Derivatives with respect to the fixed effects

First Derivatives

$$\hat{M}_i^W(\theta,\tilde{\gamma}_i(\theta)) = \frac{\partial \hat{s}_i^W(\theta,\tilde{\gamma}_i)}{\partial \tilde{\gamma}_i'} = -\begin{pmatrix} \hat{G}_{\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'\tilde{\lambda}_i(\theta) & \hat{G}_{\theta_i}(\theta,\tilde{\alpha}_i(\theta))' \end{pmatrix} \tag{2.L.2}$$

$$M_i^W = E\left[\hat{M}_i^W(\theta_0,\gamma_{i0})\right] = -\begin{pmatrix} 0 & G_{\theta_i}' \end{pmatrix} \tag{2.L.3}$$

Second Derivatives

$$\hat{M}_{i,j}^W(\theta,\tilde{\gamma}_i(\theta)) = \frac{\partial^2 \hat{s}_i^W(\theta,\tilde{\gamma}_i)}{\partial \tilde{\gamma}_{i,j}\partial \tilde{\gamma}_i'} = \begin{cases} -\begin{pmatrix} \hat{G}_{\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'\tilde{\lambda}_i(\theta) & \hat{G}_{\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))' \end{pmatrix}, & \text{if } j=1; \\ -\begin{pmatrix} \hat{G}_{\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'e_{j-1} & \hat{G}_{\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))' \end{pmatrix}, & \text{if } j>1. \end{cases} \tag{2.L.4}$$

$$M_{i,j}^W = E\left[\hat{M}_{i,j}^W(\theta_0,\gamma_{i0})\right] = \begin{cases} -\begin{pmatrix} 0 & G_{\alpha\theta_i}' \end{pmatrix}, & \text{if } j=1; \\ -\begin{pmatrix} G_{\alpha\theta_i}'e_{j-1} & 0 \end{pmatrix}, & \text{if } j>1. \end{cases} \tag{2.L.5}$$

Third Derivatives

$$\hat{M}^W_{i,jk}(\theta,\tilde{\gamma}_i(\theta)) = \frac{\partial^3 \hat{s}^W_i(\theta,\tilde{\gamma}_i)}{\partial\tilde{\gamma}_{i,k}\partial\tilde{\gamma}_{i,j}\partial\tilde{\gamma}_i'} = \begin{cases} -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'\tilde{\lambda}_i(\theta) & \hat{G}_{\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))' \end{pmatrix}, & \text{if } j=1,k=1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'e_{k-1} & 0 \end{pmatrix}, & \text{if } j=1,k>1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'e_{j-1} & 0 \end{pmatrix}, & \text{if } j>1,k=1; \\ -\begin{pmatrix} 0 & 0 \end{pmatrix}, & \text{if } j>1,k>1. \end{cases} \tag{2.L.6}$$

$$M^W_{i,jk} = E\left[\hat{M}^W_{i,jk}(\theta_0,\gamma_{i0})\right] = \begin{cases} -\begin{pmatrix} 0 & G'_{\alpha\alpha\theta_i} \end{pmatrix}, & \text{if } j=1,k=1; \\ -\begin{pmatrix} G'_{\alpha\alpha\theta_i}e_{k-1} & 0 \end{pmatrix}, & \text{if } j=1,k>1; \\ -\begin{pmatrix} G'_{\alpha\alpha\theta_i}e_{j-1} & 0 \end{pmatrix}, & \text{if } j>1,k=1; \\ -\begin{pmatrix} 0 & 0 \end{pmatrix}, & \text{if } j>1,k>1. \end{cases} \tag{2.L.7}$$

Fourth Derivatives

$$\begin{aligned} \hat{M}^W_{i,jkl}(\theta,\tilde{\gamma}_i(\theta)) &= \frac{\partial^4 \hat{s}^W_i(\theta,\tilde{\gamma}_i)}{\partial\tilde{\gamma}_{i,l}\partial\tilde{\gamma}_{i,k}\partial\tilde{\gamma}_{i,j}\partial\tilde{\gamma}_i'} = \\ &= \begin{cases} -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'\tilde{\lambda}_i(\theta) & \hat{G}_{\alpha\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))' \end{pmatrix}, & \text{if } j=1,k=1,l=1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'e_{l-1} & 0 \end{pmatrix}, & \text{if } j=1,k=1,l>1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'e_{k-1} & 0 \end{pmatrix}, & \text{if } j=1,k>1,l=1; \\ -\begin{pmatrix} 0 & 0 \end{pmatrix}, & \text{if } j=1,k>1,l>1; \\ -\begin{pmatrix} \hat{G}_{\alpha\alpha\alpha\theta_i}(\theta,\tilde{\alpha}_i(\theta))'e_{j-1} & 0 \end{pmatrix}, & \text{if } j>1,k=1,l=1; \\ -\begin{pmatrix} 0 & 0 \end{pmatrix}, & \text{if } j>1,k=1,l>1; \\ -\begin{pmatrix} 0 & 0 \end{pmatrix}, & \text{if } j>1,k>1,\forall l. \end{cases} \end{aligned} \tag{2.L.8}$$

$$M^W_{i,jkl} = E\left[\hat{M}^W_{i,jkl}(\theta_0,\gamma_{i0})\right] = \begin{cases} -\begin{pmatrix} 0 & G'_{\alpha\alpha\alpha\theta_i} \end{pmatrix}, & \text{if } j=1,k=1,l=1; \\ -\begin{pmatrix} G'_{\alpha\alpha\alpha\theta_i}e_{l-1} & 0 \end{pmatrix}, & \text{if } j=1,k=1,l>1; \\ -\begin{pmatrix} G'_{\alpha\alpha\alpha\theta_i}e_{k-1} & 0 \end{pmatrix}, & \text{if } j=1,k>1,l=1; \\ -\begin{pmatrix} 0 & 0 \end{pmatrix}, & \text{if } j=1,k>1,l>1; \\ -\begin{pmatrix} G'_{\alpha\alpha\alpha\theta_i}e_{j-1} & 0 \end{pmatrix}, & \text{if } j>1,k=1,l=1; \\ -\begin{pmatrix} 0 & 0 \end{pmatrix}, & \text{if } j>1,k=1,l>1; \\ -\begin{pmatrix} 0 & 0 \end{pmatrix}, & \text{if } j>1,k>1,\forall l. \end{cases} \tag{2.L.9}$$

2.L.3 Derivatives with respect to the common parameters

First Derivatives

$$\hat{S}^W_{i,j}(\theta, \tilde{\gamma}_i(\theta)) = \frac{\partial \hat{s}^W_i(\theta, \tilde{\gamma}_i)}{\partial \theta_j} = -\hat{G}_{\theta\theta_{j,i}}(\theta, \tilde{\alpha}_i(\theta))' \tilde{\lambda}_i(\theta) \tag{2.L.10}$$

$$S^W_{i,j} = E\left[\hat{S}^W_{i,j}(\theta_0, \gamma_{i0})\right] = 0 \tag{2.L.11}$$

2.M Second Stage Score and Derivatives: Fixed Effects

2.M.1 Score

$$\begin{aligned} \hat{t}_i(\gamma_i; \theta) &= -\frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} G_{\alpha_i}(z_{it}; \theta, \alpha_i)' \lambda_i \\ g(z_{it}; \theta, \alpha_i) + \hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i)\lambda_i \end{pmatrix} = -\begin{pmatrix} \hat{G}_{\alpha_i}(\theta, \alpha_i)' \lambda_i \\ \hat{g}_i(\theta, \alpha_i) + \Omega_i \lambda_i \end{pmatrix} - \begin{pmatrix} 0 \\ (\tilde{\Omega}_i - \Omega_i)\lambda_i \end{pmatrix} \\ &= \hat{t}^{\Omega}_i(\gamma_i; \theta) + \hat{t}^R_i(\gamma_i; \theta) \end{aligned} \tag{2.M.1}$$

Note that the formulae for the derivatives of Appendix 2.K apply for $\hat{t}^{\Omega}_i$, replacing W by Ω. Hence, we only need to derive the derivatives for $\hat{t}^R_i$.

2.M.2 Derivatives with respect to the fixed effects

First Derivatives

$$\hat{T}^R_i(\gamma_i; \theta) = \frac{\partial \hat{t}^R_i(\gamma_i; \theta)}{\partial \gamma_i'} = -\begin{pmatrix} 0 & 0 \\ 0 & \hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i) - \Omega_i \end{pmatrix} \tag{2.M.2}$$

$$T^R_i = E\left[\hat{T}^R_i(\gamma_{i0}; \theta_0)\right] = -\begin{pmatrix} 0 & 0 \\ 0 & E\left[\hat{\Omega}_i - \Omega_i\right] \end{pmatrix} \tag{2.M.3}$$

Second and Higher Order Derivatives

Since $\hat{T}^R_i(\gamma_i; \theta)$ does not depend on γ_i, the derivatives (and its expectation) of order greater than one are zero.

2.M.3 Derivatives with respect to the common parameters

First Derivatives

$$\hat{N}^R_i(\gamma_i; \theta) = \frac{\partial \hat{t}^R_i(\gamma_i; \theta)}{\partial \theta'} = 0. \tag{2.M.4}$$

2.N Second Stage Score and Derivatives: Common Parameters

2.N.1 Score

$$\hat{s}_i(\theta, \hat{\gamma}_i(\theta)) = -\frac{1}{T}\sum_{t=1}^{T} G_\theta(z_{it}; \theta, \hat{\alpha}_i(\theta))'\hat{\lambda}_i(\theta) = -\hat{G}_{\theta i}(\theta, \hat{\alpha}_i(\theta))'\hat{\lambda}_i(\theta), \tag{2.N.1}$$

Since this score does not depend explicitly on $\hat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i)$, the formulae for the derivatives in Appendix 2.L carry through replacing $\tilde{\gamma}_i$ by $\hat{\gamma}_i$.

Bibliography

[1] ABOWD, J., AND D. CARD (1989) "On the Covariance Structure of Earnings and Hour Changes," *Econometrica* 57, 441-445.

[2] ANATOLYEV, S. (2005) "GMM, GEL, Serial Correlation, and Asymptotic Bias," *Econometrica* forthcoming.

[3] ANGRIST, J. D. (2004) "Treatment effect heterogeneity in theory and practice," *The Economic Journal* 114(494), C52-C83.

[4] ANGRIST, J. D., K. GRADDY AND G. W. IMBENS (2000) "The Interpretation of Instrumental Variables Estimators in Simultaneous Equation Models with an Application to the Demand of Fish," *Review of Economic Studies* 67, 499-527.

[5] ANGRIST, J. D., AND J. HAHN (2004) "When to Control for Covariates? Panel Asymptotics for Estimates of Treatment Effects ," *Review of Economics and Statistics* 86(1), 58-72.

[6] ANGRIST, J. D., AND G. W. IMBENS (1995) "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association* 90, 431-442.

[7] ANGRIST, J. D., AND A. B. KRUEGER (1991) "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics* 106(4), 979-1014.

[8] ANGRIST, J. D., AND A. B. KRUEGER (1999) "Empirical Strategies in Labor Economics," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics, Vol. 3*, Elsevier Science.

[9] ANGRIST, J. D., AND W. K. NEWEY (1991) "Over-Identification Tests in Earnings Functions With Fixed Effects," *Journal of Business and Economic Statistics* 9, 317-323.

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

[11] ARELLANO, M. AND B. HONORÉ (2001), "Panel Data Models: Some Recent Developments," in J. J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Vol. 5*, Amsterdam: North-Holland.

[12] BUSE, A. (1992) "The Bias of Instrumental Variables Estimators," *Econometrica* 60, 173-180.

[13] CARD, D. (1996) "The Effects of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica* 64, 957-979.

[14] CARD, D., T. LEMIEUX AND W. C. RIDDELL (2004) "Unionization and Wage Inequality: A Comparative Study of the U.S., the U.K., and Canada," unpublished manuscript.

[15] CHAMBERLAIN, G. (1984), "Panel Data," in Z. GRILICHES AND M. INTRILIGATOR, eds., *Handbook of Econometrics, Vol. 2.* Amsterdam: North-Holland.

[16] Chamberlain, G. (1992), "Efficiency Bounds for Semiparametric Regression," *Econometrica* 60, 567-596.

[17] Chowdhury, G., and S. Nickell (1985), "Hourly Earnings in the United States: Another Look at Unionization, Schooling, Sickness, and Unemployment Using PSID Data," *Journal of Labor Economics* 3, 38-69.

[18] Firth, D. (2002), "Bias Reduction of Maximum Likelihood Estimates," *Biometrika* 80, 27-38.

[19] Freeman, R. B. (1984), "Longitudinal Analysis of the Effect of Trade Unions," *Journal of Labor Economics* 2, 1-26.

[20] Green, D. A.(1991), "A Comparison of Estimation Approaches for the Union-Nonunion Wage Differential," *unpublished manuscript*, The University of British Columbia.

[21] Greene, W.H. (2002), "The Bias of the Fixed Effects Estimator in Nonlinear Models," *unpublished manuscript*, New York University.

[22] Hahn, J., and J. Hausman (2002) "Notes on Bias in Estimators for Simultaneous Equation Models," *Economics Letters* 75, 237-241.

[23] Hahn, J., and G. Kuersteiner (2003), "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," unpublished manuscript.

[24] Hahn, J., G. Kuersteiner, and W. Newey (2004), "Higher Order Properties of Bootstrap and Jackknife Bias Corrections," unpublished manuscript.

[25] Hahn, J., and W. Newey (2004), "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica* 72, 1295-1319.

[26] Hansen, L. P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.

[27] Hansen, L. P., and K. J. Singleton (1982) "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50, 1269-1286.

[28] Hsiao, C., and M. H. Pesaran (2004), "Random Coefficient Panel Data Models," mimeo, University of Southern California.

[29] Holtz-Eakin, D., W. K. Newey, and H. S. Rosen (1988) "Estimating Vector Autoregressions With Panel Data," *Econometrica* 56, 1371-1396.

[30] Imbens, G. W., J. D. Angrist (1994) "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62, 467-475.

[31] Jacubson, G. (1991), "Estimation and Testing of the Union Wage Effect Using Panel Data," *Review of Economic Studies* 58, 971-991.

[32] Kane, T. J., C. E. Rouse, and D. Staiger (1999) "Estimating Returns to Schooling When Schooling is Misreported," NBER working paper series, WP7235.

[33] Kelejian, H. H. (1974) "Random Parameters in a Simultaneous Equation Framework: Identification and Estimation,"*Econometrica* 42(3), 517-528.

[34] Lancaster, T. (2002), "Orthogonal Parameters and Panel Data," *Review of Economic Studies* 69, 647-666.

[35] Lewis, H. G. (1986), *Union Relative Wage Effects: A Survey.* University of Chicago Press.

[36] LEE, L. - F. (1978), "Unionism and Wage Rates: A Simultaneous Model with Qualitative and Limited Dependent Variables," *International Economic Review* 19, 415-433.

[37] LI, H., B. G. LINDSAY, AND R. P. WATERMAN (2003), "Efficiency of Projected Score Methods in Rectangular Array Asymptotics," *Journal of the Royal Statistical Society B* 65, 191-208.

[38] MACKINNON, J. G., AND A. A. SMITH (1998), "Approximate Bias Correction in Econometrics," *Journal of Econometrics* 85, 205-230.

[39] NEWEY, W.K. (1990), "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics* 5, 99-135.

[40] NEWEY, W.K., AND D. MCFADDEN (1994), "Large Sample Estimation and Hypothesis Testing," in R.F. ENGLE AND D.L. MCFADDEN, eds., *Handbook of Econometrics, Vol. 4*. Elsevier Science. Amsterdam: North-Holland.

[41] NEWEY, W.K., AND R. SMITH (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica* 72, 219-255.

[42] NAGAR, A. L., (1959), "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica* 27, 575-595.

[43] NEYMAN, J., AND E.L. SCOTT, (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica* 16, 1-32.

[44] PHILLIPS, P. C. B., AND H. R. MOON, (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica* 67, 1057-1111.

[45] RATHOUZ, P. J., AND K.-Y. LIANG (1999), "Reducing Sensitivity to Nuisance Parameters in Semiparametric Models: A Quasiscore Method, " *Biometrika* 86, 857-869.

[46] RILSTONE, P., V. K. SRIVASTAVA, AND A. ULLAH, (1996), "The Second-Order Bias and Mean Squared Error of Nonlinear Estimators, " *Journal of Econometrics* 75, 369-395.

[47] ROBINSON, C., (1989), "The Joint Determination of Union Status and Union Wage Effect: Some Tests of Alternative Models, " *Journal of Political Economy* 97, 639-667.

[48] ROY, A., (1951), "Some Thoughts on the Distribution of Earnings, " *Oxford Economic Papers* 3, 135-146.

[49] SWAMY, P. A. V. B., (1970), "Efficient Inference in a Random Coefficient Regression Model," *Econometrica* 38, 311-323.

[50] VELLA, F., AND M. VERBEEK, (1998), "Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men, " *Journal of Applied Econometrics* 13, 163-183.

[51] WOOLDRIDGE, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.

[52] WOUTERSEN, T.M. (2002), "Robustness Against Incidental Parameters," *unpublished manuscript*, University of Western Ontario.

[53] YITZHAKI, S. (1996) "On Using Linear Regressions in Welfare Economics," *Journal of Business and Economic Statistics* 14, 478-486.

Table 1a: Coefficient θ (true = 1), T = 8, n = 100

Estimator	η = 0 Mean	Median	Std.	SE/SD	p; .05	η = .1 Mean	Median	Std.	SE/SD	p; .05	η = .5 Mean	Median	Std.	SE/SD	p; .05	η = .9 Mean	Median	Std.	SE/SD	p; .05
ρ = 0																				
OLS - FC	1.00	1.00	0.13	0.95	0.07	0.94	0.94	0.13	0.95	0.10	0.67	0.68	0.14	0.92	0.75	0.41	0.42	0.14	0.96	0.99
IV - FC	1.00	1.00	0.16	0.80	0.12	0.90	0.91	0.16	0.82	0.17	0.49	0.50	0.18	0.73	0.93	0.10	0.12	0.21	0.63	1.00
OLS - RC	1.00	1.00	0.05	0.94	0.07	1.00	1.00	0.05	0.94	0.07	1.00	1.00	0.05	0.94	0.08	1.00	1.00	0.05	0.91	0.08
BC - OLS	1.00	1.00	0.05	0.94	0.07	1.00	1.00	0.05	0.94	0.07	1.00	1.00	0.05	0.94	0.08	1.00	1.00	0.05	0.91	0.08
IBC - OLS	1.00	1.00	0.05	0.86	0.09	1.00	1.00	0.05	0.87	0.10	1.00	1.00	0.05	0.81	0.11	1.00	1.00	0.06	0.77	0.13
IV - RC	1.00	1.00	0.08	1.13	0.03	1.00	1.00	0.07	1.18	0.02	1.00	1.00	0.08	1.19	0.01	1.00	0.99	0.09	1.16	0.03
BC - IV	1.00	1.00	0.08	1.12	0.03	1.00	1.00	0.07	1.18	0.02	1.00	1.00	0.08	1.18	0.02	1.00	0.99	0.09	1.14	0.03
IBC - IV	1.00	1.01	0.08	1.05	0.04	1.00	1.00	0.08	1.11	0.02	1.00	1.00	0.08	1.10	0.03	1.00	1.00	0.09	1.06	0.03
ρ = .1																				
OLS - FC	0.96	0.97	0.12	0.97	0.07	0.90	0.90	0.13	0.93	0.15	0.64	0.65	0.13	0.94	0.86	0.38	0.39	0.15	0.91	0.99
IV - FC	0.99	1.00	0.16	0.83	0.10	0.90	0.91	0.16	0.81	0.17	0.49	0.51	0.17	0.75	0.94	0.10	0.11	0.21	0.63	1.00
OLS - RC	0.95	0.95	0.05	0.95	0.19	0.95	0.95	0.05	0.94	0.21	0.96	0.95	0.05	0.93	0.19	0.95	0.95	0.05	0.93	0.23
BC - OLS	0.95	0.95	0.05	0.95	0.20	0.95	0.95	0.05	0.93	0.21	0.96	0.96	0.05	0.92	0.19	0.95	0.95	0.05	0.93	0.22
IBC - OLS	0.95	0.96	0.05	0.86	0.22	0.95	0.95	0.05	0.84	0.23	0.96	0.96	0.05	0.83	0.21	0.95	0.95	0.06	0.77	0.26
IV - RC	0.99	0.99	0.07	1.20	0.02	0.99	0.99	0.08	1.17	0.03	1.00	1.00	0.08	1.20	0.02	0.99	0.99	0.08	1.27	0.01
BC - IV	0.99	1.00	0.07	1.18	0.02	0.99	0.99	0.08	1.15	0.03	1.00	1.00	0.08	1.19	0.02	0.99	0.99	0.08	1.26	0.01
IBC - IV	0.99	0.99	0.08	1.12	0.04	0.99	0.99	0.08	1.08	0.04	1.00	1.00	0.08	1.12	0.03	0.99	0.99	0.09	1.16	0.03
ρ = .5																				
OLS - FC	0.83	0.83	0.13	0.94	0.34	0.77	0.77	0.13	0.97	0.50	0.51	0.51	0.13	0.93	0.98	0.24	0.25	0.13	0.95	1.00
IV - FC	1.00	1.00	0.16	0.79	0.13	0.90	0.91	0.16	0.83	0.17	0.49	0.50	0.18	0.73	0.94	0.10	0.11	0.21	0.63	1.00
OLS - RC	0.77	0.77	0.05	0.90	1.00	0.77	0.77	0.05	0.89	1.00	0.77	0.77	0.04	0.93	1.00	0.77	0.77	0.05	0.88	1.00
BC - OLS	0.77	0.77	0.05	0.90	1.00	0.77	0.77	0.05	0.89	1.00	0.77	0.77	0.05	0.92	1.00	0.76	0.77	0.05	0.88	1.00
IBC - OLS	0.77	0.77	0.05	0.83	1.00	0.77	0.77	0.05	0.82	1.00	0.77	0.77	0.05	0.80	1.00	0.77	0.77	0.06	0.72	1.00
IV - RC	0.97	0.97	0.08	1.26	0.01	0.97	0.97	0.08	1.25	0.03	0.97	0.97	0.08	1.29	0.02	0.97	0.97	0.09	1.25	0.02
BC - IV	0.97	0.97	0.08	1.25	0.02	0.97	0.97	0.08	1.25	0.03	0.97	0.97	0.08	1.28	0.02	0.97	0.97	0.09	1.23	0.02
IBC - IV	0.97	0.97	0.08	1.17	0.02	0.97	0.97	0.08	1.18	0.03	0.97	0.97	0.09	1.20	0.02	0.97	0.97	0.10	1.15	0.03
ρ = .9																				
OLS - FC	0.70	0.70	0.13	0.96	0.68	0.63	0.63	0.13	0.94	0.84	0.37	0.38	0.12	0.95	1.00	0.11	0.12	0.13	0.92	1.00
IV - FC	1.00	1.00	0.15	0.84	0.09	0.90	0.91	0.16	0.82	0.17	0.49	0.50	0.17	0.76	0.96	0.10	0.11	0.20	0.62	1.00
OLS - RC	0.58	0.58	0.05	0.83	1.00	0.58	0.58	0.05	0.82	1.00	0.58	0.58	0.05	0.79	1.00	0.58	0.58	0.04	0.84	1.00
BC - OLS	0.58	0.58	0.05	0.83	1.00	0.58	0.58	0.05	0.82	1.00	0.58	0.58	0.05	0.79	1.00	0.58	0.58	0.05	0.83	1.00
IBC - OLS	0.58	0.58	0.05	0.74	1.00	0.58	0.58	0.05	0.73	1.00	0.58	0.58	0.06	0.68	1.00	0.58	0.58	0.06	0.68	1.00
IV - RC	0.95	0.95	0.08	1.34	0.03	0.95	0.95	0.08	1.30	0.03	0.95	0.95	0.09	1.34	0.02	0.95	0.94	0.09	1.36	0.02
BC - IV	0.95	0.95	0.08	1.32	0.04	0.95	0.95	0.08	1.29	0.03	0.95	0.95	0.09	1.33	0.02	0.94	0.94	0.09	1.35	0.02
IBC - IV	0.95	0.95	0.09	1.25	0.04	0.95	0.94	0.09	1.22	0.04	0.95	0.95	0.09	1.24	0.03	0.94	0.95	0.10	1.24	0.04

Notes: 1,000 replications.

Table 1b: Coefficient α (true = 1), T = 8, n = 100

Estimator	η = 0					η = .1					η = .5					η = .9				
	Mean	Median	Std.	SE/SD	p; .05	Mean	Median	Std.	SE/SD	p; .05	Mean	Median	Std.	SE/SD	p; .05	Mean	Median	Std.	SE/SD	p; .05
										ρ = 0										
OLS - FC	1.01	1.01	0.15	0.95	0.06	1.06	1.07	0.15	0.95	0.10	1.33	1.32	0.16	0.93	0.61	1.59	1.59	0.17	0.95	0.96
IV - FC	1.00	1.00	0.17	0.94	0.07	1.10	1.10	0.18	0.92	0.13	1.51	1.50	0.17	0.96	0.88	1.90	1.89	0.18	0.93	1.00
OLS - RC	1.00	1.00	0.10	1.02	0.05	1.00	1.00	0.10	1.01	0.05	1.00	1.00	0.11	0.99	0.05	1.00	1.00	0.10	1.04	0.05
BC - OLS	1.00	1.00	0.10	1.02	0.06	1.00	1.00	0.10	1.01	0.05	1.00	1.00	0.11	0.99	0.05	1.00	1.00	0.10	1.04	0.05
IBC - OLS	1.00	1.00	0.10	1.01	0.06	1.00	1.00	0.10	1.01	0.05	1.00	1.00	0.11	0.99	0.05	1.00	1.00	0.10	1.03	0.05
IV - RC	1.00	1.00	0.11	1.11	0.03	1.00	1.00	0.11	1.09	0.03	1.00	1.00	0.12	1.06	0.04	1.00	1.00	0.11	1.10	0.03
BC - IV	1.00	1.00	0.11	1.10	0.04	1.00	1.00	0.11	1.08	0.04	1.00	1.00	0.12	1.06	0.04	1.00	1.00	0.11	1.10	0.03
IBC - IV	1.00	1.00	0.11	1.10	0.03	1.00	1.00	0.11	1.08	0.04	1.00	1.00	0.12	1.04	0.04	1.00	1.00	0.11	1.08	0.04
										ρ = .1										
OLS - FC	1.03	1.04	0.15	0.94	0.08	1.10	1.10	0.14	0.96	0.12	1.36	1.36	0.15	0.94	0.71	1.62	1.62	0.18	0.90	0.96
IV - FC	1.00	1.00	0.17	0.94	0.08	1.10	1.10	0.17	0.97	0.11	1.51	1.51	0.18	0.93	0.88	1.90	1.90	0.18	0.95	1.00
OLS - RC	1.05	1.05	0.11	0.99	0.07	1.05	1.05	0.10	1.03	0.07	1.04	1.04	0.10	1.02	0.07	1.05	1.05	0.10	1.00	0.08
BC - OLS	1.05	1.05	0.10	0.99	0.08	1.05	1.05	0.10	1.03	0.07	1.04	1.04	0.10	1.02	0.07	1.05	1.05	0.10	1.00	0.07
IBC - OLS	1.05	1.05	0.11	0.99	0.08	1.05	1.05	0.10	1.03	0.08	1.04	1.04	0.10	1.02	0.07	1.05	1.05	0.10	0.99	0.07
IV - RC	1.01	1.01	0.12	1.08	0.04	1.01	1.01	0.11	1.10	0.03	1.00	1.00	0.11	1.12	0.02	1.01	1.01	0.11	1.07	0.04
BC - IV	1.01	1.01	0.12	1.08	0.03	1.01	1.01	0.11	1.10	0.03	1.00	1.00	0.11	1.11	0.03	1.01	1.01	0.11	1.07	0.05
IBC - IV	1.01	1.01	0.12	1.07	0.04	1.01	1.01	0.11	1.09	0.03	1.00	1.00	0.11	1.10	0.03	1.01	1.01	0.12	1.05	0.05
										ρ = .5										
OLS - FC	1.18	1.18	0.15	0.95	0.28	1.23	1.23	0.15	0.95	0.39	1.50	1.49	0.15	0.94	0.94	1.76	1.76	0.16	0.93	0.99
IV - FC	1.00	1.00	0.18	0.94	0.07	1.10	1.10	0.18	0.91	0.13	1.51	1.50	0.17	0.96	0.89	1.90	1.89	0.18	0.93	1.00
OLS - RC	1.24	1.24	0.10	1.01	0.63	1.23	1.24	0.10	1.00	0.61	1.24	1.24	0.10	1.02	0.66	1.23	1.23	0.09	1.06	0.68
BC - OLS	1.24	1.23	0.10	1.01	0.62	1.23	1.24	0.10	1.00	0.61	1.24	1.23	0.10	1.02	0.66	1.23	1.23	0.09	1.06	0.68
IBC - OLS	1.24	1.23	0.10	1.01	0.62	1.23	1.23	0.10	1.00	0.61	1.24	1.24	0.10	1.01	0.66	1.23	1.24	0.09	1.05	0.67
IV - RC	1.03	1.03	0.11	1.15	0.03	1.03	1.03	0.12	1.12	0.03	1.03	1.03	0.11	1.12	0.04	1.03	1.03	0.11	1.16	0.03
BC - IV	1.03	1.03	0.11	1.15	0.03	1.03	1.03	0.12	1.11	0.03	1.03	1.03	0.12	1.12	0.04	1.03	1.03	0.11	1.15	0.04
IBC - IV	1.03	1.03	0.11	1.14	0.03	1.03	1.03	0.12	1.10	0.03	1.03	1.03	0.12	1.10	0.04	1.03	1.03	0.11	1.13	0.04
										ρ = .9										
OLS - FC	1.30	1.30	0.15	0.93	0.57	1.37	1.37	0.14	0.98	0.75	1.63	1.62	0.14	0.94	0.99	1.89	1.88	0.16	0.89	1.00
IV - FC	1.00	1.00	0.17	0.94	0.08	1.10	1.10	0.17	0.98	0.10	1.50	1.49	0.17	0.94	0.90	1.90	1.89	0.17	0.95	1.00
OLS - RC	1.42	1.43	0.11	0.97	0.98	1.42	1.42	0.10	1.01	0.98	1.42	1.42	0.09	1.02	0.99	1.42	1.42	0.09	1.04	0.99
BC - OLS	1.42	1.43	0.11	0.97	0.98	1.42	1.42	0.10	1.01	0.98	1.42	1.42	0.09	1.02	0.99	1.42	1.42	0.09	1.04	0.99
IBC - OLS	1.42	1.43	0.11	0.97	0.98	1.42	1.42	0.10	1.01	0.98	1.42	1.42	0.09	1.01	0.99	1.42	1.42	0.09	1.02	1.00
IV - RC	1.05	1.05	0.12	1.12	0.04	1.05	1.05	0.12	1.15	0.04	1.05	1.05	0.11	1.17	0.03	1.05	1.06	0.11	1.14	0.05
BC - IV	1.05	1.05	0.12	1.11	0.03	1.05	1.05	0.12	1.14	0.04	1.05	1.05	0.11	1.17	0.03	1.05	1.05	0.11	1.14	0.05
IBC - IV	1.05	1.05	0.12	1.11	0.04	1.05	1.05	0.12	1.14	0.04	1.05	1.05	0.11	1.15	0.04	1.05	1.06	0.12	1.11	0.06

Notes: 1,000 replications.

Table 1c: Coefficient σ^2 (true = 1), T = 8, n = 100

	$\eta = 0$					$\eta = .1$					$\eta = .5$					$\eta = .9$				
Estimator	Mean	Median	Std.	SE/SD	p; .05	Mean	Median	Std.	SE/SD	p; .05	Mean	Median	Std.	SE/SD	p; .05	Mean	Median	Std.	SE/SD	p; .05
$\rho = 0$																				
OLS - RC	1.02	1.02	0.07	1.01	0.06	1.02	1.01	0.07	0.98	0.06	1.02	1.02	0.07	1.01	0.05	1.02	1.02	0.07	0.99	0.06
BC - OLS	1.00	1.00	0.07	1.01	0.05	1.00	1.00	0.08	0.98	0.07	1.00	1.00	0.07	1.00	0.06	1.00	1.00	0.07	0.99	0.06
IBC - OLS	1.01	1.01	0.07	1.01	0.05	1.00	1.00	0.07	0.99	0.06	1.00	1.00	0.07	1.00	0.06	1.00	1.00	0.07	0.98	0.06
IV - RC	1.10	1.09	0.09	1.14	0.08	1.09	1.09	0.09	1.11	0.07	1.08	1.07	0.08	1.11	0.08	1.06	1.06	0.08	1.08	0.07
BC - IV	1.01	1.01	0.09	1.19	0.02	1.00	1.00	0.09	1.16	0.03	1.00	1.00	0.09	1.13	0.04	1.00	1.00	0.08	1.09	0.04
IBC - IV	1.02	1.02	0.09	1.16	0.02	1.01	1.01	0.09	1.14	0.03	1.01	1.00	0.09	1.11	0.03	1.00	1.00	0.08	1.07	0.04
$\rho = .1$																				
OLS - RC	1.02	1.01	0.07	1.04	0.05	1.02	1.01	0.07	1.01	0.05	1.01	1.01	0.07	0.98	0.06	1.00	1.01	0.07	1.01	0.05
BC - OLS	1.00	1.00	0.07	1.04	0.04	1.00	1.00	0.07	1.01	0.05	0.99	0.99	0.07	0.98	0.07	0.98	0.99	0.07	1.00	0.07
IBC - OLS	1.00	1.00	0.07	1.03	0.04	1.00	1.00	0.07	1.00	0.05	0.99	0.99	0.07	0.98	0.07	0.99	0.99	0.07	1.00	0.07
IV - RC	1.10	1.09	0.09	1.15	0.08	1.10	1.09	0.09	1.16	0.08	1.07	1.08	0.08	1.07	0.08	1.06	1.06	0.08	1.07	0.08
BC - IV	1.00	1.00	0.09	1.19	0.02	1.01	1.00	0.09	1.21	0.02	0.99	1.00	0.09	1.09	0.04	0.99	0.99	0.08	1.08	0.04
IBC - IV	1.01	1.01	0.09	1.16	0.02	1.02	1.01	0.09	1.19	0.03	1.00	1.00	0.09	1.08	0.04	1.00	1.00	0.08	1.07	0.04
$\rho = .5$																				
OLS - RC	1.02	1.02	0.07	1.01	0.05	1.01	1.01	0.07	1.00	0.05	0.98	0.98	0.07	1.02	0.08	0.95	0.95	0.07	0.97	0.15
BC - OLS	1.01	1.01	0.07	1.01	0.05	0.99	0.99	0.07	1.00	0.07	0.96	0.96	0.07	1.02	0.11	0.93	0.93	0.07	0.97	0.20
IBC - OLS	1.01	1.01	0.07	1.01	0.05	1.00	0.99	0.07	0.99	0.06	0.97	0.97	0.07	1.02	0.10	0.93	0.93	0.07	0.96	0.19
IV - RC	1.12	1.11	0.09	1.14	0.11	1.11	1.10	0.09	1.13	0.11	1.09	1.08	0.08	1.13	0.09	1.06	1.06	0.08	1.08	0.06
BC - IV	1.01	1.00	0.10	1.20	0.02	1.00	1.00	0.10	1.17	0.03	0.99	0.99	0.09	1.14	0.04	0.98	0.98	0.08	1.08	0.05
IBC - IV	1.02	1.02	0.10	1.17	0.03	1.01	1.01	0.10	1.15	0.02	1.00	1.00	0.09	1.12	0.03	0.98	0.98	0.08	1.07	0.04
$\rho = .9$																				
OLS - RC	1.02	1.02	0.07	1.02	0.05	1.01	1.00	0.07	0.98	0.06	0.94	0.94	0.07	0.98	0.17	0.89	0.89	0.07	0.97	0.38
BC - OLS	1.01	1.01	0.07	1.02	0.05	1.00	0.99	0.07	0.98	0.07	0.93	0.93	0.07	0.98	0.21	0.88	0.88	0.07	0.97	0.43
IBC - OLS	1.01	1.01	0.07	1.02	0.05	1.00	1.00	0.07	0.98	0.07	0.93	0.93	0.07	0.98	0.21	0.88	0.88	0.07	0.97	0.43
IV - RC	1.14	1.13	0.10	1.14	0.13	1.13	1.12	0.09	1.20	0.12	1.09	1.09	0.09	1.11	0.09	1.05	1.05	0.08	1.08	0.06
BC - IV	1.00	1.00	0.11	1.18	0.03	1.00	0.99	0.10	1.26	0.02	0.97	0.97	0.10	1.11	0.04	0.95	0.96	0.09	1.08	0.08
IBC - IV	1.02	1.01	0.11	1.15	0.03	1.01	1.01	0.10	1.23	0.02	0.99	0.98	0.10	1.10	0.04	0.96	0.97	0.09	1.07	0.07

Notes: 1,000 replications.

Table 2: Descriptive Statistics, 1981-93

Variable	Definition	Full sample				Union		Nonunion	
		Mean	St. Dev.	Between	Within	Mean	St.Dev.	Mean	St.Dev.
SCHOOL	Years of Schooling	12.22	1.59	96	4	12.30	1.29	12.17	1.75
EXPER	Age - 6 - SCHOOL	8.41	4.33	29	71	8.40	4.10	8.41	4.47
EXPER2	Experience Squared	89.47	78.04	29	71	87.43	73.62	90.75	80.68
LEXPER	Log(1 + EXPER)	2.10	0.60	29	71	2.12	0.54	2.08	0.63
UNION	Wage set by collective bargaining	0.39	0.49	36	64	1	0	0	0
UNION1	Lag of UNION	0.39	0.49	33	67	0.70	0.46	0.19	0.39
MARRIED	Married	0.54	0.50	46	54	0.57	0.50	0.53	0.50
BLACK	Black	0.15	0.35	100	0	0.15	0.35	0.15	0.35
HISP	Hispanic	0.18	0.39	100	0	0.17	0.37	0.19	0.40
HEALTH	Has health disability	0.03	0.16	27	73	0.02	0.15	0.03	0.16
RURAL	Lives in rural area	0.18	0.39	74	26	0.18	0.38	0.19	0.39
NE	Lives in North East	0.20	0.40	95	5	0.21	0.41	0.19	0.39
NC	Lives in Northern Central	0.29	0.46	97	3	0.35	0.48	0.26	0.44
S	Lives in South	0.29	0.45	94	6	0.23	0.42	0.33	0.47
W	Lives in West	0.22	0.41	96	4	0.21	0.41	0.23	0.42
WAGE	Log of real hourly wage	1.84	0.47	51	49	1.95	0.42	1.76	0.49
Industry dummies									
AG	Agricultural	0.02	0.13	32	68	0.01	0.08	0.03	0.16
MIN	Mining	0.02	0.14	49	51	0.02	0.15	0.02	0.13
CON	Construction	0.11	0.31	53	47	0.11	0.31	0.11	0.32
TRAD	Trade	0.22	0.41	47	53	0.14	0.35	0.26	0.44
TRA	Transportation	0.11	0.31	56	44	0.17	0.37	0.07	0.26
FIN	Finance	0.02	0.14	60	40	0.01	0.09	0.03	0.17
BUS	Business and Repair Service	0.07	0.25	39	61	0.04	0.20	0.08	0.28
PER	Personal Service	0.01	0.11	23	77	0.01	0.09	0.02	0.12
ENT	Entertainment	0.01	0.10	31	69	0.00	0.06	0.01	0.12
MAN	Manufacturing	0.32	0.47	58	42	0.34	0.47	0.30	0.46
PRO	Professional and Related Service	0.05	0.21	54	46	0.06	0.24	0.04	0.20
PUB	Public Administration	0.05	0.21	52	48	0.09	0.29	0.02	0.14
Number of Observations		3822		294	13	1476		2346	

Data: NLSY Men

Table 3: Wage Regressions with Union Effects for Men, 1981-93

Variable	OLS	IV	OLS - FE				IV - FE			
	FC	FC	FC	RC	RC-BC	RC-IBC	FC	RC	RC-BC	RC-IBC
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
UNION	0.130	0.232	0.108	-	-	-	0.330	-	-	-
	(.022)	(.043)	(.018)	-	-	-	(.061)	-	-	-
Mean(UNION)	-	-	-	0.087	0.087	0.087	-	0.102	0.096	0.094
	-	-	-	(.021)	(.021)	(.021)	-	(.091)	(.09)	(.089)
Std(UNION)	-	-	-	0.335	0.311	0.311	-	1.136	0.778	0.801
	-	-	-	(.137)	(.158)	(.158)	-	(.208)	(.44)	(.419)
SCHOOL	0.085	0.084	0.107	0.098	0.098	0.098	0.115	0.106	0.107	0.107
	(.01)	(.011)	(.021)	(.014)	(.014)	(.014)	(.02)	(.015)	(.015)	(.015)
LEXPER	0.275	0.262	0.345	0.343	0.340	0.340	0.305	0.307	0.307	0.303
	(.039)	(.04)	(.043)	(.035)	(.035)	(.035)	(.047)	(.033)	(.034)	(.034)
RURAL	-0.128	-0.125	-0.047	-0.041	-0.042	-0.042	-0.052	-0.066	-0.069	-0.069
	(.038)	(.038)	(.029)	(.022)	(.022)	(.022)	(.033)	(.025)	(.025)	(.025)
MARRIED	0.089	0.087	0.054	0.048	0.047	0.047	0.040	0.047	0.046	0.046
	(.025)	(.025)	(.021)	(.015)	(.015)	(.015)	(.022)	(.015)	(.015)	(.015)
HEALTH	0.022	0.026	-0.001	-0.008	-0.004	-0.004	0.016	-0.004	-0.006	-0.011
	(.039)	(.038)	(.033)	(.032)	(.032)	(.032)	(.035)	(.038)	(.038)	(.038)
BLACK	-0.233	-0.238	-	-	-	-	-	-	-	-
	(.05)	(.05)	-	-	-	-	-	-	-	-
HISP	-0.126	-0.126	-	-	-	-	-	-	-	-
	(.045)	(.045)	-	-	-	-	-	-	-	-
R2	0.378	0.368	0.679	0.739	-	-	0.648	0.723	-	-
Obs.	3822	3822	3822	3822	3822	3822	3822	3822	3822	3822

Data: NLSY.

Notes: all regressions include industry, region and time dummies. Standard errors in parentheses. Standard errors are clustered at the individual level in columns [1], [2], [3] and [7], and robust to heteroskedasticity in columns [4], [5], [6], [8], [9] and [10].

Chapter 3

Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure

(Joint with J. D. Angrist and V. Chernozhukov)

3.1 Introduction

The Quantile Regression (QR) estimator, introduced by Koenker and Bassett (1978), is an increasingly important empirical tool, allowing researchers to fit parsimonious models to an entire conditional distribution. Part of the appeal of quantile regression derives from a natural parallel with conventional ordinary least squares (OLS) or mean regression. Just as OLS regression coefficients offer convenient summary statistics for conditional expectation functions, quantile regression coefficients can be used to make easily interpreted statements about conditional distributions. Moreover, unlike OLS coefficients, QR estimates capture changes in distribution shape and spread, as well as changes in location.

An especially attractive feature of OLS regression estimates is their robustness and interpretability under misspecification of the conditional expectation function (CEF). In addition to consistently estimating a linear conditional expectation function, OLS estimates provide the minimum mean square error linear approximation to a conditional expectation function of any shape. The approximation properties of OLS have been emphasized by White (1980), Chamberlain (1984), Goldberger (1991), and Angrist and Krueger (1999). The fact that OLS provides a meaningful and well-understood summary statistic for conditional expectations under almost all circumstances undoubtedly contributes to the primacy of OLS regression as an empirical tool. In view of the possibility of interpretation under misspecification, modern theoretical research on regression inference also expressly allows for misspecification of the regression function when deriving limiting distributions (White, 1980).

While QR estimates are as easy to compute as OLS regression coefficients, an important difference between OLS and QR is that most of the theoretical and applied work on QR postulates a correctly specified linear model for conditional quantiles. This raises the question of whether and how QR estimates can be interpreted when the linear model for conditional quantiles is misspecified (for example, QR estimates at different quantiles may imply conditional quantile functions that cross). One interpretation for QR under misspecification is that it provides the best linear predictor for a response variable under asymmetric loss. This interpretation is not very satisfying, however, since prediction under asymmetric loss is typically not the object of interest in empirical work.[1] Empirical research on quantile regression with discrete covariates suggests that QR may have an approximation property similar to that of OLS, but the exact nature of the linear approximation has remained an important unresolved question (Chamberlain, 1994, p. 181).

The first contribution of this paper is to show that QR is the best linear approximation to the conditional quantile function using a weighted mean-squared error loss function, much as OLS regression provides a minimum mean-squared loss fit to the conditional expectation function. The implied QR weighting function can be used to understand which, if any, parts of the distribution of regressors contribute disproportionately to a particular set of QR estimates. We also show how this approximation property can be used to interpret multivariate QR coefficients as partial regression coefficients and to develop an omitted variables bias formula for QR. A second contribution is to present a distribution theory for the QR process that accounts for possible misspecification of the conditional quantile function. We present the main inference results only, with proofs available in a supplementary appendix. The approximation theorems and inference results in the paper are illustrated with an analysis of wage data from recent U.S. censuses.[2] The results show a sharp change in the quantile process of schooling coefficients in the 2000 census, and an increase in conditional inequality in the upper half of the wage distribution from 1990-2000.

[1]An exception is the forecasting literature; see, e.g., Giacomini and Komunjer (2003).

[2]Quantile regression has been widely used to model changes in the wage distribution; see, e.g., Buchinsky (1994), Abadie (1997), Gosling, Machin, and Meghir (2000), Autor, Katz, and Kearney (2004).

The paper is organized as follows. Section 2 introduces assumptions and notation and presents the main approximation theorems. Section 3 presents inference theory for QR processes under misspecification. Section 4 illustrates QR approximation properties with U.S. census data. Section 5 concludes.

3.2 Interpreting QR Under Misspecification

3.2.1 Notation and Framework

Given a continuous response variable Y and a $d \times 1$ regressor vector X, we are interested in the conditional quantile function (CQF) of Y given X. The conditional quantile function is defined as:

$$Q_\tau(Y|X) := \inf \{y : F_Y(y|X) \geq \tau\},$$

where $F_Y(y|X)$ is the distribution function for Y conditional on X, which is assumed to have conditional density $f_Y(y|X)$. The CQF is also known to be a solution to the following minimization problem, assuming integrability:

$$Q_\tau(Y|X) \in \arg \min_{q(X)} E\left[\rho_\tau(Y - q(X))\right], \tag{3.2.1}$$

where $\rho_\tau(u) = (\tau - 1(u \leq 0))u$ and the minimum is over the set of measurable functions of X. This is a potentially infinite-dimensional problem if covariates are continuous, and can be high-dimensional even with discrete X. It may nevertheless be possible to capture important features of the CQF using a linear model. This motivates linear quantile regression.

The linear quantile regression (QR), introduced by Koenker and Bassett (1978), solves the following minimization problem in the population, assuming integrability and uniqueness of the solution:

$$\beta(\tau) := \arg \min_{\beta \in \mathbb{R}^d} E\left[\rho_\tau(Y - X'\beta)\right]. \tag{3.2.2}$$

If $q(X)$ is in fact linear, the QR minimand will find it, just as if the conditional expectation function is linear, OLS will find it. More generally, QR provides the best linear predictor for Y under the asymmetric loss function, ρ_τ. As noted in the introduction, however, prediction under asymmetric loss is rarely the object of empirical work. Rather, the conditional quantile function is usually of intrinsic interest. For example, labor economists are often interested in comparisons of conditional deciles as a measure of how the spread of a wage distribution changes conditional on covariates, as in Katz and Murphy (1992), Juhn, Murphy, and Pierce (1993), and Buchinsky (1994). Thus, our first goal is to establish the nature of the approximation to conditional quantiles that QR provides.

3.2.2 QR Approximation Properties

Our principal theoretical result is that the population QR vector minimizes a weighted sum of squared specification errors. This is easiest to show using notation for a quantile-specific specification error and for a quantile-specific residual. For any quantile index $\tau \in (0,1)$, we define the QR specification error as:

$$\Delta_\tau(X,\beta) := X'\beta - Q_\tau(Y|X).$$

Similarly, let ϵ_τ be a quantile-specific residual, defined as the deviation of the response variable from the conditional quantile of interest:

$$\epsilon_\tau := Y - Q_\tau(Y|X),$$

with conditional density $f_{\epsilon_\tau}(e|X)$ at $\epsilon_\tau = e$. The following theorem shows that QR is a weighted least squares approximation to the unknown CQF.

Theorem 1 (Approximation Property) *Suppose that (i) the conditional density $f_Y(y|X)$ exists a.s., (ii) $E[Y]$, $E[Q_\tau(Y|X)]$, and $E\|X\|$ are finite, and (iii) $\beta(\tau)$ uniquely solves (3.2.2). Then*

$$\beta(\tau) = \arg \min_{\beta \in \mathbb{R}^d} E\left[w_\tau(X,\beta) \cdot \Delta_\tau^2(X,\beta)\right],$$

where

$$\begin{aligned} w_\tau(X,\beta) &= \int_0^1 (1-u) \cdot f_{\epsilon_\tau}\left(u\Delta_\tau(X,\beta)|X\right)\, du \\ &= \int_0^1 (1-u) \cdot f_Y\left(u \cdot X'\beta + (1-u) \cdot Q_\tau(Y|X)|X\right) du \geq 0. \end{aligned}$$

PROOF: We have that $\beta(\tau) = \arg\min_{\beta \in \mathbb{R}^d} E[\rho_\tau(\epsilon_\tau - \Delta_\tau(X,\beta))]$, or equivalently, since $E[\rho_\tau(\epsilon_\tau)]$ does not depend on β and is finite by condition (ii),

$$\beta(\tau) = \arg \min_{\beta \in \mathbb{R}^d} \left\{E\left[\rho_\tau\left(\epsilon_\tau - \Delta_\tau(X,\beta)\right)\right] - E\left[\rho_\tau\left(\epsilon_\tau\right)\right]\right\}. \tag{3.2.3}$$

By definition of ρ_τ and the law of iterated expectations, it follows further that

$$\beta(\tau) = \arg\min_{\beta \in \mathbb{R}^d} \{E[\mathcal{A}(X,\beta)] - E[\mathcal{B}(X,\beta)]\},$$

where

$$\begin{aligned}
\mathcal{A}(X,\beta) &= E\left[(1\{\epsilon_\tau < \Delta_\tau(X,\beta)\} - \tau)\,\Delta_\tau(X,\beta)\ |X\right],\\
\mathcal{B}(X,\beta) &= E\left[(1\{\epsilon_\tau < \Delta_\tau(X,\beta)\} - 1\{\epsilon_\tau < 0\})\,\epsilon_\tau|X\right].
\end{aligned}$$

The conclusion of the theorem can then be obtained by showing that

$$\mathcal{A}(X,\beta) = \left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)du\right)\cdot \Delta_\tau^2(X,\beta), \tag{3.2.4}$$

$$\mathcal{B}(X,\beta) = \left(\int_0^1 uf_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)du\right)\cdot \Delta_\tau^2(X,\beta), \tag{3.2.5}$$

establishing that both components are density-weighted quadratic specification errors.

Consider $\mathcal{A}(X,\beta)$ first. Observe that

$$\begin{aligned}
\mathcal{A}(X,\beta) &= [F_{\epsilon_\tau}(\Delta_\tau(X,\beta)|X) - F_{\epsilon_\tau}(0|X)]\,\Delta_\tau(X,\beta)\\
&= \left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)\Delta_\tau(X,\beta)du\right)\Delta_\tau(X,\beta),
\end{aligned} \tag{3.2.6}$$

where the first statement follows by the definition of conditional expectation and noting that $E[1\{\epsilon_\tau \leq 0\}|X] = F_{\epsilon_\tau}(0|X) = \tau$ and the second follows from the fundamental theorem of calculus (for Lebesgue integrals). This verifies (3.2.4). Turning to $\mathcal{B}(X,\beta)$, suppose first that $\Delta_\tau(X,\beta) > 0$. Then, setting $u_\tau = \epsilon_\tau/\Delta_\tau(X,\beta)$, we have

$$\begin{aligned}
\mathcal{B}(X,\beta) &= E\ \left[1\{\epsilon_\tau \in [0,\Delta_\tau(X,\beta)]\}\cdot \epsilon_\tau\Big|X\right]\\
&= E\left[1\{u_\tau \in [0,1]\}\cdot u_\tau \cdot \Delta_\tau(X,\beta)\Big|X\right]\\
&= \left(\int_0^1 uf_{u_\tau}(u|X)du\right)\Delta_\tau(X,\beta)\\
&= \left(\int_0^1 uf_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)\Delta_\tau(X,\beta)du\right)\cdot \Delta_\tau(X,\beta),
\end{aligned} \tag{3.2.7}$$

which verifies (3.2.5). A similar argument shows that (3.2.5) also holds if $\Delta_\tau(X,\beta) < 0$. Finally, if $\Delta_\tau(X,\beta) = 0$, then $\mathcal{B}(X,\beta) = 0$, so that (3.2.5) holds in this case too. *Q.E.D.*

Theorem 1 states that the population QR coefficient vector $\beta(\tau)$ minimizes the expected weighted mean squared approximation error, i.e., the square of the difference between the true CQF and a linear approximation, with weighting function $w_\tau(X, \beta)$.[3] The weights are given by the average density of the response variable over a line from the point of approximation, $X'\beta$, to the true conditional quantile, $Q_\tau(Y|X)$. Pre-multiplication by the term $(1-u)$ in the integral results in more weight being applied at points on the line closer to the true CQF.

We refer to the function $w_\tau(X, \beta)$ as defining *importance weights*, since this function determines the importance the QR minimand gives to points in the support of X for a given distribution of X.[4] In addition to the importance weights, the probability distribution of X also determines the ultimate weight given to different values of X in the least squares problem. To see this, note that we can also write the QR minimand as

$$\beta(\tau) = \arg \min_{\beta \in \mathbb{R}^d} \int \Delta_\tau^2(x, \beta)\ w_\tau(x, \beta)\ d\Pi(x),$$

where $\Pi(x)$ is the distribution function of X with associated probability mass or density function $\pi(x)$. Thus, the overall weight varies in the distribution of X according to

$$w_\tau(x, \beta) \cdot \pi(x).$$

A natural question is what determines the shape of the importance weights. This can be understood using the following approximation. When Y has a smooth conditional density, we have for β in the neighborhood of $\beta(\tau)$:

$$w_\tau(X, \beta) = 1/2 \cdot f_Y\left(Q_\tau(Y|X)|X\right) + \varrho_\tau(X), \qquad |\varrho_\tau(X)| \leq 1/6 \cdot |\Delta_\tau(X, \beta)| \cdot \bar{f}'(X). \tag{3.2.8}$$

Here, $\varrho_\tau(X)$ is a remainder term and the density $f_Y\left(y|X\right)$ is assumed to have a first derivative in y bounded in absolute value by $\bar{f}'(X)$ a.s.[5] Hence in many cases the *density weights* $1/2 \cdot f_Y\left(Q_\tau(Y|X)|X\right)$ are the primary determinants of the importance weights, a point we illustrate in Section 4. It is also of interest to note that $f_Y\left(Q_\tau(Y|X)|X\right)$ is constant across X in location models, and inversely proportional to the conditional standard deviation in location-

[3] Note that if we define $\beta(\tau)$ via (3.2.3), then integrability of Y is not required in Theorem 1.

[4] This terminology should not to be confused with similar terminology from Bayesian statistics.

[5] The remainder term $\varrho_\tau(X) = w_\tau(X, \beta) - (1/2) \cdot f_{\epsilon_\tau}(0|X)$ is bounded as $|\varrho_\tau(X)| = |\int_0^1 (1-u)(f_{\epsilon_\tau}(u \cdot \Delta_\tau(X, \beta)|X) - f_{\epsilon_\tau}(0|X))du| \leq |\Delta_\tau(X, \beta)| \cdot \bar{f}'(X) \cdot \int_0^1 (1-u) \cdot u \cdot du = (1/6) \cdot |\Delta_\tau(X, \beta)| \cdot \bar{f}'(X)$.

scale models.[6]

QR has a second approximation property closely related to the first. This second property is particularly well-suited to the development of a partial regression decomposition and the derivation of an omitted variables bias formula for QR.

Theorem 2 (Iterative Approximation Property) *Suppose that (i) the conditional density $f_Y(y|X)$ exists and is bounded a.s., (ii) $E[Y]$, $E[Q_\tau(Y|X)^2]$, and $E\|X\|^2$ are finite, and (iii) $\beta(\tau)$ uniquely solves (3.2.2). Then $\bar{\beta}(\tau) = \beta(\tau)$ uniquely solves the equation*

$$\bar{\beta}(\tau) = \arg\min_{\beta \in \mathbb{R}^d} E\left[\bar{w}_\tau(X, \bar{\beta}(\tau)) \cdot \Delta_\tau^2(X, \beta)\right], \tag{3.2.9}$$

where

$$\begin{aligned} \bar{w}_\tau(X, \bar{\beta}(\tau)) &= \frac{1}{2}\int_0^1 f_{\epsilon_\tau}\left(u \cdot \Delta_\tau(X, \bar{\beta}(\tau))|X\right)\, du \\ &= \frac{1}{2}\int_0^1 f_Y\left(u \cdot X'\bar{\beta}(\tau) + (1-u)\cdot Q_\tau(Y|X)|X\right)\, du \geq 0. \end{aligned}$$

PROOF: We want show that

$$\beta(\tau) = \arg\min_{\beta \in \mathbb{R}^d} E[\rho_\tau(Y - X'\beta)], \tag{3.2.10}$$

is equivalent to the fixed point $\bar{\beta}(\tau)$ that uniquely solves

$$\bar{\beta}(\tau) = \arg\min_{\beta \in \mathbb{R}^d} E\left[\bar{w}_\tau(X, \bar{\beta}(\tau)) \cdot \Delta_\tau^2(X, \beta)\right], \tag{3.2.11}$$

where the former and the latter objective functions are finite by conditions (i) and (ii).

By convexity of (3.2.11) in β, any fixed point $\beta = \bar{\beta}(\tau)$ solves the first order condition:

$$\mathcal{F}(\beta) := \ 2 \cdot E\left[\bar{w}_\tau(X, \beta)\ \Delta_\tau(X, \beta)\ X\right] = 0.$$

By convexity of (3.2.10) in β, the quantile regression vector $\beta = \beta(\tau)$ solves the first order condition:

$$\mathcal{D}(\beta) := E\left[\mathcal{D}(X, \beta)\right] = 0,$$

[6]A location-scale model is any model of the form $Y = \mu(X) + \sigma(X) \cdot e$, where e is independent of X. The location model results from setting $\sigma(X) = \sigma$.

where

$$\mathcal{D}(X,\beta) := E\left[\left(1\{\epsilon_\tau < \Delta_\tau(X,\beta)\} - \tau\right) X | X\right].$$

An argument similar to that used to establish equation (3.2.6) yields

$$\begin{aligned}\mathcal{D}(X,\beta) &= \left(F_{\epsilon_\tau}(\Delta_\tau(X,\beta)|X) - F_{\epsilon_\tau}(0|X)\right)\cdot X \\ &= \left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X,\beta)|X)du\right)\cdot \Delta_\tau(X,\beta)\cdot X \\ &= 2\cdot \bar{w}_\tau(X,\beta)\cdot \Delta_\tau(X,\beta)\cdot X,\end{aligned}$$

where we also use the definition of $\bar{w}_\tau(X;\beta)$. The functions $\mathcal{F}(\beta)$ and $\mathcal{D}(\beta)$ are therefore identical. Since $\beta = \beta(\tau)$ uniquely satisfies $\mathcal{D}(\beta) = 0$, it also uniquely satisfies $\mathcal{F}(\beta) = 0$. As a result, $\beta = \beta(\tau) = \bar{\beta}(\tau)$ is the unique solution to both (3.2.10) and (3.2.11). *Q.E.D.*

Theorem 2 differs from Theorem 1 in that it characterizes the QR coefficient as a fixed point to an iterated minimum distance approximation. Consequently, the importance weights $\bar{w}(X,\beta(\tau))$ in this approximation are defined using the QR vector $\beta(\tau)$ itself. The weighting function $\bar{w}_\tau(X,\beta(\tau))$ is also related to the conditional density of the dependent variable. In particular, when the response variable has a smooth conditional density around the relevant quantile, we have by a Taylor approximation

$$\bar{w}_\tau(X,\beta(\tau)) = 1/2\cdot f_Y\left(Q_\tau(Y|X)|X\right) + \bar{\varrho}_\tau(X), \quad |\bar{\varrho}_\tau(X)| \le 1/4\cdot |\Delta_\tau(X,\beta(\tau))|\cdot \bar{f}'(X),$$

where $\bar{\varrho}_\tau(X)$ is a remainder term, and the density $f_Y\,(y|X)$ is assumed to have a first derivative in y bounded in absolute value by $\bar{f}'(X)$ a.s. When either $\Delta_\tau(X,\beta(\tau))$ or $\bar{f}'(X)$ is small, we then have

$$\bar{w}_\tau(X,\beta(\tau)) \approx w_\tau(X,\beta(\tau)) \approx \frac{1}{2} f_Y(Q_\tau(Y|X)|X).$$

The approximate weighting function is therefore the same as derived using Theorem 1.

3.2.3 Partial Quantile Regression and Omitted Variable Bias

Partial quantile regression is defined with regard to a partition of the regressor vector X into a variable, X_1, and the remaining variables X_2, along with the corresponding partition of QR coefficients $\beta(\tau)$ into $\beta_1(\tau)$ and $\beta_2(\tau)$. We can now decompose $Q_\tau(Y|X)$ and X_1 using orthogonal projections onto X_2 weighted by $\bar{w}_\tau(X) := \bar{w}(X; \beta(\tau))$ defined in Theorem 2:

$$Q_\tau(Y|X) = X_2'\pi_Q + q_\tau(Y|X), \quad \text{where } E[\bar{w}_\tau(X) \cdot X_2 \cdot q_\tau(Y|X)] = 0,$$
$$X_1 = X_2'\pi_1 + V_1, \quad \text{where } E[\bar{w}_\tau(X) \cdot X_2 \cdot V_1] = 0.$$

In this decomposition, $q_\tau(Y|X)$ and V_1 are residuals created by a weighted linear projection of $Q_\tau(Y|X)$ and X_1 on X_2, respectively, using $\bar{w}_\tau(X)$ as the weight.[7] Standard least squares algebra then gives

$$\beta_1(\tau) = \arg\min_{\beta_1} E\left[\bar{w}_\tau(X)\ (q_\tau(Y|X) - V_1\beta_1)^2\right]$$

and also $\beta_1(\tau) = \arg\min_{\beta_1} E\left[\bar{w}_\tau(X)\ (Q_\tau(Y|X) - V_1\beta_1)^2\right]$. This shows that $\beta_1(\tau)$ is a partial quantile regression coefficient in the sense that it can be obtained from a weighted least squares regression of $Q_\tau(Y|X)$ on X_1, once we have partialled out the effect of X_2. Both the first-step and second-step regressions are weighted by $\bar{w}_\tau(X)$.

We can similarly derive an omitted variables bias formula for QR. In particular, suppose we are interested in a quantile regression with explanatory variables $X = [X_1', X_2']'$, but X_2 is not available, e.g., a measure of ability or family background in a wage equation. We run QR on X_1 only, obtaining the coefficient vector $\gamma_1(\tau) = \arg\min_{\gamma_1} E[\rho_\tau(Y - X_1'\gamma_1)]$. The long regression coefficient vectors are given by $(\beta_1(\tau)', \beta_2(\tau)')' = \arg\min_{\beta_1,\beta_2} E[\rho_\tau(Y - X_1'\beta_1 - X_2'\beta_2)]$. Then,

$$\gamma_1(\tau) = \beta_1(\tau) + (E[\tilde{w}_\tau(X) \cdot X_1X_1'])^{-1}E[\tilde{w}_\tau(X) \cdot X_1R_\tau(X)],$$

where $R_\tau(X) := Q_\tau(Y|X) - X_1'\beta_1(\tau)$, $\tilde{w}_\tau(X) := \int_0^1 f_{\epsilon_\tau}(u \cdot \Delta_\tau(X, \gamma_1(\tau))|X)du/2$, $\Delta_\tau(X, \gamma_1) := X_1'\gamma_1 - Q_\tau(Y|X)$, and $\epsilon_\tau := Y - Q_\tau(Y|X)$.[8] Here $R_\tau(X)$ is the part of the CQF not explained by the linear function of X_1 in the long QR. If the CQF is linear, then $R_\tau(X) = X_2'\beta_2(\tau)$. The proof of this result is similar to the previous arguments and therefore omitted.

As with OLS short and long calculations, the omitted variables formula in this case shows the short QR coefficients to be equal to the corresponding long QR coefficients plus the coefficients in a weighted projection of omitted effects on included variables. While the parallel with OLS seems clear, there are two complications in the QR case. First, the effect of omitted variables appears through the remainder term, $R_\tau(X)$. In practice, it seems reasonable to think of this as being approximated by the omitted linear part, $X_2'\beta_2(\tau)$. Second, the regression of omitted

[7]Thus, $\pi_Q = E\left[\bar{w}_\tau(X)X_2X_2'\right]^{-1}E\left[\bar{w}_\tau(X)X_2Q_\tau(Y|X)\right]$ and $\pi_1 = E\left[\bar{w}_\tau(X)X_2X_2'\right]^{-1}E\left[\bar{w}_\tau(X)X_2X_1\right]$.

[8]Note that the weights in this case depend on how the regressor vector is partitioned.

variables on included variables is weighted by $\tilde{w}_\tau(X)$, while for OLS it is unweighted.[9]

3.3 Sampling Properties of QR Under Misspecification

Parallelling the interest in robust inference methods for OLS, it is also of interest to know how specification error affects inference for QR. In this case, inference under misspecification means distribution theory for quantile regressions in large samples without imposing the restriction that the CQF is linear. While not consistent for the true nonlinear CQF, quantile regression consistently estimates the approximations to the CQF given in Theorems 1 and 2. We would therefore like to quantify the sampling uncertainty in estimates of these approximations. This question can be compactly and exhaustively addressed by obtaining the large sample distribution of the sample quantile regression process, which is defined by taking all or many sample quantile regressions.

As in Koenker and Xiao (2001), the entire QR process is of interest here because we would like to either test global hypotheses about (approximations to) conditional distributions or make comparisons across different quantiles. Therefore our interest is in the QR process, and is not confined to a specific quantile. The second motivation for studying the process comes from the fact that formal statistical comparisons across quantiles, often of interest in empirical work, require the construction of simultaneous (joint) confidence regions. Process methods provide a natural and simple way of constructing these regions.

The QR process $\hat{\beta}(\cdot)$ is formally defined as

$$\hat{\beta}(\tau) \in \arg\min_{\beta \in \mathbb{R}^d} n^{-1} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta), \ \tau \in \mathcal{T} := \text{ a closed subset of } [\epsilon, 1-\epsilon] \text{ for } \epsilon > 0. \quad (3.3.1)$$

Koenker and Machado (1999) and Koenker and Xiao (2001) previously focused on QR process inference in correctly specified models, while earlier treatments of specification error discussed only pointwise inference for a single quantile coefficient (Hahn, 1997). As it turns out, the empirical results in the next section show misspecification has a larger effect on process inference than on pointwise inference. Our main theoretical result on inference is as follows:

Theorem 3 *Suppose that (i) $(Y_i, X_i, i \leq n)$ are iid on the probability space $(\Omega, \mathcal{F}, P)$ for each n, (ii) the conditional density $f_Y(y|X = x)$ exists, and is bounded and uniformly continuous in*

[9]The formula obtained above can be used to determine the bias from measurement error in regressors, by setting the error to be the omitted variable. This suggests that classical measurement error is likely to generate an attenuation bias in QR as well as OLS estimates. We thank Arthur Lewbel for pointing this out.

y, uniformly in x over the support of X, (iii) $J(\tau) := E[f_Y(X'\beta(\tau)|X)XX']$ is positive definite for all $\tau \in \mathcal{T}$, and (iv) $E\ \|X\|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$. Then, the quantile regression process is uniformly consistent, $\sup_{\tau\in\mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\| = o_p(1)$, and $J(\cdot)\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$ converges in distribution to a zero mean Gaussian process $z(\cdot)$, where $z(\cdot)$ is defined by its covariance function $\Sigma(\tau, \tau') := E\ [z(\tau)z(\tau')']$, with

$$\Sigma(\tau, \tau') = E\ [(\tau - 1\{Y < X'\beta(\tau)\})(\tau' - 1\{Y < X'\beta(\tau')\})XX']. \tag{3.3.2}$$

If the model is correctly specified, i.e. $Q_\tau(Y|X) = X'\beta(\tau)$ a.s., then $\Sigma(\tau, \tau')$ simplifies to

$$\Sigma_0(\tau, \tau') := [\min(\tau, \tau') - \tau\tau'] \cdot E\ [XX']. \tag{3.3.3}$$

Theorem 3 establishes joint asymptotic normality for the entire QR process.[10] The proof of this theorem appears in the supplementary appendix. Theorem 3 allows for misspecification and imposes little structure on the underlying conditional quantile function (e.g., smoothness of $Q_\tau(Y|X)$ in X, needed for a fully nonparametric approach, is not needed here). The result states that the limiting distribution of the QR process (and of any single QR coefficient) will in general be affected by misspecification. The covariance function that describes the limiting distribution is generally different from the covariance function that arises under correct specification.

Inference on the QR process is useful for testing basic hypotheses of the form:

$$R(\tau)'\beta(\tau) = r(\tau) \text{ for all } \tau \in \mathcal{T}. \tag{3.3.4}$$

For example, we may be interested in whether a variable or a subset of variables $j \in \{k + 1, ..., d\}$ enter the regression equations at all quantiles with zero coefficients, i.e. whether $\beta_j(\tau) = 0$ for all $\tau \in \mathcal{T}$ and $j \in \{k+1, ..., d\}$. This corresponds to $R(\tau) = \left[0_{(d-k)\times k}\ I_{d-k}\right]$ and $r(\tau) = 0_{d-k}$. Similarly, we may want to construct simultaneous (uniform) confidence intervals for linear functions of parameters

$$R(\tau)'\beta(\tau) - r(\tau) \quad \text{ for all } \tau \in \mathcal{T}.$$

Theorem 3 has a direct consequence for these confidence intervals and hypothesis testing, since it implies that $(EXX')^{-1}\Sigma(\tau, \tau') \neq [\min(\tau, \tau') - \tau\tau'] \cdot I_d$. That is, the covariance function under misspecification is not proportional to the covariance function of the standard d-dimensional

[10] A simple corollary is that any finite collection of $\sqrt{n}(\hat{\beta}(\tau_k) - \beta(\tau_k))$, $k = 1, 2, ...,$ are asymptotically jointly normal, with asymptotic covariance between the k-th and l-th subsets equal to $J(\tau_k)^{-1}\Sigma(\tau_k, \tau_l)J(\tau_l)^{-1}$. Hahn (1997) previously derived this corollary for a single quantile.

Brownian bridge arising in the correctly specified case. Hence, unlike in the correctly specified case, the critical values for confidence regions and tests are not distribution-free and can not be obtained from standard tabulations based on the Brownian bridge. However, the following corollaries facilitate both testing and the construction of confidence intervals under misspecification:

Corollary 1 *Define* $V(\tau) := R(\tau)'J(\tau)^{-1}\ \Sigma(\tau,\tau)\ J(\tau)^{-1}R(\tau)$ *and* $|x| := \max_j |x_j|$. *Under the conditions of Theorem 3, the Kolmogorov statistic* $\mathcal{K}_n := \sup_{\tau\in\mathcal{T}}\ |V(\tau)^{-1/2}\sqrt{n}(R(\tau)'\hat{\beta}(\tau) - r(\tau))|$ *for testing (3.3.4) converges in distribution to variable* $\mathcal{K} := \sup_{\tau\in\mathcal{T}} |V(\tau)^{-1/2}R(\tau)'J(\tau)^{-1}z(\tau)|$ *with an absolutely continuous distribution. The result is not affected by replacing* $J(\tau)$ *and* $\Sigma(\tau,\tau)$ *with estimates that are consistent uniformly in* $\tau \in \mathcal{T}$.

Thus, Kolmogorov-type statistics have a well-behaved limit distribution.[11] Unlike in the correctly specified case, however, this distribution is non-standard. Nevertheless, critical values and simultaneous confidence regions can be obtained as follows:

Corollary 2 *For* $\kappa(\alpha)$ *denoting the* α*-quantile of* $\mathcal{K}$ *and* $\hat{\kappa}(\alpha)$ *any consistent estimate of it, for instance the estimate defined below,* $\lim_{n\to\infty} P\{\sqrt{n}(R(\tau)'\beta(\tau) - r(\tau)) \in \widehat{I}_n(\tau)$, *for all* $\tau \in \mathcal{T}\} = \alpha$, *where* $\widehat{I}_n(\tau) = [u(\tau) : |V(\tau)^{-1/2}\sqrt{n}(R(\tau)'\hat{\beta}(\tau) - r(\tau) - u(\tau))| \leq \hat{\kappa}(\alpha)]$. *If* $R(\tau)'\beta(\tau) - r(\tau)$ *is scalar, the simultaneous confidence interval is* $\widehat{I}_n(\tau) = [R(\tau)'\hat{\beta}(\tau) - r(\tau) \pm \hat{\kappa}(\alpha)\cdot V(\tau)^{1/2}]$. *This result is not affected by replacing* $V(\tau)$ *with an estimate that is consistent uniformly in* $\tau \in \mathcal{T}$.

A consistent estimate of the critical value, $\hat{\kappa}(\alpha)$, can be obtained by subsampling. Let $j = 1,...,B$ index B randomly chosen subsamples of $((Y_i, X_i), i \leq n)$ of size b, where $b \to \infty$, $b/n \to 0$, $B \to \infty$ as $n \to \infty$. Compute the test statistic for each subsample as $K_j = \sup_{\tau\in\mathcal{T}} |\hat{V}(\tau)^{-1/2}\sqrt{b}R(\tau)'(\widehat{\beta}_j(\tau) - \widehat{\beta}(\tau))|$, where $\widehat{\beta}_j(\tau)$ is the QR estimate using j-th subsample. Then, set $\widehat{\kappa}(\alpha)$ to be the α-quantile of $\{K_1, ..., K_B\}$.

Finally, the inference procedure above requires estimators of $\Sigma(\tau,\tau')$ and $J(\tau)$ that are uniformly consistent in $(\tau,\tau') \in \mathcal{T}\times\mathcal{T}$. These are given by:

$$\begin{aligned}
\hat{\Sigma}(\tau,\tau') &= n^{-1}\sum_{i=1}^{n}(\tau - 1\{Y_i \leq X_i'\hat{\beta}(\tau)\})(\tau' - 1\{Y_i \leq X_i'\hat{\beta}(\tau')\})\cdot X_iX_i', \\
\hat{J}(\tau) &= (2nh_n)^{-1}\sum_{i=1}^{n} 1\{|Y_i - X_i'\hat{\beta}(\tau)| \leq h_n\}\cdot X_iX_i',
\end{aligned}$$

[11] In practice, by stochastic equicontinuity of QR process, in the definition of $\mathcal{K}_n$-statistics we can replace any continuum of quantile indices $\mathcal{T}$ by a finite-grid $\mathcal{T}_{K_n}$, where the distance between adjacent grid points goes to zero as $n \to \infty$.

where $\hat{\Sigma}(\tau, \tau')$ differs from its usual counterpart $\hat{\Sigma}_0(\tau, \tau') = [\min(\tau, \tau') - \tau\tau'] \cdot n^{-1} \sum_{i=1}^{n} X_i X_i'$ used in the correctly specified case; and $\hat{J}(\tau)$ is Powell's (1986) estimator of the Jacobian, with h_n such that $h_n \to 0$ and $h_n^2 n \to \infty$. Koenker (1994) suggests $h_n = C \cdot n^{-1/3}$ and provides specific choices of C. The supplementary appendix shows that these estimates are consistent uniformly in (τ, τ').

3.4 Application to U.S. Wage Data

In this section we study the approximation properties of QR in widely used U.S. Census micro data sets.[12] The main purpose of this section is to show that linear QR indeed provides a useful minimum distance approximation to the conditional distribution of wages, accurately capturing changes in the wage distribution from 1980 to 2000. We also report new substantive empirical findings arising from the juxtaposition of data from the 2000 census with earlier years. The inference methods derived in the previous section facilitate this presentation. In our analysis, Y is the real log weekly wage for U.S. born men aged 40-49, calculated as the log of reported annual income from work divided by weeks worked in the previous year, and the regressor X consists of a years-of-schooling variable and other basic controls.[13]

The nature of the QR approximation property is illustrated in Figure 1. Panels A-C plot a nonparametric estimate of the conditional quantile function, $Q_\tau(Y|X)$, along with the linear QR fit for the 0.10, 0.50, and 0.90 quantiles, where X includes only the schooling variable. Here we take advantage of the discreteness of the schooling variable and the large census sample to compare QR fits to the nonlinear CQFs computed at each point in the support of X. We focus on the 1980 data for this figure because the 1980 Census has a true highest grade completed variable, while for more recent years this must be imputed. It should be noted, however, that the approximation results for the 1990 and 2000 censuses are similar.

Our theorems establish that QR implicitly provides a weighted minimum distance approximation to the true nonlinear CQF. It is therefore useful to compare the QR fit to an explicit

[12]The data were drawn from the 1% self-weighted 1980 and 1990 samples, and the 1% weighted 2000 sample, all from the IPUMS website (Ruggles *et al.,* 2003). The sample consists of US-born black and white men of age 40-49 with 5 or more years of education, with positive annual earnings and hours worked in the year preceding the census. Individuals with imputed values for age, education, earnings or weeks worked were also excluded from the sample. The resulting sample sizes were 65,023, 86,785, and 97,397 for 1980, 1990, and 2000.

[13]Annual income is expressed in 1989 dollars using the Personal Consumption Expenditures Price Index. The schooling variable for 1980 corresponds to the highest grade of school completed. The categorical schooling variables in the 1990 and 2000 Census were converted to years of schooling using essentially the same coding scheme as in Angrist and Krueger (1999). See Angrist, Chernozhukov, and Fernandez-Val (2004) for details.

minimum distance (MD) fit similar to that discussed by Chamberlain (1994).[14] The MD estimator for QR is the sample analog of the vector $\tilde{\beta}(\tau)$ solving

$$\tilde{\beta}(\tau) = \arg\min_{\beta \in \mathbb{R}^d} E\left[(Q_\tau(Y|X) - X'\beta)^2\right] = \arg\min_{\beta \in \mathbb{R}^d} E\left[\Delta_\tau^2(X,\beta)\right].$$

In other words, $\tilde{\beta}(\tau)$ is the slope of the linear regression of $Q_\tau(Y|X)$ on X, weighted only by the probability mass function of X, $\pi(x)$. In contrast to QR, this MD estimator relies on the ability to estimate $Q_\tau(Y|X)$ in a nonparametric first step, which, as noted by Chamberlain (1994), may be feasible only when X is low dimensional, the sample size is large, and sufficient smoothness of $Q_\tau(Y|X)$ is assumed.

Figure 1 plots this MD fit with a dashed line. The QR and MD regression lines are close, as predicted by our approximation theorems, but they are not identical because the additional weighting by $w_\tau(X,\beta)$ in the QR fit accentuates quality of the fit at values of X where Y is more densely distributed near true quantiles. To further investigate the QR weighting function, panels D-F in Figure 1 plot the overall QR weights, $w_\tau(X,\beta(\tau)) \cdot \pi(X)$, against the regressor X. The panels also show estimates of the importance weights from Theorem 1, $w_\tau(X,\beta(\tau))$, and their density approximations, $f_Y(Q_\tau(Y|X)|X)$.[15] The importance weights and the actual density weights are fairly close. The importance weights are stable across X and tend to accentuate the middle of the distribution a bit more than other parts. The overall weighting function ends up placing the highest weight on 12 years of schooling, implying that the linear QR fit should be the best in the middle of the design.

Also of interest is the ability of QR to track changes in quantile-based measures of conditional inequality. The column labeled CQ in panel A of Table 1 shows nonparametric estimates of the average 90-10 quantile spread conditional on schooling, potential experience, and race. This spread increased from 1.2 to about 1.35 from 1980 to 1990, and then to about 1.43 from 1990 to 2000. QR estimates match this almost perfectly, not surprisingly since an implication of our

[14]See Ferguson (1958) and Rothenberg (1971) for general discussions of MD. Buchinsky (1994) and Bassett, Knight, and Tam (2002) present other applications of MD to quantile problems.

[15]The importance weights defined in Theorem 1 are estimated at $\beta = \hat{\beta}(\tau)$ as follows:

$$\widehat{w}_\tau(X,\widehat{\beta}(\tau)) = (1/U) \cdot \sum_{u=1}^{U} [(1 - u/U) \cdot \widehat{f}_Y((u/U) \cdot X'\widehat{\beta}(\tau) + (1 - u/U) \cdot \widehat{Q}_\tau(Y|X)|X)], \tag{3.4.1}$$

where U is set to 100; $\hat{f}_Y(y|X)$ is a kernel density estimate of $f_Y(y|X)$, which employs a Gaussian kernel and Silverman's rule for bandwidth; $\widehat{Q}_\tau(Y|X)$ is a non-parametric estimate of $Q_\tau(Y|X)$, for each cell of the covariates X; and $X'\widehat{\beta}(\tau)$ is the QR estimate. Approximate weights are calculated similarly. Weights based on Theorem 2 are similar and therefore not shown.

theorems is that QR should fit (weighted) average quantiles exactly. The fit is not as good, however, when averages are calculated for specific schooling groups, as reported in panels B and C of the table. These results highlight the fact that QR is only an approximation. Table 1 also documents two important substantive findings, apparent in both the CQ and QR estimates. First, the table shows conditional inequality increasing in both the upper and lower halves of the wage distribution from 1980 to 1990, but in the top half only from 1990 to 2000. Second, the increase in conditional inequality since 1990 has been much larger for college graduates than for high school graduates.

Figure 2 provides a useful complement to, and a partial explanation for, the patterns and changes in Table 1. In particular, Panel A of the figure shows estimates of the schooling coefficient quantile process, along with robust simultaneous 95% confidence intervals. These estimates are from quantile regressions of log-earnings on schooling, race and a quadratic function of experience, using data from the 1980, 1990 and 2000 censuses.[16] The robust simultaneous confidence intervals allow us to asses of the significance of changes in schooling coefficients across quantiles and across years. The horizontal lines in the figure indicate the corresponding OLS estimates.

The figure suggests the returns to schooling were low and essentially constant across quantiles in 1980, a finding similar to Buchinsky's (1994) using Current Population Surveys for this period. On the other hand, the returns increased sharply and became much more heterogeneous in 1990 and especially in 2000, a result we also confirmed in Current Population Survey data. Since the simultaneous confidence bands do not contain a horizontal line, we reject the hypothesis of constant returns to schooling for 1990 and 2000. The fact that there are quantile segments where the simultaneous bands do not overlap indicates statistically significant differences across years at those segments. For instance, the 1990 band does not overlap with the 1980 band, suggesting a marked and statistically significant change in the relationship between schooling and the conditional wage distribution in this period.[17] The apparent twist in the schooling coefficient process explains why inequality increased for college graduates from 1990 to 2000. In the 2000 census, higher education was associated with increased wage dispersion to a much greater extent than in earlier years.

[16]The simultaneous bands were obtained by subsampling using 500 repetitions with subsample size $b = 5n^{2/5}$ and a grid of quantiles $\mathcal{T}_n = \{.10, .11, ..., .90\}$.

[17]Due to independence of samples across Census years, the test that looks for overlapping in two 95% confidence bands has a significance level of about 10%, namely $1 - .95^2$. Alternately, an α-level test can be based on a simultaneous α-level confidence band for the difference in quantile coefficients across years, again constructed using Theorem 3.

Another view of the stylized facts laid out in Table 1 is given in Figure 2B. This figure plots changes in the approximate conditional quantiles, based on a QR fit, with covariates evaluated at their mean values for each year. The figure also shows simultaneous 95% confidence bands. This figure provides a visual representation of the finding that between 1990 and 2000 conditional wage inequality increased more in the upper half of the wage distribution than in the lower half, while between 1980 and 1990 the increase in inequality occurred in both tails. Changes in schooling coefficients across quantiles and years, sharper above the median than below, clearly contributed to the fact that recent (conditional) inequality growth has been mostly confined to the upper half of the wage distribution.

Finally, it is worth noting that the simultaneous bands differ from the corresponding pointwise bands (the latter are not plotted). Moreover, the simultaneous bands allow multiple comparisons across quantiles without compromising confidence levels. Even more importantly in our context, accounting for misspecification substantially affects the width of simultaneous confidence intervals in this application. Uniform bands calculated assuming correct specification can be constructed using the critical values for the Kolmogorov statistic $\mathcal{K}$ reported in Andrews (1993). In this case, the resulting bands for the schooling coefficient quantile process are 26%, 23%, and 32% narrower than the robust intervals plotted in Figure 2A for 1980, 1990, and 2000.[18]

3.5 Summary and conclusions

We have shown how linear quantile regression provides a weighted least squares approximation to an unknown and potentially nonlinear conditional quantile function, much as OLS provides a least squares approximation to a nonlinear CEF. The QR approximation property leads to partial quantile regression relationships and an omitted variables bias formula analogous to those for OLS. While misspecification of the CQF functional form does not affect the usefulness of QR, it does have implications for inference. We also present a misspecification-robust distribution theory for the QR process. This provides a foundation for simultaneous confidence intervals and a basis for global tests of hypotheses about distributions.

An illustration using US. census data shows the sense in which QR fits the CQF. The

[18]The simultaneous bands take the form of $\hat{\beta}(\tau) \pm \hat{\kappa}(\alpha) \cdot$ robust std. error($\hat{\beta}(\tau)$). Using the procedure described in Corollary 2, we obtain estimates for $\hat{\kappa}(0.05)$ of 3.78, 3.70, and 3.99 for 1980, 1990, and 2000. The simultaneous bands that impose correct specification take the form $\hat{\beta}(\tau) \pm \kappa_0(\alpha) \cdot$ std. error($\hat{\beta}(\tau)$), where $\kappa_0(\alpha)$ is the α-quantile of the supremum of (the absolute value of) a standardized tied-down Bessel process of order 1. For example, $\kappa_0(0.05) = (9.31)^{1/2} = 3.05$ from Table I in Andrews (1993).

empirical example also shows that QR accurately captures changes in the wage distribution from 1980 to 2000. An important substantive finding is the sharp twist in schooling coefficients across quantiles in the 2000 census. We use simultaneous confidence bands robust to misspecification, to show that this pattern is highly significant. A related finding is that most inequality growth after 1990 has been in the upper part of the wage distribution.

Appendix

3.A Proof of Theorems 3 and its Corollaries

The proof has two steps.[19] The first step establishes uniform consistency of the sample QR process. The second step establishes asymptotic Gaussianity of the sample QR process.[20] For $W = (Y, X)$, let $\mathbb{E}_n[f(W)]$ denote $n^{-1}\sum_{i=1}^n f(W_i)$ and $\mathbb{G}_n[f(W)]$ denote $n^{-1/2}\sum_{i=1}^n (f(W_i) - E[f(W_i)])$. If $\widehat{f}$ is an estimated function, $\mathbb{G}_n[\widehat{f}(W)]$ denotes $n^{-1/2}\sum_{i=1}^n (f(W_i) - E[f(W_i)])_{f=\widehat{f}}$.

3.A.1 Uniform consistency of $\hat{\beta}(\cdot)$

For each τ in $\mathcal{T}$, $\hat{\beta}(\tau)$ minimizes $Q_n(\tau, \beta) := \mathbb{E}_n[\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))]$. Define $Q_\infty(\tau, \beta) := E[\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))]$. It is easy to show that $E\|X\| < \infty$ implies that $E|\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))| < \infty$. Therefore, $Q_\infty(\tau, \beta)$ is finite, and by the stated assumptions, it is uniquely minimized at $\beta(\tau)$ for each τ in $\mathcal{T}$.

We first show the uniform convergence, namely for any compact set $\mathcal{B}$, $Q_n(\tau, \beta) = Q_\infty(\tau, \beta) + o_{p^*}(1)$, uniformly in $(\tau, \beta) \in \mathcal{T} \times \mathcal{B}$. This statement holds pointwise by the Khinchine law of large numbers. The uniform convergence follows because $|Q_n(\tau', \beta') - Q_n(\tau'', \beta'')| \leq C_{1n} \cdot |\tau' - \tau''| + C_{2n} \cdot \|\beta' - \beta''\|$, where $C_{1n} = 2 \cdot \mathbb{E}_n\|X\| \cdot \sup_{\beta \in \mathcal{B}} \|\beta\| = O_p(1)$ and $C_{2n} = 2 \cdot \mathbb{E}_n\|X\| = O_p(1)$. Hence the empirical process $(\tau, \beta) \mapsto Q_n(\tau, \beta)$ is stochastically equicontinuous, which implies the uniform convergence.

Next, we show uniform consistency. Consider a collection of closed balls $B_M(\beta(\tau))$ of radius M and center $\beta(\tau)$, and let $\beta_M(\tau) = \beta(\tau) + \delta_M(\tau) \cdot v(\tau)$, where $v(\tau)$ is a direction vector with unity norm $\|v(\tau)\| = 1$ and $\delta_M(\tau)$ is a positive scalar such that $\delta_M(\tau) \geq M$. Then uniformly in $\tau \in \mathcal{T}$, $(M/\delta_M(\tau)) \cdot (Q_n(\tau, \beta_M(\tau)) - Q_n(\tau, \beta(\tau))) \overset{(a)}{\geq} Q_n(\tau, \beta_M^*(\tau)) - Q_n(\tau, \beta(\tau)) \overset{(b)}{\geq} Q_\infty(\tau, \beta_M^*(\tau)) - Q_\infty(\tau, \beta(\tau)) + o_{p^*}(1) \overset{(c)}{>} \epsilon_M + o_{p^*}(1)$, for some $\epsilon_M > 0$; where (a) follows by convexity in β, for $\beta_M^*(\tau)$ the point of the boundary of $B_M(\beta(\tau))$ on the line connecting $\beta_M(\tau)$ and $\beta(\tau)$; (b) follows by the uniform convergence established above; and (c) follows since $\beta(\tau)$ is the unique minimizer of $Q_\infty(\beta, \tau)$ uniformly in $\tau \in \mathcal{T}$, by convexity and assumption (iii). Hence for any $M > 0$, the minimizer $\hat{\beta}(\tau)$ must be within M from $\beta(\tau)$ uniformly for all $\tau \in \mathcal{T}$, with probability approaching one.

3.A.2 Asymptotic Gaussianity of $\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$

First, by the computational properties of $\hat{\beta}(\tau)$, for all $\tau \in \mathcal{T}$, cf. Theorem 3.3 in Koenker and Bassett (1978), we have that $\|\mathbb{E}_n[\varphi_\tau(Y - X'\hat{\beta}(\tau))X]\| \leq \text{const} \cdot \sup_{i \leq n} \|X_i\|/n$, where $\varphi_\tau(u) = \tau - 1\{u \leq 0\}$. Note that $E\|X_i\|^{2+\epsilon} < \infty$ implies $\sup_{i \leq n} \|X_i\| = o_{p^*}(n^{1/2})$, since $P\left(\sup_{i \leq n} \|X_i\| > n^{1/2}\right) \leq nP(\|X_i\| > n^{1/2}) \leq nE\|X_i\|^{2+\epsilon}/n^{\frac{2+\epsilon}{2}} = o(1)$. Hence uniformly in $\tau \in \mathcal{T}$,

$$\sqrt{n}\mathbb{E}_n\left[\varphi_\tau(Y - X'\hat{\beta}(\tau))X\right] = o_p(1). \tag{3.A.1}$$

[19]Basic concepts used in the proof, including weak convergence in the space of bounded functions, stochastic equicontinuity, Donsker and Vapnik-Červonenkis (VC) classes, are defined as in van der Vaart and Wellner (1996).

[20]The step does not rely on Pollard's (1991) convexity argument, as it does not apply to the process case.

Second, $(\tau, \beta) \mapsto \mathbb{G}_n \left[\varphi_\tau \left(Y - X'\beta\right) X\right]$ is stochastically equicontinuous over $\mathcal{B} \times \mathcal{T}$, where $\mathcal{B}$ is any compact set, with respect to the $L_2(P)$ pseudometric

$$\rho((\tau', \beta'), (\tau'', \beta''))^2 := \max_{j \in 1,\ldots,d} E\left[\left(\varphi_{\tau'}\left(Y - X'\beta'\right) X_j - \varphi_{\tau''}\left(Y - X'\beta''\right) X_j\right)^2\right],$$

for $j \in 1, ..., d$ indexing the components of X. Note that the functional class $\{\varphi_\tau \left(Y - X'\beta\right) X, \tau \in \mathcal{T}, \beta \in \mathcal{B}\}$ is formed as $(\mathcal{T} - \mathcal{F})X$, where $\mathcal{F} = \{1\{Y \leq X'\beta\}, \beta \in \mathcal{B}\}$ is a VC subgraph class and hence a bounded Donsker class. Hence $\mathcal{T} - \mathcal{F}$ is also bounded Donsker, and $(\mathcal{T} - \mathcal{F})X$ is therefore Donsker with a square integrable envelope $2 \cdot \max_{j \in 1,..d} |X|_j$, by Theorem 2.10.6 in Van der Vaart and Wellner (1996). The stochastic equicontinuity then is a part of being Donsker.

Third, by stochastic equicontinuity of $(\tau, \beta) \mapsto \mathbb{G}_n \left[\varphi_\tau \left(Y - X'\beta\right) X\right]$ we have that

$$\mathbb{G}_n \left[\varphi_\tau(Y - X'\hat{\beta}(\tau))X\right] = \mathbb{G}_n \left[\varphi_\tau(Y - X'\beta(\tau))X\right] + o_{p^*}(1), \quad \text{in } \ell^\infty(\mathcal{T}), \tag{3.A.2}$$

which follows from $\sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\| = o_{p^*}(1)$, and resulting convergence with respect to the pseudometric $\sup_{\tau \in \mathcal{T}} \rho[(\tau, \hat{\beta}(\tau)), (\tau, \beta(\tau))]^2 = o_p(1)$. The latter is immediate from $\sup_{\tau \in \mathcal{T}} \rho\left[(\tau, b(\tau)), (\tau, \beta(\tau))\right]^2 \leq C_3 \cdot \sup_{\tau \in \mathcal{T}} \|b(\tau) - \beta(\tau)\|^{\frac{\varepsilon}{(2+\varepsilon)}}$, where $C_3 = (\bar{f} \cdot (E\|X\|^2)^{1/2})^{\frac{\varepsilon}{2(2+\varepsilon)}} \cdot (E\|X\|^{2+\varepsilon})^{\frac{2}{2+\varepsilon}} < \infty$ and $\bar{f}$ is the a.s. upper bound on $f_Y(Y|X)$. (This follows by the Hölder's inequality and Taylor expansion.)

Further, the following expansion is valid uniformly in $\tau \in \mathcal{T}$

$$E\left[\varphi_\tau(Y - X'\beta)X\right]\Big|_{\beta = \hat{\beta}(\tau)} = \left[J(\tau) + o_p(1)\right]\left(\hat{\beta}(\tau) - \beta(\tau)\right). \tag{3.A.3}$$

Indeed, by Taylor expansion $E\left[\varphi_\tau(Y - X'\beta)X\right]|_{\beta = \hat{\beta}(\tau)} = E\left[f_Y(X'b(\tau)|X)XX'\right]|_{b(\tau) = \beta^*(\tau)}(\hat{\beta}(\tau) - \beta(\tau))$, where $\beta^*(\tau)$ is on the line connecting $\hat{\beta}(\tau)$ and $\beta(\tau)$ for each τ, and is different for each row of the Jacobian matrix. Then, (3.A.3) follows by the uniform consistency of $\hat{\beta}(\tau)$, and the assumed uniform continuity and boundedness of the mapping $y \mapsto f_Y(y|x)$, uniformly in x over the support of X.

Fourth, since the left hand side (lhs) of (3.A.1) = lhs of $n^{1/2}$(3.A.3)+ lhs of (3.A.2), we have that

$$o_p(1) = [J(\cdot) + o_p(1)](\hat{\beta}(\cdot) - \beta(\cdot)) + \mathbb{G}_n[\varphi_\cdot(Y - X'\beta(\cdot))X]. \tag{3.A.4}$$

Therefore, using that mineig $[J(\tau)] \geq \lambda > 0$ uniformly in $\tau \in \mathcal{T}$,

$$\sup_{\tau \in \mathcal{T}} \left\|\mathbb{G}_n[\varphi_\tau(Y - X'\beta(\tau))X] + o_p(1)\right\| \geq (\sqrt{\lambda} + o_p(1)) \cdot \sup_{\tau \in \mathcal{T}} \sqrt{n}\|\hat{\beta}(\tau) - \beta(\tau)\|. \tag{3.A.5}$$

Fifth, the mapping $\tau \mapsto \beta(\tau)$ is continuous by the implicit function theorem and stated assumptions. In fact, since $\beta(\tau)$ solves $E\left[(\tau - 1\{Y \leq X'\beta\})X\right] = 0$, $d\beta(\tau)/d\tau = J(\tau)^{-1}E\left[X\right]$. Hence $\tau \mapsto \mathbb{G}_n\left[\varphi_\tau\left(Y - X'\beta(\tau)\right)X\right]$ is stochastically equicontinuous over $\mathcal{T}$ for the pseudo-metric given by $\rho(\tau', \tau'') := \rho((\tau', \beta(\tau')), (\tau'', \beta(\tau'')))$. Stochastic equicontinuity of $\tau \mapsto$

$\mathbb{G}_n\left[\varphi_\tau(Y - X'\beta(\tau))X\right]$ and a multivariate CLT imply that

$$\mathbb{G}_n\left[\varphi_\cdot(Y - X'\beta(\cdot))X\right] \Rightarrow z(\cdot) \text{ in } \ell^\infty(\mathcal{T}), \tag{3.A.6}$$

where $z(\cdot)$ is a Gaussian process with covariance function $\Sigma(\cdot,\cdot)$ specified in the statement of Theorem 3. Therefore, the lhs of (3.A.5) is $O_p(n^{-1/2})$, implying $\sup_{\tau\in\mathcal{T}} \|\sqrt{n}(\hat\beta(\tau) - \beta(\tau))\| = O_{p^*}(1)$.

Finally, the latter fact and (3.A.4)-(3.A.6) imply that in $\ell^\infty(\mathcal{T})$

$$J(\cdot)\sqrt{n}(\hat\beta(\cdot) - \beta(\cdot)) = -\mathbb{G}_n\left[\varphi_\cdot(Y - X'\beta(\cdot))\right] + o_{p^*}(1) \Rightarrow z(\cdot). \tag{3.A.7}$$

Q.E.D.

3.A.3 Proof of Corollaries

Proof of Corollary 1. The result follows by the continuous mapping theorem in $\ell^\infty(\mathcal{T})$. Absolute continuity of $\mathcal{K}$ follows from Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998). *Q.E.D.*

Proof of Corollary 2. The result follows by absolute continuity of $\mathcal{K}$. The consistency of subsampling estimator of $\hat\kappa(\alpha)$ follows from Theorem 2.2.1 and Corollary 2.4.1 in Politis, Romano and Wolf (1999), for the case when $V(\tau)$ are known. When $V(\tau)$ is estimated consistently uniformly in $\tau \in \mathcal{T}$, the result follows by an argument similar to the proof of Theorem 2.5.1 in Politis et. al. (1999). *Q.E.D.*

3.A.4 Uniform Consistency of $\widehat\Sigma(\cdot,\cdot)$ and $\widehat J(\cdot)$.

Here it is shown that under the conditions of Theorem 3 and the additional assumption that $E\|X\|^4 < \infty$, the estimates described in the main text are consistent uniformly in $(\tau, \tau') \in \mathcal{T} \times \mathcal{T}'$.[21]

First, recall that $\hat J(\tau) = [1/(2h_n)] \cdot \mathbb{E}_n[1\{|Y_i - X_i'\hat\beta(\tau)| \le h_n\} \cdot X_iX_i']$. We will show that

$$\hat J(\tau) - J(\tau) = o_{p^*}(1) \text{ uniformly in } \tau \in \mathcal{T}. \tag{3.A.8}$$

Note that $2h_n\hat J(\tau) = \mathbb{E}_n[f_i(\hat\beta(\tau), h_n)]$, where $f_i(\beta, h) = 1\{|Y_i - X_i'\beta| \le h\} \cdot X_iX_i'$. For any compact set B and positive constant H, the functional class $\{f_i(\beta,h), \beta \in B, h \in (0, H]\}$ is a Donsker class with a square-integrable envelope by Theorem 2.10.6 in Van der Vaart and Wellner (1996), since this is a product of a VC subgraph class $\{1\{|Y_i - X_i'\beta| \le h\},\ \beta \in B,\ h \in (0,H]\}$ and a square integrable random matrix X_iX_i' (recall $E\|X_i\|^4 < \infty$ by assumption). Therefore, $(\beta, h) \mapsto \mathbb{G}_n\left[f_i(\beta,h)\right]$ converges to a Gaussian process in $\ell^\infty(B \times (0,H])$, which implies that $\sup_{\beta\in B, 0<h\le H} \|\mathbb{E}_n\left[f_i(\beta,h)\right] - E\left[f_i(\beta,h)\right]\| = O_{p^*}(n^{-1/2})$. Letting B be any compact set that covers $\cup_{\tau\in\mathcal{T}}\beta(\tau)$, this implies $\sup_{\tau\in\mathcal{T}} \|\mathbb{E}_n[f_i(\hat\beta(\tau),h_n)] - E[f_i(\beta,h_n)]|_{\beta=\hat\beta(\tau)}\| = O_{p^*}(n^{-1/2})$. Hence (3.A.8) follows by using $2h_n\hat J(\tau) = \mathbb{E}_n[f_i(\hat\beta(\tau),h_n)]$, $1/(2h_n) \cdot E[f_i(\beta,h_n)]|_{\beta=\widehat\beta(\tau)} = J(\tau) + o_p(1)$, and the assumption $h_n^2 n \to \infty$.

[21] Note that the result for $\widehat J(\tau)$ is not covered by Powell (1986) because his proof applies only pointwise in τ, whereas we require a uniform result.

Second, we can write $\hat{\Sigma}(\tau,\tau') = \mathbb{E}_n[g_i(\hat{\beta}(\tau),\hat{\beta}(\tau'),\tau,\tau')X_iX_i']$, where $g_i(\beta',\beta'',\tau',\tau'') = (\tau - 1\{Y_i \leq X_i'\beta'\})(\tau' - 1\{Y_i \leq X_i'\beta''\}) \cdot X_iX_i'$. We will show that

$$\hat{\Sigma}(\tau,\tau') - \Sigma(\tau,\tau') = o_{p^*}(1) \text{ uniformly in } (\tau,\tau') \in \mathcal{T} \times \mathcal{T}. \tag{3.A.9}$$

It is easy to verify that $\{g_i(\beta',\beta'',\tau',\tau''), (\beta',\beta'',\tau',\tau'') \in B \times B \times \mathcal{T} \times \mathcal{T}\}$ is Donsker and hence a Glivenko-Cantelli class, for any compact set B, e.g., using Theorem 2.10.6 in Van der Vaart and Wellner (1996). This implies that $\mathbb{E}_n\left[g_i(\beta',\beta'',\tau',\tau'')X_iX_i'\right] - E\left[g_i(\beta',\beta'',\tau',\tau'')X_iX_i'\right] = o_{p^*}(1)$ uniformly in $(\beta',\beta'',\tau',\tau'') \in (B \times B \times \mathcal{T} \times \mathcal{T})$. The latter and continuity of $E\left[g_i(\beta',\beta'',\tau',\tau'')X_iX_i'\right]$ in $(\beta',\beta'',\tau',\tau'')$ imply (3.A.9). *Q.E.D.*

Bibliography

[1] Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, pp. 821856.

[2] Abadie, A. (1997): "Changes in Spanish Labor Income Structure during the 1980's: A Quantile Regression Approach," *Investigaciones Economicas* XXI(2), pp. 253-272.

[3] Angrist, J., and A. Krueger (1999): "Empirical Strategies in Labor Economics," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Volume 3. Amsterdam. Elsevier Science.

[4] Autor, D., L.F. Katz, and M.S. Kearney (2004): "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists," MIT Department of Economics, mimeo, August 2004.

[5] Bassett, G.W., M.-Y. S. Tam, and K. Knight (2002): "Quantile Models and Estimators for Data Analysis," *Metrika* 55, pp. 17-26.

[6] Buchinsky, M. (1994): "Changes in the US Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica* 62, pp. 405-458.

[7] Chamberlain, G. (1984): "Panel Data," in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Volume 2. North-Holland. Amsterdam.

[8] Chamberlain, G. (1994): "Quantile Regression, Censoring, and the Structure of Wages," in C. A. Sims (ed.), *Advances in Econometrics, Sixth World Congress,* Volume 1. Cambridge University Press. Cambridge.

[9] Chernozhukov, V. (2002): "Inference on the Quantile Regression Process, An Alternative," Unpublished Working Paper 02-12 (February), MIT (www.ssrn.com).

[10] Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *Annals of Statistics* 2, pp. 267-277.

[11] Davydov, Yu. A., M. A. Lifshits, and N. V. Smorodina (1998): *Local Properties of Distributions of Stochastic Functionals.* Translated from the 1995 Russian original by V. E. Nazaĭkinskiĭ and M. A. Shishkova. Translations of Mathematical Monographs, 173. American Mathematical Society, Providence, RI.

[12] Ferguson, T. S. (1958): "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *The Annals of Mathematical Statistics* 29(4), pp. 1046-1062.

[13] Giacomini, R., and I. Komunjer (2003): "Evaluation and Combination of Conditional Quantile Forecasts," Working Paper 571, Boston College and California Institute of Technology, 06/2003.

[14] Goldberger, A. S. (1991): *A Course in Econometrics.* Harvard University Press. Cambridge, MA.

[15] Gosling, A., S. Machin, and C. Meghir (2000): "The Changing Distribution of Male Wages in the U.K.," *Review of Economic Studies* 67, pp. 635-666.

[16] Gutenbrunner, C., and J. Jurečková (1992): "Regression Quantile and Regression Rank Score Process in the Linear Model and Derived Statistics," *Annals of Statistics* 20, pp. 305-330.

[17] Hahn, J. (1997): "Bayesian Bootstrap of the Quantile Regression Estimator: A Large Sample Study," *International Economic Review* 38(4), pp. 795-808.

[18] Hansen, L.-P. , J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics* 14(3), pp. 262-280.

[19] Juhn, C., K. Murphy, and B. Pierce (1993): "Wage Inequality and the Rise in Return to Skill," *Journal of Political Economy* 101, pp. 410-422.

[20] Katz, L., and D. Autor (1999): "Changes in the Wage Structure and Earnings Inequality," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Volume 3A. Elsevier Science. Amsterdam.

[21] Katz, L., and K. Murphy (1992): "Changes in the Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics* 107, pp. 35-78.

[22] Kim T.H., and H. White (2002): "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regressions," in *Advances in Econometrics*, forthcoming.

[23] Koenker, R. (1994): "Confidence Intervals for Regression Quantiles," in M.P. and M. Hušková (eds.), *Asymptotic Statistics: Proceeding of the 5th Prague Symposium on Asymptotic Statistics.* Physica-Verlag.

[24] Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica* 46, pp. 33-50.

[25] Koenker, R., and J. A. Machado (1999): "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association* 94(448), pp. 1296-1310.

[26] Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica* 70, no. 4, pp. 1583–1612.

[27] Lemieux, T. (2003): "Residual Wage Inequality: A Re-examination," Mimeo. University of British Columbia.

[28] Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, no. 2, 186–199.

[29] Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling.* Springer-Verlag. New York.

[30] Portnoy, S. (1991): "Asymptotic Behavior of Regression Quantiles in Nonstationary, Dependent Cases" *Journal of Multivariate Analysis* 38, no. 1, pp. 100-113.

[31] Powell, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics* 32, no. 1, pp. 143-155.

[32] Powell, J. L. (1994): "Estimation of Semiparametric Models," in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, Volume IV. Amsterdam. Elsevier Science.

[33] Rothenberg, T. J. (1973): "Comparing Efficient Estimation with A Priori Information," Cowles Foundation Monograph 23. Yale University.

[34] Ruggles, S., and M. Sobek *et al.* (2003): *Integrated Public Use Microdata Series: Version 3.0.* Minneapolis: Historical Census Project. University of Minesota.

[35] Van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer Series in Statistics. Springer-Verlag. New York.

[36] White, H. (1980): "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review* 21(1), pp. 149-170.

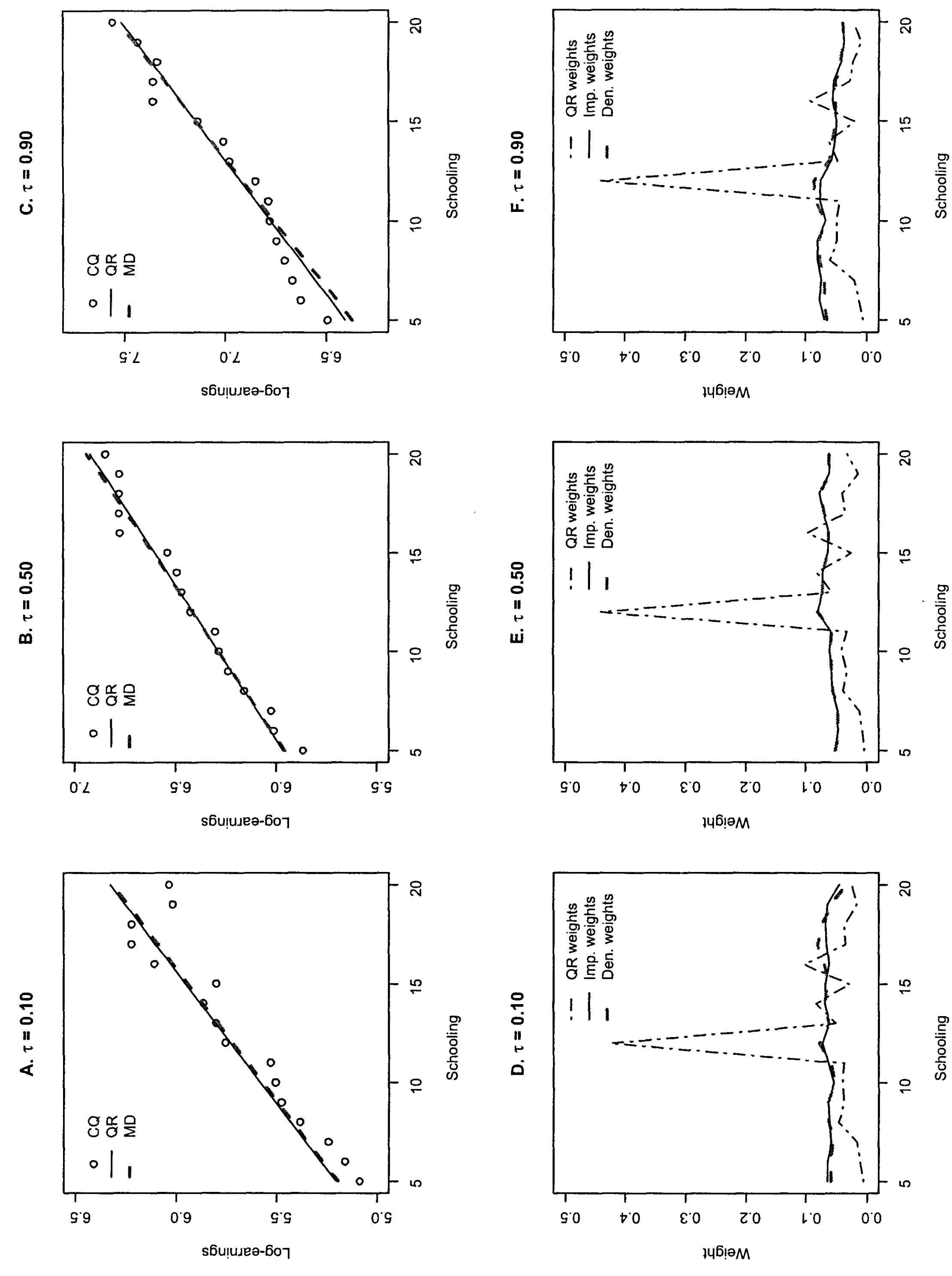

Figure 3-1: CQF and Weighting schemes in 1980 Census (US-born white and black men aged 40-49). Panels A - C plot the Conditional Quantile Function, Linear Quantile Regression fit, and Chamberlain's Minimum Distance fit for log-earnings given years of schooling. Panels D - F plot QR weighting function (histogram × importance weights), importance weights and density weights.

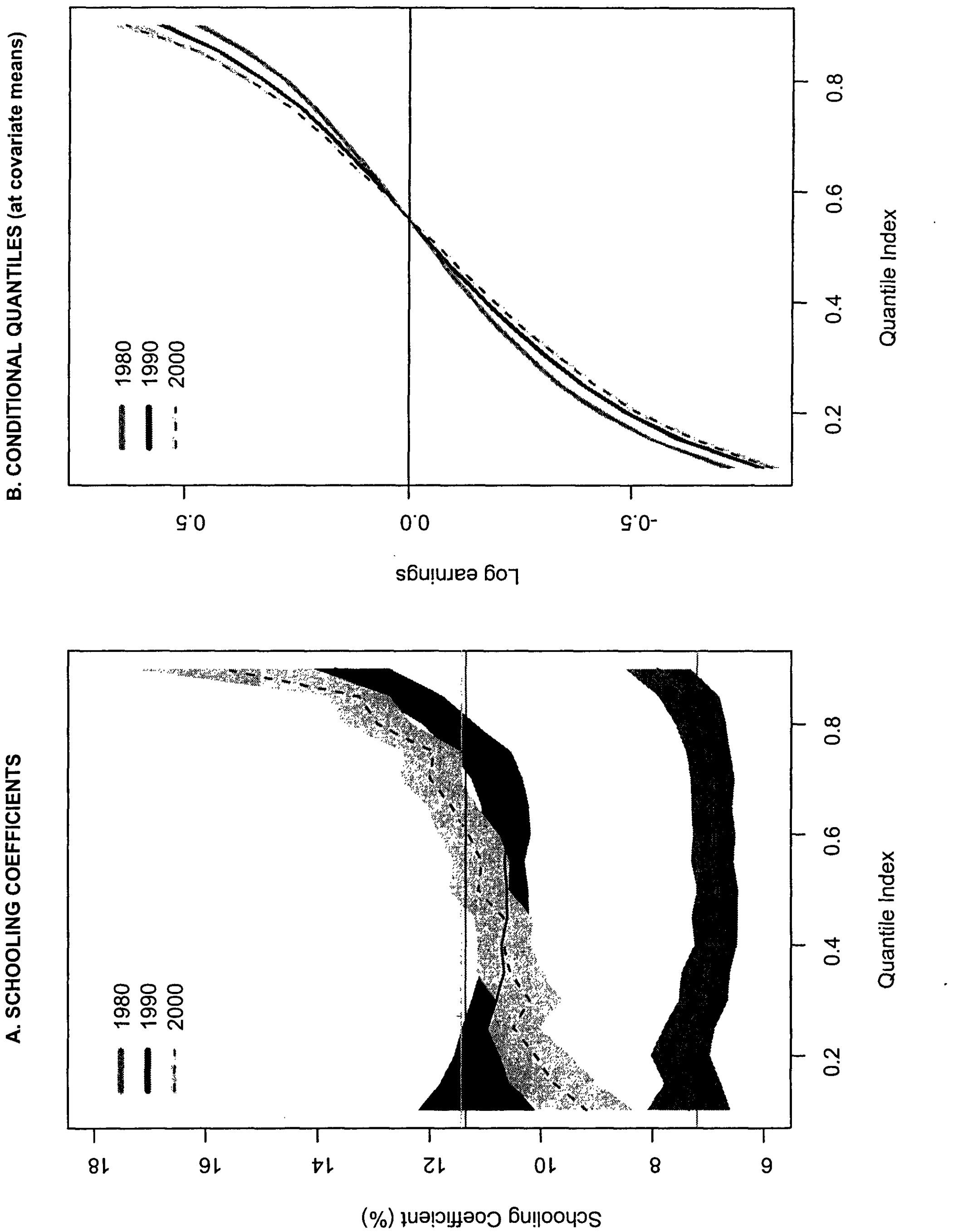

Figure 3-2: Schooling coefficients and conditional quantiles of log-earnings in 1980, 1990, and 2000 censuses (US-born white and black mean aged 40-49). Panel A plots the quantile process for the coefficient of schooling in the QR of log-earnings on years of schooling, race, and a quadratic function of experience; and robust simultaneous 95 % confidence bands. Panel B plots simultaneous 95 % confidence bands for the QR approximation to the conditional quantile function given schooling, race, and a quadratic function of experience. Horizontal lines correspond to OLS estimates of the schooling coefficients in Panel A. In Panel B, covariates are evaluated at sample mean values for each year, and distributions are centered at median earnings for each year (i.e., for each τ and year, $E[X]'(\widehat{\beta}(\tau) - \widehat{\beta}(.5))$ is plotted).

Table 1: Comparison of CQF and QR-based Interquantile Spreads

		Interquantile Spread					
		90-10		90-50		50-10	
Census	Obs.	CQ	QR	CQ	QR	CQ	QR
		A. Overall					
1980	65,023	1.20	1.19	0.51	0.52	0.68	0.67
1990	86,785	1.35	1.35	0.60	0.61	0.75	0.74
2000	97,397	1.43	1.45	0.67	0.70	0.76	0.75
		B. High School Graduates					
1980	25,020	1.09	1.17	0.44	0.50	0.65	0.67
1990	22,837	1.26	1.31	0.52	0.55	0.74	0.76
2000	25,963	1.29	1.32	0.59	0.60	0.70	0.72
		C. College Graduates					
1980	7,158	1.26	1.19	0.61	0.54	0.65	0.64
1990	15,517	1.44	1.38	0.70	0.66	0.74	0.72
2000	19,388	1.55	1.57	0.75	0.80	0.80	0.78

Notes: US-born white and black men aged 40-49. Average measures calculated using the distribution of the covariates in each year. The covariates are schooling, race and a quadratic function of experience. Sampling weights used for 2000 Census.