ECONOMETRICS ON INTERACTIONS-BASED MODELS: METHODS AND APPLICATIONS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Xiaodong Liu, M.A.

* * * * *

The Ohio State University

2007

Dissertation Committee:

Professor Lung-fei Lee, Co-Adviser

Professor John H. Kagel, Co-Adviser

Professor Stephen R. Cosslett

Professor Lixin Ye

Approved by

____

Co-Adviser

____

Co-Adviser

Graduate Program in Economics

ABSTRACT

My dissertation research emphasizes estimation methods in evaluating the extent of social, strategic and spatial interactions among economic agents. Topical applications include measuring peer group effects in experimental signaling games, structural estimation of the latent value distribution through bidder's strategic bidding behavior in empirical auctions, and GMM estimation of spatial autoregressive models.

My first essay, based on my joint research with Lung-fei Lee and John Kagel, generalizes Heckman's (1981) dynamic discrete-choice panel data models by introducing time-lagged social interactions so that the models can accommodate relationships of decision making across cross-sectional units. We derive the likelihood function for the generalized model and propose simulation based methods to implement the maximum likelihood estimation. Such dynamic social interaction models may have broad applicability, especially in interpreting experimental economics data. In this essay, we use this model to investigate learning from peers in experiments based on Milgrom and Roberts' (1982) entry limit pricing game. We find that subjects' decisions are significantly influenced by the past decisions of their peers in the experiment. Our findings are consistent with the view that the imitation of peers' strategies is an important component of one's learning how to play strategically. Similar peer group effects are likely to be present in experimental designs where subjects receive feedback on their peer's performance.

My second essay explores the robustness of Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimation procedure in auctions with a large number of risk-averse bidders. Guerre et al. show that the underlying distribution of bidders' values (or costs) is nonparametrically identified from the observation of submitted bids when the auction among risk-neutral bidders is conducted as a first-price, sealed-bid auction under the independent private value paradigm. They propose a two-step nonparametric estimation procedure for the latent value distribution based on the equilibrium bidding behavior of risk-neutral bidders. Their estimator is optimal in terms of uniform convergence rate to the true distribution. In this essay, with an asymptotic approximation of the intractable equilibrium bidding function of risk-averse bidders, I demonstrate that Guerre et al.'s two-step nonparametric estimator is still uniformly consistent even if bidders are risk-averse as long as the number of players in an auction is sufficiently large and derive the uniform convergence rate of the estimator. Furthermore, I show in Monte Carlo experiments that the two-step nonparametric estimator performs reasonably well with a moderate number of risk-averse bidders like six.

In my third essay, which is based on my joint research with Lung-fei Lee and Christopher Bollinger, we consider the GMM estimation of the regression model with spatial autoregressive disturbances and the mixed-regressive spatial autoregressive model. We derive the best GMM estimator within the class of GMM estimators that are based on linear and quadratic moment conditions. Our best GMM estimator has the merit of computational simplicity and asymptotic efficiency. We show that it is asymptotically as efficient as the conventional maximum likelihood estimator under normality and asymptotically more efficient than the quasi-maximum likelihood estimator when the normality assumption does not hold. We show in Monte Carlo studies that, with moderate sample sizes, the proposed best GMM estimator has its biggest advantage when the disturbances are asymmetrically distributed. In the event that the diagonal elements from the squared spatial weights matrix have sufficient variance, then incorporating the kurtosis of the disturbances in the moment conditions of the GMM estimator will also be valuable.

Dedicated to my parents and my wife

ACKNOWLEDGMENTS

I am deeply grateful to my advisors, Professor Lung-fei Lee and Professor John Kagel, for their guidance, inspiration and encouragement throughout my dissertation research. Their encyclopedic knowledge in economics, rigorous attitude towards research and incessant patience in mentoring me have greatly contributed to the completion of this thesis.

I thank Professor Stephen Cosslett, Professor Robert de Jong and Professor Lixin Ye for simulating discussions and valuable suggestions that led to great improvements of earlier versions of this work, Professor John Ham for sharing his expertise in empirical studies with me, and Professor Massimo Morelli, Professor James Peck, Professor Bruce Weinberg and Professor Huanxing Yang for helpful advice at various stages of my study at Ohio State. Sincerely thanks also go to Professor Christopher Bollinger, whose deep insights have inspired the third essay in this thesis. I also would like to extend my special thanks to Professor Belton Fleisher and Professor Hajime Miyazaki. Their warmhearted help makes the challenging years of PhD study so much easier. I have been truly blessed with many amazing friends over the years at Ohio State. I thank them all for their friendship.

Finally, I would like to thank my parents, Ruihua and Peng, and my beloved wife, Lexian, for their unconditional love, support and belief in me. They make all my effort worthwhile.

VITA

December 10, 1978 . . . . Born — Jinan, China

2001 . . . . B.A. Economics, Fudan University

2002 . . . . M.A. Economics, Ohio State University

FIELDS OF STUDY

Major Field: Economics

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

DYNAMIC DISCRETE CHOICE MODELS WITH LAGGED SOCIAL INTERACTIONS: WITH AN APPLICATION TO A SIGNALING GAME EXPERIMENT

1.1 Introduction

In his seminal work, Heckman (1981) has introduced a rich group of discrete choice stochastic processes that allow each cross-sectional unit's decisions to have complex dynamic economic interrelationships over time. In this chapter, we generalize the dynamic discrete choice panel data models by introducing time-lagged social interactions, so that the models can accommodate interrelationships of decisions, such as learning from peers, across cross-sectional units. This enriches the class of dynamics in Heckman (1981). As interactions across cross-sectional units carry out with a time lag, the models are well-defined without running into identification or multiple equilibria problems, which occur in some social interaction models (Manski, 1993).

Likelihood functions of dynamic discrete choice models involve multiple integrals, if explanatory variables include lagged latent dependent variables or disturbances allow for serial correlation in addition to that captured by random components. For panel data models, the dimension of integration increases with the number of periods, which makes numerical implementation impractical. To overcome the computational difficulty, simulation estimation methods have been developed. The simulator due to Geweke (1991), Borsch-Supan and Hajivassiliou (1993) and Keane (1994) is known to be practical and accurate to implement the method of simulated maximum likelihood (SML), when the time periods are not too long.

In this chapter, we show that the implementation of the Geweke-Hajivassiliou-Keane (GHK) simulator remains tractable for models with social interactions. We investigate the finite sample properties of simulated estimates for model parameters and the effects of misspecification of dynamic structures and disturbances on estimates in the Monte Carlo experiments. As the likelihood function is nonlinear, the SML estimator (SMLE) might have an asymptotic bias if the number of random draws to construct the likelihood simulator does not increase fast enough relative to the sample size. Hence special attention will be given to dominated finite sample bias (relative to standard error) of coefficient estimates due to simulation. We report some Monte Carlo results of a bias-correction procedure proposed by Lee (1995) for the estimation of dynamic models with lagged interactions.

These dynamic social interaction models may have broad applicability, in particular, for experimental economics data. Numerous experiments have been conducted with a discrete choice space, with observations obtained in consecutive rounds. One of the main concerns in experimental games is the effect of a player's learning from other players. As such a dynamic discrete choice model with lagged social interactions may fit well as a possible econometric model for the analysis of experimental data. Specifically, in this chapter, we apply our generalized dynamic models with social interactions to investigate the presence and magnitude of peer group effects in experiments based on Milgrom and Roberts' (1982) entry limit pricing game. Similar peer group effects are likely to be present in a variety of experimental designs where subjects receive feedback on their peer's performance. Empirical findings reported here may have broader economic implications. From the statistical inference point of view, the usual limited number of experimental subjects, rounds, and sessions due to feasibility or expense concerns might prevent one from determining whether peer group effects are indeed negligible or overwhelmed by estimation errors caused by insufficient sample size. So our estimation and Monte Carlo experimental results may shed some light on the sample size requirement and sample structures favorable to successfully identifying potential peer group effects in discrete choice games.

The organization of this chapter is as follows. In Section 2, we introduce a general dynamic discrete choice panel data model with lagged social interactions, derive the likelihood function and illustrate the formulation of simulators and simulated likelihood function for this model. We report Monte Carlo results for the SMLE of the Markov and Polya models with lagged social interactions in Section 3. In Section 4, we formulate empirical dynamic models to investigate the adjustment process of subjects' decisions in laboratory experiments based on an entry limit pricing game. Section 5 briefly concludes.

1.2 General Dynamic Discrete Choice Models with Social Interactions and SML Estimation

Consider a general dynamic discrete choice panel data model with lagged social interactions

$$y_{it}^* = h_{it}(y_{i,t-1}^*, \cdots, y_{i,-\infty}^*, Y_{n,t-1}, \cdots, Y_{n,-\infty}, X_{nt}, \cdots, X_{n,-\infty}, \xi_i) + v_{it}, \tag{1.1}$$

for $i = 1, \cdots, n$, where Y_{nt} is the n-dimensional vector of dichotomous indicators of the latent variables $y_{1t}^*, \cdots, y_{nt}^*$, X_{nt} is the $n \times k$-dimensional matrix of strictly exogenous variables and ξ_i is a random individual component. Suppose that the error components ξ_i are i.i.d. $N(0, \sigma^2)$ for all i and the disturbances v_{it} are i.i.d. $N(0, 1)$ for all i and t. This process is assumed to start at $t = 1$, and the initial conditions on y^*_{it}, Y_{nt} and X_{nt} for $t \leq 0$ are fixed outside the model and are assumed to be zero. The original specification of the dynamic model in Heckman (1981) does not incorporate lagged social interactions in that $y_{i,s-1}$ and x_{is} appear but not $Y_{n,s-1}$ and X_{ns} $(s \leq t)$. Depending on the specification of the function $h_{it}(\cdot)$ in terms of lagged observed or latent dependent variables, the Heckman discrete dynamic model is known to be sufficiently flexible to accommodate a wide variety of dynamic structures such as Markov models, Polya models, renewal processes, latent Markov models, with rich specifications on disturbances. It allows for unobserved heterogeneity across the n cross-sectional units and serial correlation for the remaining disturbances. The model with social interactions in (1.1) is generalized to incorporate additional dynamic effects due to peers' influence. We derive the likelihood function for (1.1) and construct the unbiased GHK simulator to implement the SML estimation for the model.

In addition to Y_{nt}, let $Y^*_{nt} = (y^*_{1t}, \cdots, y^*_{nt})'$ be the n-dimensional vector of the latent dependent variables for all the n cross-sectional units. Let X_t denote the sequence of $X_{nt}, X_{n,t-1}, \cdots$. Conditional on exogenous variables X_T and $\xi = (\xi_1, \cdots, \xi_n)'$, the joint density function of (Y^*_{nt}, Y_{nt}), $t = 1, \cdots, T$, is the product of conditional density of (Y^*_{ns}, Y_{ns}), $s = 1, \cdots, T$, over their past histories, i.e.,

$$
\begin{aligned}
& f(Y^*_{nT}, Y_{nT}, \cdots, Y^*_{n1}, Y_{n1} | X_T, \xi) \\
= & \left[\prod_{t=2}^{T} f(Y^*_{nt}, Y_{nt} | (Y^*_{ns}, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi) \right] f(Y^*_{n1}, Y_{n1} | X_1, \xi).
\end{aligned}
$$

Because v_{it} are mutually independent for $i = 1, \cdots, n$, each of the conditional densities of (Y^*_{nt}, Y_{nt}) can be further decomposed as

$$
\begin{aligned}
& f(Y^*_{nt}, Y_{nt}|(Y^*_{ns}, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi) \\
= & \prod_{i=1}^{n} f(y^*_{it}, y_{it}|(Y^*_{ns}, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi_i) \\
= & \prod_{i=1}^{n} I_{y_{it}}(y^*_{it}) g(y^*_{it}|(Y^*_{ns}, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi_i)
\end{aligned}
$$

for $t = 2, \cdots T$, and

$$
f(Y^*_{n1}, Y_{n1}|X_1, \xi) = \prod_{i=1}^{n} I_{y_{i1}}(y^*_{i1}) g(y^*_{i1}|X_1, \xi_i),
$$

where $I_{y_{it}}(y^*_{it})$ is the dichotomous indicator with $I_{y_{it}}(y^*_{it}) = 1$ if the value y^*_{it} determines the observed value y_{it}; $I_{y_{it}}(y^*_{it}) = 0$, otherwise, and g is the conditional density of y^*_{it}. Therefore, the joint probability of $Y_{nT}, \cdots, Y_{n1}$ conditional on X_T and ξ is

$$
\begin{aligned}
& P(Y_{nT}, \cdots, Y_{n1}|X_T, \xi) \\
= & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(Y^*_{nT}, Y_{nT}, \cdots, Y^*_{n1}, Y_{n1}|X_T, \xi) dvec'(Y^*_{nT}) \cdots dvec'(Y^*_{n1}) \\
= & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{t=2}^{T} \prod_{i=1}^{n} I_{y_{it}}(y^*_{it}) g(y^*_{it}|(Y^*_{ns}, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi_i) \right] \\
& \times \prod_{i=1}^{n} I_{y_{i1}}(y^*_{i1}) g(y^*_{i1}|X_1, \xi_i) dvec'(Y^*_{nT}) \cdots dvec'(Y^*_{n1}). \qquad (1.2)
\end{aligned}
$$

For (1.1), $g(y^*_{it}|(Y^*_{ns}, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi_i) = g(y^*_{it}|(y^*_{is}, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi_i)$ as interactions among different units are going through the observed Y_{ns} and X_{ns} but not Y^*_{ns} with $s < t$. Hence we have

$$
\begin{aligned}
& P(Y_{nT}, \cdots, Y_{n1}|X_T, \xi) \\
= & \prod_{i=1}^{n} \Big\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{t=2}^{T} I_{y_{it}}(y^*_{it}) g(y^*_{it}|(y^*_{is}, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi_i) \right] \\
& \times I_{y_{i1}}(y^*_{i1}) g(y^*_{i1}|X_1, \xi_i) dy^*_{iT} \cdots dy^*_{i1} \Big\}.
\end{aligned}
$$

Under the distributional assumption that v_{it} is $N(0, 1)$,

$$g(y_{it}^*|(y_{is}^*, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi_i) = \phi(y_{it}^* - h_{it}),$$

where $h_{it} = h_{it}(y_{i,t-1}^*, \cdots, y_{i,-\infty}^*, Y_{n,t-1}, \cdots, Y_{n,-\infty}, X_t, \xi_i)$ for simplicity and ϕ is the standard normal density function. Define the integral limits L_{it} and U_{it}:

$$L_{it} = \begin{cases} -h_{it} & \text{if } y_{it} = 1, \\ -\infty & \text{if } y_{it} = 0, \end{cases} \text{ and } U_{it} = \begin{cases} \infty & \text{if } y_{it} = 1, \\ -h_{it} & \text{if } y_{it} = 0. \end{cases}$$

By transformations of variables, it follows that

$$\begin{aligned} & P(Y_{nT}, \cdots, Y_{n1}|X_T, \xi) \\ = & \prod_{i=1}^n \left\{ \int_{L_{i1}}^{U_{i1}} \cdots \int_{L_{i,T-1}}^{U_{i,T-1}} \left(\int_{L_{iT}}^{U_{iT}} \phi(v_{iT}) dv_{iT} \right) \phi(v_{i,T-1}) dv_{i,T-1} \cdots \phi(v_{i1}) dv_{i1} \right\} \\ = & \prod_{i=1}^n \Big\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\Phi(U_{iT}) - \Phi(L_{iT})) \\ & \times \prod_{s=1}^{T-1} \big(\Phi(U_{i,T-s}) - \Phi(L_{i,T-s})\big) \phi_{[L_{i,T-s},\, U_{i,T-s}]}(v_{i,T-s}) dv_{i,T-s} \Big\} \\ = & \prod_{i=1}^n \Big\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi((2y_{iT} - 1) h_{iT}) \\ & \times \prod_{s=1}^{T-1} \Phi((2y_{i,T-s} - 1) h_{i,T-s}) \phi_{[L_{i,T-s},\, U_{i,T-s}]}(v_{i,T-s}) dv_{i,T-s} \Big\}, \end{aligned}$$

where $\phi_{[L_t, U_t]}$ is a truncated standard normal density function with support $[L_t, U_t]$. The probability $Y_{nT}, \cdots, Y_{n1}$ conditional on exogenous variables X_T is

$$\begin{aligned} & P(Y_{nT}, \cdots, Y_{n1}|X_T) \\ = & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(Y_{nT}, \cdots, Y_{n1}|X_T, \xi_1, \cdots, \xi_n) \phi(\xi_1) \cdots \phi(\xi_n) d\xi_1 \cdots d\xi_n \\ = & \prod_{i=1}^n \Big\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi((2y_{iT} - 1) h_{iT}) \\ & \times [\prod_{s=1}^{T-1} \Phi((2y_{i,T-s} - 1) h_{i,T-s}) \phi_{[L_{i,T-s},\, U_{i,T-s}]}(v_{i,T-s}) dv_{i,T-s}] \phi(\xi_i) d\xi_i \Big\}. \end{aligned}$$

This likelihood suggests that the GHK simulator can be recursively applied to construct a simulated likelihood. Generate u_{it} $(i = 1, \cdots, n;\ t = 1, \cdots, T-1)$ independent uniform [0, 1] random variables. Generate ξ_i $(i = 1, \cdots, n)$ independent standard normal variables. With initial conditions given, the random variables v_{it} $(i = 1, \cdots, n;\ t = 1, \cdots, T-1)$ can be generated from the following steps. For each i, from $t = 1$ to $T - 1$:

(1) Compute

$$v_{it} = -(2y_{it} - 1)\Phi^{-1}\left[u_{it}\Phi\left((2y_{it} - 1)h_{it}\right)\right].$$

(2) Generate the latent dependent variable

$$y_{it}^* = h_{it} + v_{it}.$$

With m independent simulation runs, the corresponding simulated log likelihood function is

$$\mathcal{L} = \sum_{i=1}^{n} \ln\left\{\frac{1}{m}\sum_{j=1}^{m}\prod_{t=1}^{T}\Phi((2y_{it} - 1)h_{it}^{(j)})\right\}, \tag{1.3}$$

where $h_{it}^{(j)} = h_{it}(y_{i,t-1}^{*(j)}, \cdots, y_{i0}^{*(j)}, Y_{n,t-1}, \cdots, Y_{n0}, X_t, \xi_i^{(j)})$, and the superscript (j) denotes an independent simulation run. Thus, the simulation of the likelihood for the model in (1.1), is similar to one of the conventional dynamic panel models in Lee (1997).

Asymptotic properties of the SMLE for cross-sectional or short time series panel data have been studied in Hajivassiliou and McFadden (1990), Lee (1992; 1995) and Gourieroux and Monfort (1993), among others. The SMLE can be asymptotically efficient when m increases at a rate faster than $n^{1/2}$. However, when m increases at a rate of $n^{1/2}$, as shown in Lee (1995), an asymptotic bias exists in the limiting distribution. The asymptotic bias will dominate the variance when m increases at a rate slower than $n^{1/2}$. Lee (1995) has suggested a simple bias-correction procedure to remove the leading bias term due to simulation. The asymptotic efficiency of the bias-adjusted estimator requires only that m goes to infinity at a rate faster than $n^{1/4}$.

For experimental economics, subjects are usually divided into several independent groups (experimental sessions), and games are played in several rounds within each group. Suppose that there are G groups. Within each group, there are n players and the number of rounds is T. With data from such a design, the simulated likelihood function shall be

$$\mathcal{L} = \sum_{g=1}^{G} \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi((2y_{g,it} - 1)h_{g,it}^{(j)}) \right\}, \tag{1.4}$$

where the subscript (g, it) indicates the observation is from individual i of group g at round t.

The model in (1.1) can be further generalized to allow social interactions in both observed and latent lagged dependent variables,

$$y_{it}^* = \bar{h}_{it}(Y_{n,t-1}^*, \cdots, Y_{n,-\infty}^*, Y_{n,t-1}, \cdots, Y_{n,-\infty}, X_t, \xi_i) + v_{it}, \tag{1.5}$$

where Y_{nt}^* is the n-dimensional vector of latent dependent variables and X_t is the sequence of strictly exogenous variables $X_{nt}, X_{n,t-1}, \cdots$. As in (1.1), ξ_i are i.i.d. $N(0, \sigma^2)$ for all i and v_{it} are i.i.d. $N(0, 1)$ for all i and t. The initial values for Y_{nt}^*, Y_{nt} and X_t for $t \leq 0$ are assumed to be zero. From (1.2) , the joint probability for $Y_{nT}, \cdots, Y_{n1}$ conditional on exogenous variables X_T and ξ is given by

$$\begin{aligned} & P(Y_{nT}, \cdots, Y_{n1} | X_T, \xi) \\ = & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{t=2}^{T} \prod_{i=1}^{n} I_{y_{it}}(y_{it}^*) g(y_{it}^* | (Y_{ns}^*, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi_i) \right] \\ & \times \prod_{i=1}^{n} I_{y_{i1}}(y_{i1}^*) g(y_{i1}^* | X_1, \xi_i) dvec'(Y_{nT}^*) \cdots dvec'(Y_{n1}^*). \end{aligned}$$

Under the distributional assumption of v_{it},

$$g(y_{it}^* | (Y_{ns}^*, Y_{ns}; s = 1, \cdots, t-1), X_t, \xi_i) = \phi(y_{it}^* - \bar{h}_{it}),$$

where $\bar{h}_{it} = \bar{h}_{it}(Y^*_{i,t-1}, \cdots, Y^*_{i,-\infty}, Y_{n,t-1}, \cdots, Y_{n,-\infty}, X_t, \xi_i)$ for simplicity and ϕ is the standard normal density function. Define the integral limits $\bar{L}_{it}$ and $\bar{U}_{it}$:

$$\bar{L}_{it} = \begin{cases} -\bar{h}_{it} & \text{if } y_{it} = 1, \\ -\infty & \text{if } y_{it} = 0, \end{cases} \text{ and } \bar{U}_{it} = \begin{cases} \infty & \text{if } y_{it} = 1, \\ -\bar{h}_{it} & \text{if } y_{it} = 0. \end{cases}$$

By transformations of variables, it follows that

$$\begin{aligned} & P(Y_{nT}, \cdots, Y_{n1}|X_T, \xi) \\ = & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} \Phi((2y_{iT} - 1)\bar{h}_{iT}) \right] \\ & \times \prod_{s=1}^{T-1} \left[\prod_{i=1}^{n} \Phi((2y_{i,T-s} - 1)\bar{h}_{i,T-s}) \phi_{[\bar{L}_{i,T-s},\ \bar{U}_{i,T-s}]}(v_{i,T-s}) dv_{i,T-s} \right]. \end{aligned}$$

And the probability $Y_{nT}, \cdots, Y_{n1}$ conditional on exogenous variables X_T is $P(Y_{nT}, \cdots, Y_{n1}|X_T) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(Y_{nT}, \cdots, Y_{n1}|X_T, \xi) \left[\prod_{i=1}^{n} \phi(\xi_i) d\xi_i \right]$.

In this case, with u_{it} and ξ_i generated as before, the random variables v_{it} ($i = 1, \cdots, n$; $t = 1, \cdots, T-1$) can be generated from the following steps, from $t = 1$ to $T - 1$:

(1) Compute for $i = 1, \cdots, n$

$$v_{it} = -(2y_{it} - 1)\Phi^{-1} \left[u_{it} \Phi \left((2y_{it} - 1)\bar{h}_{it} \right) \right].$$

(2) Generate the latent dependent variable

$$y^*_{it} = \bar{h}_{it} + v_{it}.$$

With m independent runs, the corresponding simulated log likelihood function shall be

$$\mathcal{L} = \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \prod_{i=1}^{n} \Phi((2y_{it} - 1)\bar{h}^{(j)}_{it}) \right\}, \tag{1.6}$$

where $\bar{h}^{(j)}_{it} = \bar{h}_{it}(Y^{*(j)}_{n,t-1}, \cdots, Y^{*(j)}_{n0}, Y_{n,t-1}, \cdots, Y_{n0}, X_t, \xi^{(j)}_i)$ for the jth simulation run.

There can be some numerical difficulties in implementing the SML estimation procedure if T and n are large, as the simulated log likelihood functions (1.3) and (1.6) involve the product consisting of many terms of small numbers that might be impossible to evaluate with computers without underflow errors. The problem is more severe in (1.6), where the simulated likelihood involves the product of cumulative probabilities of the entire history for all members in a group. Lee (2000) has suggested an algorithm that can overcome the numerical problem by interchanging the summation and product operators behind the logarithmic transformation. Here we illustrate this algorithm for (1.6). For simplicity, let $k = (t-1)n + i$, $i = 1, \cdots, n$ for each t with $t = 1, \cdots, T$, and rewrite (1.6) as

$$\mathcal{L} = \ln\left\{\frac{1}{m}\sum_{j=1}^{m}\prod_{k=1}^{T\times n}\Phi((2y_k - 1)\bar{h}_k^{(j)})\right\}.$$

Let $a_{kj} = \Phi((2y_k - 1)\bar{h}_k^{(j)})$ and let ω_{kj} be weights for $k \geq 1$, which can be computed recursively as

$$\omega_{kj} = a_{kj}\omega_{k-1,j} / \sum_{s=1}^{m} a_{ks}\omega_{k-1,s},$$

starting with $\omega_{0j} = 1/m$ for $j = 1, \cdots, m$. Then following Lee (2000), (1.6) can be rewritten as

$$\mathcal{L} = \ln\left\{\prod_{k=1}^{T\times n}\sum_{j=1}^{m} a_{kj}\omega_{kj}\right\} = \sum_{k=1}^{T\times n}\ln\left\{\sum_{j=1}^{m}\Phi((2y_k - 1)\bar{h}_k^{(j)})\omega_{kj}\right\}, \tag{1.7}$$

where the product of cumulative probabilities behind the logarithmic transformation is replaced by the weighted sum of cumulative probabilities.

Social interactions in latent lagged dependent variables are likely to appear if cross-sectional units are allowed to discuss their past preferences and choices. As we plan to apply the model to the estimation of data from lab experiments where, as is typically the case, subjects make independent decisions without communication, we focus on models that conform to (1.1) in the rest of this chapter.

1.3 Some Monte Carlo Results on SMLEs

1.3.1 A Markov Model with Lagged Social Interactions

Suppose we have observations of G independent groups, with n subjects in each group. The Markov dynamic choice model for the Monte Carlo study in this section is

$$y_{it}^{*} = \beta x_{i,t-1} + \lambda_1 y_{i,t-1} + \lambda_2 z_{i,t-1} + \sigma \xi_i + \varepsilon_{it}, \tag{1.8}$$

where $z_{i,t-1} = \sum_{j=1,j\neq i}^{n} y_{j,t-1}/(n-1)$, $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \upsilon_{it}$, and ξ_i and υ_{it} are i.i.d. $N(0,1)$. The group subscript g has been suppressed for simplicity. By replacing ε_{it} with $\rho(y_{i,t-1}^{*} - (\beta x_{i,t-2} + \lambda_1 y_{i,t-2} + \lambda_2 z_{i,t-2} + \sigma\xi_i)) + \upsilon_{it}$, i.e., by a quasi-difference transformation for (1.8), it is easy to see that (1.8) conforms to the general model (1.1).

The x_{it} are generated as $x_{it} = (1/\sqrt{2})r_{it} + \sqrt{6}s_i$ where r_{it} are independent truncated standard normal variables on $[-2, 2]$ and s_i is a uniform variable on $[-0.5, 0.5]$, so that the variance of x_{it} is about 1 and its correlation coefficient over time is about 0.5. This process of generating exogenous variables is to allow the exogenous variables to correlate over time. It is used for all the models in this chapter. The initial values of all variables for $t \leq 0$ are given as 0. Sample data are generated with $\beta = 1$, $\lambda_1 = 0.2$, $\lambda_2 = 0.4$, $\sigma^2 = 0.5$, and $\rho = 0.4$. The serial correlation of the total disturbance $\sigma\xi_i + \epsilon_{it}$ of two adjacent periods has a correlation coefficient about 0.6 and the fraction of variance due to the individual effect is about 0.3. The sample size is 200, with $G = 50$ and $n = 4$. We have experimented with small, moderate and large numbers of random draws, namely $m = 15$, $m = 50$ and $m = 100$, for the construction of the GHK simulator. The number of periods for the panel data varies from 8 to 30. For each case, the number of replications is 200. For each replication, in addition to random disturbances in the model, the set of exogenous variables is also redrawn. The maximization algorithm used is a conjugate gradient method.

For all cases and replications reported here, the algorithm converges without running into numerical problems. The initial estimate of σ is set to 1, and the initial estimates of the other parameters are set to 0. We have also tried some other starting values, with which the algorithm converges to similar solutions.

Table 1.7 reports the empirical means (Means), standard deviations (SDs) and root mean square errors (RMSEs) for both the bias unadjusted SMLE and the bias-adjusted SMLE. For all panels with periods from 8 to 30, the bias unadjusted SMLEs of β are biased downward. There are upward biases in the SMLEs of λ_1 and downward biases in the SMLEs of λ_2, σ and ρ, so the dynamic effect can be over stated, but the lagged peer group effect and the serial correlation of disturbances can be underestimated. The magnitude of bias increases with panel length, as the dimension of integration and the total number of choice alternatives are proportional to the number of periods. On the other hand, SDs of all the SMLEs decrease as panels become longer, since longer panel data provide more sample information about the stochastic process. If periods are not too long, RMSEs decrease. Biases of estimates are all substantially reduced when the number of simulated random variables m increases from 15 to 50. By increasing m to 100, biases become rather small and RMSEs can further be reduced, but the time cost is double. The issue of selecting m in practice has been addressed by Lee (1997). For small m, bias correction is valuable. Although SDs of bias-adjusted estimates are slightly larger, RMSEs of bias-adjusted estimates are smaller in general. The additional CPU cost for bias correction is negligible. However, as biases of estimates, especially for longer panels, are relatively large to begin with in this model, larger m is desirable for better improvement.[1]

[1]Results for the bias-adjusted estimates are omitted in subsequent tables to save space. The bias correction procedure for all the models in this article reduces bias and RMSE. The improvement is comparable with the gains from the bias correction procedure reported in *Table 1.7*.

Table 1.8 reports Means, SDs and RMSEs for alternative group sizes. For a given sample size $G \times n = 200$, biases, SDs and RMSEs of all the SMLEs increase when the group size n increases from 4 to 8 (by comparing results in Tables 1.7 and 1.8). As the group size becomes even larger, biases, SDs and RMSEs of the SMLEs of λ_1, λ_2 and ρ further increase, while the estimates of β and σ are not much affected. As such, other things equal, more sessions with fewer subjects are preferred to fewer sessions with more subjects in each session.

To illustrate effects of ignoring potential lagged social interactions on SMLEs, we report the restricted SMLEs under $\lambda_2 = 0$ in Table 1.9. When positive social interactions are ignored, the SMLEs of β, σ and ρ are biased downward, and the SMLEs of λ_1 are biased upward. The estimated values of λ_1 are more than double in magnitude and the estimated values of ρ are reduced almost by half, so true state dependence can be over stated but spurious state dependence can be underestimated.

Misspecified disturbances, in general, would cause parameter estimates to be inconsistent. We investigate effects of misspecification in disturbances by the following Monte Carlo experiments. First, we estimate the random component model with $\sigma\xi_i + v_{it}$, where v_{it} are serially uncorrelated, with the data samples generated by the model specified as in (1.8). For random component models, multivariate probability functions involve only single integrals, which can be effectively implemented by the Gaussian Quadrature method as suggested by Butler and Moffitt (1982). However, for the sake of easy comparison, here we report the SMLE of the random component model. The simulated log likelihood function for the random component model is

$$\sum_{i=1}^{G\times n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi \left[(2y_{it} - 1) \left(\beta x_{i,t-1} + \lambda_1 y_{i,t-1} + \lambda_2 z_{i,t-1} + \sigma \xi_i^{(j)} \right) \right] \right\}.$$

The SMLEs are reported in the upper block of Table 1.10. There are substantial downward biases in the SMLEs of β and λ_2 and upward biases in the SMLEs of λ_1. Biases are more severe for longer panels. Even with $m = 100$, the estimated values of λ_1 are three times larger than the true value; and the estimated magnitudes of λ_2 are reduced by 2/3. Hence, true state dependence tends to be overestimated and lagged social interactions tend to be underestimated when serial correlation in ϵ_{it} is ignored. Biases in the SMLEs of σ are not uniform. The lower block of Table 1.10 reports the restricted SMLEs under $\sigma = 0$, i.e., random component ξ were ignored. With this error specification, serially correlated disturbances $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \upsilon_{it}$ capture all the spurious state dependence. Ignoring random individual component biases the SMLEs of β, λ_1 downward and λ_2, ρ upward. Biases in λ_1 and λ_2 are more severe for longer panels. The magnitudes of upward bias of λ_2 are not really large. The biases of ρ are upward by 50%. But the biases of λ_1 towards zero are relatively much more severe.

1.3.2 A Polya Model with Lagged Social Interactions

In the Polya model, the entire history of the dynamic process is relevant to current decision making. The Polya model with a depreciation factor δ is specified as follows[2]:

$$y_{it}^* = \beta x_{i,t-1} + \lambda_1 \sum_{s=1}^{t} \delta^{s-1} y_{i,t-s} + \frac{\lambda_2}{\sum_{s=1}^{t} \delta^{s-1}} \sum_{s=1}^{t} \delta^{s-1} z_{i,t-s} + \sigma\xi_i + \varepsilon_{it}, \tag{1.9}$$

where $z_{i,t-s} = \sum_{j=1, j\neq i}^{n} y_{j,t-s}/(n-1)$ and $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \upsilon_{it}$ with ξ_i and υ_{it} i.i.d. $N(0,1)$. The group subscript g has been suppressed for simplicity. The initial values of all variables for $t \leq 0$ are given as 0. Substitution of $\varepsilon_{it} = \rho(y_{i,t-1}^* - (\beta x_{i,t-2} + \lambda_1 \sum_{s=1}^{t-1} \delta^{s-1} y_{i,t-s-1} + \lambda_2 \sum_{s=1}^{t-1} \delta^{s-1} z_{i,t-s-1}/\sum_{s=1}^{t-1} \delta^{s-1} + \sigma\xi_i)) + \upsilon_{it}$ in (1.9) conforms

[2]Here we specify the lagged social interactions term as the (weighted) average for observed laggd choices of peers over the entire history, so that it is not affected by the number of total observations.

it to the general model (1.1) . For comparison purpose, the discount factor δ is assumed to be a known constant and is set at 0.7. Sample data are generated with $\beta = 1$, $\lambda_1 = 0.2$, $\lambda_2 = 0.4$, $\sigma^2 = 0.5$, and $\rho = 0.4$.

The SMLEs are reported in Table 1.11. There are some downward biases in the SMLEs of β, λ_2, σ and ρ and upward bias in λ_1. Compared to estimates of the Markov model in Table 1.7, λ_1 and ρ in the Polya model can be estimated more accurately. They not only have small biases but also have much smaller SDs, due to an apparently stronger state dependence property of the Polya model. On the other hand, since we specified lagged social interactions as a weighted average of the past history instead of a weighted sum, variation in this term is reduced. So with such specification, λ_2 in the Polya model is much more difficult to estimate than in the Markov model. For small m and long panels, biases in the SMLEs of λ_2 is quite severe. By increasing m, biases in the estimates of λ_2 can be substantially reduced. For $T = 8$ or 15, the biases are smaller with $m = 50$ or 100. By comparison with the Markov model, SDs and RMSEs of the estimates of λ_2 here are two times larger.

Monte Carlo experiments are also performed to investigate effects of misspecification in dynamic structures on SMLEs. Table 1.12 reports the SMLE of the Markov model with lagged social interactions when data samples are generated by the Polya model (1.9). The SMLEs of λ_1 and σ are biased upward. And the SMLEs of λ_2 and ρ are biased downward. Hence, when the Polya dynamic structures are misspecified to be Markov, the true state dependence and the serial correlation due to unobserved heterogeneity tends to be overestimated but the lagged social interactions and the serial correlation of the remaining disturbance tend to be underestimated. The SMLEs of β are not affected very much by dynamic misspecification and their biases are not large.

1.4 An Application: Estimating Peer Group Effects in Experiments on Signaling Games

There is a large volume of literature on measuring peer group effects in field settings, while little attention has been paid to evaluating the influence of peer group effects on subjects' performance in experiments. Measuring the peer group influence in experiments is important as it affects our understanding of the evolution of subjects' behavior over time. Ignoring peer group effects potentially confounds any "sophisticated" learning process (e.g. adaptive learning) where subjects update beliefs, with the less "sophisticated" social learning where subjects simply replicate the strategy generating a better outcome. Furthermore, experimental results across diverse subject pools are much less likely to be consistent in the presence of strong peer group effects, as subjects' performance depends on the overall performance of the experimental session they were in. This section adopts dynamic discrete choice models with lagged social interactions to investigate the presence and magnitude of peer group effects in experiments on signaling games.

Following Manski (1993), similar behavior of individuals belonging to the same reference group may be due to *endogenous effects*, wherein "the propensity of an individual to behave in some way varies with the behavior of the group"; *exogenous effects*, wherein "the propensity of an individual to behave in some way varies with the exogenous characteristics of the group"; and *correlated effects*, wherein "individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments". In experimental settings, exogenous effects and correlated effects can be controlled through recruiting procedures and careful experimental designs, while endogenous effects are relatively hard to control by experimenters. We focus on measuring endogenous peer group effects in experiments in this section.

The plan of this section is as follows. Subsection 1.4.1 presents the theoretical predictions of the model of entry limit pricing. Subsection 1.4.2 outlines the experimental procedures and provides a general description of the data. Subsection 1.4.3 develops the empirical econometric models and interprets the estimation results.

1.4.1 Theoretical Considerations

Milgrom and Roberts (1982) propose a model of entry limit pricing as follows. There are two firms, an established monopolist M and a potential entrant E, in a two-stage market producing a homogeneous good. Nature decides M's cost of production along with the distribution of these costs. M's cost is his/her private information throughout the game, with the prior distribution of the cost being common knowledge. In the first stage, M chooses an output (or price) level. In the second stage, E chooses to enter or stay out in response to the observed output (or price) level. The predetermined opportunity cost to E for entering the market is common knowledge. If entry occurs, Cournot duopoly profits are realized by both M and E. If there is no entry, M receives the single period monopoly profit. Entry is profitable against M with high cost but not against M with low cost. M may have an incentive to limit pricing, which involves producing greater output (or charging lower price) in the first stage than the single period profit maximizing level in order to make entry appear unattractive.

In this game, the information sets are defined by the realized costs of M and E (c_M and c_E) and a choice of Q (quantity) by M. A (pure) strategy for M is a map s from its possible cost levels into the possible choices of Q and a (pure) strategy for E is a map t from R^2 into $\{0, 1\}$ giving its decision for each possible pair (c_E, Q), where 1 is interpreted as "enter" and 0 as "stay out". An equilibrium consists of a pair of strategies

(s^*, t^*) and a pair of conjectures $(\bar{s}, \bar{t})$ such that (i) M's pricing policy s^* is a best response to its conjectures $\bar{t}$ about E's entry rule, (ii) the strategy t^* is a best response for E to its conjecture $\bar{s}$, and (iii) the actual and conjectured strategies coincide. With two cost levels (types) for M, namely, $\underline{c}_M < \overline{c}_M$, if $s^*(\underline{c}_M) = s^*(\overline{c}_M)$, an equilibrium is called *pooling*; and if $s^*(\underline{c}_M) \neq s^*(\overline{c}_M)$ the equilibrium is *separating*. Partial pooling is a mixed-strategy equilibrium that $s^*(\underline{c}_M) = s^*(\overline{c}_M)$ with a certain probability. In a pooling equilibrium, E can infer nothing from observing Q and so enters if the expected profit is positive. In a pure-strategy separating equilibrium, the observation of Q allows the value of c_M to be inferred exactly. Depending on the cost structure, its distribution, and the market demand function, pooling equilibria and/or separating equilibria can occur (Milgrom and Roberts, 1982, pp. 446-448).

Milgrom and Roberts' model of entry limit pricing is investigated experimentally by Cooper, Garvin and Kagel (1997*a*; 1997*b*) and Cooper and Kagel (2003*a*; 2003*b*; 2004). In the experiments, the game is further simplified by adding the payoffs of the two stages together and providing Ms with a single payoff table. Payoff tables 1.1-1.3 are provided in the "quantity game" with M choosing over output levels (1-7) in payoff table 1.1, and payoff tables 1.4 and 1.5 are provided in the "price game" with M choosing over price levels in payoff table 1.4.

In the "quantity game", M is either a high-cost type (M_H) or a low-cost type (M_L) with equal probability. E's cost is common knowledge. In a given treatment of the experiment, Es are either all high cost types (E_Hs; payoff table 1.2) or all low cost types (E_Ls; payoff table 1.3). With E_Hs there exist pure-strategy pooling equilibria at output levels 1-5. There also exist two pure-strategy separating equilibria, in which M_Hs always choose 2 and are always entered on, M_Ls always choose 6 or 7 and are never entered on. Among them, only pooling at 4 or 5, and separating with M_Ls choosing 6 survive Cho-Kreps' (1987) intuitive criteria for equilibrium refinement. With E_Ls no pure-strategy pooling equilibrium exists, while the two pure-strategy separating equilibria still exist. There also exist a number of mixed-strategy equilibria. One that is of particular relevance is the partial pooling equilibrium in which M_Ls always select 5 while M_Hs mix between 2 (with probability 0.8) and 5 (with probability 0.2), and Es always enter on output levels other than 5, enter on 5 with probability 0.11. In simulations using a stochastic fictitious play learning model, this partial pooling equilibrium emerges with high frequency in the presence of E_Ls (Cooper, Garvin and Kagel, 1997*b*). Further, in practice M_Ls choose 5 with relatively high frequency as a separating equilibrium emerges (especially early on) and there is very little entry in response to it (Cooper, Garvin and Kagel, 1997*b*).

The payoffs in the "price game" are a linear transformation of payoff tables 1.1 and 1.3 in the "quantity game", (with table presentation changed as well). Hence the price game is theoretically identical to the quantity game with analogue equilibrium predictions.

1.4.2 Experimental Procedures and Data

Detailed description of the experimental procedures can be found in Cooper, Garvin and Kagel (1997*b*). The following lists some elements that are especially noteworthy, as they will be taken into account when empirically modeling the game.

1. Each experimental session employed between 12 and 16 subjects who were randomly assigned to computer terminals. Sessions typically lasted 36 periods, with the number of periods announced in advance. Subjects switched roles after every six plays, with Ms becoming Es and vice versa. Ms' types are generated each play randomly.

2. Following each play of the game the outcomes from all pairings (Ms' choice, Es' choice, and Ms' type) were revealed to all subjects. This made learning across individuals feasible, and provided the basis for potential peer group effects.

3. Subjects were randomly paired with each other for each play of the game, and subject identification numbers were suppressed when the game results were revealed. Hence there was no opportunity for reputation effects to develop. Learning, to the extent that it occurred, had to be based on own experience and observations of peer's choices and outcomes.

Experimental treatments are summarized in Table 1.6. The "Experienced Subjects" treatment recruited subjects who had participated in earlier experimental sessions with exactly the same payoff tables. The treatment "Meaningful Context" uses natural language for the instructions, and was introduced to explore the effects of context on subjects' reasoning process in signaling games (Cooper and Kagel, 2003*a*). The treatment "Crossovers from the E_H to E_L game" employed subjects with experience in the quantity game with payoff tables 1.1 and 1.2 to play the quantity game with payoff tables 1.1 and 1.3, and was devoted to investigating subjects' ability to generalize learning in one game to related games (Cooper and Kagel, 2003*b*; 2004).

1.4.3 Empirical Models and Estimation Results

According to payoff table 1.1, with full information, output levels 2 and 4 are optimal for M_Hs and M_Ls respectively. Pooling equilibria at output levels 3-5 and (partial) separating equilibria with M_Ls selecting output levels 5-7 involve strategic behavior - limit pricing - as Ms produce above (or price below) full-information levels. A "gradual, history-dependent adjustment process", starting with Ms "at their myopia maxima, followed by an attempt to pool, and then (if no pooling equilibrium exists) separation", has been observed by Cooper, Garvin and Kagel (1997*b*). Here we adopt our dynamic discrete choice models with lagged social interactions to characterize the evolution of subjects' behavior in the experiment.

We consider the estimation with two different samples: One from the experimental sessions with E_Hs (using payoff tables 1.1 and 1.2) and the other from the sessions with E_Ls (using payoff tables 1.1 and 1.3). With E_Hs, play reliably converges to a pure strategy pooling equilibrium in which M_Hs learn to imitate M_Ls. As such we model the learning process of M_Hs in this situation, treating choices of output levels 3-5 by M_Hs as limit pricing[3]. For games with E_Ls, pure-strategy pooling equilibria no longer exist, and we focus on the strategic play by M_Ls, with output levels 5-7 by M_Ls considered as limit pricing[4]. As the adjustment (learning) process for the two samples are modeled analogously, we only detail the model specifications for the estimation of games with E_Hs.

A Markov Model with Lagged Social Interactions

In a generic experimental session with $2n$ E_Hs, M_Hs have incentives to limit price. We assume that the unobservable incentives for M_Hs to limit price can be characterized by the Markov dynamic discrete choice model with lagged social interactions. To justify the Markov model as an approximation for the learning process, we assume that, besides individual characteristics, a subject's current decision only depends on his/her last decision and the feedback information from the previous round of the game. We will relax this

[3]Note that high-level outputs 6, 7 are strictly dominated by other outputs for M_Hs, according to payoff table 1.1. Among the 4576 observations in the actual experimental sample with E_Hs, only 7 choices of output 6 or 7 made by M_Hs are observed.

[4]We have also tried to estimate with alternative criterion for limit pricing. For example, we treated output levels 4, 5 by M_Hs in games with E_Hs, and output levels 6, 7 by M_Ls in games with E_Ls as limit pricing. The estimation results are similar to those reported here.

restrictive assumption and consider the estimation of a more general dynamic process in the next subsection.

By experimental design, a subject is randomly assigned turns as M_H in different plays of the game within an experimental session. At the same time, a subject can observe peers' output choices and entrants' responses from all previous rounds of the game. As such, we distinguish between a *decision period* in which a subject plays as M_H with the opportunity to limit price and a (consecutive) *calendar period*. For a subject i, let T_i be the total number of decision periods in which he/she has played as M_H. Corresponding to each decision period τ ($\tau = 1, \cdots, T_i$), there is a calender period. Let $t_i(\tau)$ be the calendar period when the subject i plays as M_H. The Markov dynamic discrete choice model with lagged social interactions for the subject i can be specified as

$$y^*_{it_i(\tau)} = \alpha + x_{i,t_i(\tau)-1}\beta + \lambda_1 y_{i,t_i(\tau-1)} + \lambda_2 w_{in} Y_{n,t_i(\tau)-1} + \gamma \ln \tau + \sigma\xi_i + \varepsilon_{it_i(\tau)}, \quad (1.10)$$

for $\tau = 1, \cdots, T_i$. We assume that $\varepsilon_{it_i(\tau)} = \rho\varepsilon_{i,t_i(\tau-1)} + \upsilon_{it_i(\tau)}$, and ξ_i and $\upsilon_{it_i(\tau)}$ are i.i.d.$N(0, 1)$. As the dynamic process starts at the first sampling period in the experiment, the initial conditions on all variables for $t \leq 0$ are zero.

If the latent dependent variable $y^*_{it_i(\tau)} > 0$, the subject i limits price in his/her τth turn as M_H, and the corresponding observed dependent variable $y_{it_i(\tau)}$ is 1; $y_{it_i(\tau)}$ is 0 otherwise. Explanatory variables are on the right hand side of (1.10). α is a constant. $x_{i,t_i(\tau)-1}$ is the perceived entry rate differential between "myopia" output choices 1-2 and strategic output choices 3-5.[5] Specifically, let $d^L_{is}(IN)$ (respectively, $d^O_{is}(IN)$) be a dummy variable indicating that the subject i chooses output level 3, 4 or 5 (respectively, output level 1 or 2) and is entered on in calendar period s. Let $d^L_{-is}(IN)$ (respectively, $d^O_{-is}(IN)$) be the

[5] The entry rates are calculated conditional on the output level selected, not on the type of M which selects the output. As Es can not observe Ms' type when making decisions of entry, the entry rate specified here can be used to approximate M's beliefs on Es' responses.

number of times in calendar period s that Ms *other than i* choose output level 3, 4 or 5 (respectively, output level 1 or 2) and observe the response IN. Define $d_{is}^{L}(OUT)$, $d_{is}^{O}(OUT)$, $d_{-is}^{L}(OUT)$, and $d_{-is}^{O}(OUT)$ in an analogous manner, where OUT involves potential Es staying out. Denote the weight a player put on entries on other Ms' choices relative to entries on his/her own in calculating entry rate differential by ω. The perceived entry rate differential is given by

$$x_{i,t_i(\tau)-1} = \frac{d_{i,t_i(\tau)-1}^{O}(IN) + \omega d_{-i,t_i(\tau)-1}^{O}(IN)}{d_{i,t_i(\tau)-1}^{O} + \omega d_{-i,t_i(\tau)-1}^{O}} - \frac{d_{i,t_i(\tau)-1}^{L}(IN) + \omega d_{-i,t_i(\tau)-1}^{L}(IN)}{d_{i,t_i(\tau)-1}^{L} + \omega d_{-i,t_i(\tau)-1}^{L}},$$

where $d_{i,t_i(\tau)-1}^{j} = d_{i,t_i(\tau)-1}^{j}(IN) + d_{i,t_i(\tau)-1}^{j}(OUT)$ for $j = L, O$.[6] This term serves as a proxy for the unobservable beliefs of Ms regarding potential entrants' responses to different output choices. $y_{i,t_i(\tau-1)}$, the time-lagged observed dependent variable, is introduced to measure the true state dependence in the dynamic process. $Y_{n,t_i(\tau)-1}$ is an n-dimensional column vector with the ith element being $y_{i,t_i(\tau)-1}$ $(i = 1, \cdots, n)$ and w_{in} is a $1 \times n$ normalized weighting vector. The coefficient on $w_{in}Y_{n,t_i(\tau)-1}$ captures the peer group effects in an experimental session, namely the influence of the other M_Hs' strategic play of limit pricing in the preceding calender period on the subject i's current choice. Given the anonymous nature of experimental design, we assume that the weighting matrix W_n, where w_{in} is its ith row, is simply $\left[\left(1_n \cdot 1_n' - I_n\right)/(n-1)\right]$, so that $w_{in}Y_{n,t_i(\tau)-1} = \sum_{j=1,j\neq i}^{n} y_{j,t_i(\tau)-1}/(n-1)$. The coefficient on $\ln \tau$, where τ is the number of decision periods that the subject i has played as M_H (the current decision period included), collects all other experience effects within an experimental session that are not captured by the other explanatory variables. An random individual component ξ_i is introduced to control unobserved heterogeneity across players. The remaining disturbances

[6] We assume that $(d_{i,t_i(\tau)-1}^{j}(IN) + \omega d_{-i,t_i(\tau)-1}^{j}(IN))/(d_{i,t_i(\tau)-1}^{j} + \omega d_{-i,t_i(\tau)-1}^{j}) = 0.5$, in the case that $d_{i,t_i(\tau)-1}^{j} + \omega d_{-i,t_i(\tau)-1}^{j} = 0$ $(j = L, O)$.

are assumed to follow an AR(1) process, as we find in the Monte Carlo experiments that flexible error specifications are favorable to identify potential peer group effects.

We model the adjustment process of M_Ls' choices in experimental sessions with E_Ls in an analogous manner, with $x_{i,t_i(\tau)-1}$ being the perceived entry rate differential between output levels 1-4 and 5-7.

By a quasi-difference transformation, i.e. by substituting in (1.10)

$$\begin{aligned}\varepsilon_{it_i(\tau)} &= \rho(y^*_{it_i(\tau-1)} - (\alpha + x_{i,t_i(\tau-1)-1}\beta + \lambda_1 y_{i,t_i(\tau-2)} + \lambda_2 w_{in} Y_{n,t_i(\tau-1)-1} \\ &\quad + \gamma \ln(\tau - 1) + \sigma\xi_i)) + \upsilon_{it_i(\tau)},\end{aligned}$$

it is easy to see that the empirical Markov model conforms to the general dynamic model (1.1). As we have shown, for the subject i, the likelihood function involves $(T_i - 1)$-dimension integrals that are analytically intractable and numerically hard to evaluate. We circumvent this computational difficulty in implementing maximum likelihood estimation by the simulation method based on the unbiased GHK simulator. Table 1.13 reports the SMLEs for the Markov model based on a simulator generated from 100 random draws using data from the experimental sessions with E_Hs and E_Ls respectively.[7]

The positive and statistically significant SMLEs of λ_1 in all cases show that a subject's current choice depends heavily on his/her choice in the previous decision period. That is, one round of strategic play substantially increases the likelihood of strategic play in the future decision periods. This indicates that subjects do not play strategically just by chance. Rather, once they began to play strategically, they are very likely to continue to do so. This is a clear evidence of learning. Interaction terms (with the dummy variable NX representing sessions with inexperienced subjects) are introduced to account for the

[7] We have tried to add more interaction terms, or remove some regressors or interaction terms with insignificant coefficients. The estimation results are trivially affected.

differences between experienced and inexperienced subjects in learning.[8] The negative and significant coefficient estimate for the interaction term of lagged choice and NX in games with E_Ls indicates that inexperienced subjects were much less confident of their choice of strategic play than their experienced selves in this more challenging game.

Learning can come about in one of two ways: social learning in which case subjects simply replicate the peers' strategies and/or (individual) adaptive learning in which case subjects update beliefs according to the opponents' responses. Positive and significant estimate of peer group effects in the dynamic model is evidence in favor of social learning, while positive and significant coefficient estimate for entry rate differential is considered as evidence in favor of adaptive learning that is independent of peers' choices.

In games with E_Ls, the SMLEs of λ_2 are positive and statistically significant in Table 1.13, indicating the existence of endogenous peer group effects in this case. For the specification without interaction terms, the average marginal impact of the peer group effect on the probability of limit pricing given exogenous variables and lagged choices is 0.054.[9] In contrast, peer group effects in games with E_Hs are not statistically significant in the Markov model.

[8]The minus two times log likelihood ratios for testing jointly the significance of interactions terms in the Markov model are, respectively, 6.56 for games with E_Hs, and 13.2 for games with E_Ls. The latter is significant at the 5 percent level with an asymptotic $\chi^2(5)$ distribution.

[9]For the general model (1.1), $E(y_{it}|(y^*_{is}, Y_{ns}, X_{ns}, s = 1, \cdots, t-1), X_{nt}, \xi_i) = \Phi(h_{it})$. The average marginal effect over time and individual of, say X_{nt} (which is assumed continuous), on the transition probability $P(y_{it} = 1|(Y_{ns}, X_{ns}, s = 1, \cdots, t-1), X_{nt})$, is given by

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\int\cdots\int\phi(h_{it})(\partial h_{it}/\partial X_{nt}) \\ \times f\left(y^*_{i1}, \cdots, y^*_{i,t-1}, \xi_i|Y_{ns}, X_{ns}, s = 1, \cdots, t-1\right)dy^*_{i1}\cdots dy^*_{i,t-1}d\xi_i$$

The multiple integrals here can be approximated by simulations. We simulate $h^{(j)}_{it}$ following the same procedure as in (1.3). With m independent simulation runs, the corresponding (sample average) simulated marginal effects is $\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{t=1}^{T}\phi(h^{(j)}_{it})(\partial h_{it}/\partial X_{nt})/mnT$. Results reported in this paper are based on a simulator generated from 1000 random draws.

In games with E_Ls, the coefficient estimates for entry rate differential are positive, statistically significant and robust to alternative specifications. For the specification without interaction terms, the average marginal effect of entry rate differential on the probability of limit pricing is 0.032. According to the adaptive learning model, when subjects update beliefs regarding entrants' responses, they should not distinguish between entries on themselves and entries on their peers given the anonymous nature of the experiment. However, the estimated ω is significantly less than 1, indicating M_L places primary weight on entries on himself/herself, with very limited weight placed on entries on other Ms. Given that λ_2 is positive and statistically significant, we think that subjects only pay attention to peers' past choices, but not the corresponding outcomes. As λ_2 captures social learning that replicates peers' strategies, while ω captures adaptive learning that updates beliefs based on peers' experience, this result is quite reasonable given the sophisticated nature of adaptive learning compared to social learning.[10] On the other hand, the coefficient estimates for entry rate differential are not statistically significant in experiments with E_Hs. As will be reported in the next subsection, in games with E_Hs, peer group effects are identifiable in the Polya model (our preferred specification), but the coefficient on entry rate differential continues to be statistically insignificant.

The proportion of M_Hs attempting to pool by choosing output levels 3 and 4 in the previous round is introduced as an additional explanatory variable in games with E_Ls because an increase in this proportion makes separation at output levels 5-7 more attractive for M_Ls. Although positive in sign, its coefficient fails to achieve statistical significance in either Markov specifications.

[10] We also consider alternative specifications where own lagged choices and peers' lagged choices are interacted with entrants' responses. This is discussed briefly in the next section where we report on the polya model specification, our preferred specification.

The positive and significant estimates of the coefficient on $\ln \tau$ pick up other experience effects that fail to be captured in the Markov model. In experiments with E_Hs, the results on its interaction effect with NX indicate that this positive impact is confined to experimental sessions employing inexperienced subjects only. It motivates us to develop a more general empirical model to characterize the remaining experience effect in the next subsection.

The dummies for experienced players are large, positive and statistically significant in games with E_Ls and E_Hs indicating that in both cases experienced subjects start out with much higher levels of strategic play than inexperienced subjects. In games with E_Ls, dummies for experiments with crossovers are positive and statistically significant, which is consistent with the findings in Cooper and Kagel (2004) that there exists positive transfer of learning across related games. The negative and significant estimates for the constants (α) indicate the slow emergence of strategic play in all cases. The larger absolute value for α in games with E_Ls suggests that strategic play is much slower to emerge in this case.

Though the overall correlation in the disturbances captured by $\sigma\xi_i + \varepsilon_{it_i(\tau)}$ is positive, the negative sign of ρ suggests the presence of some fluctuations not captured by the dynamic structure. Hence we generalize the Markov model to a more general dynamic process in the next subsection.

A Polya Model with Lagged Social Interactions

As subjects have access to all previous outcomes in an experimental session, the entire history of past plays should be relevant to the current decision making. In this subsection, we model the influence of all past plays on a subject's current decision by a Polya process with lagged social interactions. Similarly to the Markov model, we assume that the unobservable incentives to limit price can be characterized by

$$\begin{aligned} y^*_{it_i(\tau)} &= \alpha + \bar{x}_{i,t_i(\tau)-1}\beta + \lambda_1 \sum_{s=1}^{\tau} \delta_1^{s-1} y_{i,t_i(\tau-s)} + \lambda_2 \sum_{s=1}^{t_i(\tau)} \frac{\delta_2^{s-1} w_{in} Y_{n,t_i(\tau)-s}}{\sum_{s=1}^{t_i(\tau)} \delta_2^{s-1}} \\ &\quad + \gamma \ln \tau + \sigma \xi_i + \varepsilon_{it_i(\tau)}, \end{aligned} \tag{1.11}$$

and

$$\varepsilon_{it_i(\tau)} = \rho \varepsilon_{i,t_i(\tau-1)} + v_{it_i(\tau)},$$

where ξ_i, $v_{it_i(\tau)}$ are i.i.d.$N(0,1)$. The initial conditions on all variables for $t \leq 0$ are set to be zero, as we observe the data generating process from the very beginning in the experiment. Most variables in (1.11) are defined as in the Markov model (1.10), while there are some changes in the specification of the entry rate differential as follows. Let $c^j_{i,t_i(\tau)-1}(R) = \sum_{s=1}^{t_i(\tau)-1} d^j_{is}(R)$ and $c^j_{-i,t_i(\tau)-1}(R) = \sum_{s=1}^{t_i(\tau)-1} d^j_{-is}(R)$ for $j = L, O$ and $R = IN, OUT$, with $d^j_{is}(R)$ given as before. Let the weight a player puts on the experience of other Ms relative to his/her own in calculating entry rate differential be ω. The perceived cumulative entry rate differential between "myopia" output choices and strategic output choices is given by

$$\bar{x}_{i,t_i(\tau)-1} = \frac{c^O_{i,t_i(\tau)-1}(IN) + \omega c^O_{-i,t_i(\tau)-1}(IN)}{c^O_{i,t_i(\tau)-1} + \omega c^O_{-i,t_i(\tau)-1}} - \frac{c^L_{i,t_i(\tau)-1}(IN) + \omega c^L_{-i,t_i(\tau)-1}(IN)}{c^L_{i,t_i(\tau)-1} + \omega c^L_{-i,t_i(\tau)-1}},$$

where $c^j_{i,t_i(\tau)-1} = c^j_{i,t_i(\tau)-1}(IN) + c^j_{i,t_i(\tau)-1}(OUT)$ for $j = L, O$. Analogous to $x_{i,t_i(\tau)-1}$ in the Markov model (1.10), $\bar{x}_{i,t_i(\tau)-1}$ here represents the payoff incentive for M to limit price. The depreciation factors δ_1 and δ_2 measure the influence of past plays on the current choice. The coefficient on the weighted average $\sum_{s=1}^{t_i(\tau)} \delta_2^{s-1} w_{in} Y_{n,t_i(\tau)-s} / \sum_{s=1}^{t_i(\tau)} \delta_2^{s-1}$ captures the cumulative peer group effects on the subject i's current decision. As in the Markov model (1.10), we specify the row-normalized weighting matrix W_n, with its ith row being w_{in}, as $\left[\left(1_n \cdot 1'_n - I_n\right)/(n-1)\right]$. Thus, in the Polya model, a subject's current decision is assumed to be influenced by the (weighted) average of the peers' observed choices over the entire history. Based on the GHK simulator generated with 100 random draws, the SMLEs of the Polya model using data from the experimental sessions with E_Hs and with E_Ls are reported in Table 1.14 and Table 1.15 respectively.

In games with E_Hs and E_Ls, the positive and significant estimates of λ_1 on own lagged choices imply that previous strategic plays substantially increase the likelihood of current strategic play for any given M.

The coefficient on the interaction term between lagged choices and NX (dummy for sessions with inexperience subjects) is not statistically significant in games with E_Hs but is negative and statistically significant in games with E_Ls.[11] This is similar to what we found in the estimation of the Markov model. In games with E_Ls, inexperienced subjects are less confident in their choices of strategic plays than their more experienced counterparts, hence are more likely to revert back to non-strategic play.

In games with E_Hs, the cumulative peer group effects captured by the estimated λ_2 are positive and statistically significant for inexperienced subjects. In contrast, the coefficient estimate of the cumulative entry rate differential is not, and the statistical insignificance of β makes the estimate of ω extremely imprecise.[12]

In games with E_Ls, cumulative peer group effects are positive and statistically significant overall, with even stronger peer group effects for inexperienced subjects (indicated by the positive coefficient estimate for the interaction term between peer group effects and NX, with t-ratio 1.484). Thus, inexperienced subjects are influenced more by the peer

[11] The interaction terms in the Polya model are jointly significant at the 5 percent level with the minus two times log likelihood ratios being 11.76 for the games with E_Hs, and significant at the 1 percent level with the minus two times log likelihood ratios being 24.4 for games with E_Ls.

[12] Note that ω would not be identifiable if the coefficient of x_{it} were zero. The value of this estimate may reflect the insignificance of the coefficient estimate of x_{t-1} in this case.

group than experienced subjects in games with E_Ls. But unlike games with E_Hs, experienced subjects continue to be influenced by their peers, due to the fact that it takes longer for a separating equilibrium to emerge than a pooling equilibrium in the experiment. The coefficient on cumulative entry rate differential is statistically significant in games with E_Ls, but less so for inexperienced than experienced subjects. And the marginal effect of entry rate differential is smaller than that of peer group effects.[13] Furthermore, similar to what we found in the Markov model, subjects place much *less* weight on entries on other Ms than entries on themselves in calculating entry rate differential as indicated by the estimated ω.

We believe that the learning results reported on above come about for three reasons: (1) Adaptive learning is more demanding than social learning, as it requires that subjects form expectations regarding opponents' responses based on past outcomes as compared to social learning where subjects simply imitate peers' strategies. As such, social learning is likely to be more prominent in the early stages of the learning process. (2) Because Es' responses are less stable in inexperienced subject sessions (especially with respect to output choice 5, 6 or 7 in games with E_Ls) than experienced subject sessions, the entry rate differential serves as a poor proxy for M's beliefs in those sessions. Hence the coefficient estimate of entry rate differential is less significant in sessions with inexperienced subjects. (3) Strategic play of M_Ls in games with E_Ls requires innovation, whereas strategic play by M_Hs in games with E_Hs simply requires imitating M_Ls choices. As such there must be some element of adaptive learning in games with E_Ls, while this is likely to be superseded

[13] For the specification without interaction terms, in games with E_Hs, the average marginal cumulative peer group effect on the probability of limit pricing conditional on the exogenous variables and lagged choices is 0.058, and the average marginal effect of cumulative entry rate differential is 0.009. In games with E_Ls, the average marginal cumulative peer group effect is 0.113, and the average marginal effect of the cumulative entry rate differential is 0.085.

by social learning from M_Ls' choices (and responses to same) in games with E_Hs where such innovation is not required.

The SMLEs for both δ_1 and δ_2 are positive and statistically significant in games with E_Hs and E_Ls, indicating that a subject's current decision making is influenced by all past plays of the game. Note, δ_1 and δ_2 are not directly comparable, as the depreciation factor δ_1 for own lagged choices is defined on the decision period whereas the depreciation factor δ_2 for peer group effects is defined on the (consecutive) calendar period. On average a subject has one decision period (with a chance to limit price) every 4 calendar periods, because a subject plays as M only half time, and the type of M is randomly decided with equal probability. Take the games with E_Ls for instance, for the specification with no interaction terms, a generic M_L discounts peers' lagged choices $\delta_1/\delta_2^4 \approx 2$ times as fast as own lagged choices.

As in the Markov model, for games with E_Ls, we introduce the proportion of M_Hs attempting to pool by choosing output levels 3 and 4 as an additional explanatory variable. Different from the Markov model where this value is calculated based on M_Hs' choices in the previous round only, we calculate its cumulative counterpart in the Polya model. The positively significant estimates of its coefficient in experimental sessions with inexperienced subjects is consistent with the observation made by Cooper, Garvin and Kagel (1997*b*) that the adjustment process is history-dependent. It is M_Hs' attempt to pool that raises the entry rate on output level 4 and gives M_Hs incentive to separate. In both the Markov and Polya models, other experience effects represented by $\ln \tau$ are not statistically significant at conventional levels. And analogous to what happens in the Monte Carlo study, the weird negative sign of the SMLEs for ρ in the Markov model for games with E_Hs can now be explained by model misspecification.

As the Polya model does not nest the Markov model because of our different specifications of entry rate differential and the proportion of M_Hs attempting to pool, we address the issue of model selection by the well known Akaike information criterion (AIC) given as

$$AIC = -\frac{2}{\#obs} \log L + \frac{2\#p}{\#obs}, \tag{1.12}$$

where $\#obs$ is the sample size, $\#p$ is the number of parameters and $\log L$ is log likelihood of a model. According to (1.12), the Polya model is a better model than the Markov model as the former gives smaller value of AIC.[14]

One element that has been left out of the analysis reported on so far involves distinguishing between attempts at limit pricing and opposed to successful attempts at limit pricing. We introduce two new variables into the regressions: Individual M's own success in limit pricing and the percentage of successful limit pricing by peers. We view these new regressors as, essentially, additional interaction terms, the results of which are reported in Appendix 1.6.2. Introduction of these variables has essentially no effect on the log likelihood function for games with E_Hs so that the distinction has no impact on the results reported in this case.[15] In games with E_Ls own success at limit pricing plays a statistically significant role in promoting limit pricing (and diminishes the effect of cumulative entry rate differential). The percentage of limit pricing by peers (as opposed to attempts at limit pricing) plays no statistically significant, independent role in promoting limit pricing in games with E_Ls. This probably comes about because attempts at limit pricing were usually

[14]For the specification without interaction terms, in games with E_Hs, the AIC of the Markov model is 0.4807 and the AIC of the Polya model is 0.4788; and in games with E_Ls, the AIC of the Markov model is 0.3673 and the AIC of the Polya model is 0.3610.

[15]The minus two times log likelihood ratios for testing jointly the significance of new regressors in the Polya model without interactions with NX are, respectively, 1.86 for games with E_Hs, and 41.14 for games with E_Ls. The latter is significant at the 1 percent level with an asymptotic χ^2 (2) distribution.

successful so that imitators only needed innovators actions to promote limit pricing in this case.

Finally, the last column in Tables 1.14 and 1.15 look at the impact of neglecting the peer group effects. The reference specification against which to compare these estimates is the first one reported in each case. For games with E_Hs there is little if any effect on any of the coefficient values estimated and only a small change in the log likelihood function. This is not surprising as peer group effects are only significant at the 10% level in this case. For games with E_Ls, the SMLEs for the coefficient (β) on the entry rate differential and for the weight (ω) on entries on the peers are most affected by dropping the peer group effects, with both coefficients biased upwards. This is not too surprising since it is the increased entry differential in response to choices 5-7 versus other output levels (especially output levels 3 and 4) that drive M_Ls to limit price in the first place. In this context what the introduction of peer group effects does is to clarify the behavioral mechanism under which these increased entry rates operate. It is only partly related to what individual subjects have experienced themselves. Rather, much of the impact is related to what others have experienced and *their responses to same.* It is the latter that is largely missing by ignoring peer group or session level effects in the data in this case.

1.5 Conclusions

This chapter has generalized Heckman's (1981) dynamic discrete choice panel data models by introducing lagged social interactions. The likelihood function for a general model has been derived and simulation method based on the unbiased GHK simulator has been proposed to implement the SML estimation. Monte Carlo experiments have been conducted to investigate the finite sample performance of the SMLEs for the Markov and

Polya model with lagged social interactions. Some clear patterns have emerged from the Monte Carlo results.

- The true state dependence tends to be overestimated, and the lagged social interactions tend to be underestimated in the Markov and Polya models when T is long. The biases are small for $T = 8$ and 15, and $m = 50$ or 100. The lagged social interactions are relatively more difficult to estimate precisely in the Polya model than in the Markov model.

- Overall, the SMLEs of serial correlation in the disturbances have small downward biases in the Markov and Polya model.

- Given a fixed sample size, biases and SDs of all the SMLEs increase with group size, given the corresponding reduction in the number of groups. The estimates of state dependence and lagged social interactions are more sensitive to group size than the other estimates.

- The bias correction procedure reduces bias and RMSE, but the improvements are generally small. For further improvement, a larger number of random draws are desirable.

- In the Markov model, when positive lagged social interactions are ignored in the estimated model, the estimate of true state dependence is upward biased and the estimate of serial correlation in disturbances is downward biased. These biases can be severe.

- In the Markov model, when the data generating process incorporates both the random individual component and serial correlation but the estimated model only allows for the random individual component, the estimate of state dependence is upward biased and the estimate of lagged social interactions is downward biased. These biases can be severe. On the other hand, when the estimated model only allows for AR(1) serial correlation with the random individual component ignored, the estimate of state dependence can be severely downward biased and the estimate of lagged social interactions is moderately upward biased.

- When the data generating process is the Polya model but the estimated model is the Markov model, the estimate of state dependence can be severely biased upward and the estimate of lagged social interactions has some downward bias.

We have applied the model to investigate learning and peer group effects in laboratory experiments based on Milgrom and Roberts' (1982) entry limit pricing game. We employed the Markov and Polya processes with lagged social interactions to characterize the adjustment process of subjects' behavior over time. The Polya model is superior to the Markov model as it has a more natural justification and provides a better fit to the data. We obtained a number of important insights on this adjustment process.

First, the dynamic panel data model allows us to study the adjustment process with better detail. Past studies typically use the static panel data model with a time dummy to investigate the evolution of subjects' behavior over time. This model allows us to determine the correlation between aggregate frequency of strategic play and subjects' experience but says nothing about an individual player's persistency in strategic play. On the other hand, a positive statistically significant estimate of the true state dependence in the dynamic model can help to predict an individual's future strategic play conditional on his/her current and previous decisions, hence is stronger evidence for existence of learning. Furthermore, by introducing some interaction terms with the state dependence, we found that inexperienced subjects are less confident about what they learned than their experienced selves in more challenging games (i.e. games with E_Hs).

To distinguish between two different sources of learning, we introduced peers' past decisions and perceived entry rate differential into the dynamic model. We found that subjects' decisions are influenced by the past decisions of their peers in the limit pricing game experiment. These time-lagged peer group effects are more evident in the experiments employing subjects with no experience of the same or related games. These results suggest that the imitation of peers' strategies plays an important role in learning to play strategically.

On the other hand, perceived entry rate differential between "myopia" output choices and strategic output choices serves as a proxy for Ms' beliefs on Es' responses. Hence a positive estimate of its coefficient indicates existence of more intelligence-demanding (individual) adaptive learning. After controlling social learning, only in more challenging games with E_Ls, we found evidence that subjects' decision are affected by opponents' past responses. And evidence is less substantial in experimental sessions with inexperienced subjects, which, we believe, is partially due to the sophisticated nature of adaptive learning and unstable responses of inexperienced E_Ls. Furthermore, we found that subjects tend to overweigh entry on his/her own output choice relative to entries on other Ms' output choices in calculation of the perceived entry rate differential. As subjects only pay attention to peers' choices but largely ignore the outcomes of these choices, we should not be too optimism about the individual-intelligence implications of adaptive learning.

Multiple equilibria of the entry limit pricing game allows us to design related games with different equilibrium predictions and study learning processes converging to different types of equilibria. In games with E_Hs, where strategic behavior reliably converges to pure-strategy pooling equilibria, evidence of social learning is dominant. In games with E_Ls, where no pure-strategy pooling equilibria exist, evidence of social learning and (individual) adaptive learning coexists. One plausible explanation for the "inconsistency" of adjustment processes in related games is as follows. As the adjustment process in entry limit pricing game experiment has the feature of history dependence (Cooper, Garvin and Kagel, 1997*b*), the emergence of strategic play in games with E_Ls is much slower than in games with E_Hs. Only a relatively small proportion of Ms learn to play strategically at the end of the experiments with E_Ls. This slow process selects the more sophisticated subjects who are updating their beliefs in accordance with the observed responses of the opponents. On the other hand, in experiments where pure-strategy pooling equilibria exist and strategic play prevails at the end, the imitation behavior of "followers" becomes overwhelming in the population and makes other effects much harder to identify. An alternative explanation is that the nature of strategic play when pure-strategy pooling equilibria exist is imitation, with "sophisticated" M_Hs imitating the output choices of M_Ls, and the other M_Hs ("followers") imitating the choice of "sophisticated" M_Hs. When no pure-strategy pooling equilibria exist, M_Ls have no one to imitate in the first place. Hence M_Ls' strategic play may have some elements of more sophisticated learning.

Finally, we investigated the consequences of neglecting the positive significant peer group effects in estimating the Polya model. We found that in games with E_Ls, where the time-lagged peer group effects are substantial, ignoring such effects in the estimation causes the estimates of entry rate differential and weight a subject put on entries on others' output choices upward biased. As such, an individual's intelligence tend to be overstated.

1.6 Appendices

1.6.1 Payoff Schedules and Experimental Treatments

M_H (High Cost M)			M_L (Low Cost M)		
Your Choice	X (In)	Y (Out)	X (In)	Y (Out)	Your Choice
1	150	426	250	542	1
2	168	444	276	568	2
3	150	426	330	606	3
4	132	408	352	628	4
5	56	182	334	610	5
6	−188	−38	316	592	6
7	−292	−126	213	486	7

Source: Cooper, Garvin and Kagel (1997*b*).

Table 1.1: A Monopolist's Payoffs in the Quantity Game

	M Player's Type		
Your Action Choice	M_H (High Cost M) Your Payoff	M_L (Low Cost M) Your Payoff	Expected Value[a]
X (In)	300	74	187
Y (Out)	250	250	250

[a] Based on prior distribution (50% M_H, 50% M_L) of M types.

Source: Cooper, Garvin and Kagel (1997*b*).

Table 1.2: An Entrant's Payoffs in the Quantity Game

	M Player's Type		
	M_H	M_L	
Your Action	(High Cost *M*)	(Low Cost *M*)	
Choice	Your Payoff	Your Payoff	Expected Value[a]
X (In)	500	200	350
Y (Out)	250	250	250

[a] Based on prior distribution (50% M_H, 50% M_L) of M types.

Source: Cooper, Garvin and Kagel (1997*b*).

Table 1.3: An Entrant's Payoffs in the Quantity Game

M_H (High Cost *M*)			M_L (Low Cost *M*)		
Your Choice	X (In)	Y (Out)	X (In)	Y (Out)	Your Choice
1	−428	−220	204	545	1
2	−298	−110	333	678	2
3	8	165	355	700	3
4	103	448	378	723	4
5	125	470	350	695	5
6	148	493	283	648	6
7	125	470	250	615	7

Source: Cooper and Kagel (2004).

Table 1.4: A Monopolist's Payoffs in the Price Game

	M Player's Type		
	M_H	M_L	
Your Action	(High Cost *M*)	(Low Cost *M*)	
Choice	Your Payoff	Your Payoff	Expected Value[a]
X (In)	219	594	406.5
Y (Out)	281	281	281

[a] Based on prior distribution (50% M_H, 50% M_L) of M types.

Source: Cooper and Kagel (2004).

Table 1.5: An Entrant's Payoffs in the Price Game

1.6.2 Alternative Specifications for Empirical Models

In the main content of the application, we focus on disentangling the influence on incumbents' current decisions from the entrants (captured by the entry rate differential) and from the other incumbents of the same type (captured by the peer group effects). Here we also consider some interactions between them in the Polya model. Let the dichotomous indicator o_{is} be 1 if incumbent i is not entered on in calendar period s, and 0 otherwise. Let $\overline{Y}_{ns}$ be an n-dimensional vector with the ith element being $\overline{y}_{is} = y_{is}o_{is}$. We consider an alternative specification of the Polya model (1.11) as follows

$$\begin{aligned} y^*_{it_i(\tau)} &= \alpha + \overline{x}_{i,t_i(\tau)-1}\beta + \sum_{s=1}^{\tau} \delta_1^{s-1}\left(\lambda_1 y_{i,t_i(\tau-s)} + \lambda_1' \overline{y}_{i,t_i(\tau-s)}\right) \\ &\quad + \sum_{s=1}^{t_i(\tau)} \frac{\delta_2^{s-1} w_{in}\left(\lambda_2 Y_{n,t_i(\tau)-s} + \lambda_2' \overline{Y}_{n,t_i(\tau)-s}\right)}{\sum_{s=1}^{t_i(\tau)} \delta_2^{s-1}} + \gamma \ln \tau + \sigma \xi_i + \varepsilon_{it_i(\tau)}, \end{aligned}$$

and

$$\varepsilon_{it_i(\tau)} = \rho \varepsilon_{i,t_i(\tau-1)} + v_{it_i(\tau)},$$

where ξ_i, $v_{it_i(\tau)}$ are i.i.d.$N(0, 1)$. The coefficients λ_1' and λ_2' capture the influence of own successful limit pricing and peers' successful limit pricing on M's current choice respectively.

Based on the GHK simulator generated with 100 random draws, the SMLEs of the Polya model with samples from the experimental sessions with E_Hs and with E_Ls are reported in Table 1.16 and Table 1.17 respectively. In those tables, we also include the SMLEs of the original specification from Table 1.14 and Table 1.15 for ease of comparison.

The SMLEs of λ_1' are positively significant in games with E_L, indicating that, in the case where strategic play requires innovation, subjects are more confident in their decision to play strategically when limit pricing generates higher payoffs in the previous rounds.

However, the SMLEs of λ_1' are insignificant in games with E_H, which is consistent with the insignificant estimates of β in the original specification. The SMLEs of λ_2' are insignificant, which is consistent with the insignificant estimates of ω in the original specification, indicating subjects tend to ignore previous entries on their peers when making current decisions.

1.6.3 Monte Carlo and Empirical Results

Payoff Tables (Type of Entrants)	Prior Experience	Number of Sessions GC^a	MC
1.1 & 1.2 (High cost type E_Hs)	None or same game	7	9 (2)[b]
1.1 & 1.3 or 1.4 & 1.5 (Low cost type E_Ls)	None or same game	15 (9)	12 (7)
	Game with high cost Es[c]	5	7 (2)

[a] GC: generic context; MC: meaningful context

[b] number of inexperienced-subject sessions (number of experienced-subject sessions)

[c] crossover after the 1st 12-period cycle

Table 1.6: Experimental Treatments

T	m	$\beta = 1$	$\lambda_1 = 0.2$	$\lambda_2 = 0.4$	$\sigma = \sqrt{0.5}$	$\rho = 0.4$
Bias unadjusted SMLE						
8	15	0.969 (.067) [.073]	.246 (.132) [.139]	.349 (.141) [.150]	.628 (.136) [.157]	.379 (.123) [.125]
15	15	0.965 (.052) [.063]	.296 (.105) [.142]	.318 (.103) [.132]	.672 (.077) [.084]	.323 (.079) [.111]
30	15	0.957 (.035) [.056]	.345 (.071) [.161]	.292 (.068) [.127]	.686 (.058) [.061]	.286 (.045) [.123]
8	50	0.991 (.066) [.067]	.207 (.124) [.123]	.387 (.137) [.137]	.675 (.111) [.115]	.398 (.108) [.108]
15	50	0.988 (.052) [.053]	.236 (.091) [.097]	.371 (.102) [.106]	.692 (.071) [.073]	.369 (.061) [.069]
30	50	0.982 (.034) [.039]	.264 (.068) [.093]	.354 (.066) [.080]	.698 (.053) [.053]	.349 (.042) [.066]
8	100	0.996 (.065) [.065]	.203 (.124) [.124]	.393 (.135) [.135]	.684 (.110) [.112]	.397 (.104) [.104]
15	100	0.994 (.052) [.052]	.219 (.088) [.090]	.384 (.098) [.099]	.697 (.070) [.070]	.382 (.059) [.061]
30	100	0.990 (.035) [.037]	.235 (.065) [.074]	.370 (.065) [.072]	.702 (.052) [.052]	.372 (.040) [.049]
Bias adjusted SMLE						
8	15	0.991 (.069) [.069]	.217 (.131) [.132]	.380 (.143) [.144]	.681 (.135) [.138]	.385 (.127) [.127]
15	15	0.982 (.053) [.056]	.259 (.107) [.122]	.352 (.106) [.116]	.699 (.078) [.078]	.347 (.082) [.098]
30	15	0.970 (.035) [.046]	.307 (.072) [.129]	.326 (.069) [.101]	.704 (.060) [.060]	.314 (.046) [.097]
8	50	1.000 (.067) [.067]	.194 (.123) [.123]	.400 (.138) [.137]	.692 (.107) [.108]	.403 (.109) [.109]
15	50	0.996 (.052) [.052]	.214 (.092) [.092]	.390 (.103) [.104]	.701 (.072) [.072]	.386 (.062) [.063]
30	50	0.991 (.035) [.036]	.232 (.068) [.075]	.377 (.067) [.071]	.708 (.054) [.054]	.374 (.043) [.050]
8	100	1.001 (.066) [.066]	.195 (.124) [.124]	.400 (.135) [.135]	.692 (.109) [.109]	.401 (.104) [.104]
15	100	0.999 (.052) [.052]	.205 (.088) [.088]	.395 (.098) [.098]	.701 (.070) [.070]	.394 (.059) [.060]
30	100	0.996 (.036) [.036]	.212 (.066) [.067]	.386 (.066) [.067]	.708 (.053) [.053]	.391 (.040) [.041]

The values are given as Means (SDs) [RMSEs]

Table 1.7: Markov model with unobserved heterogeneity and serially correlated disturbances (sample size: G=50, n=4)

T	m	$\beta = 1$	$\lambda_1 = 0.2$	$\lambda_2 = 0.4$	$\sigma = \sqrt{0.5}$	$\rho = 0.4$
$G = 25, n = 8$						
8	15	0.961 (.067) [.078]	.261 (.157) [.168]	.321 (.180) [.197]	.604 (.174) [.202]	.372 (.140) [.142]
8	50	0.986 (.066) [.067]	.215 (.141) [.142]	.374 (.168) [.169]	.665 (.126) [.133]	.392 (.119) [.119]
8	100	0.993 (.063) [.063]	.195 (.125) [.124]	.388 (.166) [.165]	.675 (.100) [.105]	.398 (.106) [.106]
$G = 10, n = 20$						
8	15	0.958 (.069) [.080]	.285 (.182) [.200]	.288 (.218) [.245]	.605 (.155) [.185]	.356 (.154) [.160]
8	50	0.983 (.068) [.070]	.227 (.160) [.162]	.357 (.203) [.207]	.663 (.119) [.126]	.384 (.131) [.131]
8	100	0.989 (.067) [.068]	.218 (.160) [.161]	.370 (.199) [.200]	.676 (.107) [.111]	.386 (.126) [.127]
$G = 4, n = 50$						
8	15	0.957 (.069) [.081]	.300 (.193) [.217]	.271 (.238) [.270]	.606 (.150) [.181]	.345 (.158) [.167]
8	50	0.985 (.068) [.070]	.231 (.161) [.164]	.354 (.209) [.213]	.666 (.113) [.120]	.381 (.130) [.131]
8	100	0.990 (.068) [.068]	.221 (.164) [.165]	.368 (.208) [.210]	.675 (.107) [.112]	.386 (.127) [.127]
$G = 1, n = 200$						
8	15	0.958 (.071) [.082]	.302 (.200) [.224]	.268 (.252) [.284]	.606 (.152) [.182]	.345 (.163) [.172]
8	50	0.985 (.070) [.071]	.234 (.171) [.174]	.352 (.221) [.226]	.665 (.116) [.123]	.380 (.136) [.137]
8	100	0.991 (.069) [.070]	.220 (.170) [.171]	.370 (.217) [.219]	.674 (.110) [.114]	.388 (.130) [.131]

Table 1.8: Markov model with unobserved heterogeneity and serially correlated disturbances (alternative group sizes)

T	m	$\beta = 1$	$\lambda_1 = 0.2$	$\lambda_2 = 0.4$	$\sigma = \sqrt{0.5}$	$\rho = 0.4$
8	15	0.939 (.065)	0.478 (.116)	-	0.617 (.118)	0.236 (.121)
15	15	0.941 (.054)	0.493 (.097)	-	0.643 (.074)	0.214 (.080)
8	50	0.961 (.066)	0.469 (.111)	-	0.662 (.090)	0.228 (.109)
15	50	0.962 (.054)	0.459 (.097)	-	0.657 (.068)	0.241 (.075)
8	100	0.966 (.064)	0.466 (.114)	-	0.665 (.096)	0.229 (.110)
15	100	0.967 (.054)	0.444 (.100)	-	0.661 (.068)	0.254 (.079)

Table 1.9: Model Misspecification (sample size: G=50, n=4)

T	m	$\beta = 1$	$\lambda_1 = 0.2$	$\lambda_2 = 0.4$	$\sigma = \sqrt{0.5}$	$\rho = 0.4$
Random components						
8	15	0.935 (.066)	0.592 (.090)	0.091 (.122)	0.685 (.085)	-
15	15	0.919 (.051)	0.639 (.058)	0.075 (.090)	0.652 (.069)	-
8	50	0.961 (.067)	0.567 (.091)	0.123 (.120)	0.713 (.081)	-
15	50	0.939 (.051)	0.626 (.057)	0.101 (.084)	0.655 (.063)	-
8	100	0.965 (.068)	0.561 (.090)	0.132 (.118)	0.717 (.079)	-
15	100	0.943 (.052)	0.623 (.057)	0.105 (.085)	0.654 (.058)	-
AR(1) correlated errors						
8	15	0.932 (.061)	0.103 (.103)	0.434 (.124)	-	0.624 (.055)
15	15	0.940 (.051)	0.102 (.081)	0.436 (.095)	-	0.612 (.043)
8	50	0.945 (.060)	0.072 (.100)	0.462 (.122)	-	0.649 (.050)
15	50	0.962 (.052)	0.056 (.078)	0.474 (.096)	-	0.649 (.039)
8	100	0.948 (.061)	0.065 (.100)	0.468 (.122)	-	0.655 (.049)
15	100	0.968 (.051)	0.044 (.076)	0.484 (.096)	-	0.658 (.038)

Table 1.10: Error Misspecification (sample size: G=50, n=4)

T	m	$\beta = 1$	$\lambda_1 = 0.2$	$\lambda_2 = 0.4$	$\sigma = \sqrt{0.5}$	$\rho = 0.4$
8	15	0.981 (.071) [.073]	.218 (.085) [.087]	.337 (.249) [.256]	.647 (.177) [.187]	.399 (.102) [.101]
15	15	0.977 (.056) [.060]	.228 (.050) [.057]	.301 (.179) [.204]	.679 (.088) [.092]	.366 (.057) [.066]
30	15	0.964 (.036) [.051]	.255 (.038) [.067]	.201 (.134) [.240]	.676 (.067) [.074]	.339 (.033) [.069]
8	50	0.995 (.070) [.070]	.200 (.083) [.082]	.394 (.245) [.245]	.690 (.134) [.135]	.399 (.083) [.083]
15	50	0.995 (.055) [.055]	.204 (.051) [.051]	.389 (.181) [.181]	.697 (.079) [.080]	.390 (.054) [.055]
30	50	0.988 (.037) [.038]	.219 (.040) [.044]	.329 (.132) [.149]	.699 (.062) [.062]	.381 (.037) [.041]
8	100	0.997 (.068) [.068]	.198 (.082) [.082]	.400 (.246) [.245]	.691 (.118) [.119]	.402 (.083) [.083]
15	100	0.997 (.056) [.056]	.201 (.052) [.052]	.398 (.185) [.185]	.696 (.081) [.081]	.393 (.053) [.054]
30	100	0.994 (.037) [.037]	.211 (.042) [.043]	.358 (.133) [.139]	.703 (.064) [.064]	.391 (.036) [.037]

Table 1.11: Polya model with unobserved heterogeneity and serially correlated disturbances (sample size: G=50, n=4)

T	m	$\beta = 1$	$\lambda_1 = 0.2$	$\lambda_2 = 0.4$	$\sigma = \sqrt{0.5}$	$\rho = 0.4$
8	15	0.978 (.069)	0.429 (.198)	0.289 (.176)	0.710 (.142)	0.250 (.179)
15	15	0.968 (.057)	0.515 (.149)	0.304 (.128)	0.733 (.080)	0.196 (.116)
8	50	1.004 (.069)	0.408 (.188)	0.321 (.170)	0.767 (.099)	0.242 (.161)
15	50	0.993 (.057)	0.436 (.157)	0.369 (.133)	0.751 (.077)	0.252 (.120)
8	100	1.009 (.068)	0.392 (.192)	0.334 (.168)	0.770 (.096)	0.253 (.166)
15	100	0.998 (.058)	0.416 (.169)	0.382 (.137)	0.754 (.079)	0.267 (.131)

Table 1.12: Model Misspecification (true model: Polya; estimated model: Markov; sample size: G=50, n=4)

	with High-Cost Type Entrants*		with Low-Cost Type Entrants**	
		w/ interactions		w/ interactions
entry rate differential (β)	0.109 (0.085)	0.179 (0.176)	0.318^c (0.069)	0.383^c (0.089)
entry rate differential $\times NX$	-	−0.151 (0.187)	-	−0.110 (0.101)
weight on others' experience (ω)	0.003 (0.254)	0.523 (2.536)	0.028 (0.061)	0.022 (0.062)
lagged choice (λ_1)	0.730^c (0.184)	0.725^c (0.240)	1.730^c (0.094)	1.961^c (0.113)
lagged choice $\times NX$	-	0.153 (0.205)	-	-0.417^c (0.119)
peer group effects (λ_2)	0.023 (0.125)	−0.109 (0.286)	0.541^c (0.106)	0.516^c (0.131)
peer group effects $\times NX$	-	0.146 (0.320)	-	0.011 (0.214)
% of M_H choosing 3,4	-	-	0.019 (0.089)	−0.053 (0.131)
% of M_H choosing 3,4 $\times NX$	-	-	-	0.171 (0.179)
experience within a session (γ)	0.256^c (0.083)	−0.019 (0.188)	0.099^b (0.045)	0.058 (0.067)
experience within a session $\times NX$	-	0.304^a (0.176)	-	0.039 (0.077)
constant (α)	-0.545^c (0.162)	-0.619^c (0.160)	-1.871^c (0.101)	-1.838^c (0.108)
random effects (σ)	1.164^c (0.103)	1.118^c (0.098)	0.888^c (0.062)	0.867^c (0.063)
serial correlation (ρ)	−0.080 (0.103)	−0.148 (0.100)	-0.244^c (0.048)	-0.236^c (0.052)
Dummies:				
w/ experience of the same game	0.702^c (0.127)	1.026^c (0.224)	0.827^c (0.073)	0.794^c (0.118)
meaningful context	−0.079 (0.168)	−0.078 (0.164)	−0.035 (0.093)	−0.035 (0.092)
crossovers from E_H to E_L games	-	-	0.640^c (0.108)	0.611^c (0.141)
Log Likelihood	−1089.88	−1086.60	−2106.82	−2100.22

NX is a dummy variable for experimental sessions employing subjects with no experience of the same or related games

* 266 subjects, 4576 observations; ** 568 subjects, 11536 observations

[a] significantly different from 0 at the 10 percent level

[b] significantly different from 0 at the 5 percent level

[c] significantly different from 0 at the 1 percent level

standard errors in parentheses

Table 1.13: SMLEs for the Markov Model

	Model w/ peer group effects		Model w/o
		w/ interactions	peer group effects
cumulative entry rate differential (β)	0.069 (0.139)	0.279 (0.218)	0.097 (0.138)
cumulative entry rate differential $\times NX$	-	-0.271 (0.264)	-
weight on others' experience (ω)	0.552 (6.227)	0.563 (3.865)	0.528 (4.216)
lagged choices (λ_1)	0.510^c (0.159)	0.579^c (0.211)	0.490^c (0.161)
lagged choices $\times NX$	-	0.001 (0.154)	-
depreciation factor (δ_1)	0.594^c (0.158)	0.528^c (0.173)	0.601^c (0.161)
peer group effects (λ_2)	0.441^a (0.250)	-0.254 (0.526)	-
peer group effects $\times NX$	-	1.082^a (0.620)	-
depreciation factor (δ_2)	0.932^c (0.338)	0.992^c (0.210)	-
experience within a session (γ)	0.069 (0.113)	-0.161 (0.229)	0.128 (0.110)
experience within a session $\times NX$	-	0.259 (0.202)	-
constant (α)	-0.542^c (0.165)	-0.739^c (0.170)	-0.419^c (0.150)
random effects (σ)	1.011^c (0.124)	0.991^c (0.125)	1.032^c (0.125)
serial correlation (ρ)	0.102 (0.106)	0.051 (0.119)	0.119 (0.107)
Dummies:			
w/ experience of the same game	0.555^c (0.149)	1.234^c (0.313)	0.638^c (0.135)
meaningful context	-0.093 (0.161)	-0.104 (0.158)	-0.083 (0.161)
Log Likelihood	-1083.41	-1077.53	-1085.18

standard errors in parentheses

Table 1.14: SMLEs for the Polya Model (Experiments with High-Cost Type Entrants)

	Model w/ peer group effects		Model w/o
		w/ interactions	peer group effects
cumulative entry rate differential (β)	0.563^c (0.133)	0.878^c (0.203)	0.885^c (0.122)
cumulative entry rate differential $\times NX$	-	-0.528^b (0.231)	-
weight on others' experience (ω)	0.098 (0.078)	0.120 (0.085)	0.219^b (0.103)
lagged choices (λ_1)	0.870^c (0.134)	1.048^c (0.146)	0.911^c (0.130)
lagged choices $\times NX$	-	-0.325^c (0.096)	-
depreciation factor (δ_1)	0.608^c (0.083)	0.605^c (0.076)	0.624^c (0.076)
peer group effects (λ_2)	1.094^c (0.194)	0.837^c (0.234)	-
peer group effects $\times NX$	-	0.580 (0.391)	-
depreciation factor (δ_2)	0.742^c (0.099)	0.769^c (0.103)	-
% of M_H choosing 3,4 (cumulative)	0.221 (0.176)	−0.052 (0.199)	0.351^b (0.173)
% of M_H choosing 3,4 $\times NX$	-	0.868^b (0.366)	-
experience within a session (γ)	−0.053 (0.055)	−0.120 (0.082)	−0.004 (0.054)
experience within a session $\times NX$	-	0.056 (0.097)	-
constant (α)	-1.754^c (0.113)	-1.886^c (0.129)	-1.731^c (0.111)
random effects (σ)	0.821^c (0.075)	0.785^c (0.077)	0.747^c (0.072)
serial correlation (ρ)	0.241^c (0.088)	0.252^c (0.087)	0.241^b (0.085)
Dummies:			
w/ experience of the same game	0.589^c (0.094)	0.794^c (0.144)	0.740^c (0.089)
meaningful context	−0.098 (0.094)	−0.110 (0.092)	−0.093 (0.087)
crossovers from E_H to E_L games	0.466^c (0.117)	0.705^c (0.165)	0.554^c (0.112)
Log Likelihood	−2068.10	−2055.90	−2088.56

standard errors in parentheses

Table 1.15: SMLEs for the Polya Model (Experiments with Low-Cost Type Entrants)

	Original Specification		Alternative Specification	
		w/ interactions		w/ interactions
cumulative entry rate differential (β)	0.069 (0.139)	0.279 (0.218)	0.173 (0.182)	0.287 (0.317)
cumulative entry rate differential $\times NX$	-	-0.271 (0.264)	-	-0.144 (0.396)
weight on others' experience (ω)	0.552 (6.227)	0.563 (3.865)	0.947 (5.152)	0.929 (5.027)
lagged choices (λ_1)	0.510^c (0.159)	0.579^c (0.211)	0.452^c (0.175)	0.543^a (0.305)
lagged choices $\times NX$	-	0.001 (0.154)	-	-0.052 (0.329)
successful limit pricing (λ_1')	-	-	0.099 (0.107)	0.057 (0.350)
successful limit pricing $\times NX$	-	-	-	0.098 (0.394)
depreciation factor (δ_1)	0.594^c (0.158)	0.528^c (0.173)	0.580^c (0.161)	0.498^c (0.179)
peer group effects (λ_2)	0.441^a (0.250)	-0.254 (0.526)	0.748^b (0.364)	-0.169 (0.872)
peer group effects $\times NX$	-	1.082^a (0.620)	-	1.383 (1.045)
positive peer group effects (λ_2')	-	-	-0.501 (0.442)	-0.118 (0.923)
positive peer group effects $\times NX$	-	-	-	-0.566 (1.188)
depreciation factor (δ_2)	0.932^c (0.338)	0.992^c (0.210)	0.905^c (0.283)	0.991^c (0.192)
experience within a session (γ)	0.069 (0.113)	-0.161 (0.229)	0.072 (0.114)	-0.150 (0.238)
experience within a session $\times NX$	-	0.259 (0.202)	-	0.268 (0.205)
constant (α)	-0.542^c (0.165)	-0.739^c (0.170)	-0.556^c (0.165)	-0.759^c (0.174)
random effects (σ)	1.011^c (0.124)	0.991^c (0.125)	1.009^c (0.126)	1.002^c (0.130)
serial correlation (ρ)	0.102 (0.106)	0.051 (0.119)	0.099 (0.107)	0.041 (0.124)
Dummies:				
w/ experience of the same game	0.555^c (0.149)	1.234^c (0.313)	0.567^c (0.146)	1.251^c (0.325)
meaningful context	-0.093 (0.161)	-0.104 (0.158)	-0.097 (0.160)	-0.106 (0.161)
Log Likelihood	-1083.41	-1077.53	-1082.48	-1076.23

standard errors in parentheses

Table 1.16: Alternative Specifications for the Polya Model (Experiments with High-Cost Type Entrants)

	Original Specification		Alternative Specification	
		w/ interactions		w/ interactions
cumulative entry rate differential (β)	0.563^c (0.133)	0.878^c (0.203)	0.368^b (0.155)	0.824^c (0.250)
cumulative entry rate differential $\times NX$	-	-0.528^b (0.231)	-	-0.797^b (0.320)
weight on others' experience (ω)	0.098 (0.078)	0.120 (0.085)	0.811 (1.515)	0.804 (1.148)
lagged choices (λ_1)	0.870^c (0.134)	1.048^c (0.146)	0.423^c (0.153)	0.541^c (0.187)
lagged choices $\times NX$	-	-0.325^c (0.096)	-	-0.175 (0.164)
successful limit pricing (λ_1')	-	-	0.680^c (0.111)	0.785^c (0.164)
successful limit pricing $\times NX$	-	-	-	-0.207 (0.210)
depreciation factor (δ_1)	0.608^c (0.083)	0.605^c (0.076)	0.603^c (0.085)	0.593^c (0.078)
peer group effects (λ_2)	1.094^c (0.194)	0.837^c (0.234)	0.992^c (0.360)	1.238^b (0.487)
peer group effects $\times NX$	-	0.580 (0.391)	-	-0.623 (0.854)
positive peer group effects (λ_2')	-	-	0.206 (0.418)	-0.464 (0.568)
positive peer group effects $\times NX$	-	-	-	1.649 (1.056)
depreciation factor (δ_2)	0.742^c (0.099)	0.769^c (0.103)	0.746^c (0.095)	0.793^c (0.093)
% of M_H choosing 3,4 (cumulative)	0.221 (0.176)	-0.052 (0.199)	0.213 (0.177)	-0.106 (0.206)
% of M_H choosing 3,4 $\times NX$	-	0.868^b (0.366)	-	0.947^b (0.375)
experience within a session (γ)	-0.053 (0.055)	-0.120 (0.082)	-0.046 (0.055)	-0.122 (0.083)
experience within a session $\times NX$	-	0.056 (0.097)	-	0.066 (0.097)
constant (α)	-1.754^c (0.113)	-1.886^c (0.129)	-1.733^c (0.115)	-1.851^c (0.130)
random effects (σ)	0.821^c (0.075)	0.785^c (0.077)	0.830^c (0.076)	0.792^c (0.078)
serial correlation (ρ)	0.241^c (0.088)	0.252^c (0.087)	0.221^b (0.094)	0.223^b (0.092)
Dummies:				
w/ experience of the same game	0.589^c (0.094)	0.794^c (0.144)	0.597^c (0.095)	0.771^c (0.147)
meaningful context	-0.098 (0.094)	-0.110 (0.092)	-0.062 (0.095)	-0.069 (0.093)
crossovers from E_H to E_L games	0.466^c (0.117)	0.705^c (0.165)	0.482^c (0.118)	0.683^c (0.167)
Log Likelihood	-2068.10	-2055.90	-2047.53	-2034.57

standard errors in parentheses

Table 1.17: Alternative Specifications for the Polya Model (Experiments with Low-Cost Type Entrants)

CHAPTER 2

NONPARAMETRIC ESTIMATION OF LARGE AUCTIONS WITH RISK AVERSE BIDDERS

2.1 Introduction

In this chapter, we explore the robustness of Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimation procedure in first-price, sealed-bid auctions with a large number of risk averse bidders.

The seminal work by Guerre, Perrigne and Vuong (2000) has shown that the underlying distribution of bidders' values is nonparametrically identified from the observations of submitted bids in first-price, independent private value (FP-IPV) auctions with risk neutral bidders. Based on the equilibrium bidding behavior, they propose a two-step kernel-based estimator for the latent density of bidders' private values wherein the unobserved private values are estimated in the first step. The proposed two-step estimator is optimal in terms of the uniform convergence rate. As the private values are estimated from submitted bids, the best uniform convergence rate of this "indirect estimation" problem (Groeneboom, 1996) is slower than the best uniform convergence rate given by Stone (1982) when the private values are observable. However, when bidders are potentially risk averse, Campo et al. (2006) have shown that the distribution of bidders' private values and bidders' utility functions in FP-IPV auctions cannot be nonparametrically identified from observed bids. To estimate the latent density of bidders' private values, it is necessary to specify the utility function parametrically. They propose a multi-step semiparametric estimation procedure wherein the utility function is recovered parametrically in the initial steps. In deriving asymptotic properties, both works assume that the number of bidders n in each auction is fixed and the number of observed auctions L approaches infinity.

On the other hand, as n goes to infinity, it has been shown that the discrepancy between risk averse bidding behavior and risk neutral bidding behavior is of order $O(n^{-2})$ (Fibich, Gavious and Sela, 2004) and the discrepancy between strategic bidding behavior and perfectly competitive behavior, wherein bidders simply bid their value, is of order $O(n^{-1})$. In other words, as the size of an auction increases, the effect of risk aversion diminishes much faster than the rate at which the strategic bidding behavior degenerates to the price-taking behavior in perfect competition. Hence when the size of auction is large, Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimator based on strategic bidding behavior may possess some robust properties against potential risk aversion. In this chapter, we study the asymptotic properties of Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimator allowing both the number of bidders n and the number of auctions L to approach infinity. We show that when n increases not too slowly relative to L, the two-step nonparametric estimator of the latent density of private values is consistent and attains the best uniform convergence rate given by Stone (1982) as if bidders' private values are observable.

Allowing both n and L to diverge to infinity introduces some extra complications in the analysis. Since the unknown private values are recovered from the observations of submitted bids and the estimated bid density, the smoothness of bid density and the uniform convergence rate of its estimator are crucial in determining the convergence rate of Guerre, Perrigne and Vuong's (2000) two-step estimator. As the equilibrium bid density depends on n, the derivatives of bid density that are bounded with fixed n could be unbounded as $n \to \infty$, and there is no standard result on the best uniform convergence rate for the nonparametric estimation of a density that is shifting with sample size as the bid density is here. Furthermore, when there exists observed heterogeneity across auctions, we need to estimate the density of private values conditional on the "fixed effects" characterizing heterogeneity across auctions. However, the best uniform convergence rate of the estimator for a conditional (or joint) density with observations in such a panel structure, where private values are of order $O(nL)$ and "fixed effects" variables are of order $O(L)$, has seldom been addressed in the literature. We show that the kernel estimator for the conditional density of private values given the "fixed effects" can attain the best uniform convergence rate at which the marginal density of "fixed effects" can be estimated.[16]

We conduct a Monte Carlo experiment to study the finite sample performance of Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimator and get some interesting results. The two-step nonparametric estimator performs reasonably well in the presence of significant risk aversion when the number of bidders is six. In other words, an auction with six bidders can be considered as a large auction. In addition, the two-stage nonparametric estimation procedure sometimes outperforms the multi-step semiparametric estimation procedure when the utility function is misspecified.

This rest of the chapter is organized as follows. Section 2 presents the first-price, sealed-bid auction model with risk averse bidders and derives the asymptotic approximation of the equilibrium bidding function. Section 3 establishes the uniform consistency with the convergence rate of Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimator

[16]We assume that the marginal density of "fixed effects" is as smooth as the conditional density of the private values.

in large auctions with risk averse bidders. Section 4 specifies Monte Carlo experiments and reports the results. Section 5 briefly concludes.

2.2 Large Auctions with Risk Averse Bidders

Suppose there are a large number of potential buyers competing for a single, indivisible item. The number of potential buyers n ($n \gg 1$) is common knowledge[17]. In the first-price, sealed-bid auction under the independent private value (IPV) paradigm, the buyers simultaneously submit bids, and the highest bidder wins and pays his own bid to the seller. Buyer p's value v_p ($p = 1, \cdots, n$) for the auctioned item is his private information, while it is commonly known that the values are independently distributed on $\left[\underline{v}, \bar{v}\right] \subset \mathbb{R}^+$ according to a common distribution $F(\cdot)$, which is absolutely continuous with density $f(\cdot) > 0$. Each bidder is potentially risk averse with utility given by a common von Neumann-Morgenstern utility function $U(\cdot)$, which is twice continuously differentiable with $U'(\cdot) > 0$ and $U''(\cdot) \leq 0$. The seller is assumed to be risk neutral. Moreover, we assume each bidder's initial wealth $w > 0$ is the same and commonly known.

Suppose the equilibrium bid for the pth bidder with private value v_p in an auction with n bidders is $b_p = s_n\left(v_p\right)$. Following Maskin and Riley (2000; 2003), and Athey (2001), the unique symmetric Bayesian Nash equilibrium of the corresponding game is characterized by the following differential equation in $s_n(\cdot)$

$$s_n'\left(v_p\right) = (n-1)\frac{f\left(v_p\right)}{F\left(v_p\right)}\lambda\left(v_p - s_n\left(v_p\right)\right), \tag{2.1}$$

where $\lambda(\cdot) = \left(U(w+\cdot) - U(w)\right)/U'(w+\cdot)$. The boundary condition is given by $s_n\left(\underline{v}\right) = \underline{v}$.

[17]We assume in this paper that the reservation price is nonbinding, hence the number of potential bidders is equal to the number of actual bidders.

In general, the equilibrium strategy is intractable without specification of a functional form for $U(\cdot)$. However, analytical approximations to the equilibrium strategy $s_n(\cdot)$ can be derived. To proceed, we need some regularity assumptions on $U(\cdot)$ and $F(\cdot)$ following Campo et al. (2006) as summarized in the following definitions. Throughout we denote the support of $*$ by $S(*)$, and the rth derivative of $*$ by $*^{(r)}$ $(r \geq 0)$ with $*^{(0)} = *$.

Definition 1 *For $R \geq 1$, let $\mathcal{U}_R$ be the set of van Neumann-Morgenstern utility functions $U(\cdot)$ with initial wealth $w > 0$ such that:*

(i) $U : [0, \infty) \to [0, \infty)$*;*

(ii) $U(\cdot)$ *is continuous on* $S(U)$*, and admits up to* $R+2$ *continuous bounded derivatives on* $(0, \infty)$ *with* $U'(\cdot) > 0$ *and* $U''(\cdot) \leq 0$ *on* $(0, \infty)$.

Definition 2 *For $R \geq 1$, let $\mathcal{F}_R$ be the set of distributions $F(\cdot)$ such that:*

(i) $S(F) = \left\{v : v \in \left[\underline{v}, \bar{v}\right]\right\}$*, with* $0 \leq \underline{v} < \bar{v} < \infty$*;*

(ii) $f(v) \geq c_f > 0$ *for* $v \in S(F)$*;*

(iii) $F(\cdot)$ *admits up to* $R+1$ *continuous bounded derivatives on* $S(F)$.

Except for the additional assumption that $w > 0$, $\mathcal{U}_R$ and $\mathcal{F}_R$ are defined similar to Campo et al. (2006) and thus have similar implications. Definition 1 requires that $\lambda(x)$ admits $R+1$ continuous bounded derivatives on $[0, \infty)$, and Definition 2 specifies the smoothness of $F(\cdot)$ and requires the corresponding density $f(v)$ to be bounded away from zero on $S(F)$. These regularity assumptions are quite weak. The additional assumption on initial wealth is to guarantee proper behavior of the utility function at the initial wealth level. To relax this assumption so that $w \geq 0$, Definition 1(ii) needs to be replaced by the stronger assumption that "$U(\cdot)$ is continuous and admits up to $R+2$ continuous bounded derivatives on $S(U)$ with $U'(\cdot) > 0$ and $U''(\cdot) \leq 0$ on $S(U)$". The assumption on initial wealth is necessary for analytical approximation of the equilibrium bidding behavior in large auctions. Furthermore, we assume that the private values and the number of bidders are independent so that $f(v|n) = f(v)$. As noted by Guerre, Perrigne and Vuong (2000), this assumption is justified by the economic model. Otherwise, endogenous entry to the auction should be considered, which is outside the scope of this chapter.

It is well known that, as the number of bidders n approaches infinity, the equilibrium bid approaches the bidder's private value under quite general conditions. Applying repeated integration by parts and the Laplace approximation (Copson, 1965) to the integral form of the differential equation (2.1),

$$\lambda\left(v_p - s_n\left(v_p\right)\right) = \frac{1}{F^{n-1}\left(v_p\right)} \int_{\underline{v}}^{v_p} F^{n-1}(u)\, d\left(s_n(u) + \lambda\left(u - s_n(u)\right)\right),$$

we can derive the leading order deviation of the equilibrium bid from the private value. This is formally stated in the following proposition.[18] Another contribution of Proposition 2.1 is to characterize the implied smoothness of the equilibrium bidding function as $n \to \infty$, which is used to derive the uniform convergence rate of the two-step nonparametric estimator in the next section. Let $\varsigma_n(v) = v - s_n(v)$ be the consumer surplus conditional on winning.

Proposition 2.1 *In a first-price IPV auction with n $(n \gg 1)$ bidders, if $F(\cdot) \in \mathcal{F}_R$ and $U(\cdot) \in \mathcal{U}_R$ for $R \geq 1$, the equilibrium bid in the symmetric Bayesian Nash equilibrium is*

[18] Fibich, Gavious and Sela (2004) have shown (2.2) based on the unproved claim that $s_n'(v) = 1 + O\left(n^{-1}\right)$, which, in general, is not directly implied by the (uniform) convergence of $s_n(v)$. Here we take a different approach to derive the leading order deviation of $s_n(v)$ from v. The approach presented here is more rigorous as $s_n'(v) = 1 + O\left(n^{-1}\right)$ is proved instead of assumed and more general as it allows us to express $s_n(v)$ as its asymptotic expansion with precision of $O\left(n^{-(R+1)}\right)$ instead of just the leading order deviation.

given by

$$s_n(v) = v - \frac{1}{n}\frac{F(v)}{f(v)} + O(n^{-2}).^{19} \tag{2.2}$$

Furthermore, we have $\varsigma_n^{(r)}(v) = O(n^{-1})$ *for* $1 \leq r \leq R$.

Let $G_n(\cdot)$ denote the distribution of equilibrium bids. We have $G_n(b) = F(v)$ with support $S(G_n) = \{b : b \in [\underline{v}, s_n(\bar{v})]\}$ and density $g_n(b) = f(v)/s_n'(v) = f(v) + O(n^{-1})$ by Proposition 2.1, where $v = s_n^{-1}(b)$. It follows from (2.2) that

$$v = s_n^{-1}(b) = b + \frac{1}{n}\frac{G_n(b)}{g_n(b)} + O(n^{-2}), \tag{2.3}$$

which represents the unobserved private value as a function of the observed bid with an error of order $O(n^{-2})$. This allows us to employ Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimation procedure to recover the underlying distribution of risk averse bidders' private values with satisfactory precision when n is large.

2.3 Nonparametric Estimation and Robustness

2.3.1 Estimation Procedure and Asymptotic Properties

To clarify conceptual issues, we first consider L homogeneous auctions with n bidders in each auction. In order to implement Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimation procedure, we first need to estimate the distribution of equilibrium bids $G_n(\cdot)$, which depends on the number of bidders. Hence it is important to study the implied smoothness of $G_n(\cdot)$ as $n \to \infty$. The following proposition summarizes the properties of $G_n(\cdot)$ relevant to the asymptotic properties of the nonparametric estimator.

Proposition 2.2 *If* $F(\cdot) \in \mathcal{F}_R$ *and* $U(\cdot) \in \mathcal{U}_R$ *for* $R \geq 1$*, the distribution* $G_n(\cdot)$ *satisfies:*

[19] Throughout $f_n(x) = g_n(x) + O(n^p)$ or $f_n(x) = g_n(x) + o(n^p)$ means $\sup_x |f_n(x) - g_n(x)| = O(n^p)$ or $\sup_x |f_n(x) - g_n(x)| = o(n^p)$ respectively, for a pair of functions $f_n(\cdot)$ and $g_n(\cdot)$ and a constant p.

(i) its support is $S(G_n) = \{b : b \in [\underline{v}, s_n(\bar{v})]\}$*, with* $\inf_{n\in\{2,3,\cdots\}} (s_n(\bar{v}) - \underline{v}) > 0$*. Moreover,* $S(G_n) \subset S(G_{n+1})$ *for all* $n \in \{2, 3, \cdots\}$*, and* $\lim_{n\to\infty} S(G_n) = S(F)$*;*

(ii) for $b \in S(G_n)$*,* $g_n(b) \geq c_g > 0$ *as* $n \to \infty$*;*

(iii) if C *is a closed subset of the interior of* $S(G_\infty)$*, then* $g_n(\cdot)$ *is bounded and admits up to* R *continuous bounded derivatives on* C *as* $n \to \infty$*.*

Contrary to its counterpart with fixed n derived in Campo et al. (2006) where $g_n(\cdot)$ is smoother than $f(\cdot)$ with $R + 1$ continuous bounded derivatives, Proposition 2.2(iii) shows that as $n \to \infty$, the uniform boundedness of the $(R + 1)$th derivative of $g_n(\cdot)$ cannot be implied from the existing assumptions on the structure $[U, F]$.

Following Guerre, Perrigne and Vuong (2000), with the observations $\{B_{pl}; p = 1, \cdots, n, l = 1, \cdots, L\}$, the bid distribution $G_n(\cdot)$ and density $g_n(\cdot)$ can be nonparametrically estimated respectively by the empirical distribution and the kernel density estimator of the form

$$\tilde{G}_n(b) = \frac{1}{nL} \sum_{l=1}^{L} \sum_{p=1}^{n} \mathbf{1}(B_{pl} \leq b), \tag{2.4}$$

$$\tilde{g}_n(b) = \frac{1}{nLh_R} \sum_{l=1}^{L} \sum_{p=1}^{n} K_R\left(\frac{B_{pl} - b}{h_R}\right), \tag{2.5}$$

where h_R is a bandwidth such that $h_R = \lambda(\log(nL)/nL)^{1/(2R+1)}$ with λ being a strictly positive constant, and $K_R(\cdot)$ is a symmetric kernel of order R with a compact support and twice continuous bounded derivatives satisfying $\int K_R(b)\,db = 1$ and $\int K_R^2(b)\,db < \infty$. Note that classical asymptotic results regarding the empirical distribution and kernel estimator based on the i.i.d. assumption of observations do not apply to the current model as $n \to \infty$, because the equilibrium bid and hence its distribution depend on the number of bidders n. The uniform consistency of $\tilde{G}_n$ and $\tilde{g}_n$ with the convergence rate based on a triangular array of random variables that are independent but not identically distributed as we have here is derived in the appendix.

Because the kernel estimator is asymptotically biased at the boundaries of the support, Guerre, Perrigne and Vuong (2000) suggest trimming the observations B_{pl} that are too close to the boundaries of $S(G_n)$. However, in our case, as n increases, $S(G_n)$ is expanding such that $\lim_{n\to\infty} S(G_n) = S(F)$. Hence the kernel estimator is asymptotically biased at the boundaries of the support of $F(\cdot)$. Denote the length of the support of $K_R(\cdot)$ by ρ. For $b = \bar{v} - \lambda\rho h_R/2$ with $\lambda \in [0, 1)$, it follows that $E[\tilde{g}_n(\bar{v} - \lambda\rho h_R/2)] = \int_{(\underline{b}_n-\bar{v})/h_R+\lambda\rho/2}^{(\bar{b}_n-\bar{v})/h_R+\lambda\rho/2} K_R(u)\, g_n(\bar{v} - \lambda\rho h_R/2 + h_R u)du \to g_n(\bar{v} - \lambda\rho h_R/2) \int_{-\infty}^{\lambda\rho/2} K_R(u)\, du$ as n and L approach infinity. As $\int_{-\infty}^{\lambda\rho/2} K_R(u)\, du \neq 1$, the density estimator is asymptotically biased for $b \in (\bar{v} - \rho h_R/2, \bar{v}]$ and similarly for $b \in [\underline{v}, \underline{v} + \rho h_R/2)$. Let $B_{\min}$ and $B_{\max}$ be the minimum and maximum of the nL observed bids. The trimmed pseudo-private value is defined as

$$\hat{V}_{pl} = \begin{cases} B_{pl} + \tilde{G}_n\left(B_{pl}\right) / (n-1)\, \tilde{g}_n\left(B_{pl}\right), \\ \quad \text{if } B_{pl} \in \left[B_{\min} + \rho h_R/2, B_{\max} - \rho h_R/2\right], \\ \infty \ \text{ otherwise,} \end{cases} \tag{2.6}$$

for $p = 1, \cdots, n$ and $l = 1, \cdots, L$. The following proposition gives the rate at which the trimmed pseudo-private value converges to the true value on a closed inner subset of its support. The result will be used to derive the uniform convergence rate of the two-step estimator. Let $r = (nL/\log(nL))^{R/(2R+1)}$.

Proposition 2.3 *Suppose $F(\cdot) \in \mathcal{F}_R$ and $U(\cdot) \in \mathcal{U}_R$ for $R \geq 1$. Then, for any closed inner subset $C(V)$ of $S(F)$, we have almost surely*

$$\sup_{pl} \mathbf{1}_{C(V)}(V_{pl}) \left|\hat{V}_{pl} - V_{pl}\right| = O\left(\max(n/r, 1)n^{-2}\right).$$

Basically, the error of pseudo-private value $\hat{V}_{pl}$ comes from two sources: estimation error from $\tilde{G}_n(\cdot)/\tilde{g}_n(\cdot)$ and approximation error from ignoring the utility structure. So the uniform convergence rate of the pseudo-private value is determined by the slower convergence rate of these two types of errors. Suppose $R = 1$, then $n/r \approx n^2/L$ by ignoring the relatively small $\log(nL)$ term. So if n increases much slower than L such that $n^2/L \to 0$, then the approximation error dominates. The estimation error dominates otherwise.

With the trimmed pseudo-private values, the private value density $f(\cdot)$ can be estimated by the kernel density estimator

$$\hat{f}(v) = \frac{1}{nLh_R} \sum_{l=1}^{L} \sum_{p=1}^{n} K_R\left(\frac{\hat{V}_{pl} - v}{h_R}\right). \tag{2.7}$$

The following result establishes the uniform consistency of Guerre, Perrigne and Vuong's (2000) two-step estimator with its rate of convergence in homogenous auctions with risk averse bidders.

Proposition 2.4 *Suppose $F(\cdot) \in \mathcal{F}_R$ and $U(\cdot) \in \mathcal{U}_R$ for $R \geq 1$. Then, for any closed inner subset $C(V)$ of $S(F)$,*

(i) if $L \to \infty$ and $(nh_R)^{-1} \to 0$, $(r/n)(nh_R)^{-1} \to 0$ as $n \to \infty$, we have almost surely

$$\sup_{v \in C(V)} \left| \hat{f}(v) - f(v) \right| = O(r^{-1});$$

(ii) if $L \to \infty$ and $(nh_R)^{-1} \to 0$, $(r/n)(nh_R)^{-1} \to \infty$ as $n \to \infty$, we have almost surely

$$\sup_{v \in C(V)} \left| \hat{f}(v) - f(v) \right| = O(n^2 h_R)^{-1};$$

(iii) if $L \to \infty$ and $(nh_R)^{-1} \to \infty$ as $n \to \infty$, we have almost surely

$$\sup_{v \in C(V)} \left| \hat{f}(v) - f(v) \right| = O(n^4 h_R^3)^{-1}.$$

Proposition 2.4(iii) shows that, when n does not diverge fast enough relative to L, Guerre, Perrigne and Vuong's (2000) two-step estimator may not be consistent in the presence of risk aversion given our choice of $K_R(\cdot)$ and h_R because of the overwhelming approximation error. A sufficient condition for the two-step estimator to be consistent is that $(nh_R)^{-1} \to 0$, which imposes a lower bound of the divergence rate of n in terms of L. By ignoring the relatively small $\log(nL)$ term, we have $(nh_R)^{-1} \approx L/n^{2R}$. Hence the constraint on the divergence rate of n is quite weak, especially for a smooth private value density (with larger R). On the other hand, when n goes to infinity fast enough relative to L, it is possible for the two-step nonparametric estimator to attain the uniform convergence rate $r = (nL/\log(nL))^{R/(2R+1)}$, which is the best uniform convergence rate when private values are observable. The intuition for the result is as follows. As $f(v) = g_n(s_n(v))\, s_n'(v)$, to estimate the private value density, $g_n(\cdot), s_n(\cdot)$ and $s_n'(\cdot)$ need to be estimated. When n is fixed, $s_n'(\cdot)$ is the hardest to estimate as it requires estimating $g_n'(\cdot)$. In fact, the best uniform convergence rate for estimating $s_n'(\cdot)$ determines the best rate for estimating $f(\cdot)$. However, when $n \to \infty$, it follows from Proposition 2.1 that $s_n'(v) = 1 + O(n^{-1})$. So if n diverges fast enough, $f(\cdot)$ can be estimated at the same best rate as $g_n(\cdot)$, which is r.

As in Guerre, Perrigne and Vuong (2000), asymptotic normality of the two-step estimator is not derived. This is because the first and second order terms in the expansion of $\hat{f}(v) - f(v)$ may be close (see the proof of Proposition 2.4), so the classical asymptotic normality result that relies only on the leading order term in the Taylor expansion may be imprecise. Guerre, Perrigne and Vuong (2000) suggest circumventing this drawback by establishing an exponential-type inequality, and that approach also applies to the current model. Interested readers may refer to that paper for more details.

2.3.2 Auctions with Heterogeneity

Now we can extend the above analysis to a more realistic model allowing heterogeneity. Heterogeneity across auctions is characterized by a vector of observed variables X_l and the number of bidders nI_l ($l = 1, \cdots, L$), where the I_l's are strictly positive constants.[20] We assume n, but not I_l, approaches infinity for asymptotic properties. Let $\mathcal{I}$ be the set of possible values for I_l. Following Guerre, Perrigne and Vuong (2000), the latent joint distribution of $\left(V_{pl}, X_l, I_l\right)$ for $p = 1, \cdots, nI_l$ and $l = 1, \cdots, L$ satisfies the following regularity assumptions:

Assumption A1

(i) *The $(d+1)$-dimensional vectors (X_l, I_l), $l = 1, \cdots, L$, are independently and identically distributed as $F_m(\cdot, \cdot)$ with density $f_m(\cdot, \cdot)$.*

(ii) *For each l, the variables V_{pl}, $p = 1, \cdots, nI_l$, are independently and identically distributed conditionally upon X_l as $F(\cdot|\cdot)$ with density $f(\cdot|\cdot)$.*

Assumption A2 *For $\mathcal{I}$ a bounded countable subset of $\mathbb{R}^+$ and $R \geq 1$,*

(i) $S(F) = \{(v, x) : x \in \left[\underline{x}, \bar{x}\right], v \in [\underline{v}(x), \bar{v}(x)]\}$, *with* $\underline{x} < \bar{x}$;

(ii) *for* $(v, x) \in S(F)$, $f(v|x) \geq c_f > 0$, *and, for* $(x, i) \in S(F_m)$, $f_m(x, i) \geq c_f > 0$;

(iii) *for each* $i \in \mathcal{I}$, $f(\cdot|\cdot)$ *and* $f_m(\cdot, i)$ *admit up to R continuous bounded partial derivatives on* $S(F)$ *and* $S(F_m(\cdot, i))$.

[20] Empirically, we can decompose the number of bidders of the lth auction arbitrarily into $n \in \{2, 3, \cdots\}$ and $I_l \in \mathbb{R}^+$. Say, let $n = \min_l\{nI_l\}$.

As argued by Guerre, Perrigne and Vuong (2000), we can assume that $\underline{x}$ and $\bar{x}$ are known as they can be readily estimated. X is assumed to be a vector of continuous variables.[21] The economic model implies that the private values and the number of bidders are independent conditional on X so that $f(v|x, ni) = f(v|x)$. With the smoothness of $F(\cdot|\cdot)$ specified in Assumption A2, the next proposition studies the implied smoothness of bid density $g_n(\cdot|\cdot,\cdot)$.

Proposition 2.5 *Suppose $U(\cdot) \in \mathcal{U}_R$ for $R \geq 1$. Given A1 and A2, the conditional distribution $G_n(\cdot|\cdot,\cdot)$ satisfies:*

(i) its support $S(G_n)$ is such that $S(G_n(\cdot|\cdot, i)) = \{(b, x) : x \in [\underline{x}, \bar{x}], b \in [\underline{b}_n(x, i), \bar{b}_n(x, i)]\}$, with $\inf(\bar{b}_n(x, i) - \underline{b}_n(x, i)) > 0$. Moreover, $\bar{b}_n(x, i) \geq \bar{b}_m(x, i)$ for $n \geq m$, $\underline{b}_n(\cdot, i) = \underline{v}(\cdot)$, and $\lim_{n\to\infty} \bar{b}_n(\cdot, i) = \bar{v}(\cdot)$;

(ii) for $(b, x, i) \in S(G_n)$, $g_n(b|x, i) \geq c_g > 0$ as $n \to \infty$;

(iii) if C is a closed subset of the interior of $S(G_\infty)$, then, for each $i \in \mathcal{I}$, $g_n(\cdot|\cdot, i)$ is bounded and admits up to R continuous bounded derivatives on C as $n \to \infty$.

Proposition 2.5 extends Proposition 2.2 by allowing possible heterogeneity across auctions and has similar implications. Specially, item (iii) characterizes the uniform boundedness of g_n's derivatives as $n \to \infty$, which is used to derive asymptotic properties of the nonparametric estimator.

[21] If some X's are discrete, the following results hold with d replaced by the number of continuous variables in X.

Following Guerre, Perrigne and Vuong (2000), using the observations $\{(B_{pl}, X_l, I_l);\ p = 1, \cdots, nI_l, l = 1, \cdots, L\}$, we can nonparametrically estimate $G_n(\cdot,\cdot,\cdot)$ and $g_n(\cdot,\cdot,\cdot)$ respectively by

$$\tilde{G}_n(b,x,i) = \frac{1}{nLh_G^d}\sum_{l=1}^{L}\frac{1}{I_l}\sum_{p=1}^{nI_l}\mathbf{1}\left(B_{pl}\le b\right)K_G\left(\frac{X_l-x}{h_G},\frac{I_l-i}{h_{GI}}\right), \tag{2.8}$$

$$\tilde{g}_n(b,x,i) = \frac{1}{nLh_g^{d+1}}\sum_{l=1}^{L}\frac{1}{I_l}\sum_{p=1}^{nI_l}K_g\left(\frac{B_{pl}-b}{h_g},\frac{X_l-x}{h_g},\frac{I_l-i}{h_{gI}}\right), \tag{2.9}$$

where h_G, h_{GI}, h_g, and h_{gI} are bandwidths and K_G and K_g are kernels with a compact support.

Similar to the case with homogeneous auctions, the asymptotic results of nonparametric estimators based on i.i.d. assumptions do not apply to $\tilde{G}_n$ and $\tilde{g}_n$ as $n\to\infty$ due to the dependence of the equilibrium bid distribution on n. We derive the uniform consistency with the convergence rate of $\tilde{G}_n$ and $\tilde{g}_n$ in the appendix. On the other hand, since the number of B_{pl} is of order $O(nL)$ while the number of observed auctions and hence (X_l, I_l) (which are analogous to fixed effects in a panel data model) are of order $O(L)$, the best uniform convergence rate for the nonparametric estimation of the joint density of (B_{pl}, X_l, I_l) as both n and L approach infinity has seldom been addressed in the literature. The following analysis sheds light on whether and to what extent $n\to\infty$ speeds up the convergence of the joint density estimator.

Since the kernel density estimator is biased at the boundaries of the support of $S(F)$ as we discussed in the case with homogenous auctions, we trim the observations that are too close to the boundary of $S(F)$. To this end, we need to estimate the unknown $S(F) = \{(v,x) : x\in\left[\underline{x},\bar{x}\right], v\in[\underline{v}(x),\bar{v}(x)]\}$. Since $\left[\underline{x},\bar{x}\right]$ is known, we only need to estimate the support $[\underline{v}(x),\bar{v}(x)]$. Let $h_\partial > 0$. Following Guerre, Perrigne and Vuong (2000), we consider the following partition of $\mathbf{R}^d$ with a generic hypercube of side h_∂:

$$\pi_{k_1,\cdots,k_d} = [k_1 h_\partial, (k_1+1)\, h_\partial) \times \cdots \times [k_d h_\partial, (k_d+1)\, h_\partial),$$

where $(k_1, \cdots, k_d)$ runs over $\mathbb{Z}^d$. The support $[\underline{v}(x), \bar{v}(x)]$ can be estimated as

$$\widehat{\bar{v}}(x) = \sup\{B_{pl}, p = 1, \cdots, nI_l, l = 1, \cdots, L; X_l \in \pi_{k_1,\cdots,k_d}\}, \tag{2.10}$$

$$\widehat{\underline{v}}(x) = \inf\{B_{pl}, p = 1, \cdots, nI_l, l = 1, \cdots, L; X_l \in \pi_{k_1,\cdots,k_d}\}, \tag{2.11}$$

where $\pi_{k_1,\cdots,k_d}$ is the hypercube containing x. And the estimator for $S(F)$ is $\hat{S}(F) \equiv \{(v, x) : x \in [\underline{x}, \bar{x}], v \in [\widehat{\bar{v}}(x), \widehat{\underline{v}}(x)]\}$.

Note that (2.3) can be rewritten as

$$V_{pl} = B_{pl} + \frac{1}{nI_l} \frac{G_n\left(B_{pl}, X_l, I_l\right)}{g_n\left(B_{pl}, X_l, I_l\right)} + O(n^{-2}),$$

where $G_n(b, x, i) = G_n(b|x, i) f_m(x, i)$. Guerre, Perrigne and Vuong's (2000) pseudo-private value is estimated by

$$\hat{V}_{pl} = B_{pl} + \frac{1}{nI_l - 1} \hat{\psi}\left(B_{pl}, X_l, I_l\right),$$

where

$$\hat{\psi}(b, x, i) \equiv \begin{cases} \tilde{G}_n(b, x, i) / (nI_l - 1)\, \tilde{g}_n(b, x, i), \\ \quad \text{if } (b, x) + S(2h_G) \subset \hat{S}(F) \text{ and} \\ \quad (b, x) + S\left(2h_g\right) \subset \hat{S}(F), \\ \infty \ \text{otherwise}, \end{cases}$$

with $S(h_G)$ and $S\left(h_g\right)$ being the supports of $\{0 \times K_G(\cdot/h_G, 0)\}$ and $K_g\left(\cdot/h_g, \cdot/h_g, 0\right)$ respectively.

In the second step of Guerre, Perrigne and Vuong's (2000) two-step estimation approach, the density $f(v|x)$ is estimated nonparametrically by $\hat{f}(v|x) = \hat{f}(v, x) / \hat{f}(x)$

using the pseudo-sample $\{(\hat{V}_{pl}, X_l), p = 1, \cdots, nI_l, l = 1, \cdots, L\}$, where

$$\hat{f}(v, x) = \frac{1}{nLh_f^{d+1}} \sum_{l=1}^{L} \frac{1}{I_l} \sum_{p=1}^{nI_l} K_f\left(\frac{\hat{V}_{pl} - v}{h_f}, \frac{X_l - x}{h_f}\right), \tag{2.12}$$

$$\hat{f}(x) = \frac{1}{Lh_X^d} \sum_{l=1}^{L} K_X\left(\frac{X_l - x}{h_X}\right), \tag{2.13}$$

h_f and h_X are bandwidths, and K_f and K_X are kernels with compact supports. The choice of kernels and bandwidths in the definition of the two-step nonparametric estimator are summarized in the following two assumptions:

Assumption A3

(i) *The kernels* $K_G(\cdot, \cdot)$, $K_g(\cdot, \cdot, \cdot)$, $K_f(\cdot, \cdot)$ *and* $K_X(\cdot)$ *are symmetric with bounded hypercube supports and twice continuous bounded (uniformly in I) derivatives with respect to their continuous arguments.*

(ii) $\int K_G(x, 0)\, dx = 1$, $\int K_g(b, x, 0)\, db dx = 1$, $\int K_f(v, x)\, dv dx = 1$, *and* $\int K_X(x)\, dx = 1$.

(iii) $K_G(x, 0)$ is of order $R + 1$, *and* $K_g(b, x, 0)$, $K_f(v, x)$ *and* $K_X(x)$ *are of order* R.

Assumption A4

(i) *As* $L \to \infty$, *the "discrete" bandwidths* h_{GI} *and* h_{gI} *vanish.*

(ii) *The "continuous" bandwidths* h_G, h_g, h_f, *and* h_X *are of the form:*

$$h_G = \lambda_G(\log L/L)^{1/(2R+d+2)}, \; h_g = \lambda_g(\log L/L)^{1/(2R+d)},$$
$$h_f = \lambda_f(\log L/L)^{1/(2R+d)}, \; h_X = \lambda_X(\log L/L)^{1/(2R+d)},$$

where the λ*'s are strictly positive constants.*

(iii) *The "boundary" bandwidth is of the form* $h_\partial = \lambda_\partial (\log L/L)^{1/(d+1)}$ *with* $\lambda_\partial > 0$*, if* $d > 0$.

It follows from Hardle (1991) that h_G and h_X given in A4(ii) are optimal bandwidths given Proposition 2.5 and A2(iii). Hence $G_n(\cdot,\cdot,\cdot)$ and $f(\cdot)$ are optimally estimated in terms of the uniform convergence rate. If n were fixed and private values were observed, the optimal bandwidth for estimating $f(\cdot,\cdot)$ would be of order $(\log L/L)^{1/(2R+d+1)}$, which is asymptotically larger than the rate for h_f given in A4(ii). Similarly, the rate for h_g given in A4(ii) is asymptotically smaller than the optimal bandwidth with fixed n. However, our choices of h_f and h_g are optimal when n approaches infinity fast enough relative to L as shown below.

The following results establish the uniform consistency of the nonparametric estimators of $S(F)$ and $f(v|x)$ in large auctions with risk averse bidders.

Proposition 2.6 *Let* $r_\partial = (L/\log L)^{1/(d+1)}$. *Given A1, A2 and A4(iii), we have almost surely*

$$\sup_{x\in[\underline{x},\bar{x}]} \left|\widehat{\bar{v}}(x) - \bar{v}(x)\right| = O(r_\partial^{-1}), \text{ and } \sup_{x\in[\underline{x},\bar{x}]} \left|\widehat{\underline{v}}(x) - \underline{v}(x)\right| = O(r_\partial^{-1}).$$

We have shown in the case with homogeneous auctions that a sufficient condition for Guerre, Perrigne and Vuong's (2000) two-step estimator to be uniformly consistent is that n goes to infinity fast enough relative to L so that $(nh_R)^{-1} \to 0$. So the next result on the uniform convergence rate focuses on the case with $(nh_f)^{-1} \to 0$. Let $r_f = (L/\log L)^{R/(2R+d)}$.

Proposition 2.7 *Suppose* $U(\cdot) \in \mathcal{U}_R$ *for* $R \geq 1$. *Given A1-A4, for any closed inner subset* $C(V)$ *of* $S(F)$, *(i) if $L \to \infty$ and $(nh_f)^{-1} \to 0$, $(r_f/n)(nh_f)^{-1} \to 0$ as $n \to \infty$, we have almost surely*

$$\sup_{v \in C(V)} \left| \hat{f}(v|x) - f(v|x) \right| = O(r_f^{-1});$$

(ii) if $L \to \infty$ and $(nh_f)^{-1} \to 0$, $(r_f/n)(nh_f)^{-1} \to \infty$ as $n \to \infty$, we have almost surely

$$\sup_{v \in C(V)} \left| \hat{f}(v|x) - f(v|x) \right| = O(n^2 h_f)^{-1}.$$

So when n approaches infinity fast enough relative to L, the two-step estimator of $f(v|x)$ can attain the best rate at which $f(x)$ can be estimated. Even though $f(v|x)$ is as smooth as $f(x)$ given A2(iii), one would expect $f(v|x)$ to be estimated with a convergence rate slower than $f(x)$ because private values are unobservable and the vector (V, X) has one more dimension than X. The counterintuitive result in Proposition 2.7 can be understood as follows. First, since unknown private values can be approximated by observed bids with precision of order $O(n^{-1})$ by Proposition 2.1, the approximation error may be trivial compared to the estimation error of the kernel estimator when n goes to infinity fast enough relative to L. Hence, the information loss from not observing V may be negligible given the conditions in Proposition 2.7(i). Second, because there are $(n-1)L$ more (pseudo) observations of V than X, the noise from estimating the extra dimension of random variables in $f(v, x)$ and hence $f(v|x)$ reduces dramatically as $n \to \infty$. We show in the appendix that, when n diverges fast enough so that $(nh_g)^{-1} \to 0$ and $(nh_f)^{-1} \to 0$, kernel estimators of $g_n(b, x, i)$ and $f(v, x)$ can attain the best rate at which $\hat{f}(x)$ uniformly converges to $f(x)$.

2.4 Monte Carlo Experiments

We conduct the Monte Carlo experiments with 1000 replications, each consisting of three sets of observations. In set 1, we consider $L = 300$ auctions, each with $n = 3$ bidders. In set 2, we consider $L = 150$ auctions, each with $n = 6$ bidders. In set 3, we consider $L = 75$ auctions, each with $n = 12$ bidders. The total number of observations of submitted bids is 900 for each set. Bidders' private values for each replication are generated from the log-normal distribution F with parameters (0, 1), truncated at 0.055 and 2.5. The true utility takes the functional form $U(x) = 1 - \exp(-\theta x)$, where $\theta = 0.8$. The equilibrium bids are computed numerically by

$$b = \frac{1}{\theta} \log \frac{\int^{v} \exp(\theta t)\, dF(t)^{n-1}}{F(v)^{n-1}}. \tag{2.14}$$

We consider four different estimation procedures for each replication. Method 1 serves as the basis for comparison. We specify the functional form of utility as the true $U(\cdot)$ and adopt the semiparametric approach proposed by Campo et al. (2006). To estimate θ, we pool the observations from all 3 sets. Let $G_n(b)$ denote the distribution of bids in auctions with n bidders, v_α denote the αth percentile of F, and b_α^n denote the αth percentile of G_n. For $n \neq m$, (2.1) gives

$$\begin{aligned} v_\alpha - b_\alpha^n &= \frac{1}{\theta} \log[\frac{\theta}{n-1} \frac{G_n\left(b_\alpha^n\right)}{g_n\left(b_\alpha^n\right)} + 1], \\ v_\alpha - b_\alpha^m &= \frac{1}{\theta} \log[\frac{\theta}{m-1} \frac{G_m\left(b_\alpha^m\right)}{g_m\left(b_\alpha^m\right)} + 1]. \end{aligned}$$

Taking difference gives

$$b_\alpha^m - b_\alpha^n = \frac{1}{\theta} \log[\frac{\theta}{n-1} \frac{G_n\left(b_\alpha^n\right)}{g_n\left(b_\alpha^n\right)} + 1] - \frac{1}{\theta} \log[\frac{\theta}{m-1} \frac{G_m\left(b_\alpha^m\right)}{g_m\left(b_\alpha^m\right)} + 1] \tag{2.15}$$

With a large number of percentiles α, we can estimate θ using the empirical analogue of (2.15) by nonlinear least squares. Given an estimate $\hat{\theta}$ of θ, we then estimate f using the two-step kernel-based estimation procedure described above for each set of observations separately.

Method 2 investigates the consequences of model misspecification by assuming the utility is CRRA with $U(x) = x^{1-\theta}$ using the semiparametric approach proposed by Campo et al. (2006). Analogous to Model 1, we identify θ through the heterogeneity of the bid distributions across auctions with different number of bidders. With CRRA utility, for $n \neq m$, (2.1) gives

$$\begin{aligned} v_\alpha - b_\alpha^n &= \frac{1-\theta}{n-1}\frac{G_n\left(b_\alpha^n\right)}{g_n\left(b_\alpha^n\right)}, \\ v_\alpha - b_\alpha^m &= \frac{1-\theta}{m-1}\frac{G_m\left(b_\alpha^m\right)}{g_m\left(b_\alpha^m\right)}. \end{aligned}$$

Taking the difference gives

$$b_\alpha^m - b_\alpha^n = (1-\theta)\left(\frac{1}{n-1}\frac{G_n\left(b_\alpha^n\right)}{g_n\left(b_\alpha^n\right)} - \frac{1}{m-1}\frac{G_m\left(b_\alpha^m\right)}{g_m\left(b_\alpha^m\right)}\right). \tag{2.16}$$

Evaluating the empirical analogue of (2.1) at a finite number of percentiles, we can recover θ using least squares. Then we estimate f nonparametrically for each set of observations separately.

Method 3 recovers f using Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimation procedure without imposing any restrictions on the functional form of $U(\cdot)$. Method 4 is a one-step nonparametric estimation method using the observed bids as the pseudo-private values to estimate f directly, based on the fact that $\lim_{n\to\infty} s_n(v) = v$. Method 4 can only be justified when the number of bidders in each auction is very large and strategic bidding behavior is overwhelmed by the price-taking behavior in perfection competition. We compare the estimates from Methods 3 and 4 to understand the gains from incorporating strategic bidding behavior in the structural estimation.

Following Guerre, Perrigne and Vuong (2000), in nonparametric estimations we choose the triweight kernel $(35/32)\left(1-u^2\right)^3 \mathbf{1}(|u| \leq 1)$ for $K_g(\cdot)$ and $K_f(\cdot)$ so that $\rho_g = \rho_f = 2$. We also choose $h_g = 1.06\hat{\sigma}_b (nL)^{-1/5}$ and $h_f = 1.06\hat{\sigma}_v (nL_T)^{-1/5}$, where $\hat{\sigma}_b$ and $\hat{\sigma}_v$ are the standard deviations of the observed bids and the trimmed pseudo-private values, nL_T are the number of observations left after trimming, and 1.06 follows the rule of thumb.[22]

Figures 2.1-2.4 display the true density of the private values with solid line and the 5th, 50th and 95th percentiles of the 1000 estimates of $\hat{f}(v)$ with dash-dot lines evaluated at 500 equally spaced points on $[0.055, 2.5]$. When the utility functional form is correctly specified, the mean of the semiparametric estimates in Figure 1 perfectly matches the true density on the 25-75th percentile of the distribution and the empirical pointwise 90% confidence interval becomes narrower as n increases. In the case that the utility function is misspecified, the semiparametric estimates in Figure 2 are biased upwards for small private values and biased downwards for large private values when $n = 3$. The bias reduces as n increases. Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimates in Figure 3 are slightly downward biased when $n = 3$. The bias reduces as n increases to 6. The one-step nonparametric estimates in Figure 4 are very imprecise as a large part of the true density lies outside the empirical 90% confidence interval when $n = 3$. The performance of the one-step nonparametric estimates improves when $n = 12$.

To compare Methods 1-3 with higher precision, we report the integrated absolute bias evaluated respectively on the 5-95th percentile and the 25-75th percentile of the value distribution in Table 2.1. We use the integrated absolute bias instead of the integrated mean squared error as a measure of discrepancy because the semiparametric estimates may have

[22] Our choices of kernel functions and bandwidths do not follow Assumptions A3 and A4 because the gains of high order kernels in terms of a lower MISE are trivial with this sample size. (Fan and Marron, 1992)

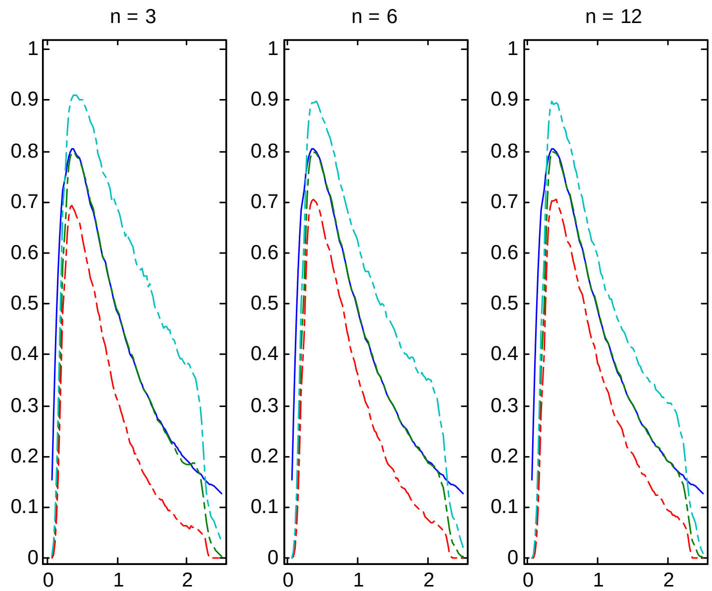

Figure 2.1: True and Estimated Densities of Private Values (Method 1)

larger standard error than the two-step nonparametric estimates as the former involves an additional step to estimate unknown parameters in the utility function. The integrals are evaluated by simulations. The two-step nonparametric estimates have smaller integrated absolute bias relative to the semiparametric estimates with misspecified utility function when $n = 6$ and 12. The bias of the two-step nonparametric estimator reduces much faster than the semiparametric estimates with misspecified utility function on the 25-75th percentile of the value distribution as n increases.

There are two important lessons to draw from the Monte Carlo experimental results. First, Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimation procedure is quite robust with respect to risk aversion in auctions with a moderate number of bidders,

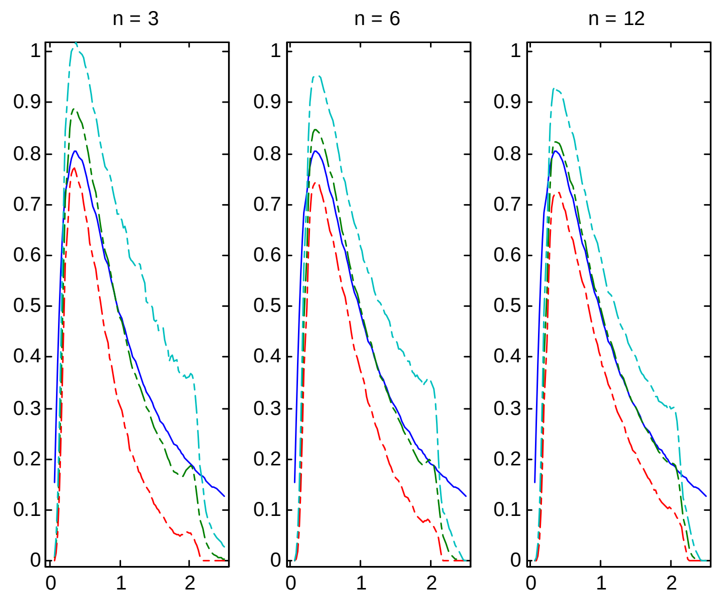

Figure 2.2: True and Estimated Densities of Private Values (Method 2)

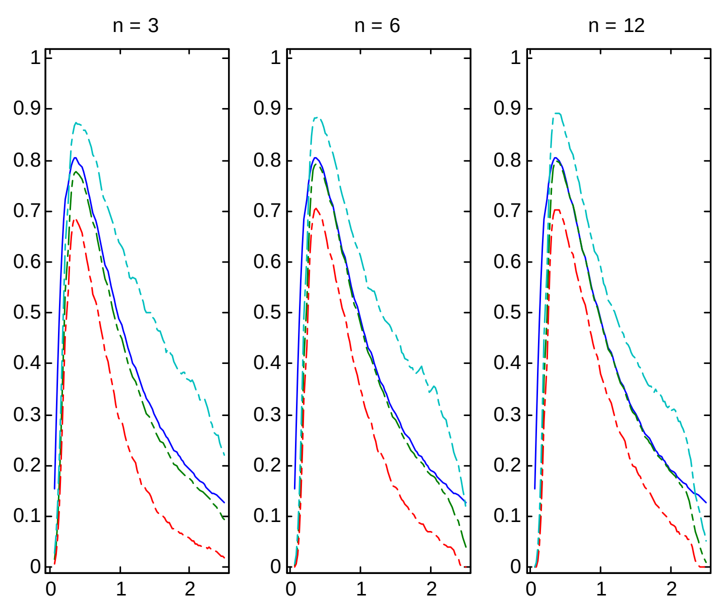

Figure 2.3: True and Estimated Densities of Private Values (Method 3)

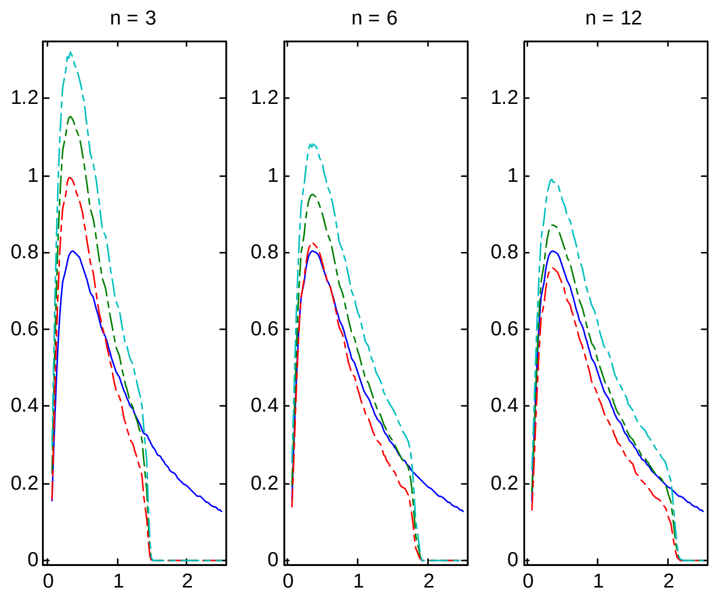

Figure 2.4: True and Estimated Densities of Private Values (Method 4)

	5-95th percentile			25-75th percentile		
	$n = 3$	$n = 6$	$n = 12$	$n = 3$	$n = 6$	$n = 12$
Method 1	0.0258	0.0279	0.0320	0.0023	0.0019	0.0017
Method 2	0.0757	0.0528	0.0427	0.0232	0.0163	0.0117
Method 3	0.0700	0.0442	0.0366	0.0242	0.0076	0.0022

Method 1: mean(θ)=0.7682, std(θ)=0.1934;

Method 2: mean(θ)=0.3835, std(θ)=0.0916.

Table 2.1: Intregrated Absolute Bias of Estimated Densities

and inclusion of the term $G_n(b)/(n-1)\,g_n(b)$ in the approximated bidding function substantially improves the performance of the estimator, as illustrated by Figures 3 and 4. Second, though the CRRA structure $U(x) = x^{1-\theta}$ is popularly assumed in the literature for many reasons, the semiparametric specification does not necessarily help to improve the fitting of $\hat{f}$. This can be understood as follows. As we discussed in Section 2, if $w = 0$, Definition 1(ii) needs to be replaced by the stronger assumption that "$U(\cdot)$ is continuous and admits up to $R+2$ continuous bounded derivatives on $S(U)$ with $U'(\cdot) > 0$ and $U''(\cdot) \leq 0$ on $S(U)$". However, $U(x) = x^{1-\theta} \notin \mathcal{U}_R$ as $w = 0$ and $U'(0)$ is not bounded. Hence the effects of model misspecification on the equilibrium bids are $O(n^{-1})$ in this case, which dominates the errors incurred by totally ignoring risk aversion.

2.5 Concluding Remarks

We study the robustness of Guerre, Perrigne and Vuong's (2000) two-step nonparametric estimation procedure in large auctions with risk averse bidders. With an asymptotic approximation of the equilibrium bidding function, we show that when the number of bidders in each auction diverges not too slowly relative to the number of observed auctions, Guerre, Perrigne and Vuong's (2000) two-step kernel-based estimator is uniformly consistent on an arbitrary closed inner subset of the support of the true density and attains the best uniform convergence rate as if latent private values are observable. Monte Carlo experiments show that the two-step estimator performs reasonably well with a moderate number of bidders such as six.

One possible extension of the current work is to allow bidders to have different attitude towards risk captured by heterogeneous utility functions and initial wealths. Campo (2004) has shown that in such a model the utility functions and latent distribution of bidders' private values cannot be nonparametrically identified jointly from observed bids, and, to recover the private value distribution, it is necessary to specify the asymmetric utility structure parametrically. On the other hand, when the number of bidders is large, the effects of asymmetric risk aversion on equilibrium bids diminish. Hence asymptotic approximation of the equilibrium bidding function may provide a feasible way to implement nonparametric estimation methods in large auctions with asymmetric risk averse bidders as well, which could be of interest for future research.

2.6 Appendices

2.6.1 Proofs of Mathematical Properties

Proof of Proposition 2.1. (i) Since the equilibrium solution is symmetric in nature, we can drop the individual subscript in (2.1). Let $s_{RN}(\cdot)$ be the solution of the following first-order differential equation

$$s'_{RN,n}(v) = (n-1)\frac{f(v)}{F(v)}\left(v - s_{RN,n}(v)\right), \tag{2.17}$$

with boundary condition $s_{RN,n}\left(\underline{v}\right) = \underline{v}$. Fibich, Gavious and Sela (2004) have shown that $s_{RN,n}(v) = v + O(n^{-1})$. As $0 \leq v - s_n(v) \leq v - s_{RN,n}(v)$ for all $v \in S(F)$ (Riley and Samuelson, 1981), we can extend $\varsigma_n(v)$ to the following form

$$\varsigma_n(v) = v - s_n(v) = \frac{1}{n-1}\varsigma_{1n}(v) + o(n^{-1}), \tag{2.18}$$

where $\varsigma_{1n}(v) = O(1)$. As $\lambda(0) = 0$ and $\lambda'(0) = 1$, a Taylor expansion of $\lambda\left(\varsigma_n(v)\right) = \lambda(v - s_n(v))$ around 0 gives

$$\lambda\left(\varsigma_n(v)\right) = \lambda(0) + \lambda'(0)\varsigma_n(v) + \frac{1}{2}\lambda''(\tilde{x})\varsigma_n^2(v) = \varsigma_n(v) + \frac{1}{2}\lambda''(\tilde{x})\varsigma_n^2(v), \tag{2.19}$$

for $\tilde{x} \in \left[0, \varsigma_n(v)\right]$. Since $\lambda''(\tilde{x})$ is bounded as $n \to \infty$ by Definition 1, substitution of (2.18) into (2.19) gives

$$\lambda\left(\varsigma_n(v)\right) = \frac{1}{n-1}\left(\varsigma_{1n}(v) + o(1)\right), \tag{2.20}$$

which implies $s_n'(v) = O(1)$ by (2.1). Multiplying both sides of the differential equation (2.1) by $F^{n-1}(v)$ and taking integrals gives

$$\begin{aligned} s_n'(v) F^{n-1}(v) &= (n-1) f(v) F^{n-2}(v) \lambda(v - s_n(v)) \\ \int_{\underline{v}}^{v} F^{n-1}(u)\, ds_n(u) &= \int_{\underline{v}}^{v} \lambda(u - s_n(u))\, dF^{n-1}(u). \end{aligned} \tag{2.21}$$

Applying integration by parts to the right hand side of (2.21), and rearranging terms yields the integral form of the first order condition

$$\lambda(v - s_n(v)) = \frac{1}{F^{n-1}(v)} \int_{\underline{v}}^{v} F^{n-1}(u)\, d\left(s_n(u) + \lambda(u - s_n(u))\right). \tag{2.22}$$

Let $\phi_n(v) = s_n(v) + \lambda(v - s_n(v))$, we have

$$\begin{aligned} \phi_n'(v) &= s_n'(v) + \lambda'(v - s_n(v))\left(1 - s_n'(v)\right) \\ &= s_n'(v) + \left(\lambda'(0) + O(v - s_n(v))\right)\left(1 - s_n'(v)\right) \\ &= 1 + \varsigma_{1n}(v)\, O(n^{-1})\left(1 - s_n'(v)\right) \\ &= 1 + O(n^{-1}), \end{aligned}$$

where the second equality holds by the mean value theorem and boundedness of $\lambda''(\cdot)$, the third equality holds because $\lambda'(0) = 1$, and the last equality holds because $\varsigma_{1n}(v) = O(1)$ and $s_n'(v) = O(1)$. We rewrite (2.22) in the format of Laplace integral and apply the

Laplace approximation (Copson, 1965)

$$\begin{aligned}
\lambda\left(v-s_n(v)\right) &= \frac{1}{F^{n-1}(v)}\int_{\underline{v}}^{v} F^{n-1}(u)\,\phi_n'(u)\,du = \frac{1}{F^{n-1}(v)}\int_{\underline{v}}^{v} e^{(n-1)\ln F(u)}\phi_n'(u)\,du \\
&= \frac{\phi_n'(v)}{F^{n-1}(v)}\frac{e^{(n-1)\ln F(v)}}{(n-1)\,d\ln F(v)/dv}+o(n^{-1}) \\
&= \frac{1}{n-1}\frac{F(v)}{f(v)}+o(n^{-1}), \qquad (2.23)
\end{aligned}$$

where the last equality holds because $\phi_n'(v)=1+O(n^{-1})$. Matching leading order terms in (2.20) and (2.23) gives $\varsigma_{1n}(v)=F(v)/f(v)$, which, together with (2.20) and (2.1), implies that $s_n'(v)=1+o(1)$ and $\phi_n'(v)=1+o(n^{-1})$. So we can further extend $\varsigma_n(v)$ to the following form

$$\varsigma_n(v)=v-s_n(v)=\frac{1}{n}\frac{F(v)}{f(v)}+\frac{1}{n^2}\varsigma_{2n}(v)+o(n^{-2}). \qquad (2.24)$$

Substitution of (2.24) in (2.19) gives

$$\lambda\left(\varsigma_n(v)\right)=\frac{1}{n}\frac{F(v)}{f(v)}+\frac{1}{n^2}\varsigma_{2n}(v)+\frac{1}{2}\lambda''(\tilde{x})\left(\frac{1}{n}\frac{F(v)}{f(v)}\right)^2+o(n^{-2}), \qquad (2.25)$$

Taking derivatives on both sides of (2.1) gives

$$s_n''(v)=(n-1)\left[\frac{d}{dv}\left(\frac{f(v)}{F(v)}\right)\lambda\left(v-s_n(v)\right)+\frac{f(v)}{F(v)}\lambda'\left(v-s_n(v)\right)\left(1-s_n'(v)\right)\right].$$

Taylor approximations of $\lambda\left(v-s_n(v)\right)$ and $\lambda'\left(v-s_n(v)\right)$ around 0 yield $s_n''(v)=o(n)$, which implies that

$$\begin{aligned}
\phi_n''(v) &= s_n''(v)-\lambda'\left(v-s_n(v)\right)s_n''(v)+\lambda''\left(v-s_n(v)\right)\left(1-s_n'(v)\right)^2 \\
&= s_n''(v)-\left(\lambda'(0)+O(n^{-1})\right)s_n''(v)+\lambda''\left(v-s_n(v)\right)o(1)=o(1).
\end{aligned}$$

It follows from applying integration by parts and the Laplace approximation to (2.22) that

$$\begin{aligned}
\lambda(v - s_n(v)) &= \frac{1}{F^{n-1}(v)}\int_{\underline{v}}^{v} F^{n-1}(u)\,\phi_n'(u)\,du = \frac{1}{nF^{n-1}(v)}\int_{\underline{v}}^{v}\frac{\phi_n'(u)}{f(u)}dF^n(u) \\
&= \frac{1}{n}\frac{F(v)\,\phi_n'(v)}{f(v)} - \frac{1}{nF^{n-1}(v)}\int_{\underline{v}}^{v} F^n(u)\,d\frac{\phi_n'(u)}{f(u)} \\
&= \frac{1}{n}\frac{F(v)\,\phi_n'(v)}{f(v)} - \frac{1}{nF^{n-1}(v)}\int_{\underline{v}}^{v} e^{n\ln F(u)}\frac{\phi_n''(u)\,f(u) - \phi_n'(u)\,f'(u)}{f^2(u)}du \\
&= \frac{1}{n}\frac{F(v)\,\phi_n'(v)}{f(v)} - \frac{F^2(v)}{n^2}\frac{\phi_n''(v)\,f(v) - \phi_n'(v)\,f'(v)}{f^3(v)} + o(n^{-2}) \\
&= \frac{1}{n}\frac{F(v)}{f(v)} + \frac{F^2(v)}{n^2}\frac{f'(v)}{f^3(v)} + o(n^{-2}), \qquad (2.26)
\end{aligned}$$

where the last equality holds because $\phi_n'(v) = 1 + o(n^{-1})$ and $\phi_n''(v) = o(1)$. Matching leading terms of (2.25) and (2.26) yields that

$$\varsigma_{2n}(v) = \left(\frac{f'(v)}{f(v)} - \frac{1}{2}\lambda''(\tilde{x})\right)\left(\frac{F(v)}{f(v)}\right)^2 = O(1),$$

which implies (2.2). Substitution of (2.26) into (2.1) gives $\varsigma_n'(v) = 1 - s_n'(v) = \frac{1}{n}\frac{d}{dv}\left(\frac{F(v)}{f(v)}\right) + o(n^{-1})$.

(ii) First, we show by mathematical induction that

$$\lambda\left(\varsigma_n(v)\right) = \lambda(v - s_n(v)) = \frac{\alpha_1(v)}{n} + \cdots + \frac{\alpha_{r+1}(v)}{n^{r+1}} + o\left(n^{-(r+1)}\right), \qquad (2.27)$$

$$\varsigma_n(v) = v - s_n(v) = \frac{\beta_1(v)}{n} + \cdots + \frac{\beta_{r+1}(v)}{n^{r+1}} + o\left(n^{-(r+1)}\right), \qquad (2.28)$$

where $\alpha_1(v), \cdots, \alpha_{r+1}(v)$ and $\beta_1(v), \cdots, \beta_{r+1}(v)$ are known functions invariant with n, and

$$\varsigma_n^{(r)}(v) = \frac{1}{n}\frac{d^r}{dv^r}\left(\frac{F(v)}{f(v)}\right) + o(n^{-1}), \qquad (2.29)$$

for $0 \le r \le R - 1$. We have already shown in (i) that (2.27)-(2.29) hold for $r = 0$, so we only need to show that (2.27)-(2.29) holding for $0 \le r \le k - 1$ implies (2.27)-(2.29) hold for $r=k\leq R-1$. A Taylor expansion of $\lambda\left(\varsigma_n(v)\right)$ with an integral remainder gives

$$\lambda\left(\varsigma_n(v)\right)=\lambda(0)+\lambda'(0)\,\varsigma_n(v)+\cdots+\frac{1}{R!}\lambda^{(R)}(0)\,\varsigma_n^R(v)+\int_0^{\varsigma_n(v)}\lambda^{(R+1)}(t)\,\frac{\left(\varsigma_n(v)-t\right)^R}{R!}dt. \tag{2.30}$$

Analogously, we have for $1\leq r\leq R$

$$\lambda^{(r)}\left(\varsigma_n(v)\right)=\lambda^{(r)}(0)+\cdots+\frac{1}{(R-r)!}\lambda^{(R)}(0)\,\varsigma_n^{R-r}(v)+\int_0^{\varsigma_n(v)}\lambda^{(R+1)}(t)\,\frac{\left(\varsigma_n(v)-t\right)^{R-r}}{(R-r)!}dt. \tag{2.31}$$

From (2.1), we have

$$1-s_n'(v)=1-(n-1)\frac{f(v)}{F(v)}\lambda\left(\varsigma_n(v)\right).$$

For $r\geq 2$, taking the $(r-1)$th derivatives on both sides gives

$$\begin{aligned}\varsigma_n^{(r)}(v) &= -(n-1)\frac{d^{r-1}}{dv^{r-1}}\left(\frac{f(v)}{F(v)}\lambda\left(\varsigma_n(v)\right)\right)\\ &= -(n-1)\left\{\frac{d^{r-1}}{dv^{r-1}}\left(\frac{f(v)}{F(v)}\right)\lambda\left(\varsigma_n(v)\right)+\cdots\right.\\ &\quad +\binom{r-1}{l}\frac{d^{r-1-l}}{dv^{r-1-l}}\left(\frac{f(v)}{F(v)}\right)\frac{d^l}{dv^l}\lambda\left(\varsigma_n(v)\right)+\cdots\\ &\quad \left.+\frac{f(v)}{F(v)}\frac{d^{r-1}}{dv^{r-1}}\lambda\left(\varsigma_n(v)\right)\right\},\end{aligned} \tag{2.32}$$

where by Faà di Bruno's formula,

$$\begin{aligned}\frac{d^l}{dv^l}\lambda\left(\varsigma_n(v)\right) &= \sum\left\{\frac{l!}{m_1!+m_2!+\cdots+m_l!}\lambda^{(m_1+m_2+\cdots+m_l)}\left(\varsigma_n(v)\right)\right.\\ &\quad \left.\times\prod_{j:m_j\neq 0}\left(\frac{1}{j!}\varsigma_n^{(l)}(v)\right)^{m_j}\right\},\end{aligned} \tag{2.33}$$

where the sum is over all l-tuples $(m_1,\cdots,m_l)$ satisfying the constraint $1m_1+2m_2+\cdots+lm_l=l$. By plugging (2.30) and (2.31) into (2.32) and substituting $\varsigma_n(v)$ by (2.28), i.e.,

$$\varsigma_n(v)=\frac{\beta_1(v)}{n}+\frac{\beta_2(v)}{n^2}+\cdots+\frac{\beta_k(v)}{n^k}+o\left(n^{-k}\right),$$

where $\beta_1(v), \cdots, \beta_k(v)$ are known and invariant with n by induction assumptions, we can derive from $r = 1$ to $k-1$ that

$$\varsigma_n^{(r)}(v) = \frac{\gamma_{r1}(v)}{n} + \cdots + \frac{\gamma_{r,k-r}(v)}{n^{k-r}} + o\left(n^{r-k}\right), \tag{2.34}$$

where $\gamma_{r1}(v), \cdots, \gamma_{r,k-r}(v)$ are known functions invariant with n and $\varsigma_n^{(r)}(v)$ is of order $O(n^{-1})$ by the induction assumptions (2.29). From (2.33), it follows for $l \leq k-1$

$$\begin{aligned}
\frac{d^l}{dv^l}\lambda\left(\varsigma_n(v)\right) &= \lambda'\left(\varsigma_n(v)\right)\varsigma_n^{(l)}(v) + O(n^{-2}) \\
&= \frac{1}{n}\lambda'\left(\varsigma_n(v)\right)\frac{d^l}{dv^l}\left(\frac{F(v)}{f(v)}\right) + o(n^{-1}) \\
&= \frac{1}{n}\frac{d^l}{dv^l}\left(\frac{F(v)}{f(v)}\right) + o(n^{-1}),
\end{aligned}$$

where the second equality holds because of (2.29), and the last equality follows from a Taylor approximation of $\lambda'\left(\varsigma_n(v)\right)$. From (2.30), we have $\lambda\left(\varsigma_n(v)\right) = F(v)/(nf(v)) + o(n^{-1})$. Hence if $k = 1$

$$\varsigma_n^{(k)}(v) = 1 - (n-1)\frac{f(v)}{F(v)}\lambda\left(\varsigma_n(v)\right) = o(1),$$

and, if $k \geq 2$, substitution of $\lambda\left(\varsigma_n(v)\right)$ and $d^l\lambda\left(\varsigma_n(v)\right)/dv^l$ into (2.32) gives

$$\begin{aligned}
\varsigma_n^{(k)}(v) &= -\frac{n-1}{n}\Big\{\frac{d^{k-1}}{dv^{k-1}}\left(\frac{f(v)}{F(v)}\right)\frac{F(v)}{f(v)} + \cdots \\
&\quad + \binom{k-1}{l}\frac{d^{k-1-l}}{dv^{k-1-l}}\left(\frac{f(v)}{F(v)}\right)\frac{d^l}{dv^l}\left(\frac{F(v)}{f(v)}\right) + \cdots \\
&\quad + \frac{f(v)}{F(v)}\frac{d^{k-1}}{dv^{k-1}}\left(\frac{F(v)}{f(v)}\right)\Big\} + o(1) \\
&= -\frac{n-1}{n}\frac{d^{k-1}}{dv^{k-1}}\left(\frac{f(v)}{F(v)}\frac{F(v)}{f(v)}\right) + o(1) = o(1),
\end{aligned}$$

which implies that

$$\frac{d^k}{dv^k}\lambda\left(\varsigma_n(v)\right) = \lambda'\left(\varsigma_n(v)\right)\varsigma_n^{(k)}(v) + O(n^{-2}) = o(1).$$

It follows that

$$\begin{aligned}
\varsigma_n^{(k+1)}(v) &= -(n-1)\frac{d^k}{dv^k}\left(\frac{f(v)}{F(v)}\lambda\left(\varsigma_n(v)\right)\right) \\
&= -(n-1)\left\{\frac{d^k}{dv^k}\left(\frac{f(v)}{F(v)}\right)\lambda\left(\varsigma_n(v)\right)+\cdots\right. \\
&\quad +\binom{k}{l}\frac{d^{k-l}}{dv^{k-l}}\left(\frac{f(v)}{F(v)}\right)\frac{d^l}{dv^l}\lambda\left(\varsigma_n(v)\right)+\cdots \\
&\quad \left.+\frac{f(v)}{F(v)}\frac{d^k}{dv^k}\lambda\left(\varsigma_n(v)\right)\right\} \\
&= o(n).
\end{aligned}$$

As

$$\phi_n(v)=s_n(v)+\lambda\left(\varsigma_n(v)\right)=s_n(v)+\frac{1}{n-1}\frac{F(v)}{f(v)}s_n^{(1)}(v),$$

where the second equality follows from (2.1), we have

$$\begin{aligned}
\phi_n^{(r)}(v) &= s_n^{(r)}(v)+\frac{1}{n-1}\left\{\frac{d^r}{dv^r}\left(\frac{f(v)}{F(v)}\right)s_n^{(1)}(v)+\cdots\right. \\
&\quad \left.+\binom{r}{l}\frac{d^{r-l}}{dv^{r-l}}\left(\frac{f(v)}{F(v)}\right)s_n^{(l+1)}(v)+\cdots+\frac{f(v)}{F(v)}s_n^{(r+1)}(v)\right\}.
\end{aligned}$$

By substituting (2.34), it can be rewritten in the following form

$$\phi_n^{(r)}(v)=\frac{\delta_{r1}(v)}{n}+\cdots+\frac{\delta_{r,k-r}(v)}{n^{k-r}}+o\left(n^{r-k}\right), \tag{2.35}$$

where $\delta_{r1}(v),\cdots,\delta_{r,k-r}(v)$ are known functions invariant with n for $r\leq k-1$. Furthermore, we have

$$\begin{aligned}
\phi_n^{(k)}(v) &= s_n^{(k)}(v)+\frac{d^k}{dv^k}\lambda\left(\varsigma_n(v)\right) \\
&= s_n^{(k)}(v)+\lambda'\left(\varsigma_n(v)\right)\varsigma_n^{(k)}(v)+O(n^{-2}) \\
&= \mathbf{1}(k=1)+o(n^{-1}),
\end{aligned}$$

where $\mathbf{1}(k=1)$ is an indicator of $k=1$, and

$$\begin{aligned}
\phi_n^{(k+1)}(v) &= s_n^{(k+1)}(v) + \frac{d^{k+1}}{dv^{k+1}}\lambda\left(\varsigma_n(v)\right) \\
&= s_n^{(k+1)}(v) + \lambda'\left(\varsigma_n(v)\right)\varsigma_n^{(k+1)}(v) + o(1) = o(1),
\end{aligned}$$

since $\lambda'\left(\varsigma_n(v)\right) = \lambda'(0) + O\left(\varsigma_n(v)\right) = 1 + O(n^{-1})$, $\varsigma_n^{(k)}(v) = o(1)$, and $\varsigma_n^{(k+1)}(v) = o(n)$. Repeated integration by parts of (2.22) gives that

$$\begin{aligned}
\lambda\left(\varsigma_n(v)\right) &= \frac{1}{n}\psi_1(v)F(v) + \cdots + \frac{(-1)^{k-1}}{n(n+1)\cdots(n+k-1)}\psi_k(v)F^k(v) \\
&\quad + \frac{(-1)^k}{n(n+1)\cdots(n+k-1)F^{n-1}(v)}\int_{\underline{v}}^{v} F^{n+k-1}(u)\psi_{k+1}(u)f(u)\,du,
\end{aligned}$$

where $\psi_1(v) = \phi_n'(v)/f(v)$, $\psi_2(v) = \psi_1'(v)/f(v), \cdots, \psi_{k+1}(v) = \psi_k'(v)/f(v)$. As $\psi_l(v)$ is a polynomial of $\phi_n'(v), \cdots, \phi_n^{(l)}(v)$, by substitution of (2.35), we have for $l \le k-1$

$$\psi_l(v) = \zeta_{l0}(v) + \frac{\zeta_{l1}(v)}{n} + \cdots + \frac{\zeta_{l,k-l}(v)}{n^{k-l}} + o\left(n^{l-k}\right), \tag{2.36}$$

$\psi_k(v) = \zeta_{k0}(v) + o(n^{-1})$, and $\psi_{k+1}(v) = \zeta_{k+1,0}(v) + o(1)$, where $\zeta_{l0}(v), \cdots, \zeta_{l,k-l}(v)$ are known functions invariant with n. The Laplace approximation gives

$$\begin{aligned}
\lambda\left(\varsigma_n(v)\right) &= \frac{1}{n}\psi_1(v)F(v) + \cdots + \frac{(-1)^{k-1}}{n(n+1)\cdots(n+k-1)}\psi_k(v)F^k(v) \\
&\quad + \frac{(-1)^k}{n(n+1)\cdots(n+k-1)F^{n-1}(v)}\int_{\underline{v}}^{v} e^{(n+k-1)\ln F(u)}\psi_{k+1}(u)f(u)\,du \\
&= \frac{1}{n}\psi_1(v)F(v) + \cdots + \frac{(-1)^{k-1}}{n(n+1)\cdots(n+k-1)}\psi_k(v)F^k(v) \\
&\quad + \frac{(-1)^k}{n(n+1)\cdots(n+k-1)^2}\psi_{k+1}(v)F^{k+1}(v) + o\left(n^{-(k+1)}\right).
\end{aligned}$$

By substitution of (2.36), it can be rewritten in the form of (2.27)

$$\lambda\left(\varsigma_n(v)\right) = \frac{\alpha_1(v)}{n} + \frac{\alpha_2(v)}{n^2} + \cdots + \frac{\alpha_{k+1}(v)}{n^{k+1}} + o\left(n^{-(k+1)}\right), \tag{2.37}$$

with $\alpha_1, \cdots, \alpha_k(v)$ known by the induction assumptions and $\alpha_{k+1}(v)$ explicitly derived. Let

$$\varsigma_n(v) = \frac{\beta_1(v)}{n} + \frac{\beta_2(v)}{n^2} + \cdots + \frac{\beta_{k+1}(v)}{n^{k+1}} + o\left(n^{-(k+1)}\right), \tag{2.38}$$

where $\beta_1, \cdots, \beta_k(v)$ are known by the induction assumptions. By substituting (2.38) into (2.30) and matching leading order terms with (2.37), we can solve for $\beta_{k+1}(v)$. Substitute (2.30) and (2.31) into (2.32) and replace $\varsigma_n(v)$ by (2.38). Now we can derive from $r = 1$ to k that

$$\varsigma_n^{(r)}(v) = \frac{\gamma_{r1}(v)}{n} + \cdots + \frac{\gamma_{r,k+1-r}(v)}{n^{k+1-r}} + o\left(n^{r-k-1}\right),$$

where $\gamma_{r1}(v), \cdots, \gamma_{r,k+1-r}(v)$ are known functions invariant with n. Specifically, $\varsigma_n^{(k)}(v) = \gamma_{k1}(v)/n + o(n^{-1})$. This, together with the induction assumption (2.29), implies that $\gamma_{k1}(v) = d^k(F(v)/f(v))/dv^k$.

Lastly, we can show with analogous arguments that (2.27)-(2.29) holding for $r \leq R-1$ implies (2.29) holds for $r = R$. ■

Corollary 2.1 *In a first-price IPV auction with n ($n \gg 1$) bidders, suppose $U(\cdot) \in \mathcal{U}_R$ for $R \geq 1$. Given A1 and A2, the equilibrium bid in the symmetric Bayesian Nash equilibrium is given by*

$$s_n(v, x) = v - \frac{1}{n}\frac{F(v|x)}{f(v|x)} + O(n^{-2}).$$

Furthermore, we have $\varsigma_n^{(r)}(v, x) = O(n^{-1})$ for $1 \leq r \leq R$.

Proof of Proposition 2.2. Let $h(v) = s_n(v) - s_m(v)$, with $n > m \geq 2$. $h(v) = 0$ implies that

$$\begin{aligned} h'(v) &= s_n'(v) - s_m'(v) \\ &= (n-1)\frac{f(v)}{F(v)}\lambda(v - s_n(v)) - (m-1)\frac{f(v)}{F(v)}\lambda(v - s_m(v)) \\ &= (n-m)\frac{f(v)}{F(v)}\lambda(v - s_n(v)) > 0, \end{aligned}$$

where the inequality holds because $n > m$ and $U(\cdot)$ is monotonically increasing. Since $h(0) = 0$, by the single crossing lemma $s_n(v) > s_m(v)$ for $\bar{v} \geq v > \underline{v}$. As $\lim_{n\to\infty} s_n(\bar{v}) = \bar{v}$ by Proposition 2.1, (i) follows. Next, $g_n(b) = f(v)/s_n'(v)$ with $b = s_n(v)$. Because $f(v)$ is bounded away from zero by assumption and $s_n'(v)$ is bounded with $\lim_{n\to\infty} s_n'(v) = 1$ by Proposition 2.1, (ii) follows. To prove (iii), we note that substitution of $g_n(b) = f(v)/s_n'(v)$ into (2.1) gives

$$g_n(b) = \frac{F(v)}{(n-1)(v - s_n(v))}, \tag{2.39}$$

with $b = s_n(v)$. Since $(n-1)(v - s_n(v)) = F(v)/f(v) + O(n^{-1})$, it follows from Proposition 2.1 that $\sup_{b\in C} |g_n(b)|$ is bounded as $n \to \infty$. Similarly, $g_n^{(r)}(b)$ $(r = 1, \cdots, R)$ can be derived by taking rth differentiation on both sides of (2.39). Using mathematical induction, the desired result follows from Proposition 2.1. ■

Proof of Proposition 2.5. Trivial extension of Proposition 2.2 based on Corollary 2.1. ■

2.6.2 Proofs of Statistical Properties

To prove Propositions 2.3 and 2.4 we need two auxiliary lemmas on the uniform convergence of $\tilde{G}_n(\cdot)$ and $\tilde{g}_n(\cdot)$ defined by (2.4) and (2.5). Throughout $|\cdot|_{r,*}$ denotes the sup-norm of the rth derivatives of $\cdot$ on the set $*$.

Lemma 2.1 *Suppose for $R \geq 1$, $F(\cdot) \in F_R$, $U(\cdot) \in U_R$, and $\tilde{G}_n(\cdot)$ is given by* (2.4)*, we have almost surely*

$$\left|\tilde{G}_n(b) - G_n(b)\right|_{0,C} = O\left(1/\sqrt{nL}\right),$$

where C is an arbitrary closed inner subset of $S(G_\infty)$.

Proof. It follows from Proposition 2.2 that

$$\begin{aligned}\tilde{G}_n(b) &= \frac{1}{nL}\sum\nolimits_{l=1}^{L}\sum\nolimits_{p=1}^{n}\mathbf{1}\left(B_{pl}\le b\right)\\ &= \frac{1}{nL}\sum\nolimits_{l=1}^{L}\sum\nolimits_{p=1}^{n}\mathbf{1}\left(G_n\left(B_{pl}\right)\le G_n(b)\right)\\ &= \frac{1}{nL}\sum\nolimits_{l=1}^{L}\sum\nolimits_{p=1}^{n}\mathbf{1}\left(u_{pl}\le G_n(b)\right),\end{aligned}$$

where $u_{pl} = G_n\left(B_{pl}\right)$ is uniformly distributed on $[0,1]$ since $B_{pl} \sim G_n(\cdot)$. Let $u = G_n(b) \in [0,1]$, and $n_C = \min\left\{n : C \subset S_n^o\right\}$ where S_n^o is the interior of $S(G_n)$. n_C exists because of Proposition 2.2(i). Then for $n > n_C$,

$$\begin{aligned}\left|\tilde{G}_n(b) - G_n(b)\right|_{0,C} &= \left|\frac{1}{nL}\sum\nolimits_{l=1}^{L}\sum\nolimits_{p=1}^{n}\mathbf{1}\left(u_{pl}\le G_n(b)\right) - G_n(b)\right|_{0,C}\\ &= \left|\frac{1}{nL}\sum\nolimits_{l=1}^{L}\sum\nolimits_{p=1}^{n}\mathbf{1}\left(u_{pl}\le u\right) - u\right|_{0,C}\\ &= O(1/\sqrt{nL}),\end{aligned}$$

where the last step holds because the empirical distribution of uniform distribution (which does not depend on n) converges uniformly to the true distribution at the rate of $\sqrt{nL}$. ■

Lemma 2.2 *Suppose for* $R \ge 1$, $F(\cdot) \in F_R$, $U(\cdot) \in U_R$, *and* $\tilde{g}_n(\cdot)$ *as given by* (2.5)*, we have almost surely*

$$|\tilde{g}_n(b) - g_n(b)|_{0,C} = O(1/r),$$

where C *is an arbitrary closed inner subset of* $S(G_\infty)$ *and* $r = (nL/\log(nL))^{R/(2R+1)}$.

Proof. The proof relies on the argument of Guerre, Perrigne and Vuong's (2000) proof for the case of fixed n. However, the problem is different because, as we allow both n and L to approach infinity, the observations are from a triangular array of random variables shifting with sample size. Hence the standard consistency results based on the i.i.d. assumption of observations do not apply directly. We divide the proof into three steps. The first step studies the uniform bias of $\tilde{g}_n(\cdot)$, the second step studies its uniform variance bound, and the last step establishes the exponential-type inequality. We simplify notation by omitting the subscript R in h_R and K_R in this proof. The sup-norm is taken over the whole support of the function unless otherwise indicated.

Step 1: Uniform Bias

For any closed inner subset C of $S(G_\infty)$, let $n_C = \min\{n : C \subset S_n^o\}$ where S_n^o is the interior of $S(G_n)$. n_C exists because of Proposition 2.2(i). For $n > n_C$,

$$\begin{aligned} E\tilde{g}_n(b) &= E\frac{1}{nLh}\sum\nolimits_{l=1}^{L}\sum\nolimits_{p=1}^{n} K\left(\frac{B_{pl}-b}{h}\right) \\ &= \int K(u)\, g_n(b+hu)\, du. \end{aligned}$$

Without loss of generality, suppose $u \geq 0$. Then for $b \in C$ and L sufficiently large, $\tilde{b} \in [b, b+hu] \subset C'$, where C' is a closed inner subset of $S(G_n)$. Since $g_n(\cdot)$ admits up to R continuous bounded derivatives on any closed inner subset of $S(G_n)$, a Taylor expansion gives

$$g_n(b+hu) - g_n(b) \leq hug_n^{(1)}(b) + \cdots + \frac{(hu)^{R-1}}{(R-1)!}g_n^{(R-1)}(b) + \frac{|hu|^R}{R!}|g_n|_{R,C'}.$$

As $K(\cdot)$ is of order R, moments of order strictly smaller than R vanish. So we have

$$\begin{aligned} |E\tilde{g}_n(b) - g_n(b)|_{0,C} &= \sup\nolimits_{b\in C}\left|\int K(u)\left(g_n(b+hu) - g_n(b)\right) du\right| \\ &\leq h^R |g_n|_{R,C'} \frac{1}{R!}\left(\int |u|^R K(u)\, du\right) \\ &= h^R |g_n|_{R,C'} M^R, \end{aligned}$$

where $M^R = (1/R!)\int |u|^R K(u)\, du$. It follows from the definition of r and h that

$$r\,|E\tilde{g}_n(b) - g_n(b)|_{0,C} \leq \lambda^R M^R |g_n|_{R,C'}. \tag{2.40}$$

Step 2: Uniform Variance

For $b \in C$, we have

$$\begin{aligned} Var\left(\tilde{g}_n(b)\right) &= Var\left(\frac{1}{nLh}\sum_{l=1}^{L}\sum_{p=1}^{n} K\left(\frac{B_{pl}-b}{h}\right)\right) \\ &= \frac{1}{nLh^2} Var\left(K\left(\frac{B-b}{h}\right)\right) \leq \frac{1}{nLh^2} E\left(K\left(\frac{B-b}{h}\right)\right)^2 \\ &= \frac{1}{nLh}\int K^2(u)\, g_n(b+hu)\, du. \end{aligned}$$

Let $Q = \int K^2(u)\, du$, it follows that

$$|Var\left(\tilde{g}_n(b)\right)|_{0,C} \leq \frac{Q\,|g_n|_0}{nLh} = \frac{Q\,|g_n|_0}{\lambda r^2 \log(nL)}. \tag{2.41}$$

Step 3: Exponential-type Inequality

In this step, we establish the exponential-type inequality for the probability of deviation of $\tilde{g}_n(b) - g_n(b)$ in sup-norm over C. Let C be covered by T inner intervals of the form

$$B_t \equiv B(b_t, \Delta) = \{b \in S(G_\infty) : b \in [b_t - \Delta, b_t + \Delta]\},$$

where $b_t \in C$ and $\Delta > 0$. Moreover, we consider minimal coverings for C, i.e., coverings for which T is the smallest number denoted by $T(C, \Delta)$. Let

$$\begin{aligned} e(\iota, \tau) &= \iota + 2\tau|K|_1 + \lambda^R M^R |g_n|_{R,C'}, \\ P(\iota, \tau) &= 2T\left(C, \tau h^2/r\right)\exp\left(-\frac{\lambda \iota^2 \log(nL)}{2Q\,|g_n|_0 + 4\iota\,|K|_0/(3r)}\right), \end{aligned}$$

where ι, τ are strictly positive constants.

Step 3(a): From (2.40) and the triangular inequality, we obtain

$$\begin{aligned} &\Pr\left(r\,|\tilde{g}_n(b) - g_n(b)|_{0,C} > e(\iota,\tau)\right) \\ \leq\ &\Pr\left(r\,|\tilde{g}_n(b) - E\tilde{g}_n(b)|_{0,C} + r\,|E\tilde{g}_n(b) - g_n(b)|_{0,C} > e(\iota,\tau)\right) \\ \leq\ &\Pr\left(r\,|\tilde{g}_n(b) - E\tilde{g}_n(b)|_{0,C} > e(\iota,\tau) - \lambda^R M^R |g_n|_{R,C'}\right). \end{aligned} \tag{2.42}$$

Let $\tilde{g}_n(b) - E\tilde{g}_n(b) = (1/nL)\sum_{i=1}^{nL} \zeta_{i,nL}(b)$, where

$$\zeta_{i,nL}(b) = \frac{1}{h}\left[K\left(\frac{B_i - b}{h}\right) - EK\left(\frac{B - b}{h}\right)\right].$$

As the $\zeta_{i,nL}$'s are independent zero-mean variables, it follows from (2.41)

$$Var\left(r\zeta_{i,nL}\right) = nLr^2 Var\left(\tilde{g}_n\right) \le \frac{nLQ\left|g_n\right|_0}{\lambda \log(nL)}.$$

By the triangular inequality we have

$$\left|r\zeta_{i,nL}\right| \le \frac{2r\left|K\right|_0}{h} = \frac{2nL\left|K\right|_0}{\lambda r \log(nL)}.$$

Hence the Bernstein inequality gives

$$\begin{aligned}
& \Pr\left(r\left|\tilde{g}_n(b) - E\tilde{g}_n(b)\right| > \iota\right) \\
= \; & \Pr\left(\left|\sum_{i=1}^{nL} r\zeta_{i,nL}(b) - \sum_{i=1}^{nL} E\left(r\zeta_{i,nL}(b)\right)\right| > nL\iota\right) \\
\le \; & 2\exp\left(-\frac{n^2L^2\iota^2}{2\sum_{i=1}^{nL} Var\left(r\zeta_{i,nL}\right) + 4n^2L^2\iota\left|K\right|_0/(3\lambda r\log(nL))}\right) \\
\le \; & 2\exp\left(-\frac{\lambda\iota^2\log(nL)}{2Q\left|g_n\right|_0 + 4\iota\left|K\right|_0/(3r)}\right) \\
= \; & \frac{P(\iota,\tau)}{T\left(C, \tau h^2/r\right)},
\end{aligned}$$

for any $b \in C$, ι, n, and L.

Step 3(b): Consider a minimal covering of C for some $\Delta > 0$. For any $b \in B_t$, we have by the triangular inequality

$$r\left|\tilde{g}_n(b) - E\tilde{g}_n(b)\right| \le \sup_{1\le t\le T}\left|\frac{r}{nL}\sum_{i=1}^{nL}\zeta_{i,nL}(b_t)\right| + \sup_{1\le t\le T}\sup_{b\in B_t}\left|\frac{r}{nL}\sum_{i=1}^{nL}\left(\zeta_{i,nL}(b_t) - \zeta_{i,nL}(b)\right)\right|,$$

which implies that

$$\begin{aligned}
& \Pr\left(r\sup_{b\in C}\left|\tilde{g}_n(b) - E\tilde{g}_n(b)\right| > e(\iota,\tau) - \lambda^R M^R\left|g_n\right|_{R,C'}\right) \\
\le \; & \Pr\left(\sup_{1\le t\le T}\sup_{b\in B_t}\left|\frac{r}{nL}\sum_{i=1}^{nL}\left(\zeta_{i,nL}(b_t) - \zeta_{i,nL}(b)\right)\right| > e(\iota,\tau) - \iota - \lambda^R M^R\left|g_n\right|_{R,C'}\right) \\
& + \Pr\left(r\sup_{1\le t\le T}\left|\tilde{g}_n(b_t) - E\tilde{g}_n(b_t)\right| > \iota\right). \qquad (2.43)
\end{aligned}$$

Since

$$\left|\frac{1}{h}K\left(\frac{B-b_t}{h}\right)-\frac{1}{h}K\left(\frac{B-b}{h}\right)\right|\leq\frac{\Delta\left|K\right|_1}{h^2},$$

by the mean value theorem, we have by the triangular inequality

$$\left|\zeta_{i,nL}\left(b_t\right)-\zeta_{i,nL}\left(b\right)\right|\leq\frac{\Delta\left|K\right|_1}{h^2}+E\frac{\Delta\left|K\right|_1}{h^2}=\frac{2\Delta\left|K\right|_1}{h^2}.$$

Step 3(c): Let $\Delta=\tau h^2/r$, it follows

$$\sup_{1\leq t\leq T}\sup_{b\in B_t}\left|\frac{r}{nL}\sum\nolimits_{i=1}^{nL}\left(\zeta_{i,nL}\left(b_t\right)-\zeta_{i,nL}\left(b\right)\right)\right|\leq\frac{2r\Delta\left|K\right|_1}{h^2}=2\tau\left|K\right|_1.$$

Hence

$$\begin{aligned}&\Pr\left(\sup_{1\leq t\leq T}\sup_{b\in B_t}\left|\frac{r}{nL}\sum\nolimits_{i=1}^{nL}\left(\zeta_{i,nL}\left(b_t\right)-\zeta_{i,nL}\left(b\right)\right)\right|>e\left(\iota,\tau\right)-\iota-\lambda^R M^R\left|g_n\right|_{R,C'}\right)\\ =\;&\Pr\left(\sup_{1\leq t\leq T}\sup_{b\in B_t}\left|\frac{r}{nL}\sum\nolimits_{i=1}^{nL}\left(\zeta_{i,nL}\left(b_t\right)-\zeta_{i,nL}\left(b\right)\right)\right|>2\tau\left|K\right|_1\right)=0.\end{aligned}\tag{2.44}$$

Then it follows from (2.42), (2.43), (2.44), and the Bernstein inequality that

$$\begin{aligned}&\Pr\left(r\left|\tilde{g}_n\left(b\right)-g_n\left(b\right)\right|_{0,C}>e\left(\iota,\tau\right)\right)\\ \leq\;&\Pr\left(r\left|\tilde{g}_n\left(b\right)-E\tilde{g}_n\left(b\right)\right|_{0,C}>e\left(\iota,\tau\right)-\lambda^R M^R\left|g_n\right|_{R,C'}\right)\\ \leq\;&\Pr\left(r\sup\nolimits_{1\leq t\leq T}\left|\tilde{g}_n\left(b_t\right)-E\tilde{g}_n\left(b_t\right)\right|>\iota\right)\\ \leq\;&\sum\nolimits_{t=1}^{T}\Pr\left(r\left|\tilde{g}_n\left(b_t\right)-E\tilde{g}_n\left(b_t\right)\right|>\iota\right)\\ \leq\;&P\left(\iota,\tau\right).\end{aligned}$$

The covering number $T\left(C,\Delta\right)$ is of order Δ^{-1}, as the covered set C is an interval. Hence $T\left(C,\tau h^2/r\right)=O((nL/\log\left(nL\right))^{(R+2)/(2R+1)})$. By taking ι sufficiently large, $P\left(\iota,\tau\right)$ converges as $nL\to\infty$. The desired result follows from the Borel-Cantelli lemma and the fact that $e\left(\iota,\tau\right)=O\left(1\right)$. ■

Proof of Proposition 2.3. The proof presented here follows Guerre, Perrigne and Vuong's (2000) proof for the risk neutrality case. Let

$$\bar{V}_{pl} = B_{pl} + \frac{1}{n-1}\psi_n\left(B_{pl}\right),$$

with $\psi_n(\cdot) = G_n(\cdot)/g_n(\cdot)$. Let $\tilde{\psi}_n(\cdot) = \tilde{G}_n(\cdot)/\tilde{g}_n(\cdot)$, with $\tilde{G}_n(\cdot)$ and $\tilde{g}_n(\cdot)$ given by (2.4) and (2.5) respectively. Since $C(V)$ is a closed inner subset of $S(F)$ and $s_n(\cdot)$ is a strictly increasing continuous function, $C(B) = C(s_n(V))$ is a closed inner subset of $S(G_n)$. From (2.3), we have

$$\begin{aligned}
& \mathbf{1}_{C(V)}\left(V_{pl}\right)\left|\hat{V}_{pl} - V_{pl}\right| \\
\leq\ & \mathbf{1}_{C(V)}\left(V_{pl}\right)\left(\left|\hat{V}_{pl} - \bar{V}_{pl}\right| + \left|\bar{V}_{pl} - V_{pl}\right|\right) \\
=\ & \frac{\mathbf{1}_{C(B)}\left(B_{pl}\right)}{n-1}\left|\tilde{\psi}_n\left(B_{pl}\right) - \psi_n\left(B_{pl}\right)\right| + O(n^{-2}) \\
=\ & \frac{\mathbf{1}_{C(B)}\left(B_{pl}\right)\mathbf{1}(\hat{V}_{pl} \neq \infty)}{n-1}\left|\tilde{\psi}_n\left(B_{pl}\right) - \psi_n\left(B_{pl}\right)\right| \\
& + \frac{\mathbf{1}_{C(B)}\left(B_{pl}\right)\left(1 - \mathbf{1}(\hat{V}_{pl} \neq \infty)\right)}{n-1}\left|\tilde{\psi}_n\left(B_{pl}\right) - \psi_n\left(B_{pl}\right)\right| + O(n^{-2}).
\end{aligned}$$

It is easy to see that $1_{C(B)}\left(B_{pl}\right)(1 - 1(\hat{V}_{pl} \neq \infty)) = 0$ almost surely for any p and l as $n, L \to \infty$. Since $G_n(\cdot) \leq 1$ and $g_n(\cdot)$ has a positive lower bound c_g by Proposition 2.2(ii), we have

$$\begin{aligned}
& \mathbf{1}_{C(B)}\left(B_{pl}\right)\mathbf{1}(\hat{V}_{pl} \neq \infty)\left|\tilde{\psi}_n\left(B_{pl}\right) - \psi_n\left(B_{pl}\right)\right| \\
=\ & \frac{\mathbf{1}_{C(B)}\left(B_{pl}\right)\mathbf{1}(\hat{V}_{pl} \neq \infty)}{g_n\left|\tilde{g}_n\right|}\left|(\tilde{G}_n - G_n)g_n + (g_n - \tilde{g}_n)\,G_n\right| \\
\leq\ & \frac{\mathbf{1}_{C(B)}\left(B_{pl}\right)\mathbf{1}(\hat{V}_{pl} \neq \infty)}{c_g\hat{c}_g}\left|(\tilde{G}_n - G_n)\left|g_n\right|_0 + (g_n - \tilde{g}_n)\right|.
\end{aligned}$$

where $\hat{c}_g = \min\left\{\left|\tilde{g}_n\left(B_{pl}\right)\right|\right\} \to c_g > 0$. It follows from Lemma 2.1 and 2.2 that

$$\sup \mathbf{1}_{C(B)}\left(B_{pl}\right)\mathbf{1}(\hat{V}_{pl} \neq \infty)\left|\tilde{\psi}\left(B_{pl}\right) - \psi\left(B_{pl}\right)\right| = O\left(1/r\right).$$

Thus if $L \to \infty$ and $r/n \to 0$ as $n \to \infty$, we have almost surely for any closed inner subset $C(V)$ of $S(F)$,

$$\sup_{pl} \mathbf{1}_{C(V)}(V_{pl}) \left| \hat{V}_{pl} - V_{pl} \right| = O(1/nr), \tag{2.45}$$

and, if $L \to \infty$ and $r/n \to \infty$ as $n \to \infty$, we have almost surely for any closed inner subset $C(V)$ of $S(F)$,

$$\sup_{pl} \mathbf{1}_{C(V)}(V_{pl}) \left| \hat{V}_{pl} - V_{pl} \right| = O\left(1/n^2\right). \tag{2.46}$$

■

Proof of Proposition 2.4. Following Guerre, Perrigne and Vuong (2000), let

$$\tilde{f}(v) = \frac{1}{nL} \sum_{l=1}^{L} \sum_{p=1}^{n} K_R \left(\frac{V_{pl} - v}{h_R} \right) \tag{2.47}$$

be the "infeasible" nonparametric estimator of f using the true private values V_{pl}. Let $C'(V)$ be an inner closed subset of $S(F)$ containing all hypercubes of size δ (small enough) centered at v in $C(V)$. Define $C''(V)$ analogously with respect to $C'(V)$. Hence $C(V) \subset C'(V) \subset C''(V) \subset S(F)$. For $v \in C(V)$ and n, L large enough, $\hat{f}(v)$ uses at most observations $\hat{V}_{pl}$ in $C'(V)$ and hence at most V_{pl} in $C''(V)$ by the uniform convergence of pseudo-private values $\hat{V}_{pl}$ to V_{pl} in Proposition 2.3. Similarly, $\tilde{f}(v)$ uses at most V_{pl} in

$C''(V)$ for any v in $C(V)$. Hence we have almost surely for n, L large enough,

$$
\begin{aligned}
& \left|\hat{f}(v)-\tilde{f}(v)\right| \\
= & \left|\frac{1}{nLh_R}\sum_{l=1}^{L}\sum_{p=1}^{n}\mathbf{1}_{C''(V)}(V_{pl})\left[K_R\left(\frac{v-\hat{V}_{pl}}{h_R}\right)-K_R\left(\frac{v-V_{pl}}{h_R}\right)\right]\right| \\
\leq & \left|\frac{1}{nLh_R}\sum_{l=1}^{L}\sum_{p=1}^{n}\mathbf{1}_{C''(V)}(V_{pl})\frac{(\hat{V}_{pl}-V_{pl})}{h_R}\frac{\partial K_R}{\partial v}\left(\frac{v-V_{pl}}{h_R}\right)\right| \\
& +\frac{1}{2nLh_R}\sum_{l=1}^{L}\sum_{p=1}^{n}\mathbf{1}_{C''(V)}(V_{pl})\frac{(\hat{V}_{pl}-V_{pl})^2}{h_R^2}\left|\frac{\partial^2 K_R}{\partial v^2}(v)\right|_0 \\
\leq & \frac{\sup_{p,l}\mathbf{1}_{C''(V)}(V_{pl})\left|\hat{V}_{pl}-V_{pl}\right|}{h_R}\frac{1}{nLh_R}\sum_{l=1}^{L}\sum_{p=1}^{n}\left|\frac{\partial K_R}{\partial v}\left(\frac{v-V_{pl}}{h_R}\right)\right| \\
& +\frac{\sup_{p,l}\mathbf{1}_{C''(V)}(V_{pl})\left|\hat{V}_{pl}-V_{pl}\right|^2}{2h_R^3}\left|\frac{\partial^2 K_R}{\partial v^2}(v)\right|_0 .
\end{aligned}
$$

Let

$$\tilde{K}(x)=\left|\frac{\partial K_R}{\partial v}(x)\right|/\int\left|\frac{\partial K_R}{\partial v}(u)\right|du.$$

Thus we have almost surely, as $\tilde{K}(x)$ is a well defined kernel,

$$\left|\frac{1}{nLh_R}\sum_{l=1}^{L}\sum_{p=1}^{n}\tilde{K}\left(\frac{v-V_{pl}}{h_R}\right)-f(v)\right|_0 \to 0,$$

which implies $\frac{1}{nLh_R}\sum_{l=1}^{L}\sum_{p=1}^{n}\left|\partial K_R\left(\left(v-V_{pl}\right)/h_R\right)/\partial v\right|$ converges uniformly on $C(V)$ to

$$f(v)\int\left|\frac{\partial K_R}{\partial v}(u)\right|du.$$

Hence $\frac{1}{nLh_R}\sum_{l=1}^{L}\sum_{p=1}^{n}\left|\partial K_R\left(\left(v-V_{pl}\right)/h_R\right)/\partial v\right|$ is bounded almost surely. $\left|\partial^2 K_R(v)/\partial v^2\right|_0$ is bounded by the definition of $K_R(\cdot)$. We consider the following two cases:

(i) $L\to\infty$, and $r/n\to 0$ as $n\to\infty$

From (2.45), we have almost surely

$$\left|\hat{f}(v)-\tilde{f}(v)\right|_{0,C(V)}=O\left(\frac{1}{n}\left(\frac{\log nL}{nL}\right)^{(R-1)/(2R+1)}\right)+O\left(\frac{1}{n^2}\left(\frac{\log nL}{nL}\right)^{(2R-3)/(2R+1)}\right).$$

If $R=1$, then $r/n\to 0$ implies that

$$\left|\hat{f}(v)-\tilde{f}(v)\right|_{0,C(V)}=O\left(\frac{1}{n}\left(\frac{\log nL}{nL}\right)^{(R-1)/(2R+1)}\right).$$

If $R\geq 2$, then $(2R-3)/(2R+1)\geq(R-1)/(2R+1)$, which also implies that

$$\left|\hat{f}(v)-\tilde{f}(v)\right|_{0,C(V)}=O\left(\frac{1}{n}\left(\frac{\log nL}{nL}\right)^{(R-1)/(2R+1)}\right)=O\left(\frac{1}{nrh_R}\right).$$

Since $r/n\to 0$ implies $1/(nh_R)\to 0$, we have almost surely

$$\begin{aligned}\left|\hat{f}(v)-f(v)\right|_{0,C(V)} &\leq \left(\left|\hat{f}(v)-\tilde{f}(v)\right|_{0,C(V)}+\left|\tilde{f}(v)-f(v)\right|_{0,C(V)}\right)\\ &= O\left(\frac{1}{nrh_R}\right)+O\left(\frac{1}{r}\right)=O\left(r^{-1}\right),\end{aligned}$$

where $|\tilde{f}(v)-f(v)|_{0,C(V)}=O\left(r^{-1}\right)$ follows from analogous arguments used in the proof for Lemma 2.2.

(ii) $L\to\infty$, and $r/n\to\infty$ as $n\to\infty$

From (2.46), we have almost surely

$$|\hat{f}(v)-\tilde{f}(v)|_{0,C(V)}=O\left(n^2h_R\right)^{-1}+O(n^4h_R^3)^{-1}.$$

If $(nh_R)^{-1}\to 0$, then $|\hat{f}(v)-\tilde{f}(v)|_{0,C(V)}=O\left(n^2h_R\right)^{-1}$. Hence, if $(r/n)(nh_R)^{-1}\to 0$, we have almost surely that $|\hat{f}(v)-f(v)|_{0,C(V)}=O\left(n^2h_R\right)^{-1}+O\left(r^{-1}\right)=O\left(r^{-1}\right)$; and if $(r/n)(nh_R)^{-1}\to\infty$, we have almost surely that $|\hat{f}(v)-f(v)|_{0,C(V)}=O\left(n^2h_R\right)^{-1}+O\left(r^{-1}\right)=O\left(n^2h_R\right)^{-1}$. On the other hand, if $(nh_R)^{-1}\to\infty$, then $|\hat{f}(v)-\tilde{f}(v)|_{0,C(V)}=$

$O(n^4 h_R^3)^{-1}$. We have almost surely that $|\hat{f}(v) - f(v)|_{0,C(V)} = O(n^4 h_R^3)^{-1} + O\left(r^{-1}\right) = O(n^4 h_R^3)^{-1}$. ■

Proof of Proposition 2.6. Trivial extension of Proposition 2 in Guerre, Perrigne and Vuong (2000). ■

To prove Proposition 2.7 we need an auxiliary lemma on the uniform convergence of $\tilde{G}_n(b, x, i)$ $\tilde{g}_n(b, x, i)$ and $\tilde{f}(v, x)$ defined in (2.8) (2.9) and (2.55).

Lemma 2.3 *Suppose A1-A4 hold, and $L \to \infty$, $(nh_g)^{-1} \to 0$ as $n \to \infty$. We have almost surely*

$$\begin{aligned} |\tilde{G}_n(b, x, i) - G_n(b, x, i)|_{0,C} &= O(1/r_G), \\ |\tilde{g}_n(b, x, i) - g_n(b, x, i)|_{0,C} &= O\left(1/r_g\right) \\ |\tilde{f}(v, x) - f(v, x)|_{0,C} &= O\left(1/r_f\right), \end{aligned}$$

where C is an arbitrary closed inner subset of $S(G_\infty)$, $r_G = (L/\log L)^{(R+1)/(2R+d+2)}$, and $r_g = r_f = (L/\log L)^{R/(2R+d)}$.

Proof. The proof relies on the argument of Guerre, Perrigne and Vuong's (2000) proof for the case of fixed n. However the problem is different as we are interested in the asymptotic properties of the estimators allowing both n and L to approach infinity. The arguments are more involved here, because $G_n(\cdot, \cdot, \cdot)$ shifts with sample size, and (B_{pl}, X_l, I_l) and (V_{pl}, X_l) observed in the same auction are correlated. We divide the proof into three steps. The first step studies the uniform bias, the second step studies the uniform variance bound, and the last step establishes exponential-type inequality. As the proofs are similar, we only detail the proof for $\tilde{g}_n(\cdot, \cdot, \cdot)$, as it is the most different from Guerre, Perrigne and Vuong's (2000) proof. The sup-norm is taken over the whole support of the function unless otherwise indicated.

Step 1: Uniform Bias

For any closed inner subset C of $S(G_\infty)$, let $n_C = \min\{n : C \subset S_n^o\}$ where S_n^o is the interior of $S(G_n)$. n_C exists because of Proposition 2.5(i). For $n > n_C$,

$$\begin{aligned} E\tilde{g}_n(b,x,i) &= E\left[\frac{1}{h_g^{d+1}} K_g\left(\frac{B_p - b}{h_g}, \frac{X-x}{h_g}, 0\right) \mathbf{1}(I=i)\right] \\ &= \int\int K_g(u,y,0)\, g_n\left(b + h_g u, x + h_g y, i\right) dudy. \end{aligned}$$

Define $\gamma(t) = g_n\left(b + th_g u, x + th_g y, i\right) - g_n(b,x,i)$ for $t \in [0,1]$. For L large enough, $\left(b + th_g u, x + th_g y\right) \in (b,x) + S\left(h_g\right) \subset C_i'$ for $(b,x,i) \in C$ and $t \in [0,1]$, where C_i' is a closed inner subset of $S(G_n(\cdot,\cdot,i))$. Since $g_n(\cdot,\cdot,\cdot)$ admits up to R continuous bounded derivatives with

$$|\gamma|_{R,[0,1]} \leq h_g^R \|(u,y)\|^R |g_n|_{R,C'}.$$

Thus a Taylor expansion gives

$$\gamma(1) - \gamma(0) \leq \gamma^{(1)}(0) + \cdots + \frac{1}{(R-1)!}\gamma^{(R)}(0) + \frac{1}{R!}|\gamma|_{R,[0,1]},$$

where $\gamma^{(r)}(0)$ is a polynomial of order r in (u,y). As $K_g(\cdot,\cdot)$ is of order R, moments of order strictly smaller than R vanish. It follows that

$$\begin{aligned} & |E\tilde{g}_n(b,x,i) - g_n(b,x,i)|_{0,C} \\ = & \left|\int K_g(u,y,0)(\gamma(1) - \gamma(0))\, dudy\right| \leq h_g^R \frac{1}{R!} |g_n|_{R,C'} \int \|(u,y)\|^R \left|K_g(u,y,0)\right| dudy \\ = & h_g^R M_g^R |g_n|_{R,C'} = \lambda_g^R M_g^R |g_n|_{R,C'} / r_g, \end{aligned} \tag{2.48}$$

where $M_g^R = (1/R!) \int \|(u,y)\|^R \left|K_g(u,y,0)\right| dudy$.

Step 2: Uniform Variance

For $(b, x, i) \in C$, we have

$$
\begin{aligned}
& Var\left(\tilde{g}_n(b, x, i)\right) \\
= & Var\left(\frac{1}{Lh_g^{d+1}} \sum_{l=1}^{L} \frac{\mathbf{1}(I_l = i)}{ni} \sum_{p=1}^{ni} K_g\left(\frac{B_{pl} - b}{h_g}, \frac{X_l - x}{h_g}, 0\right)\right) \\
= & \frac{1}{Lh_g^{2(d+1)}} Var\left(\frac{\mathbf{1}(I = i)}{ni} \sum_{p=1}^{ni} K_g\left(\frac{B_{pl} - b}{h_g}, \frac{X_l - x}{h_g}, 0\right)\right) \\
\leq & \frac{1}{Lh_g^{2(d+1)}} E\left(\frac{\mathbf{1}(I = i)}{ni} \sum_{p=1}^{ni} K_g\left(\frac{B_{pl} - b}{h_g}, \frac{X_l - x}{h_g}, 0\right)\right)^2 \\
= & \frac{1}{Lh_g^{2(d+1)}} E\left(\frac{\mathbf{1}(I = i)}{(ni)^2} \sum_{p=1}^{ni} K_g^2\left(\frac{B_{pl} - b}{h_g}, \frac{X_l - x}{h_g}, 0\right)\right) \\
& + \frac{1}{Lh_g^{2(d+1)}} E\left(\frac{\mathbf{1}(I = i)}{(ni)^2} \sum_{p=1}^{ni} \sum_{q=1, q \neq p}^{ni} K_g\left(\frac{B_{pl} - b}{h_g}, \frac{X_l - x}{h_g}, 0\right) K_g\left(\frac{B_{ql} - b}{h_g}, \frac{X_l - x}{h_g}, 0\right)\right) \\
= & \frac{1}{(ni)Lh_g^{d+1}} \int K_g^2(u, y, 0)\, g_n\left(b + h_g u, x + h_g y, i\right) du dy \\
& + \frac{ni - 1}{(ni)Lh_g^d} \int K_g(u_1, y, 0)\, K_g(u_2, y, 0)\, g_{n,B|(X,I)}\left(b + h_g u_1 | x + h_g y, i\right) \\
& \times g_{n,B|(X,I)}\left(b + h_g u_2 | x + h_g y, i\right) g_{n,(X,I)}\left(x + h_g y, i\right) du_1 du_2 dy.
\end{aligned}
$$

Let $Q_{g1} = \int K_g^2(u, y, 0)\, dudy$ and $Q_{g2} = \int K_g(u_1, y, 0)\, K_g(u_2, y, 0)\, du_1 du_2 dy$.

$$
\left|Var\left(\tilde{g}_n(b, x, i)\right)\right|_{0,C} \leq \frac{Q_{g1} |g_n|_0}{(ni)Lh_g^{d+1}} + \frac{(ni - 1) Q_{g2} \left|g_{n,B|(X,I)}^2 g_{n,(X,I)}\right|_0}{(ni)Lh_g^d}. \tag{2.49}
$$

Step 3: Exponential-type Inequality

In this step, we establish the exponential-type inequalities for the probabilities of deviations of $\tilde{g}_n(b, x, i) - g_n(b, x, i)$ in sup-norm over C_i, where $C_i = \{(b, x) : (b, x, i) \in C\}$. Let C_i be covered by T inner "balls" of the form

$$
B_{it} \equiv B_i\left((b_t, x_t); \Delta\right) = \left\{(b, x) \in S\left(G_\infty\right) : b \in \left[b_t - \Delta, b_t + \Delta\right], x \in \left[x_t - \Delta, x_t + \Delta\right]\right\},
$$

where $(b_t, x_t) \in C_i$, and $\Delta > 0$ for $t = 1, \cdots, T$. Moreover, we consider minimal coverings for C_i, i.e., coverings for which T is the smallest number denoted by $T(C_i, \Delta)$.

Let

$$e_g(\iota,\tau) = \iota + 2(d+1)\tau \left|K_g\right|_1 + \lambda_g^R M_g^R \left|g_n\right|_{R,C'},$$

$$P_g(\iota,\tau) = 2T\left(C_i, \tau h_g^{d+2}/r_g\right)$$

$$\times \exp\left(-\frac{\lambda_g^d \iota^2 \log L}{2Q_{g1}\left|g_n\right|_0/(nih_g) + 2(1+1/ni)Q_{g2}\left|g_{n,B|(X,I)}^2 g_{n,(X,I)}\right|_0 + 4\iota\left|K_g\right|_0/\left(3r_g h_g\right)}\right),$$

where ι and τ are strictly positive constants.

Step 3(a): From (2.48) and the triangular inequality, we obtain

$$\begin{aligned}
& \Pr\left(r_g\left|\tilde{g}_n - g_n\right|_{0,C} > e_g(\iota,\tau)\right) \\
\leq\ & \Pr\left(r_g\left|\tilde{g}_n - E\tilde{g}_n\right|_{0,C} + r_g\left|E\tilde{g}_n - g_n\right|_{0,C} > e_g(\iota,\tau)\right) \\
\leq\ & \Pr\left(r_g\left|\tilde{g}_n - E\tilde{g}_n\right|_{0,C} > e_g(\iota,\tau) - \lambda_g^R M_g^R \left|g_n\right|_{R,C'}\right). \qquad (2.50)
\end{aligned}$$

Let $\tilde{g}_n(b,x,i) - E\tilde{g}_n(b,x,i) = (1/L)\sum_{m=1}^{L}\zeta_{mL}(b,x,i)$, where

$$\begin{aligned}
\zeta_{mL}(b,x,i) &= \frac{1}{nih_g^{d+1}}\sum_{p=1}^{ni}\left\{K_g\left(\frac{B_{pm}-b}{h_g}, \frac{X_m - x}{h_g}, 0\right)\mathbf{1}(I_m = i)\right. \\
&\quad \left. - E\left(K_g\left(\frac{B_p - b}{h_g}, \frac{X-x}{h_g}, 0\right)\mathbf{1}(I=i)\right)\right\}.
\end{aligned}$$

As the ζ_{mL}'s are independent zero-mean variables for $m = 1, \cdots, L$, it follows from (2.49)

$$\begin{aligned}
Var\left(r_g\zeta_{mL}(b,x,i)\right) &= Lr_g^2 Var(\tilde{g}_n) \leq \frac{Lr_g^2 Q_{g1}\left|g_n\right|_0}{(nih_g)Lh_g^d} + \frac{(ni-1)Lr_g^2 Q_{g2}\left|g_{n,B|(X,I)}^2 g_{n,(X,I)}\right|_0}{(ni)Lh_g^d} \\
&= \frac{LQ_{g1}\left|g_n\right|_0}{(nih_g)\lambda_g^d \log L} + \frac{(ni-1)LQ_{g2}\left|g_{n,B|(X,I)}^2 g_{n,(X,I)}\right|_0}{(ni)\lambda_g^d \log L}.
\end{aligned}$$

By the triangular inequality we have

$$\left|r_g\zeta_{mL}(b,x,i)\right| \leq \frac{2r_g}{h_g^{d+1}}\left|K_g\right|_0 = \frac{2L\left|K_g\right|_0}{\lambda_g^d r_g h_g \log L}.$$

Hence the Bernstein inequality gives

$$
\begin{aligned}
& \Pr\left(r_g \left|\tilde{g}_n(b,x,i) - E\tilde{g}_n(b,x,i)\right| > \iota\right) \\
= & \Pr\left(\left|\sum_{m=1}^{L} r_g \zeta_{mL}(b,x,i) - \sum_{m=1}^{L} E\left(r_g \zeta_{mL}(b,x,i)\right)\right| > L\iota\right) \\
\leq & 2\exp\left(-\frac{L^2\iota^2}{2\sum_{m=1}^{L} Var\left(r_g\zeta_{mL}\right) + 4L^2\iota \left|K_g\right|_0 / \left(3\lambda_g^d r_g h_g \log L\right)}\right) \\
\leq & 2\exp\left(-\frac{\lambda_g^d \iota^2 \log L}{2Q_{g1}\left|g_n\right|_0/(nih_g) + 2(1+1/ni)Q_{g2}\left|g^2_{n,B|(X,I)} g_{n,(X,I)}\right|_0 + 4\iota \left|K_g\right|_0 / \left(3r_g h_g\right)}\right) \\
= & \frac{P_g(\iota,\tau)}{T\left(C_i, \tau h_g^{d+2}/r_g\right)},
\end{aligned}
$$

for any $(b,x,i) \in C$, ι, n, and L.

Step 3(b): Consider a minimal covering of C for some $\Delta > 0$. For any $b \in B_t$, we have

$$
\begin{aligned}
r_g \left|\tilde{g}_n(b,x,i) - E\tilde{g}_n(b,x,i)\right|_{0,C_i} \leq & \sup_{1\leq t\leq T}\left|\frac{r_g}{L}\sum_{m=1}^{L}\zeta_{mL}(b_t,x_t,i)\right| \\
& + \sup_{1\leq t\leq T}\sup_{(b,x)\in B_{it}}\left|\frac{r_g}{L}\sum_{m=1}^{L}\left(\zeta_{mL}(b_t,x_t,i) - \zeta_{mL}(b,x,i)\right)\right|.
\end{aligned}
$$

This gives

$$
\begin{aligned}
& \Pr\left(r_g\left|\tilde{g}_n(b,x,i) - E\tilde{g}_n(b,x,i)\right|_{0,C_i} > e(\iota,\tau) - \lambda_g^R M_g^R \left|g_n\right|_{R,C'}\right) \\
\leq & \Pr\left(\sup_{1\leq t\leq T}\sup_{(b,x)\in B_{it}}\left|\frac{r_g}{L}\sum_{m=1}^{L}\left(\zeta_{mL}(b_t,x_t,i) - \zeta_{mL}(b,x,i)\right)\right| > e(\iota,\tau) - \iota - \lambda_g^R M_g^R \left|g_n\right|_{R,C'}\right) \\
& + \Pr\left(r_g \sup_{1\leq t\leq T}\left|\tilde{g}_n(b_t,x_t,i) - E\tilde{g}_n(b_t,x_t,i)\right| > \iota\right). \qquad (2.51)
\end{aligned}
$$

For any $(b,x) \in B_{it}$, it follows from the mean value theorem

$$
\left|\frac{1}{h_g^{d+1}}K_g\left(\frac{B-b_t}{h_g}, \frac{X-x_t}{h_g}, 0\right) - \frac{1}{h_g^{d+1}}K_g\left(\frac{B-b}{h_g}, \frac{X-x}{h_g}, 0\right)\right| \leq \frac{(d+1)\Delta}{h_g^{d+2}}\left|K_g\right|_1.
$$

The triangular inequality gives

$$
\left|\left(\zeta_{mL}(b_t,x_t,i) - \zeta_{mL}(b,x,i)\right)\right| \leq \frac{(d+1)\Delta}{h_g^{d+2}}\left|K_g\right|_1 + E\frac{(d+1)\Delta}{h_g^{d+2}}\left|K_g\right|_1 = \frac{2(d+1)\Delta}{h_g^{d+2}}\left|K_g\right|_1.
$$

Step 3(c): Let $\Delta = \tau h_g^{d+2}/r_g$, it follows that

$$\sup_{1\le t\le T}\sup_{b\in B_t}\left|\frac{r_g}{L}\sum_{m=1}^{L}\left(\zeta_{mL}\left(b_t,x_t,i\right)-\zeta_{mL}\left(b,x,i\right)\right)\right| \le \frac{2r_g(d+1)\Delta}{h_g^{d+2}}\left|K_g\right|_1 = 2(d+1)\tau\left|K_g\right|_1.$$

Hence

$$\Pr\left(\sup_{1\le t\le T}\sup_{b\in B_t}\left|\frac{r_g}{L}\sum_{m=1}^{L}\left(\zeta_{mL}\left(b_t,x_t,i\right)-\zeta_{mL}\left(b,x,i\right)\right)\right| > e\left(\iota,\tau\right)-\iota-\lambda_g^R M_g^R\left|g_n\right|_{R,C'}\right)=0. \tag{2.52}$$

Then it follows from (2.50), (2.51), (2.52), and the Bernstein inequality that

$$\begin{aligned}
&\Pr\left(r_g\left|\tilde{g}_n-g_n\right|_{0,C} > e_g\left(\iota,\tau\right)\right)\\
\le\ &\Pr\left(r_g\left|\tilde{g}_n\left(b,x,i\right)-E\tilde{g}_n\left(b,x,i\right)\right|_{0,C_i} > e\left(\iota,\tau\right)-\lambda_g^R M_g^R\left|g_n\right|_{R,C'}\right)\\
\le\ &\Pr\left(r_g\sup_{1\le t\le T}\left|\tilde{g}_n\left(b_t,x_t,i\right)-E\tilde{g}_n\left(b_t,x_t,i\right)\right| > \iota\right)\\
\le\ &\sum_{t=1}^{T}\Pr\left(r_g\left|\tilde{g}_n\left(b_t,x_t,i\right)-E\tilde{g}_n\left(b_t,x_t,i\right)\right| > \iota\right) \le P\left(\iota,\tau\right).
\end{aligned}$$

As the dimension of the covered set C is $d+1$, the covering number $T\left(C,\Delta\right)$ is of order $\Delta^{-(d+1)}$. Hence $T\left(C,\tau h_g^{d+2}/r_g\right) = O(L/\log L)^{(d+1)(R+d+2)/(2R+d)}$. By taking ι sufficiently large, $P\left(\iota,\tau\right)$ converges as $L\to\infty$. The desired result follows from the Borel-Cantelli lemma and the fact that $e\left(\iota,\tau\right)=O\left(1\right)$. ■

Proof of Proposition 2.7. First, the uniform consistency of pseudo-private values follows from similar arguments as used in the proof of Proposition 2.3. If $L\to\infty$, $r_g/n\to 0$ as $n\to\infty$, we have almost surely for any closed inner subset $C\left(V\right)$ of $S\left(F\right)$,

$$\sup_{pl}\mathbf{1}_{C(V)}(V_{pl},X_l)\left|\hat{V}_{pl}-V_{pl}\right| = O\left(1/nr_g\right), \tag{2.53}$$

and, if $L\to\infty$, $r_g/n\to\infty$ as $n\to\infty$, we have almost surely for any closed inner subset $C\left(V\right)$ of $S\left(F\right)$,

$$\sup_{pl}\mathbf{1}_{C(V)}(V_{pl},X_l)\left|\hat{V}_{pl}-V_{pl}\right| = O\left(1/n^2\right). \tag{2.54}$$

To establish the uniform consistency of the two-step estimator, let

$$\tilde{f}(v,x) = \frac{1}{nLh_f^{d+1}} \sum_{l=1}^{L} \frac{1}{I_l} \sum_{p=1}^{nI_l} K_f\left(\frac{V_{pl}-v}{h_f}, \frac{X_l - x}{h_f}\right) \tag{2.55}$$

be the "infeasible" nonparametric estimator of f using the true private values V_{pl}. Let $C'(V)$ be an inner closed subset of $S(F)$ containing all hypercubes of size δ (small enough) centered at (v,x) in $C(V)$. Define $C''(V)$ analogously with respect to $C'(V)$. Hence $C(V) \subset C'(V) \subset C''(V) \subset S(F)$. For $(v,x) \in C(V)$ and n, L large enough, $\hat{f}(v,x)$ uses at most observations $(\hat{V}_{pl}, X_l)$ in $C'(V)$ and hence at most (V_{pl}, X_l) is in $C''(V)$ by the uniform convergence of pseudo-private values $\hat{V}_{pl}$ to V_{pl}. Similarly, $\tilde{f}(v,x)$ uses at most (V_{pl}, X_l) in $C''(V)$ for any (v,x) in $C(V)$. Hence we have almost surely for n, L large enough,

$$\begin{aligned}
& \left|\hat{f}(v,x) - \tilde{f}(v,x)\right| \\
= & \left|\frac{1}{nLh_f^{d+1}} \sum_{l=1}^{L} \frac{1}{I_l} \sum_{p=1}^{nI_l} \mathbf{1}_{C''(V)}(V_{pl}, X_l)\left[K_f\left(\frac{\hat{V}_{pl}-v}{h_f}, \frac{X_l-x}{h_f}\right) - K_f\left(\frac{V_{pl}-v}{h_f}, \frac{X_l-x}{h_f}\right)\right]\right| \\
\leq & \left|\frac{1}{nLh_f^{d+1}} \sum_{l=1}^{L} \frac{1}{I_l} \sum_{p=1}^{nI_l} \mathbf{1}_{C''(V)}(V_{pl}, X_l) \frac{(\hat{V}_{pl}-V_{pl})}{h_f} \frac{\partial K_f}{\partial v}\left(\frac{V_{pl}-v}{h_f}, \frac{X_l-x}{h_f}\right)\right| \\
& + \frac{1}{2nLh_f^{d+1}} \sum_{l=1}^{L} \frac{1}{I_l} \sum_{p=1}^{nI_l} \mathbf{1}_{C''(V)}(V_{pl}, X_l) \frac{(\hat{V}_{pl}-V_{pl})^2}{h_f^2} \left|\frac{\partial^2 K_f}{\partial v^2}\left(v, \frac{X_l-x}{h_f}\right)\right|_0 \\
\leq & \frac{\sup_{p,l} \mathbf{1}_{C''(V)}(V_{pl}, X_l)\left|\hat{V}_{pl}-V_{pl}\right|}{h_f} \frac{1}{nLh_f^{d+1}} \sum_{l=1}^{L} \frac{1}{I_l} \sum_{p=1}^{nI_l} \left|\frac{\partial K_f}{\partial v}\left(\frac{V_{pl}-v}{h_f}, \frac{X_l-x}{h_f}\right)\right| \\
& + \frac{\sup_{p,l} \mathbf{1}_{C''(V)}(V_{pl}, X_l)\left|\hat{V}_{pl}-V_{pl}\right|^2}{2h_f^3} \frac{1}{Lh_f^d} \sum_{l=1}^{L} \left|\frac{\partial^2 K_f}{\partial v^2}\left(v, \frac{X_l-x}{h_f}\right)\right|_0 .
\end{aligned}$$

The two sums may be viewed as kernel estimators and hence uniformly bounded on $C(V)$. We consider the following two cases:

(i) $L \to \infty$, and $r_f/n \to 0$ as $n \to \infty$

From (2.53), we have almost surely

$$\left|\hat{f}(v,x)-\tilde{f}(v,x)\right|_{0,C(V)}=O\left(\frac{1}{n}\left(\frac{\log L}{L}\right)^{(R-1)/(2R+d)}\right)+O\left(\frac{1}{n^2}\left(\frac{\log L}{L}\right)^{(2R-3)/(2R+d)}\right).$$

If $R=1$, then $r_f/n\to 0$ implies that

$$\left|\hat{f}(v,x)-\tilde{f}(v,x)\right|_{0,C(V)}=O\left(\frac{1}{n}\left(\frac{\log nL}{nL}\right)^{(R-1)/(2R+1)}\right).$$

If $R\geq 2$, then $(2R-3)/(2R+d)\geq (R-1)/(2R+d)$, which also implies that

$$\left|\hat{f}(v,x)-\tilde{f}(v,x)\right|_{0,C(V)}=O\left(\frac{1}{n}\left(\frac{\log nL}{nL}\right)^{(R-1)/(2R+d)}\right)=O\left(\frac{1}{nr_fh_f}\right).$$

Since $r_f/n\to 0$ implies $1/\left(nh_f\right)\to 0$, we have almost surely

$$\begin{aligned}\left|\hat{f}(v,x)-f(v,x)\right|_{0,C(V)} &\leq \left(\left|\hat{f}(v,x)-\tilde{f}(v,x)\right|_{0,C(V)}+\left|\tilde{f}(v,x)-f(v,x)\right|_{0,C(V)}\right)\\ &= O\left(\frac{1}{nr_fh_f}\right)+O\left(\frac{1}{r_f}\right)=O(r_f^{-1}),\end{aligned}$$

where $|\tilde{f}(v,x)-f(v,x)|_{0,C(V)}=O(r_f^{-1})$ follows from Lemma 2.3.

(ii) $L\to\infty$, and $r_f/n\to\infty$ as $n\to\infty$

From (2.54), we have almost surely

$$|\hat{f}(v,x)-\tilde{f}(v,x)|_{0,C(V)}=O\left(n^2h_f\right)^{-1}+O(n^4h_f^3)^{-1}=O\left(n^2h_f\right)^{-1},$$

as $\left(nh_f\right)^{-1}\to 0$. Hence, if $(r_f/n)\left(nh_f\right)^{-1}\to 0$, we have almost surely that $|\hat{f}(v,x)-f(v,x)|_{0,C(V)}=O\left(n^2h_f\right)^{-1}+O(r_f^{-1})=O(r_f^{-1})$; and if $(r_f/n)\left(nh_f\right)^{-1}\to\infty$, we have almost surely that $|\hat{f}(v,x)-f(v,x)|_{0,C(V)}=O\left(n^2h_f\right)^{-1}+O(r_f^{-1})=O\left(n^2h_f\right)^{-1}$.

■

CHAPTER 3

IMPROVED EFFICIENT QUASI MAXIMUM LIKELIHOOD ESTIMATOR OF SPATIAL AUTOREGRESSIVE MODELS

3.1 Introduction

In this chapter, we derive the best generalized method of moments estimators (BGMME) for the regression model with spatial autoregressive (SAR) disturbances and the mixed regressive spatial autoregressive (MRSAR) model, within the class of generalized method of moments estimators (GMME) based on linear and quadratic moment conditions. The BGMME proposed here has the merit of computational simplicity and asymptotic efficiency. It is asymptotically as efficient as the maximum likelihood estimator (MLE) when the disturbances are normally distributed, and asymptotically more efficient than the quasi maximum likelihood estimator (QMLE) otherwise.

The generalized method of moments (GMM) by Hansen (1982) has been noted for its possible use for the estimation of spatial autoregressive (SAR) models in the presence of exogenous regressors, e.g., Anselin (1988; 1990), Land and Deane (1992), Kelejian and Robinson (1993), Kelejian and Prucha (1997; 1998), and Lee (2003), among others. Those GMM methods are 2SLS methods as their moment conditions are based on exogenous regressors (and spatial weights matrices) in the model and all the instrumental variables (IV) used are generated from them. The 2SLS estimators have been shown to be consistent and asymptotically normally distributed (Kelejian and Prucha, 1998), but not efficient relative to MLE when errors are normally distributed. And the 2SLS methods would not be consistent when all the exogenous regressors in the MRSAR model are really irrelevant.

Kelejian and Prucha (1999) propose a method of moments (MOM) for the regression model with SAR disturbances. Their MOM estimator is consistent but inefficient as compared to the MLE. Lee (2001*a*) generalizes the MOM procedure for the estimation of the regression model with SAR disturbances into a systematic GMM framework and shows the existence of BGMME in the case of normally distributed disturbances. The GMM framework is further extended for the estimation of the MRSAR model in Lee (2006), based on a combination of the moments in the 2SLS framework with some modified moment functions originated from the estimation of the regression model with SAR disturbances. Lee (2006) shows that the proposed GMME can be asymptotically more efficient than the 2SLS estimators, and the BGMME exists in the case with normally distributed disturbances. As Lee's (2006) BGMME has the same limiting distribution as the MLE under normality, it is unlikely to be efficient when the disturbances are not normally distributed. Here, we show the existence of distributionally free BGMME within the class of GMME based on the linear and quadratic moments of the disturbances.[23]

This chapter is organized as follows. In Section 2, we consider the GMM estimation of the regression model with SAR disturbances and the MRSAR model respectively. The best selection of moment functions and optimal IVs will be discussed and the possible efficiency property is derived. All the proofs of the results are collected in the appendices. Section 3 provides some Monte Carlo results for the comparison of finite sample properties of estimators. Section 4 briefly concludes.

[23] A preliminary investigation of possible BGMME which may improve upon the QMLE is in Bollinger (2001).

3.2 GMM Estimation and the BGMME

3.2.1 GMM Estimation of the Regression Model with SAR Disturbances

The regression model with SAR disturbances is specified as

$$\begin{aligned} Y_n &= X_n\beta + u_n, \\ u_n &= \lambda W_n u_n + \epsilon_n, \end{aligned} \tag{3.1}$$

where n is the total number of spatial units, X_n is an $n \times k$ dimensional matrix of nonstochastic exogenous variables, W_n is an $n \times n$ dimensional spatial weights matrix of known constants with a zero diagonal, and the disturbances $\epsilon_{n1}, \cdots, \epsilon_{nn}$ of the n-dimensional vector ϵ_n are i.i.d. $(0, \sigma^2)$. The $W_n u_n$ in (3.1) is called a spatial lag and its coefficient is supposed to represent the spatial effect due to the influence of neighboring units on a single spatial unit. In order to distinguish the true parameters from other possible values in the parameter space, we denote β_0, λ_0, and σ_0^2 as the true parameters that generate the observed sample. For any possible value λ, denote $S_n(\lambda) = I_n - \lambda W_n$. At λ_0, $S_n = S_n(\lambda_0)$ for simplicity. This model is supposed to be an equilibrium model.

Equation (3.1) implies that

$$Y_n = X_n\beta_0 + S_n^{-1}\epsilon_n. \tag{3.2}$$

The regression model is a generalized linear model with variance $\sigma_0^2 S_n^{-1} S_n'^{-1}$ for the disturbance vector u_n. A possible estimator of β_0 is the feasible generalized least squares estimator (GLSE) $\hat{\beta}_{FG} = (X_n'\hat{S}_n'\hat{S}_n X_n)^{-1} X_n'\hat{S}_n'\hat{S}_n Y_n$ with a consistently estimated weighting matrix. In order to estimate $S_n'S_n$, one needs to estimate the unknown parameter λ_0 in the SAR disturbance process.

Let $\hat{\beta}_L = \left(X_n' X_n\right)^{-1} X_n' Y_n$ be the ordinary least square estimator (OLSE). The disturbance vector u_n can be estimated by the estimated residual $\hat{u}_n = Y_n - X_n \hat{\beta}_L$. And following Lee (2001*a*), λ_0 can then be estimated by the GMM:

$$\min_{\lambda} g_n' (\lambda) a_n' a_n g_n (\lambda), \tag{3.3}$$

based on the quadratic moment conditions of ϵ_n

$$g_n (\lambda) = [P_{1n} S_n (\lambda) \hat{u}_n, \cdots, P_{mn} S_n (\lambda) \hat{u}_n]' S_n (\lambda) \hat{u}_n, \tag{3.4}$$

where P_{jn}'s are $n \times n$ dimensional constant matrices such that $tr\left(P_{jn}\right) = 0$ $(j = 1, \cdots, m)$.[24] For rigorous analysis, the following regularity assumptions for the GMM estimation are specified in Lee (2001*a*; 2006).

Assumption 1 *The* ϵ_{ni}*'s are i.i.d. with zero mean, variance* σ_0^2 *and that a moment of order higher than the fourth exists.*

Assumption 2 *The elements of* X_n *are uniformly bounded constants,* X_n *has the full rank k, and* $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ *exists and is nonsingular.*

Assumption 3 *The spatial weights matrices* $\{W_n\}$ *and* $\{S_n^{-1}\}$ *are uniformly bounded in both row and column sums in absolute value.*[25]

Assumption 4 *The matrices* P_n*'s with* $tr\,(P_n) = 0$ *are uniformly bounded in both row and column sums in absolute value.*

The higher than the fourth moment condition in Assumption 1 is needed in order to apply a central limit theorem due to Kelejian and Prucha (2001). In general, denote μ_3 and

[24] By selecting P_j such that $tr\left(P_j\right) = 0$, σ^2 is concentrated out from the objective function (3.3), so that the dimension of the parameter space is reduced and the estimate of σ^2 (in a subsequent step) is guaranteed to be positive. Furthermore, by comparing the asymptotic covariance matrix with that from joint estimation of λ and σ^2, no efficiency loss in the estimation of λ is incurred by concentrating σ^2 out. Detailed discussion is given in Appendix 3.5.4.

[25] A sequence of square matrices $\{A_n\}$, where $A_n = [a_{n,ij}]$, is said to be uniformly bounded in row sums (column sums) in absolute value if the sequence of row sum matrix norm $||A_n||_\infty = \max_{i=1,\cdots,n} \sum_{j=1}^{n} |a_{n,ij}|$ (column sum matrix norm $||A_n||_1 = \max_{j=1,\cdots,n} \sum_{i=1}^{n} |a_{n,ij}|$) are bounded. (Horn and Johnson, 1985)

μ_4 be respectively, the third and fourth moments of ϵ_{ni}'s. Assumption 3 limits the spatial dependences among the units to a tractable degree and is originated by Kelejian and Prucha (1999). It rules out the unit root case (in time series as a special case). Uniform boundedness conditions for X_n and P_n's in Assumption 2 and 4 are for analytical tractability. Let $G_n = W_n S_n^{-1}$, and $A^s = A + A'$ for any square matrix A. (A list of special notations used for this chapter has been collected in the Appendix for convenient reference.) Assumption 5 summarizes some sufficient conditions for the identification of λ_0.

Assumption 5 $\lim_{n\to\infty} \frac{1}{n} tr(P_{jn}G_n) \neq 0$ *for some* $j = 1, \cdots, m$, *and*

$$\lim_{n\to\infty} \frac{1}{n}(tr(P_{1n}^s G_n), \cdots, tr(P_{mn}^s G_n))'$$

is linearly independent of $\lim_{n\to\infty} \frac{1}{n}(tr(G_n' P_{1n} G_n), \cdots, tr(G_n' P_{mn} G_n))'$.

Let $\mathbf{\Omega}_n = var\left(g_n\left(\lambda_0\right)\right)$. The variance matrix $\mathbf{\Omega}_n$ is assumed to satisfy some conventional regularity conditions in Assumption 6. And the parameter space is assumed to be a compact set as usual for nonlinear estimation.

Assumption 6 *The limit of* $\frac{1}{n}\mathbf{\Omega}_n$ *exists and is a nonsingular matrix.*

Assumption 7 *The* λ_0 *is in the interior of the parameter space* Λ, *which is a compact subset of the real line.*

Interested readers may refer to Lee (2001*a*; 2006) for detailed discussions on the regularity assumptions.[26] Lee (2001*a*) shows the GMME $\hat{\lambda}_P$ is $\sqrt{n}$-consistent and it has the limiting distribution of the corresponding GMME of the SAR process for u_n as if u_n is observable. Furthermore, with a consistent estimator of λ_0, the feasible GLSE $\hat{\beta}_{FG} = (X_n' \hat{S}_n' \hat{S}_n X_n)^{-1} X_n' \hat{S}_n' \hat{S}_n Y_n$ is asymptotically equivalent to the exact GLSE

$$\hat{\beta}_G = \left(X_n' S_n' S_n X_n\right)^{-1} X_n' S_n' S_n Y_n.$$

[26]Assumptions 5 and 6 exclude the case of large (group) interactions in Lee (2004). These can simplify the presentation of our results. The cases under our assumptions here are relevant to spatial scenario, where interactions are usually among a few neighbors.

First, we consider the case that u_n is observable, and we will discuss the feasibility issue later. The optimal choice of the weighting matrix $a_n' a_n$ in (3.3) is, as usual, the inverse of a matrix proportional to the variance matrix of $g_n(\lambda_0)$. Let $\mathcal{M}_n = \{\hat{\lambda}_O\}$ be the class of optimal GMMEs derived from $\min_{\lambda \in \Lambda} g_n'(\lambda)\,\Omega_n^{-1} g_n(\lambda)$, where $g_n(\lambda) = [P_{1n} S_n(\lambda) u_n, \cdots, P_{mn} S_n(\lambda) u_n]' S_n(\lambda) u_n$ is a vector of moment functions with P_n's satisfying Assumption 4. We are interested in the BGMME within the class of optimal GMMEs $\mathcal{M}_n$.

Following Lee (2001*a*), the limiting variance of the consistent GMME $\hat{\lambda}_P$ based on the quadratic moment $\epsilon_n' P_n \epsilon_n$ with $tr(P_n) = 0$ is

$$\Sigma_P^{-1} = \lim_{n \to \infty} \left[(\eta_4 - 3)\frac{\sum_{i=1}^n p_{n,ii}^2}{\frac{1}{n} tr^2(P_n^s G_n)} + \frac{tr(P_n P_n^s)}{\frac{1}{n} tr^2(P_n^s G_n)}\right],$$

with $\eta_4 = \mu_4/\sigma_0^4$ being the kurtosis of the disturbance. The search of a best quadratic moment is to find the P_n with $tr(P_n) = 0$ which minimizes the variance Σ_P^{-1}. Equivalently, one may maximize the corresponding precision measure, i.e., consider

$$\max_{P_n} \frac{\frac{1}{n} tr^2(P_n^s G_n)}{(\eta_4 - 3)\sum_{i=1}^n p_{n,ii}^2 + tr(P_n P_n^s)}.$$

Let $D(A)$ be a diagonal matrix with diagonal elements being A if A is a vector, or diagonal elements of A if A is a square matrix. Note that

$$tr(P_n^s P_n) - tr[(P_n - D(P_n))^s P_n] = 2tr[D(P_n) \cdot P_n] = 2\sum_{i=1}^n p_{n,ii}^2.$$

Hence,

$$\begin{aligned}(\eta_4 - 3)\sum_{i=1}^n p_{n,ii}^2 + tr(P_n^s P_n) &= (\eta_4 - 1)\sum_{i=1}^n p_{n,ii}^2 + tr[(P_n - D(P_n))^s P_n] \\ &= \frac{1}{2}\{2(\eta_4 - 1)\sum_{i=1}^n p_{n,ii}^2 + tr[(P_n - D(P_n))^s (P_n - D(P_n))^s]\}.\end{aligned}$$

By Jensen's inequality with concave function, it is known that $\eta_4 > 1$. Define a modified matrix $P_n^+ = P_n + (\sqrt{\frac{\eta_4-1}{2}} - 1)D(P_n)$. The P_n^+ is constructed from P_n by the multiplication of the diagonal of P_n by the factor $\sqrt{\frac{\eta_4-1}{2}}$. As $tr(P_n) = 0$, $tr(P_n^+) = 0$. The square of the Euclidean norm of $(P_n^+)^s$ is

$$tr[(P_n^+)^s(P_n^+)^s] = 2(\eta_4 - 1)\sum_{i=1}^{n} p_{n,ii}^2 + tr([(P_n - D(P_n))^s(P_n - D(P_n))^s]).$$

The P_n and its modified matrix P_n^+ have a one-to-one relation. Given P_n^+, P_n can be recovered as $P_n = P_n^+ + (\sqrt{\frac{2}{\eta_4-1}} - 1)D(P_n^+)$. Because $tr(P_n^s G_n) = tr(P_n^s(G_n - \frac{tr(G_n)}{n}I_n)) = \frac{1}{2}tr(P_n^s(G_n - \frac{tr(G_n)}{n}I_n)^s)$, the maximization search is thus equivalent to

$$\max_{P_n^+} \frac{\frac{1}{n}tr^2\{[P_n^+ + (\sqrt{\frac{2}{\eta_4-1}} - 1)D(P_n^+)]^s(G_n - \frac{tr(G_n)}{n}I_n)^s\}}{tr[(P_n^+)^s(P_n^+)^s]}.$$

To make this optimization operationable, we shall look for the possible existence of a matrix A_n such that

$$tr\{[P_n^+ + (\sqrt{\frac{2}{\eta_4 - 1}} - 1)D(P_n^+)]^s(G_n - \frac{tr(G_n)}{n}I_n)^s\} = tr\{(P_n^+)^s[(G_n - \frac{tr(G_n)}{n}I_n) + A_n]^s\}.$$

This identity is equivalent to

$$(\sqrt{\frac{2}{\eta_4 - 1}} - 1)tr[D(P_n^{+s})(D(G_n) - \frac{tr(G_n)}{n}I_n)^s] = tr(P_n^{+s}A_n^s).$$

If A_n is taken to be a diagonal matrix, then $tr(P_n^{+s}A_n) = tr(D(P_n^{+s})A_n^s)$. One sees that the possible A_n is $A_n = (\sqrt{\frac{2}{\eta_4-1}} - 1)(D(G_n) - \frac{tr(G_n)}{n}I_n)$, which is a function determined by G_n alone. Thus the optimization becomes

$$\max_{P_n^+} \frac{\frac{1}{n}tr^2[(P_n^+)^s[(G_n - \frac{tr(G_n)}{n}I_n) + A_n]^s]}{tr[(P_n^+)^s(P_n^+)^s]}.$$

For any square confirmable matrices B and C, $tr^2(BC) \leq tr(B^2)tr(C^2)$ is a version of the Cauchy inequality. Hence the optimum P_n^+ is

$$P_n^{+*} = (G_n - \frac{tr(G_n)}{n}I_n) + A_n = (G_n - \frac{tr(G_n)}{n}I_n) + (\sqrt{\frac{2}{\eta_4 - 1}} - 1)(D(G_n) - \frac{tr(G_n)}{n}I_n).$$

In terms of the original P_n^*, one has

$$P_n^* = P_n^{+*} + (\sqrt{\frac{2}{\eta_4 - 1}} - 1)D(P_n^{+*}) = (G_n - \frac{tr(G_n)}{n} I_n) - \frac{\eta_4 - 3}{\eta_4 - 1}(D(G_n) - \frac{tr(G_n)}{n} I_n),$$

because $D(P_n^{+*}) = \sqrt{\frac{2}{\eta_4 - 1}}(D(G_n) - \frac{tr(G_n)}{n} I_n)$.

The above analysis motivates the existence of the BGMME and can be generalized to derive analytically the best P_n^*. An alternative approach can be based on the characterization of best moments in terms of any additional moments being redundant in Breusch et al. (1999). The following proposition summarizes the main results of the BGMME of λ_0 for the SAR disturbance process, which may not be normally distributed. We demonstrate the validity of the best moments with both the optimization of variance approach and the characterization of Breusch et al. (1999) in its proof.

Proposition 3.1 *Under Assumptions 1-6, within the class of optimal GMMEs $\mathcal{M}_n$, the consistent root $\hat{\lambda}_B$ derived from* $\min_{\lambda \in \Lambda} \left[u_n' S_n' (\lambda) P_n^* S_n (\lambda) u_n\right]^2$*, where*

$$P_n^* = (G_n - \frac{tr(G_n)}{n} I_n) - \frac{\eta_4 - 3}{\eta_4 - 1}(D(G_n) - \frac{tr(G_n)}{n} I_n),$$

is the BGMME with the limiting distribution $\sqrt{n}(\hat{\lambda}_B - \lambda_0) \xrightarrow{D} N\left(0, \Sigma_B^{-1}\right)$ and $\Sigma_B =$ $\lim_{n \to \infty} \frac{1}{n} tr\left(P_n^{*s} G_n\right)$.

Let $\mathcal{P}_{1n}$ be the class of constant $n \times n$ matrices P_n's satisfying Assumption 4. A subclass $\mathcal{P}_{2n}$ of $\mathcal{P}_{1n}$ consisting of P_n's with a zero diagonal is also interesting, as the corresponding GMME is robust against unknown heteroskedasticity (Lin and Lee, 2006) and distributional assumptions. Lee (2001*a*) has shown best selection of P_n from $\mathcal{P}_{2n}$ is $(G_n - D(G_n))$, and when ϵ_n is normally distributed, the best selection of P_n from $\mathcal{P}_{1n}$ is $(G_n - \frac{tr(G_n)}{n} I_n)$, which is a special case of P_n^* in Proposition 3.1 with $\eta_4 = 3$.

The asymptotic distribution of the GMME $\hat{\lambda}_{G1}$ based on the quadratic moment $\epsilon_n'(G_n - \frac{tr(G_n)}{n}I_n)\epsilon_n$ has been derived in Lee (2001a) with limiting variance being $\Sigma_{G1}^{-1} = \left(\lim_{n\to\infty}\frac{1}{n}\Sigma_{G1,n}\right)^{-1}$, where

$$\Sigma_{G1,n} = \frac{tr^2[(G_n - \frac{tr(G_n)}{n}I_n)^s G_n]}{(\eta_4 - 3)\sum_{i=1}^n (G_{n,ii} - \frac{tr(G_n)}{n})^2} + tr[(G_n - \frac{tr(G_n)}{n}I_n)^s G_n].$$

One the other hand, the limiting variance of the BGMME $\hat{\lambda}_B$ in Proposition 3.1 is $(\lim_{n\to\infty}\frac{1}{n}\Sigma_{B,n})^{-1}$ where

$$\Sigma_{B,n} = tr(P_n^{*s}G_n) = tr[(G_n - \frac{tr(G_n)}{n}I_n)^s G_n] - 2(\frac{\eta_4 - 3}{\eta_4 - 1})tr[(D(G_n) - \frac{tr(G_n)}{n}I_n)G_n].$$

To simplify notations, denote

$$v_G^2 = \frac{1}{n}\sum_{i=1}^n (G_{n,ii} - \frac{tr(G_n)}{n})^2 = \frac{1}{n}\sum_{i=1}^n (G_{n,ii} - \frac{\sum_{j=1}^n G_{n,jj}}{n})^2 \tag{3.5}$$

the empirical variance formed by the diagonal elements of G_n. Furthermore, denote

$$l_{G,1}^2 = \frac{1}{n}tr[(G_n - \frac{tr(G_n)}{n}I_n)^s G_n] = \frac{1}{2n}tr[(G_n - \frac{tr(G_n)}{n}I_n)^s (G_n - \frac{tr(G_n)}{n}I_n)^s], \tag{3.6}$$

and

$$l_{G,2}^2 = \frac{1}{n}tr[(G_n - D(G_n))^s G_n] = \frac{1}{2n}tr[(G_n - D(G_n))^s (G_n - D(G_n))^s], \tag{3.7}$$

which are, respectively, $\frac{1}{2n}$ of the square of the Euclidean norm of $(G_n - \frac{tr(G_n)}{n}I_n)^s$ and $(G_n - D(G_n))^s$.

Instead of comparing the limiting variances of these two estimates, it is desirable to compare the limiting precision measures $\frac{1}{n}\Sigma_{G1,n}$ and $\frac{1}{n}\Sigma_{B,n}$, which are the inverses of variances. One has $\frac{1}{n}\Sigma_{G1,n} = l_{G,1}^4/[(\eta_4 - 3)v_G^2 + l_{G,1}^2]$ and $\frac{1}{n}\Sigma_{B,n} = l_{G,1}^2 - 2(\frac{\eta_4-3}{\eta_4-1})v_G^2$. It follows that

$$\begin{aligned}\frac{1}{n}\Sigma_{B,n} - \frac{1}{n}\Sigma_{G1,n} &= -2(\frac{\eta_4 - 3}{\eta_4 - 1})v_G^2 + \frac{(\eta_4 - 3)v_G^2 l_{G,1}^2}{(\eta_4 - 3)v_G^2 + l_{G,1}^2} \\ &= \frac{(\eta_4 - 3)^2 v_G^2 (l_{G,1}^2 - 2v_G^2)}{(\eta_4 - 1)[(\eta_4 - 3)v_G^2 + l_{G,1}^2]} = \frac{(\eta_4 - 3)^2 v_G^2 l_{G,2}^2}{(\eta_4 - 1)[(\eta_4 - 1)v_G^2 + l_{G,2}^2]},\end{aligned}$$

because $l_{G,1}^2 - l_{G,2}^2 = 2v_G^2$. Note that $\eta_4 > 1$ by Jensen's inequality with concave functions, and $l_{G,2}^2 > 0$, Hence, $\frac{1}{n}\Sigma_{B,n} \geq \frac{1}{n}\Sigma_{G1,n}$. This verifies the efficiency of $\hat{\lambda}_B$ relative to $\hat{\lambda}_{G1}$. The percentage loss of asymptotic efficiency of $\hat{\lambda}_{G1}$ can be evaluated as

$$1 - \frac{\Sigma_{G1,n}}{\Sigma_{B,n}} = \frac{(\eta_4 - 3)^2 v_G^2 l_{G,2}^2}{[(\eta_4 - 1)v_G^2 + l_{G,2}^2] \cdot [4v_G^2 + (\eta_4 - 1)l_{G,2}^2]}, \tag{3.8}$$

when $\eta_4 \neq 3$. Note that the variance is the inverse of the precision measure. So, $1 - \frac{\Sigma_{G1,n}}{\Sigma_{B,n}} = 1 - \frac{\Sigma_{B,n}^{-1}}{\Sigma_{G1,n}^{-1}} = \frac{\Sigma_{G1,n}^{-1} - \Sigma_{B,n}^{-1}}{\Sigma_{G1,n}^{-1}}$ is also the percentage of reduction in asymptotic variance of $\hat{\lambda}_B$ relative to $\hat{\lambda}_{G1}$.

Similarly, we can compare the efficiency gain of $\hat{\lambda}_B$ relative to $\hat{\lambda}_{G2}$ derived from the quadratic moment $\epsilon_n'(G_n - D(G_n))\epsilon_n$. Following Lee (2001*a*), with limiting variance of $\hat{\lambda}_{G2}$ is $\Sigma_{G2}^{-1} = \left(\lim_{n\to\infty} \frac{1}{n}\Sigma_{G2,n}\right)^{-1}$, where

$$\frac{1}{n}\Sigma_{G2,n} = \frac{1}{n} tr[(G_n - D(G_n))^s G_n] = l_{G,2}^2.$$

It follows that

$$\frac{1}{n}\Sigma_{B,n} - \frac{1}{n}\Sigma_{G2,n} = l_{G,1}^2 - 2(\frac{\eta_4 - 3}{\eta_4 - 1})v_G^2 - l_{G,2}^2 = \frac{4}{\eta_4 - 1}v_G^2,$$

because $l_{G,1}^2 - l_{G,2}^2 = 2v_G^2$. As $\eta_4 > 1$ by Jensen's inequality with concave functions, $\frac{1}{n}\Sigma_{B,n} \geq \frac{1}{n}\Sigma_{G2,n}$. The percentage loss of asymptotic efficiency of $\hat{\lambda}_{G2}$ can be evaluated as

$$1 - \frac{\Sigma_{G2,n}}{\Sigma_{B,n}} = \frac{4v_G^2}{4v_G^2 + (\eta_4 - 1)l_{G,2}^2}, \tag{3.9}$$

which is also the percentage of reduction in asymptotic variance of $\hat{\lambda}_B$ relative to $\hat{\lambda}_{G2}$. From this, $\hat{\lambda}_B$ is more precise as it takes into account the variance of the diagonal elements of G_n.

The BGMME associated with P_n^* involves the unknown λ_0 and η_4. In practice, the unknown λ_0 can be estimated with some P_n's from $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$ within the GMM framework, and η_4 can be replaced by the empirical moments of the estimated residuals. With initial consistent estimates $\hat{\lambda}_n$ and $\hat{\eta}_4$, G_n can be estimated by $\hat{G}_n = G_n(\hat{\lambda}_n) = W_n S_n^{-1}(\hat{\lambda}_n)$ and P_n^* can be estimated by

$$\hat{P}_n^* = (\hat{G}_n - \frac{tr(\hat{G}_n)}{n} I_n) - \frac{\hat{\eta}_4 - 3}{\hat{\eta}_4 - 1}(D(\hat{G}_n) - \frac{tr(\hat{G}_n)}{n} I_n). \tag{3.10}$$

The following proposition shows that the feasible BGMME with P_n^* replaced by $\hat{P}_n^*$ in the moment functions has the same limiting distribution as the corresponding BGMME in Proposition 3.1. Let $M_n = X_n \left(X_n' X_n\right)^{-1} X_n'$.

Proposition 3.2 *Under Assumptions 1-7, suppose $\hat{\lambda}_n$ and $\hat{\eta}_4$ are $\sqrt{n}$-consistent estimates of λ_0 and η_4, and $\hat{P}_n^*$ is given by* (3.10). *Then* $\min_{\lambda \in \Lambda} \left[\hat{u}_n' S_n' (\lambda) \hat{P}_n^* S_n (\lambda) \hat{u}_n\right]^2$, *with $\hat{u}_n = (I_n - M_n) Y_n$, has a consistent root $\hat{\lambda}_{FB}$ which has the same limiting distribution of $\hat{\lambda}_B$ derived from* $\min_{\lambda \in \Lambda} \left[u_n' S_n' (\lambda) P_n^* S_n (\lambda) u_n\right]^2$.

3.2.2 GMM Estimation of the MRSAR Model

The MRSAR model is specified as

$$Y_n = X_n \beta + \lambda W_n Y_n + \epsilon_n. \tag{3.11}$$

where ϵ_n is an n-dimensional vector of i.i.d. disturbances with zero mean and finite variance σ^2. Let $\delta_0 = (\beta_0', \lambda_0, \sigma_0^2)'$ be the true parameter vector. The equilibrium vector Y_n is

$$Y_n = S_n^{-1} \left(X_n \beta_0 + \epsilon_n\right). \tag{3.12}$$

It follows that $W_n Y_n = G_n X_n \beta_0 + G_n \epsilon_n$ where $G_n = W_n S_n^{-1}$, and $W_n Y_n$ is correlated with ϵ_n because, in general, $E((G_n \epsilon_n)' \epsilon_n) = \sigma_0^2 tr(G_n) \neq 0$.

Let Q_n be an $n \times k'$ matrix of IVs constructed as functions of X_n and W_n in a 2SLS approach. Denote $\epsilon_n(\theta) = S_n(\lambda) Y_n - X_n \beta$, where $\theta = \left(\beta', \lambda\right)'$. Thus, $\epsilon_n = \epsilon_n(\theta_0)$. The moment functions correspond to the orthogonality conditions of X_n and ϵ_n are $Q_n'\epsilon_n(\theta)$. Lee (2006) suggests the use of the moment functions $\epsilon_n'(\theta)P_n\epsilon_n(\theta)$ with P_n's satisfying Assumption 4 in addition to $Q_n'\epsilon_n(\theta)$ for the estimation of (3.11) in the GMM framework. With the selected matrices P_{jn}'s ($j = 1, \cdots, m$) and IV matrix Q_n, the set of moment functions form a vector

$$g_n(\theta) = (Q_n, P_{1n}\epsilon_n(\theta), \cdots, P_{mn}\epsilon_n(\theta))'\epsilon_n(\theta). \tag{3.13}$$

At θ_0, $g_n(\theta_0) = (Q_n, P_{1n}\epsilon_n, \cdots, P_{mn}\epsilon_n)'\epsilon_n$, which has a zero mean because $E(Q_n'\epsilon_n) = Q_n'E(\epsilon_n) = 0$ and $E(\epsilon_n' P_{jn}'\epsilon_n) = \sigma_0^2 tr(P_{jn}) = 0$ for $j = 1, \cdots, m$.

Regularity assumptions 1-7 specified in the regression model with SAR disturbances are adopted for the GMM estimation of the MRSAR model with proper modifications. Assumption 5' summarizes some sufficient identification conditions of θ_0 from the moment equations $E\,(g_n(\theta_0)) = 0$. In the case that $G_nX_n\beta_0$ and X_n are linearly dependent, which includes the case that all exogenous variables X_n are irrelevant, Assumption 5' (ii) assures the identification of λ_0 from the quadratic moment functions $\epsilon_n'(\theta)P_n\epsilon_n(\theta)$. Assumption 7' extends the parameter space to a compact convex subset of R^{k+1}.

Assumption 4' *The matrices P_n's with $tr\,(P_n) = 0$ are uniformly bounded in both row and column sums in absolute value, and elements of Q_n are uniformly bounded.*

Assumption 5' *Either (i) $\lim_{n\to\infty} \frac{1}{n} Q_n'(G_nX_n\beta_0, X_n)$ has the full rank $(k+1)$, or (ii) $\lim_{n\to\infty} \frac{1}{n} Q_n'X_n$ has the full rank k, $\lim_{n\to\infty} \frac{1}{n} tr(P_{jn}G_n) \neq 0$ for some j, and*

$$\lim_{n\to\infty} \frac{1}{n}(tr(P_{1n}^sG_n), \cdots, tr(P_{mn}^sG_n))'$$

is linearly independent of $\lim_{n\to\infty} \frac{1}{n}(tr(G_n'P_{1n}G_n), \cdots, tr(G_n'P_{mn}G_n))'$.

Assumption 7' *The θ_0 is in the interior of the parameter space Θ, which is a compact convex subset of R^{k+1}.*

Lee (2006) has shown the GMME $\hat{\theta}_P$ from $\min_{\theta\in\Theta} g_n'(\theta)\, a_n' a_n g_n(\theta)$ is $\sqrt{n}$-consistent and derived its limiting distribution. The optimal choice of the weighting matrix $a_n' a_n$ is $\left(\frac{1}{n}\Omega_n\right)^{-1}$, where $\Omega_n = var\left(g_n(\theta_0)\right)$, by the generalized Schwartz inequality. Let $\mathcal{M}_n = \{\hat{\theta}_O\}$ be the class of optimal GMMEs derived from $\min_{\theta\in\Theta} g_n'(\theta)\, \Omega_n^{-1} g_n(\theta)$, where $g_n(\theta)$ is a vector of moment functions given by (3.13). Within the class of optimal GMMEs $\mathcal{M}_n$, Lee (2006) has shown that the best selection of Q_n shall be $(X_n, G_n X_n \beta_0)$, the best selection of P_n from the subclass $\mathcal{P}_{2n}$ shall be $(G_n - D(G_n))$, and in the event that ϵ_n is normally distributed, $(G_n - \frac{tr(G_n)}{n} I_n)$ shall be the best selection of P_n from the broader class $\mathcal{P}_{1n}$.

In the following proposition, we show the existence of the BGMME within the class $\mathcal{M}_n$, when the disturbances are not normally distributed. To show this result, we adopt Breusch et al. (1999) in demonstrating that additional moment conditions are redundant to the best selection of moment conditions. If an intercept appears in X_n, we have $X_n = \left[X_n^*, l_n\right]$, where l_n is an n-dimensional vector of ones. Otherwise $X_n^* \equiv X_n$. Suppose there are k^* columns in X_n^*. Let X_{nj} be the jth column of X_n, and X_{nj}^* be the jth column of X_n^*. Denote $X_{nj}^{*d} = X_{nj}^* - \frac{1}{n} l_n l_n' X_{nj}^*$ the deviation of observation X_{nj}^* from its sample mean. Let

$$G_n^* = G_n - \frac{(\eta_4 - 3) - \eta_3^2}{(\eta_4 - 1) - \eta_3^2} D(G_n) - \frac{\eta_3}{\sigma_0[(\eta_4 - 1) - \eta_3^2]} D\left(G_n X_n \beta_0\right), \tag{3.14}$$

with $\eta_3 = \mu_3/\sigma_0^3$ being the skewness of the disturbance. Let $vec_D(A)$ be a column vector formed by the diagonal elements of a square matrix A.

Proposition 3.3 *Suppose Assumptions 1-3, 4', 5', and 6 are satisfied. Let* $P_{1n}^* = G_n^* - \frac{1}{n} tr\left(G_n^*\right) I_n$*, and* $P_{j+1,n}^* = D(X_{nj}^{*d})$ *for* $j = 1, \cdots, k^*$*. Let* $Q_n^* = \left[Q_{1n}^*, Q_{2n}^*\right]$ *with*

$$Q_{1n}^* = X_n + \frac{\eta_3^2}{(\eta_4 - 1) - \eta_3^2}\left(X_n - \frac{1}{n} l_n l_n' X_n\right), \tag{3.15}$$

and

$$
\begin{aligned}
Q_{2n}^{*} &= G_n X_n \beta_0 + \frac{\eta_3^2}{(\eta_4 - 1) - \eta_3^2}\left(G_n X_n \beta_0 - \frac{1}{n} l_n l_n' G_n X_n \beta_0\right) \\
&\quad - \frac{2\sigma_0 \eta_3}{(\eta_4 - 1) - \eta_3^2}\left(vec_D\left(G_n\right) - \frac{1}{n} tr\left(G_n\right) l_n\right).
\end{aligned} \tag{3.16}
$$

Within the class of optimal GMMEs $\mathcal{M}_n$, the consistent root $\hat{\theta}_B$ derived from

$$
\min_{\theta \in \Theta} g_n^{*\prime}(\theta)\, \Omega_n^{*-1} g_n^{*}(\theta),
$$

where $\Omega_n^{*} = var\left(g_n^{*}(\theta_0)\right)$ and $g_n^{*}(\theta) = (Q_n^{*}, P_{1n}^{*}\epsilon_n(\theta), \cdots, P_{k^{*}+1,n}^{*}\epsilon_n(\theta))'\epsilon_n(\theta)$, is the BGMME with the limiting distribution $\sqrt{n}(\hat{\theta}_B - \theta_0) \xrightarrow{D} N(0, \Sigma_B^{-1})$ and

$$
\Sigma_B = \lim_{n \to \infty} \frac{1}{n}\begin{pmatrix} \sigma_0^{-2} X_n' Q_{1n}^{*} & \sigma_0^{-2} X_n' Q_{2n}^{*} \\ \sigma_0^{-2} Q_{2n}^{*\prime} X_n & \sigma_0^{-2}\left(G_n X_n \beta_0\right)' Q_{2n}^{*} + tr\left(P_{1n}^{*s} G_n\right) \end{pmatrix}.
$$

When ϵ_n is normally distributed, $\eta_3 = 0$ and $\eta_4 = 3$. Hence, the best selection of $Q_n = (X_n, G_n X_n \beta_0)$ and $P_n = (G_n - \frac{tr(G_n)}{n} I_n)$ under normality are the degenerated Q_n^{*} and P_{1n}^{*} in Proposition 3.3. Based on the characterization of best moments in Breusch et al. (1999), it can be shown that moment functions $\epsilon_n'(\theta) P_{j+1,n}^{*} \epsilon_n(\theta)$ ($j = 1, \cdots, k^{*}$) are redundant given $(X_n, G_n X_n \beta_0, (G_n - \frac{tr(G_n)}{n} I_n)'\epsilon_n(\theta))'\epsilon_n(\theta)$ under normality, with similar arguments used in the proof of Proposition 3.3. Furthermore, Lee (2006) has shown that the BGMME derived from the set of moments $(X_n, G_n X_n \beta_0, (G_n - \frac{tr(G_n)}{n} I_n)'\epsilon_n(\theta))'\epsilon_n(\theta)$ has the same limiting distribution as MLE under normality.

In practice, with initial consistent estimates $\hat{\delta}_n$, $\hat{\mu}_3$ and $\hat{\mu}_4$, P_{1n}^{*} and Q_n^{*} can be replaced by their empirical counterparts $\hat{P}_{1n}^{*} = P_{1n}^{*}(\hat{\delta}_n', \hat{\mu}_3, \hat{\mu}_4)$ and $\hat{Q}_n^{*} = Q_n^{*}(\hat{\delta}_n', \hat{\mu}_3, \hat{\mu}_4)$. The corresponding variance matrix Ω_n^{*} of the best moment functions can be estimated as $\hat{\Omega}_n^{*} = \Omega_n^{*}(\hat{\delta}_n', \hat{\mu}_3, \hat{\mu}_4)$. The following proposition shows that the feasible BGMME with the moment functions

$$
\hat{g}_n^{*}(\theta) = (\hat{Q}_n^{*}, \hat{P}_{1n}^{*}\epsilon_n(\theta), P_{2n}^{*}\epsilon_n(\theta), \cdots, P_{k^{*}+1,n}^{*}\epsilon_n(\theta))'\epsilon_n(\theta) \tag{3.17}
$$

has the same limiting distribution as the corresponding BGMME in Proposition 3.3.

Proposition 3.4 *Under Assumptions 1-3, 4', 5', 6, and 7', suppose $\hat{\delta}_n$, $\hat{\mu}_3$ and $\hat{\mu}_4$ are $\sqrt{n}$-consistent estimates of δ_0, μ_3 and μ_4. Then $\min_{\theta\in\Theta} \hat{g}_n^{*\prime}(\theta)\,\hat{\Omega}_n^{*-1}\hat{g}_n^{*}(\theta)$, with $\hat{g}_n^{*}(\theta)$ given by (3.17) and $\hat{\Omega}_n^{*} = \Omega_n^{*}(\hat{\delta}_n, \hat{\mu}_3, \hat{\mu}_4)$, has a consistent root $\hat{\theta}_{FB}$ which has the same limiting distribution of $\hat{\theta}_B$ derived from $\min_{\theta\in\Theta} g_n^{*\prime}(\theta)\,\Omega_n^{*-1}g_n^{*}(\theta)$.*

3.3 Monte Carlo Study

In the Monte Carlo study, the regression model with SAR disturbances is specified as

$$
\begin{aligned}
Y_n &= X_{n1}\beta_1 + u_n, \\
u_n &= \lambda W_n u_n + \epsilon_n,
\end{aligned}
$$

and the MRSAR model is specified as

$$
Y_n = X_{n1}\beta_1 + X_{n2}\beta_2 + X_{n3}\beta_3 + \lambda W_n Y_n + \epsilon_n,
$$

where x_{i1}, x_{i2} and x_{i3} are three independently generated standard normal variables and are i.i.d. for all i, and ϵ_{ni}'s are independently generated from the following 5 distributions, all of which are scaled to have mean 0 and variance 2:

(a) normal, $\epsilon_{ni} \sim N(0, 2)$,

(b) student t, $\epsilon_{ni} = \sqrt{6/5}u$ where $u \sim t(5)$,

(c) symmetric bimodal mixture normal, $\epsilon_{ni} = u/\sqrt{5}$ where $u \sim .5N(-3, 1) + .5N(3, 1)$,

(d) asymmetric bimodal mixture normal, $\epsilon_{ni} = u/2\sqrt{2}$ where $u \sim .5N(-3, 1) + .5N(3, 13)$,

(e) gamma, $\epsilon_{ni} = u - 2$ where $u \sim gamma(2, 1)$.

To facilitate comparison, skewness (η_3) and kurtosis (η_4) for these distributions are correspondingly: (a) $\eta_3 = 0$, $\eta_4 = 3$; (b) $\eta_3 = 0$, $\eta_4 = 9$; (c) $\eta_3 = 0$, $\eta_4 = 1.38$; (d) $\eta_3 \approx 0.84$,

$\eta_4 \approx 2.79$; and (e) $\eta_3 = \sqrt{2}$, $\eta_4 = 6$. Normal distribution is considered as the basis for comparison. When the disturbances are normally distributed, both MLE and BGMME are asymptotically efficient. However, the finite sample performance of BGMME may not be as good as MLE, because the moment functions involve some unknown parameters need to be estimated in an initial step. Student t and symmetric bimodal mixture normal distribution are introduced to explore the effects of, respectively, leptokurtic ($\eta_4 > 3$) and platykurtic ($\eta_4 < 3$) disturbances on the small sample performance of various estimates. Asymmetric bimodal mixture normal and gamma distributions are introduced to study the effects of skewness. To be specific, asymmetric bimodal mixture normal specified here corresponds to the case where disturbances is slightly platykurtic and has a moderate skewness, and gamma corresponds to the case where disturbances is leptokurtic and has a relatively large skewness. Asymptotically, BGMME is more efficient than QMLE under the distributions (b)-(e). For the regression model with SAR disturbances, the proposed BGMME improves upon QMLE as the quadratic moment function incorporates kurtosis of the distribution. On the other hand, similar to QMLE, the BGMME does not involve skewness in their formulations and, hence, are robust against skewness. Therefore, in the Monte Carlo experiments for the regression model with SAR disturbances, we focus only on distributions with (a)-(c).

The estimators considered are (i) QMLE (ii) OGMME and (iii) BGMME. For the regression model with SAR disturbances, OGMME refers to the feasible optimal GMME using $\hat{G}_n - D(\hat{G}_n)$ for the quadratic moment, with the inverse of their (estimated) variance matrix as the distance matrix, and BGMME refers to the feasible BGMME described in Proposition 3.2. For the MRSAR model, OGMME refers to the feasible optimal GMME using X_n and $\hat{G}_n X_n \hat{\beta}_n$ for the linear moments and $\hat{G}_n - D(\hat{G}_n)$ for the quadratic moment, with the inverse of their (estimated) variance matrix as the distance matrix, and BGMME refers to the feasible BGMME described in Proposition 3.4. For the Monte Carlo results reported here, we use QMLEs[27] as the initial estimators to get feasible OGMMEs and BGMMEs.[28]

The number of repetitions is $1,000$ for each case in the Monte Carlo experiment. The regressors are randomly redrawn for each repetition. In each case, we report the mean 'Mean' and standard deviation 'SD' of the empirical distributions of the estimates. To facilitate the comparison of various estimators, their root mean square errors 'RMSE' are also reported. In all the cases of this study, the true λ_0 is set to 0.6, and $\beta_{10} = 1.0, \beta_{20} = 0, \beta_{30} = -1.0$. For the MRSAR model, the variance ratio of $x\beta_0$ with the sum of variances of $x\beta_0$ and ϵ is 0.5. If one ignores the interaction term, this ratio would represent $R^2 = 0.5$ in a regression equation. The smallest sample size is $n = 49$, and the moderate sample sizes are 245 and 490.

When the sample size is $n = 49$, the spatial weights matrix W_n corresponds to the weights matrix for the study of crimes across 49 districts in Columbus, Ohio in Anselin (1988). For moderate sample sizes of $n = 245$ and 490, the corresponding spatial weights matrices are block diagonal matrices with the preceding 49×49 matrix as their diagonal blocks. These correspond to the pooling, respectively, of five and ten separate districts with

[27] The QMLEs are calculated using sar.m in Econometrics Toolbox (version 7) by James P. Lesage. Function option $info.lflag = 0$ for full computation (instead of approximation), and other options are set to the default values.

[28] Using QMLEs as initial estimates can be justified as BGMMEs are adopted for the purpose of improving QMLEs. We have also run the Monte Carlo of feasible OGMMEs and BGMMEs with initial estimates derived from optimal GMM based on quadratic moments W_n and $W_n^2 - D(W_n^2)$ for the regression model with SAR disturbances, and initial estimates from the 2SLS approach in (Kelejian and Prucha, 1998) for the MRSAR model. For small sample size $n = 49$, the feasible OGMMEs and BGMMEs have large variances than those reported here. For moderate sample size $n = 490$, the results are largely the same as those reported here.

similar neighboring structures in each district. We use this spatial weights matrix in the Monte Carlo experiments for the MRSAR model.

For the regression model with SAR disturbances, with the W_n in Anselin (1988) and distribution (c), the percentage of reduction in asymptotic variance of BGMME of λ_0 relative to QMLE is 2.59%, and relative to OGMME is 3.95%, following (3.8) and (3.9). With distribution (b), the percentage of reduction in asymptotic variance is even smaller. These are caused by the small empirical variance formed by the diagonal elements of G_n. With the W_n in Anselin (1988), $\upsilon_G^2 = 0.005$, by (3.5). In order to have larger υ_G^2, we construct a weights matrix as follows. With $\lambda_0 = 0.6$, G_n can be expanded as $G_n = W_n(I_n - \lambda_0 W_n)^{-1} = W_n + \lambda_0 W_n^2 + \cdots$. As $D(W_n) = 0$, the empirical variance of the diagonal elements of G_n is largely determined by that of W_n^2. When W_n is row-normalized, diagonal elements of W_n^2 are weighted average for each column of W_n. We generate 7×7 upper triangular matrices A_n's, whose non-zero elements in each column are either all $(200+u)$'s or all u's ($u \sim U[0, 1]$) with equal probability. We calculate υ_G^2 for the row-normalized B_n, with $B_n = A_n + A_n'$ as the weights matrix W_n in (3.5).[29] We generate 1000 such B_n's, and pick 7 of them with the largest υ_G^2. In this way, we get a 49×49 block diagonal matrix W_n with the 7 row-normalized B_n's being the diagonal blocks. This W_n gives $\upsilon_G^2 = 0.134$, when $\lambda_0 = 0.6$. And with distribution (c), the percentage of reduction in asymptotic variance of BGMME relative to QMLE is 23.17%, and relative to OGMME is 36.03%, following (3.8) and (3.9). For large sample sizes of $n = 245$ and 490, the corresponding spatial weights matrices are block diagonal matrices with the preceding 49×49 matrix as their diagonal blocks. We use the constructed spatial weights matrix in the Monte Carlo experiments for the regression model with SAR disturbances.

[29]As W_n is symmetric before row normalization, we can apply the approach in Ord (1975) to implement QML estimation.

For the regression model with SAR disturbances, QMLEs and various GMMEs of λ_0 are reported in Tables 3.1-3.3. For small sample size $n = 49$, the various estimates of λ_0 are biased down. Among them, QMLE has the largest bias. The magnitude of the bias is about $5 \sim 6\%$. The bias reduces as sample size increases, and for sample size $n = 490$, various estimates are essentially unbiased. When the disturbances are normally distributed, MLE (QMLE) is efficient, and the finite sample performance of BGMME is as good as MLE (QMLE) in terms of SD and RMSE. SD and RMSE of OGMME is slightly larger when the sample size is small. When the disturbances follow student t distribution, BGMME has the smallest SD and RMSE for all sample size considered. For sample size $n = 490$, the percentage reduction in RMSE of BGMME relative to QMLE is about 12.5%. This is also the case when the disturbances follow bimodal mixture normal distribution. For sample size $n = 490$, the percentage reduction in RMSE of BGMME relative to OGMME and QMLE is about 13.8%.

Tables 3.4-3.8 report QMLEs and various GMMEs of λ_0 and β_0 for the MRSAR model. For small sample size $n = 49$, QMLEs of λ_0 are biased downwards by $4 \sim 5\%$ for all specifications of disturbances. And OGMMEs of λ_0 are biased downwards by 2% when the disturbances follow bimodal mixture normal distributions. The two GMMEs of β_{10} are biased downwards and β_{30} are biased upwards, with the largest bias being about 2% for all disturbance specifications. The other estimates are essential unbiased. The bias disappears as sample size increases to $n = 249$. When the disturbances are normally distributed, MLEs (QMLEs) are efficient. For small sample size $n = 49$, the finite sample performance of MLEs (QMLEs) are better than the two GMMEs in terms of smaller SD and RMSE. And BGMMEs have the largest SD and RMSE because the feasible best moment functions involve initial estimation of several unknown parameters. For moderate sample size, the finite sample performance of BGMME is as good as MLEs (QMLEs). When the disturbances are not normally distributed, for the two disturbance specifications with $\eta_3 = 0$, due to the small empirical variance in the diagonal elements of G_n, BGMMEs are not better than QMLEs for moderate sample sizes, even though the latter is not asymptotically efficient. When the disturbances follow the distributions with $\eta_3 \neq 0$, BGMMEs of λ_0 are better than OGMMEs and QMLEs for moderate sample sizes, and BGMMEs of β_0 are better than OGMMEs and QMLEs for all sample sizes considered, in terms of SD and RMSE. When the disturbances follow gamma distribution and $n = 490$, the percentage reduction in SD of BGMMEs of λ_0, β_{10}, β_{20}, and β_{30} relative to QMLEs are, respectively, 11.8%, 23.1%, 21.9%, and 21.2%.

In summary, BGMME improves on QMLE as the former incorporates correlation between linear and quadratic moment conditions when the disturbances are skewed. Both the BGMMEs of the spatial effect λ_0 and coefficients of other explanatory variables β_0 have smaller SD and RMSE relative to QMLE and OGMME. On the other hand, for cases with $\eta_3 = 0$, gains of BGMME by including measure of kurtosis is relatively small, and can be insignificant when the diagonal elements of G_n do not vary enough.

3.4 Conclusion

In this chapter, we consider the GMM estimation of the regression models with SAR disturbances and MRSAR models. The MLE approach is efficient when the disturbances is normally distributed, and Lee (2006) has shown the existence of GMME based on linear and quadratic moment conditions that can attain the same limiting distribution of the MLE under normal disturbances. This chapter improves upon the QMLE approach by incorporating potential skewness and kurtosis of the disturbances into the moment conditions used in the GMM framework. The proposed BGMME is asymptotically as efficient as MLE under normality, and more efficient than the QMLE when the disturbances are not normally distributed. Monte Carlo studies show that the potential inefficiency of the QMLE in finite sample for the MRSAR model mainly comes from the possible correlation between linear and quadratic moment conditions in the likelihood function. Hence, the proposed BGMME has its biggest advantage when the skewness of the disturbances is nonzero. In the event that the diagonal elements of G_n have good variance, then, taking into account kurtosis will also be valuable.

3.5 Appendices

3.5.1 Summary of Notations

$D(A) = Diag(A)$ is a diagonal matrix with diagonal elements being A if A is a vector, or diagonal elements of A if A is a square matrix.

$vec_D(A)$ is a column vector formed by the diagonal elements of a square matrix A.

$A^s = A + A'$ where A is a square matrix.

$A^d = A - \frac{1}{n} tr(A) I_n$ where A is an $n \times n$ matrix.

A^L is a linearly transformed square matrix of A which preserves the uniform boundedness property.

$S_n(\lambda) = I_n - \lambda W_n$; $S_n = S_n(\lambda_0)$.

$G_n(\lambda) = W_n S_n^{-1}(\lambda) = W_n (I_n - \lambda W_n)^{-1}$; $G_n = G_n(\lambda_0)$.

$\delta = (\beta', \lambda, \sigma^2)'$; $\delta_0 = (\beta_0', \lambda_0, \sigma_0^2)'$; $\theta = \left(\beta', \lambda\right)'$. $\theta_0 = \left(\beta_0', \lambda_0\right)'$.

$M_n = X_n \left(X_n' X_n\right)^{-1} X_n'$.

$G_n^* = G_n - \frac{(\eta_4 - 3) - \eta_3^2}{(\eta_4 - 1) - \eta_3^2} D(G_n) - \frac{\eta_3}{\sigma_0((\eta_4 - 1) - \eta_3^2)} D\left(G_n X_n \beta_0\right)$.

l_n is an $n \times 1$ vector of ones.

e_{kj} is the jth unit vector in R^k.

If an intercept appears in X_n, we have $X_n = \left[X_n^*, l_n\right]$. Otherwise $X_n^* \equiv X_n$.

$X_{nj}^{*d} = X_{nj}^* - \frac{1}{n} l_n l_n' X_{nj}^*$ is the deviation of observation X_{nj}^* from its sample mean.

3.5.2 Some Useful Lemmas

In this appendix, we list some lemmas which are useful for the proofs of the results in the text.

Lemma 3.1 *Suppose that the elements of the sequences of n-dimensional column vectors $\{z_{1n}\}$ and $\{z_{2n}\}$ are uniformly bounded. If $\{A_n\}$ are uniformly bounded in either row or column sums in absolute value, then $\left|z_{1n}' A_n z_{2n}\right| = O(n)$.*

Proof. Trivial. ■

Lemma 3.2 *Suppose that $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. random variables with zero mean and finite variance σ^2 and finite fourth moment μ_4. Then, for any two $n \times n$ matrices A_n and B_n,*

$$E\left(\epsilon_n' A_n \epsilon_n \cdot \epsilon_n' B_n \epsilon_n\right) = \left(\mu_4 - 3\sigma^4\right) vec_D\left(A_n\right) vec_D\left(B_n\right) + \sigma^4 \left[tr\left(A_n\right) tr\left(B_n\right) + tr\left(A_n B_n^s\right)\right],$$

where $B_n^s = B_n + B_n'$.

Proof. See Lee (2001a). ■

Lemma 3.3 *Suppose that $\{A_n\}$ are uniformly bounded in both row and column sums in absolute value. $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. with zero mean and finite fourth moment. Then, $E(\epsilon_n' A_n \epsilon_n) = O(n)$, $var(\epsilon_n' A_n \epsilon_n) = O(n)$, $\epsilon_n' A_n \epsilon_n = O_p(n)$, and $\frac{1}{n}\epsilon_n' A_n \epsilon_n - \frac{1}{n} E(\epsilon_n' A_n \epsilon_n) = o_p(1)$.*

Proof. See Lee (2001*a*). ■

Lemma 3.4 *Suppose that A_n is an $n \times n$ matrix with its column sums being uniformly bounded in absolute value, elements of the $n \times k$ matrix C_n are uniformly bounded, and $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. with zero mean and finite variance σ^2. Then, $\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n = O_p(1)$ and $\frac{1}{n} C_n' A_n \epsilon_n = o_p(1)$. Furthermore, if the limit of $\frac{1}{n} C_n' A_n A_n' C_n$ exists and is positive definite, then $\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n \xrightarrow{D} N(0, \sigma^2 \lim_{n\to\infty} \frac{1}{n} C_n' A_n A_n' C_n)$.*

Proof. See Lee (2004). ■

Lemma 3.5 *Suppose that $\{A_n\}$ is a sequence of symmetric $n \times n$ matrices with row and column sums uniformly bounded in absolute value and $b_n = (b_{n1}, \cdots, b_{nn})'$ is an n-dimensional vector such that $\sup_n \frac{1}{n} \sum_{i=1}^n |b_{ni}|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$. $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. random variables with zero mean and finite variance σ^2, and its moment $E(|\epsilon|^{4+2\delta})$ for some $\delta > 0$ exists. Let $\sigma^2_{Q_n}$ be the variance of Q_n where $Q_n = \epsilon_n' A_n \epsilon_n + b_n' \epsilon_n - \sigma^2 tr(A_n)$. Assume that the variance $\sigma^2_{Q_n}$ is bounded away from zero at the rate n. Then, $\frac{Q_n}{\sigma_{Q_n}} \xrightarrow{D} N(0,1)$.*

Proof. See Kelejian and Prucha (2001). ■

Lemma 3.6 *Suppose that $\frac{1}{n}(g_n(\lambda) - \bar{g}_n(\lambda))$ converges in probability to zero uniformly in $\lambda \in \Lambda$, which is a compact set, and $\lim_{n\to\infty} \frac{1}{n}\bar{g}_n(\lambda) = 0$ has a unique root at λ_0 in Λ. The $\hat{\lambda}_n$ and $\hat{\lambda}_n^*$ are, respectively, the roots of the moment equations $g_n(\lambda) = 0$ and $g_n^*(\lambda) = 0$. If $\frac{1}{n}(g_n^*(\lambda) - g_n(\lambda)) = o_p(1)$ uniformly in $\lambda \in \Lambda$, then both $\hat{\lambda}_n$ and $\hat{\lambda}_n^*$ converge in probability to λ_0.*

In addition, suppose that $\frac{1}{n}\frac{\partial g_n(\lambda)}{\partial \lambda}$ converges in probability to a well defined nonzero limit function uniformly in $\lambda \in \Lambda$, and $\frac{1}{\sqrt{n}} g_n(\lambda_0) = O_p(1)$. If $\frac{1}{n}(\frac{\partial g_n^*(\lambda)}{\partial \lambda} - \frac{\partial g_n(\lambda)}{\partial \lambda}) = o_p(1)$

uniformly in $\lambda \in \Lambda$, *and* $\frac{1}{\sqrt{n}}(g_n^*(\lambda_0) - g_n(\lambda_0)) = o_p(1)$, *then both* $\sqrt{n}(\hat{\lambda}_n - \lambda_0)$ *and* $\sqrt{n}(\hat{\lambda}_n^* - \lambda_0)$ *have the same limiting distribution.*

Proof. See Lee (2001*a*). ■

Lemma 3.7 *Let* $\hat{\theta}_n$ *and* $\hat{\theta}_n^*$ *be, respectively, the minimizers of* $\mathcal{F}_n(\theta)$ *and* $\mathcal{F}_n^*(\theta)$ *in* Θ. *Suppose that* $\frac{1}{n}(\mathcal{F}_n(\theta) - \bar{\mathcal{F}}_n(\theta))$ *converges in probability to zero uniformly in* $\theta \in \Theta$, *which is a compact set, and* $\left\{\frac{1}{n}\bar{\mathcal{F}}_n(\theta)\right\}$ *satisfies the uniqueness identification condition at* θ_0. *If* $\frac{1}{n}(\mathcal{F}_n^*(\theta) - \mathcal{F}_n(\theta)) = o_p(1)$ *uniformly in* $\theta \in \Theta$, *then both* $\hat{\theta}_n$ *and* $\hat{\theta}_n^*$ *converge in probability to* θ_0.

In addition, suppose that $\frac{1}{n}\frac{\partial^2 \mathcal{F}_n(\theta)}{\partial\theta\partial\theta'}$ *converges in probability to a well defined limiting matrix, uniformly in* $\theta \in \Theta$, *which is nonsingular at* θ_0, *and* $\frac{1}{\sqrt{n}}\frac{\partial \mathcal{F}_n(\theta_0)}{\partial\theta} = O_p(1)$. *If* $\frac{1}{n}(\frac{\partial^2 \mathcal{F}_n^*(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \mathcal{F}_n(\theta)}{\partial\theta\partial\theta'}) = o_p(1)$ *uniformly in* $\theta \in \Theta$ *and* $\frac{1}{\sqrt{n}}(\frac{\partial \mathcal{F}_n^*(\theta_0)}{\partial\theta} - \frac{\partial \mathcal{F}_n(\theta_0)}{\partial\theta}) = o_p(1)$, *then* $\sqrt{n}(\hat{\theta}_n^* - \theta_0)$ *and* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *have the same limiting distribution.*

Proof. See Lee (2006). ■

Lemma 3.8 *Under Assumption 2, the projectors* M_n *and* $I_n - M_n$, *where* $M_n = X_n(X_n'X_n)^{-1}X_n'$, *are uniformly bounded in both row and column sums in absolute value.*

Proof. See Lee (2004). ■

Lemma 3.9 *Suppose that* $\{||W_n||\}$ *and* $\{||S_n^{-1}||\}$, *where* $||\cdot||$ *is a matrix norm, are bounded. Then* $\{||S_n(\lambda)^{-1}||\}$, *where* $S_n(\lambda) = I_n - \lambda W_n$, *are uniformly bounded in a neighborhood of* λ_0.

Proof. See Lee (2004). ■

Lemma 3.10 *Suppose that z_{1n} and z_{2n} are n-dimensional column vectors of constants which are uniformly bounded, the $n \times n$ constant matrix A_n is uniformly bounded in column sums in absolute value, and B_{1n} and B_{2n} are uniformly bounded in both row and column sums in absolute value, and $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. random variables with zero mean and finite second moment. $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_p(1)$ where α_0 is a p-dimensional vector in the interior of its convex parameter space. The matrix $C_n(\hat{\alpha}_n)$ has the expansion that*

$$\begin{aligned} C_n(\hat{\alpha}_n) - C_n(\alpha_0) &= \sum_{i=1}^{m-1} \sum_{j_1=1}^{p} \cdots \sum_{j_i=1}^{p} (\hat{\alpha}_{nj_1} - \alpha_{j_1 0}) \cdots (\hat{\alpha}_{nj_i} - \alpha_{j_i 0}) K_{in}(\alpha_0) \\ &+ \sum_{j_1=1}^{p} \cdots \sum_{j_m=1}^{p} (\hat{\alpha}_{nj_1} - \alpha_{j_1 0}) \cdots (\hat{\alpha}_{nj_m} - \alpha_{j_m 0}) K_{mn}(\hat{\alpha}_n), \end{aligned} \quad (3.18)$$

for some $m \geq 2$, where $C_n(\alpha_0)$ and $K_{in}(\alpha_0)$ are uniformly bounded in both row and column sums in absolute value for $i = 1, \cdots, m-1$, and $K_{mn}(\alpha)$ is uniformly bounded in both row and column sums in absolute value, uniformly in a small neighborhood of α_0. Then,

(a) $\frac{1}{n} z'_{1n}(C_n(\hat{\alpha}_n) - C_n(\alpha_0)) z_{2n} = o_p(1)$;

(b) $\frac{1}{\sqrt{n}} z'_{1n}(C_n(\hat{\alpha}_n) - C_n(\alpha_0)) A_n \epsilon_n = o_p(1)$;

(c) $\frac{1}{n} \epsilon'_n B'_{1n}(C_n(\hat{\alpha}_n) - C_n(\alpha_0)) B_{2n} \epsilon_n = o_p(1)$, *if* (3.18) *holds for $m > 2$; and*

(d) $\frac{1}{\sqrt{n}} \epsilon'_n(C_n(\hat{\alpha}_n) - C_n(\alpha_0)) \epsilon_n = o_p(1)$, *if* (3.18) *holds for $m > 3$ with $tr(K_{in}(\alpha_0)) = 0$ for $i = 1, \cdots, m-1$.*

Proof. Let $T_n = \frac{1}{n} z'_{1n}(C_n(\hat{\alpha}_n) - C_n(\alpha_0)) z_{2n}$. With (3.18), $T_n = T_{n1} + T_{n2}$, where

$$T_{n1} = \sum_{i=1}^{m-1} \sum_{j_1=1}^{p} \cdots \sum_{j_i=1}^{p} (\hat{\alpha}_{nj_1} - \alpha_{j_1 0}) \cdots (\hat{\alpha}_{nj_i} - \alpha_{j_i 0}) \frac{1}{n} z'_{1n} K_{in}(\alpha_0) z_{2n},$$

and

$$T_{n2} = \sum_{j_1=1}^{p} \cdots \sum_{j_m=1}^{p} (\hat{\alpha}_{nj_1} - \alpha_{j_1 0}) \cdots (\hat{\alpha}_{nj_m} - \alpha_{j_m 0}) \frac{1}{n} z'_{1n} K_{mn}(\hat{\alpha}_n) z_{2n}.$$

$T_{n1} = o_p(1)$ because $\frac{1}{n} z'_{1n} K_{in}(\alpha_0) z_{2n} = O(1)$ by Lemma 3.1, and $\hat{\alpha}_n - \alpha_0 = o_p(1)$. Similarly, as $K_{mn}(\alpha)$ is uniformly bounded in both row and column sums in absolute value, uniformly in a small neighborhood of α_0, and $\hat{\alpha}_n - \alpha_0 = o_p(1)$, it follows that $K_{mn}(\hat{\alpha}_n)$ is uniformly bounded in both row and column sums in absolute value with probability one. Hence $\frac{1}{n} z'_{1n} K_{mn}\left(\hat{\alpha}_n\right) z_{2n} = O_p(1)$ by Lemma 3.1, which implies $T_{n2} = o_p(1)$ because $\hat{\alpha}_n - \alpha_0 = o_p(1)$. This proves (a).

Similarly, let $U_n = \frac{1}{\sqrt{n}} z'_{1n}(C_n(\hat{\alpha}_n) - C_n(\alpha_0)) A_n \epsilon_n$. Then, with (3.18), $U_n = U_{n1} + U_{n2}$ where

$$U_{n1} = \sum_{i=1}^{m-1} \sum_{j_1=1}^{p} \cdots \sum_{j_i=1}^{p} (\hat{\alpha}_{nj_1} - \alpha_{j_1 0}) \cdots (\hat{\alpha}_{nj_i} - \alpha_{j_i 0}) \frac{1}{\sqrt{n}} z'_{1n} K_{in}(\alpha_0) A_n \epsilon_n = o_p(1),$$

because $\frac{1}{\sqrt{n}} z'_{1n} K_{in}(\alpha_0) A_n \epsilon_n = O_p(1)$ by Lemma 3.4 and $\hat{\alpha}_n - \alpha_0 = o_p(1)$; and

$$U_{n2} = \sum_{j_1=1}^{p} \cdots \sum_{j_m=1}^{p} (\hat{\alpha}_{nj_1} - \alpha_{j_1 0}) \cdots (\hat{\alpha}_{nj_m} - \alpha_{j_m 0}) \frac{1}{\sqrt{n}} z'_{1n} K_{mn}(\hat{\alpha}_n) A_n \epsilon_n.$$

Let $\|\cdot\|_1$ be the maximum column sum norm. Because the product of matrices uniformly bounded in the maximum column sum norm is uniformly bounded in the maximum column sum norm, $||K_{mn}(\hat{\alpha}_n) A_n||_1 \leq c_1$ for some constant c_1 for all n. As elements of z_{1n} are uniformly bounded, there exists a constant c_2 such that $\left\| z'_{1n} \right\|_1 \leq c_2$. It follows that

$$\begin{aligned} ||U_{n2}||_1 &\leq n^{(1-m)/2} \sum_{j_1=1}^{p} \cdots \sum_{j_m=1}^{p} |\sqrt{n}(\hat{\alpha}_{nj_1} - \alpha_{j_1 0})| \cdots |\sqrt{n}(\hat{\alpha}_{nj_m} - \alpha_{j_m 0})| \\ &\quad \times ||z'_{1n}||_1 \cdot ||K_{mn}(\hat{\alpha}_n) A_n||_1 \cdot \frac{1}{n} ||\epsilon_n||_1 \\ &\leq c_1 c_2 n^{(1-m)/2} \sum_{j_1=1}^{p} \cdots \sum_{j_m=1}^{p} |\sqrt{n}(\hat{\alpha}_{nj_1} - \alpha_{j_1 0})| \cdots |\sqrt{n}(\hat{\alpha}_{nj_m} - \alpha_{j_m 0})| \cdot (\frac{1}{n} \sum_{i=1}^{n} |\epsilon_{ni}|). \end{aligned}$$

Hence $U_{n2} = o_p(1)$ for $m \geq 2$ because $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_p(1)$ and $\frac{1}{n}\sum_{i=1}^{n} |\epsilon_{ni}| = O_p(1)$ by the strong law of large numbers. These prove (b).

For (c), let $R_n = \frac{1}{n}\epsilon_n' B_{1n}'(C_n(\hat{\alpha}_n) - C_n(\alpha_0))B_{2n}\epsilon_n$. With (3.18), $R_n = R_{n1} + R_{n2}$, where

$$R_{n1} = \sum_{i=1}^{m-1}\sum_{j_1=1}^{p}\cdots\sum_{j_i=1}^{p}(\hat{\alpha}_{nj_1} - \alpha_{j_1 0})\cdots(\hat{\alpha}_{nj_i} - \alpha_{j_i 0})\frac{1}{n}\epsilon_n' B_{1n}' K_{in}(\alpha_0)B_{2n}\epsilon_n,$$

and

$$R_{n2} = \sum_{j_1=1}^{p}\cdots\sum_{j_m=1}^{p}(\hat{\alpha}_{nj_1} - \alpha_{j_1 0})\cdots(\hat{\alpha}_{nj_m} - \alpha_{j_m 0})\frac{1}{n}\epsilon_n' B_{1n}' K_{mn}(\hat{\alpha}_n)B_{2n}\epsilon_n.$$

$R_{n1} = o_p(1)$, because $\frac{1}{n}\epsilon_n' B_{1n}' K_{in}(\alpha_0) B_{2n}\epsilon_n = O_p(1)$ by Lemma 3.3, and $\hat{\alpha}_n - \alpha_0 = o_p(1)$. On the other hand,

$$\begin{aligned}
||R_{n2}||_1 &\leq n^{-m/2}\sum_{j_1=1}^{p}\cdots\sum_{j_m=1}^{p}|\sqrt{n}(\hat{\alpha}_{nj_1} - \alpha_{j_1 0})|\cdots|\sqrt{n}(\hat{\alpha}_{nj_m} - \alpha_{j_m 0})| \\
&\quad \times\frac{1}{n}||\epsilon_n||_1 \cdot ||\epsilon_n||_1 \cdot ||B_{1n}' K_{mn}(\hat{\alpha}_n)B_{2n}||_1 \\
&\leq cn^{1-m/2}\sum_{j_1=1}^{p}\cdots\sum_{j_m=1}^{p}|\sqrt{n}(\hat{\alpha}_{nj_1} - \alpha_{j_1 0})|\cdots|\sqrt{n}(\hat{\alpha}_{nj_m} - \alpha_{j_m 0})| \cdot (\frac{1}{n}\sum_{i=1}^{n}|\epsilon_{ni}|)^2,
\end{aligned}$$

for some constant c. Hence $R_{n2} = o_p(1)$ for $m > 2$ because $\frac{1}{n}\sum_{i=1}^{n}|\epsilon_{ni}|$ converges in probability to the absolute first moment of ϵ_{ni} and $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_p(1)$. These prove (c).

For (d), let $V_n = \frac{1}{\sqrt{n}}\epsilon_n'(C_n(\hat{\alpha}_n) - C_n(\alpha_0))\epsilon_n$. Then, $V_n = V_{n1} + V_{n2}$ where

$$V_{n1} = \sum_{i=1}^{m-1}\sum_{j_1=1}^{p}\cdots\sum_{j_i=1}^{p}(\hat{\alpha}_{nj_1} - \alpha_{j_1 0})\cdots(\hat{\alpha}_{nj_i} - \alpha_{j_i 0})\frac{1}{\sqrt{n}}\epsilon_n' K_{in}(\alpha_0)\epsilon_n = o_p(1),$$

because $\frac{1}{\sqrt{n}}\epsilon_n' K_{in}(\alpha_0)\epsilon_n = O_p(1)$ by Lemma 3.5; and

$$V_{n2} = \frac{1}{\sqrt{n}}\sum_{j_1=1}^{p}\cdots\sum_{j_m=1}^{p}(\hat{\alpha}_{nj_1} - \alpha_{j_1 0})\cdots(\hat{\alpha}_{nj_m} - \alpha_{j_m 0})\epsilon_n' K_{mn}(\hat{\alpha}_n)\epsilon_n.$$

The term $V_{n2} = o_p(1)$ for $m > 3$ because

$$||V_{n2}||_1 \le cn^{(3-m)/2}\sum_{j_1=1}^{p}\cdots\sum_{j_m=1}^{p}|\sqrt{n}(\hat{\alpha}_{nj_1}-\alpha_{j_10})|\cdots|\sqrt{n}(\hat{\alpha}_{nj_m}-\alpha_{j_m0})|\cdot(\frac{1}{n}\sum_{i=1}^{n}|\epsilon_{ni}|)^2.$$

The desired results follow. ■

Lemma 3.11 *Suppose that z_{1n} and z_{2n} are n-dimensional column vectors of constants which are uniformly bounded, the $n \times n$ constant matrix A_n is uniformly bounded in column sums in absolute value, $n \times n$ matrices B_{1n} and B_{2n} are uniformly bounded in both row and column sums in absolute value, and $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. with zero mean and finite fourth moment. Let $\hat{\delta}_n$, $\hat{\mu}_3$ and $\hat{\mu}_4$ be $\sqrt{n}$-consistent estimates of δ_0, μ_3 and μ_4. Then, under Assumption 3,*

(a) $\frac{1}{n}z'_{1n}(\hat{G}'_n - G'_n)^L z_{2n} = o_p(1)$, $\frac{1}{\sqrt{n}}z'_{1n}(\hat{G}'_n - G'_n)^L A_n\epsilon_n = o_p(1)$,

$\frac{1}{n}\epsilon'_n B'_{1n}(\hat{G}_n - G_n)^L B_{2n}\epsilon_n = o_p(1)$, $\frac{1}{\sqrt{n}}\epsilon'_n(\hat{G}_n - G_n)^d\epsilon_n = o_p(1)$;

(b) $\frac{1}{n}z'_{1n}(\hat{G}^{*\prime}_n - G^{*\prime}_n)^L z_{2n} = o_p(1)$, $\frac{1}{\sqrt{n}}z'_{1n}(\hat{G}^{*\prime}_n - G^{*\prime}_n)^L A_n\epsilon_n = o_p(1)$,

$\frac{1}{n}\epsilon'_n B'_{1n}(\hat{G}^*_n - G^*_n)^L B_{2n}\epsilon_n = o_p(1)$, $\frac{1}{\sqrt{n}}\epsilon'_n(\hat{G}^*_n - G^*_n)^d\epsilon_n = o_p(1)$;

(c) $\frac{1}{n}vec'_D(\hat{G}^*_n - G^*_n)^L z_{2n} = o_p(1)$, $\frac{1}{n}tr[A'_n(\hat{G}^*_n - G^*_n)^L] = o_p(1)$; and

(d) $\frac{1}{n}\epsilon'_n B'_{1n}(\hat{P}^*_n - P^*_n)^L B_{2n}\epsilon_n = o_p(1)$, $\frac{1}{\sqrt{n}}\epsilon'_n(\hat{P}^*_n - P^*_n)^d\epsilon_n = o_p(1)$.

Proof. As $S_n - S_n(\hat{\lambda}_n) = (\hat{\lambda}_n - \lambda_0)W_n$, it follows that $\hat{G}_n - G_n = W_n[S_n^{-1}(\hat{\lambda}_n) - S_n^{-1}] = W_nS_n^{-1}(\hat{\lambda}_n)[S_n - S_n(\hat{\lambda}_n)]S_n^{-1} = (\hat{\lambda}_n - \lambda_0)\hat{G}_nG_n$. By induction,

$$\hat{G}_n - G_n = \sum_{i=1}^{m-1}(\hat{\lambda}_n - \lambda_0)^i G_n^{i+1} + (\hat{\lambda}_n - \lambda_0)^m\hat{G}_nG_n^m,$$

which implies

$$(\hat{G}_n - G_n)^L = \sum_{i=1}^{m-1}(\hat{\lambda}_n - \lambda_0)^i(G_n^{i+1})^L + (\hat{\lambda}_n - \lambda_0)^m(\hat{G}_nG_n^m)^L, \qquad (3.19)$$

for any positive integer $m \geq 2$. (3.19) conforms to the expansion (3.18) with $p = 1$, $K_{in}(\lambda_0) = (G_n^{i+1})^L$, and $K_{mn}(\hat{\lambda}_n) = (\hat{G}_n G_n^m)^L$. As G_n^L, $(G_n^{i+1})^L$, and $(\hat{G}_n(G_n^m))^L$ satisfy assumptions in Lemma 3.10, and $tr((G_n^{i+1})^d) = 0$, (a) follows from Lemma 3.10.

For (b), as

$$\begin{aligned} G_n^* &= G_n - \frac{(\eta_4 - 3) - \eta_3^2}{(\eta_4 - 1) - \eta_3^2} D(G_n) - \frac{\eta_3}{\sigma_0[(\eta_4 - 1) - \eta_3^2]} D\left(G_n X_n \beta_0\right) \\ &= G_n - \frac{\sigma_0^2(\mu_4 - 3\sigma_0^4) - \mu_3^2}{\sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2} D(G_n) - \frac{\sigma_0^2 \mu_3}{\sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2} D\left(G_n X_n \beta_0\right), \end{aligned}$$

it follows that

$$\begin{aligned} \hat{G}_n^* - G_n^* &= (\hat{G}_n - G_n) - (1 - \frac{2(\hat{\sigma}_n^2)^3}{\hat{\kappa}}) D(\hat{G}_n - G_n) - (\frac{2\sigma_0^6}{\kappa} - \frac{2(\hat{\sigma}_n^2)^3}{\hat{\kappa}}) D(G_n) \\ &\quad - \frac{\hat{\sigma}_n^2 \hat{\mu}_3}{\hat{\kappa}} D(\hat{G}_n X_n \hat{\beta}_n - G_n X_n \beta_0) - (\frac{\hat{\sigma}_n^2 \hat{\mu}_3}{\hat{\kappa}} - \frac{\sigma_0^2 \mu_3}{\kappa}) D(G_n X_n \beta_0) \quad (3.20) \end{aligned}$$

where $\kappa = \sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2$, with $\hat{\kappa}$ being its empirical counterpart. As $D((\hat{G}_n - G_n)X_n\beta_0)^L = \sum_{i=1}^{m-1}(\hat{\lambda}_n - \lambda_0)^i D(G_n^{i+1}X_n\beta_0)^L + (\hat{\lambda}_n - \lambda_0)^m D(\hat{G}_n G_n^m X_n \beta_0)^L$ conforms to the expansion (3.18) with $p = 1$, $K_{in}(\lambda_0) = D(G_n^{i+1}X_n\beta_0)^L$, and $K_{mn}(\hat{\lambda}_n) = D(\hat{G}_n G_n^m X_n \beta_0)^L$. It is obvious that assumptions in Lemma 3.10 are satisfied. Hence we have $\frac{1}{n} z_{1n}' D'((\hat{G}_n - G_n)X_n\beta_0)^L z_{2n} = o_p(1)$ by Lemma 3.10. On the other hand, let e_{kj} be the jth unit vector in R^k, then $\frac{1}{n} z_{1n}' D'(\hat{G}_n X_n(\hat{\beta}_n - \beta_0))^L z_{2n} = \frac{1}{n}\sum_{i=1}^n z_{1n,i} z_{2n,i} e_{ni}' \hat{G}_n X_n(\hat{\beta}_n - \beta_0) = o_p(1)$ because $\frac{1}{n}\sum_{i=1}^n z_{1n,i} z_{2n,i} e_{ni}' \hat{G}_n X_n = O_p(1)$ as $\hat{\lambda}_n$ is a consistent estimate, and $\hat{\beta}_n - \beta_0 = o_p(1)$. Hence $\frac{1}{n} z_{1n}' D'(\hat{G}_n X_n \hat{\beta}_n - G_n X_n \beta_0)^L z_{2n} = \frac{1}{n} z_{1n}' D'((\hat{G}_n - G_n)X_n\beta_0)^L z_{2n} + \frac{1}{n} z_{1n}' D'(\hat{G}_n X_n(\hat{\beta}_n - \beta_0))^L z_{2n} = o_p(1)$. And the remaining terms in $\frac{1}{n} z_{1n}'(\hat{G}_n^{*\prime} - G_n^{*\prime})^L z_{2n}$ are $o_p(1)$ by (a) and Lemma 3.1. Therefore, $\frac{1}{n} z_{1n}'(\hat{G}_n^{*\prime} - G_n^{*\prime})^L z_{2n} = o_p(1)$. And with similar arguments, the other results in (b) follow.

For (c), as $vec_D'(\hat{G}_n^* - G_n^*)^L = l_n' D(\hat{G}_n^* - G_n^*)^L$, it follows from (b) that $\frac{1}{n} vec_D'(\hat{G}_n^* - G_n^*)^L z_{2n} = o_p(1)$. Since $\hat{G}_n - G_n = (\hat{\lambda}_n - \lambda_0) G_n^2 + (\hat{\lambda}_n - \lambda_0)^2 \hat{G}_n G_n^2$, it follows

$$\frac{1}{n} tr[A_n'(\hat{G}_n - G_n)^L] = (\hat{\lambda}_n - \lambda_0) \frac{1}{n} tr(A_n' G_n^{2L}) + (\hat{\lambda}_n - \lambda_0)^2 \frac{1}{n} tr[A_n'(\hat{G}_n G_n^2)^L] = o_p(1),$$

because $\frac{1}{n} tr(A_n' G_n^{2L}) = O(1)$, $\frac{1}{n} tr(A_n'(\hat{G}_n G_n^2)^L) = O_p(1)$ and $(\hat{\lambda}_n - \lambda_0) = o_p(1)$. Similarly,

$$\begin{aligned}
& \frac{1}{n} tr[A_n' D(\hat{G}_n X_n \hat{\beta}_n - G_n X_n \beta_0)] \\
= & \frac{1}{n} tr[A_n' D((\hat{G}_n - G_n) X_n \beta_0 + \hat{G}_n X_n (\hat{\beta}_n - \beta_0))] \\
= & (\hat{\lambda}_n - \lambda_0) \frac{1}{n} tr[A_n' D(G_n^2 X_n \beta_0)] + (\hat{\lambda}_n - \lambda_0)^2 \frac{1}{n} tr[A_n' D(\hat{G}_n G_n^2 X_n \beta_0)] \\
& + \frac{1}{n} tr[A_n' D(\hat{G}_n X_n (\hat{\beta}_n - \beta_0))] \\
= & o_p(1).
\end{aligned}$$

As $tr[A_n' D(G_n)] = O(1)$, $tr[A_n' D(G_n X_n \beta_0)] = O(1)$, and $\hat{\sigma}_n^2, \hat{\mu}_3, \hat{\kappa}$ are consistent estimates, it follows that $\frac{1}{n} tr[A_n'(\hat{G}_n^* - G_n^*)^L] = o_p(1)$.

For (d), explicitly,

$$\begin{aligned}
\hat{P}_n^* - P_n^* &= \left(\hat{G}_n - \frac{\hat{\eta}_4 - 3}{\hat{\eta}_4 - 1} D(\hat{G}_n)\right)^d - \left(G_n - \frac{\eta_4 - 3}{\eta_4 - 1} D\left(G_n\right)\right)^d \\
&= (\hat{G}_n - G_n)^d - \frac{\hat{\eta}_4 - 3}{\hat{\eta}_4 - 1} D(\hat{G}_n - G_n)^d - \left(\frac{\hat{\eta}_4 - 3}{\hat{\eta}_4 - 1} - \frac{\eta_4 - 3}{\eta_4 - 1}\right) D\left(G_n^d\right).
\end{aligned}$$

As $\cdot^d$ and $D(\cdot)$ are linear transformations that preserve the uniform boundedness property of the original matrix, and $(\hat{\eta}_4 - 3)/(\hat{\eta}_4 - 1)$ is a consistent estimate of $(\eta_4 - 3)/(\eta_4 - 1)$, the desired result follows from (a) and Lemmas 3.3 and 3.5. ■

Lemma 3.12 *Suppose that z_n is an n-dimensional column vector of constants which are uniformly bounded, the $n \times n$ constant matrix A_n is uniformly bounded in column sums in absolute value, and $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. with zero mean and finite fourth moment. Let $\hat{\delta}_n$,*

$\hat{\mu}_3$ and $\hat{\mu}_4$ be $\sqrt{n}$-consistent estimates of δ_0, μ_3 and μ_4. Let $Q_n^* = [Q_{1n}^*, Q_{2n}^*]$ be given by (3.15) *and* (3.16)*, with $\hat{Q}_n^* = [\hat{Q}_{1n}^*, \hat{Q}_{2n}^*]$ being their empirical counterparts. Then, under Assumption 3,*

(a) $\frac{1}{n}(\hat{Q}_n^* - Q_n^*)' z_n = o_p(1)$, *and*

(b) $\frac{1}{\sqrt{n}}(\hat{Q}_n^* - Q_n^*)' A_n \epsilon_n = o_p(1)$.

Proof. Let $\kappa = \sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2$, with $\mu_3 = \eta_3\sigma_0^3$, $\mu_4 = \eta_4\sigma_0^4$, and $\hat{\kappa}$ being the empirical counterpart.

$$\hat{Q}_{1n}^* - Q_{1n}^* = -(\frac{\hat{\mu}_3^2}{\hat{\kappa}} - \frac{\mu_3^2}{\kappa})\frac{1}{n} l_n l_n' X_n, \tag{3.21}$$

and

$$\begin{aligned}
\hat{Q}_{2n}^* - Q_{2n}^* &= (\hat{G}_n X_n \hat{\beta}_n - G_n X_n \beta_0) + (I_n - \frac{1}{n} l_n l_n')\left(\frac{\hat{\mu}_3^2}{\hat{\kappa}} \hat{G}_n X_n \hat{\beta}_n - \frac{\mu_3^2}{\kappa} G_n X_n \beta_0\right) \\
&\quad - \left(\frac{2(\hat{\sigma}_n^2)^2 \hat{\mu}_3}{\hat{\kappa}} vec_D(\hat{G}_n^d) - \frac{2\sigma_0^4 \mu_3}{\kappa} vec_D(G_n^d)\right) \\
&= [I_n + \frac{\hat{\mu}_3^2}{\hat{\kappa}}(I_n - \frac{1}{n} l_n l_n')](\hat{G}_n X_n \hat{\beta}_n - G_n X_n \beta_0) + R_{n1} \\
&\quad - \frac{2(\hat{\sigma}_n^2)^2 \hat{\mu}_3}{\hat{\kappa}} vec_D(\hat{G}_n - G_n)^d - R_{n2},
\end{aligned} \tag{3.22}$$

where $R_{n1} = (\frac{\hat{\mu}_3^2}{\hat{\kappa}} - \frac{\mu_3^2}{\kappa})(I_n - \frac{1}{n} l_n l_n') G_n X_n \beta_0$ and $R_{n2} = (\frac{2(\hat{\sigma}_n^2)^2 \hat{\mu}_3}{\hat{\kappa}} - \frac{2\sigma_0^4 \mu_3}{\kappa}) vec_D(G_n^d)$.

As $\frac{1}{n} l_n l_n'$ is uniformly bounded in both row and column sums in absolute value, Lemma 3.1 implies that $\frac{1}{n}(\frac{1}{n} l_n l_n' X_n)' z_n = O(1)$. Hence, $\frac{1}{n}(\hat{Q}_{1n}^* - Q_{1n}^*)' z_n = o_p(1)$ as $\hat{\mu}_3$ and $\hat{\kappa}$ are consistent estimates.

$$\begin{aligned}
\frac{1}{n}(\hat{Q}_{2n}^* - Q_{2n}^*)' z_n &= [I_n + \frac{\hat{\mu}_3^2}{\hat{\kappa}}(I_n - \frac{1}{n} l_n l_n')]\frac{1}{n}(\hat{G}_n X_n \hat{\beta}_n - G_n X_n \beta_0)' z_n + \frac{1}{n} R_{n1}' z_n \\
&\quad - \frac{2(\hat{\sigma}_n^2)^2 \hat{\mu}_3}{\hat{\kappa}} \frac{1}{n} vec_D'(\hat{G}_n - G_n)^d z_n - \frac{1}{n} R_{n2}' z_n.
\end{aligned}$$

Since $\frac{1}{n}[(\hat{G}_n - G_n)X_n]'z_n = o_p(1)$ by Lemma 3.11, and $(\hat{\beta}_n - \beta_0)\frac{1}{n}(G_nX_n)'z_n = o_p(1)$ as $\hat{\beta}_n - \beta_0 = o_p(1)$, it follows that $\frac{1}{n}(\hat{G}_nX_n\hat{\beta}_n - G_nX_n\beta_0)'z_n = \frac{1}{n}[(\hat{G}_n - G_n)X_n\hat{\beta}_n + G_nX_n(\hat{\beta}_n - \beta_0)]'z_n = o_p(1)$. For the remaining terms, $\frac{1}{n}vec_D'(\hat{G}_n - G_n)^d z_n = \frac{1}{n}l_n'D(\hat{G}_n - G_n)^d z_n = o_p(1)$ by Lemma 3.11, and $\frac{1}{n}R_{n1}'z_n = o_p(1)$, $\frac{1}{n}R_{n2}'z_n = o_p(1)$ as $\hat{\delta}_n$, $\hat{\mu}_3$ and $\hat{\mu}_4$ are consistent estimates. Hence $\frac{1}{n}(\hat{Q}_{2n}^* - Q_{2n}^*)'z_n = o_p(1)$. This proves (a).

Lemma 3.4 implies that $\frac{1}{n}(\frac{1}{n}l_nl_n'X_n)'A_n\epsilon_n = o_p(1)$. Hence

$$\frac{1}{\sqrt{n}}(\hat{Q}_{1n}^* - Q_{1n}^*)'A_n\epsilon_n = -\sqrt{n}(\frac{\hat{\mu}_3^2}{\hat{\kappa}} - \frac{\mu_3^2}{\kappa})\frac{1}{n}(\frac{1}{n}l_nl_n'X_n)'A_n\epsilon_n = o_p(1),$$

as $\hat{\mu}_3$ and $\hat{\kappa}$ are $\sqrt{n}$-consistent estimates. The first term in $\frac{1}{\sqrt{n}}(\hat{Q}_{2n}^* - Q_{2n}^*)'A_n\epsilon_n$ is $o_p(1)$ because

$$\begin{aligned} & \frac{1}{\sqrt{n}}(\hat{G}_nX_n\hat{\beta}_n - G_nX_n\beta_0)'A_n\epsilon_n \\ = & \frac{1}{\sqrt{n}}\hat{\beta}_n'X_n'(\hat{G}_n - G_n)'A_n\epsilon_n + \sqrt{n}(\hat{\beta}_n - \beta_0)'\frac{1}{n}X_n'G_n'A_n\epsilon_n = o_p(1), \end{aligned}$$

where $\frac{1}{\sqrt{n}}\hat{\beta}_n'X_n'(\hat{G}_n - G_n)'A_n\epsilon_n = o_p(1)$ by Lemma 3.11, $\frac{1}{n}X_n'G_n'A_n\epsilon_n = o_p(1)$ by Lemma 3.4, and $\sqrt{n}(\hat{\beta}_n - \beta_0) = O_p(1)$. Similarly, the remaining terms in $\frac{1}{\sqrt{n}}(\hat{Q}_{2n}^* - Q_{2n}^*)'A_n\epsilon_n$ are also $o_p(1)$. The desired results follow. ■

Lemma 3.13 *Suppose that the elements of the $n \times k$ matrix C_n are uniformly bounded, the $n \times n$ matrix A_n is uniformly bounded in column sums in absolute value, and $\hat{\lambda}_n$ is a $\sqrt{n}$-consistent estimator. Then, $\frac{1}{\sqrt{n}}C_n'\hat{G}_n^LA_n\epsilon_n = O_p(1)$, and $\frac{1}{\sqrt{n}}C_n'\hat{P}_n^*A_n\epsilon_n = O_p(1)$, where $\hat{P}_n^* = \hat{G}_n^d - \frac{\hat{\eta}_4-3}{\hat{\eta}_4-1}D(\hat{G}_n^d)$.*

Proof. $\frac{1}{\sqrt{n}}C_n'\hat{G}_n^LA_n\epsilon_n = O_p(1)$ is a case of Lee (2001*a*) Lemma A.11. As $(\hat{\eta}_4 - 3)/(\hat{\eta}_4 - 1) = O_p(1)$, it follows that

$$\frac{1}{\sqrt{n}}C_n'\hat{P}_n^*A_n\epsilon_n = \frac{1}{\sqrt{n}}C_n'\hat{G}_n^dA_n\epsilon_n - \frac{\hat{\eta}_4 - 3}{\hat{\eta}_4 - 1}\frac{1}{\sqrt{n}}C_n'D(\hat{G}_n^d)A_n\epsilon_n = O_p(1).$$

■

Lemma 3.14 *Suppose that A_n, B_n and C_n are matrices uniformly bounded in column sums in absolute value, X_n satisfies Assumption 2, and $\widehat{\lambda}_n$ is a $\sqrt{n}$-consistent estimator of λ_0. Then, $\epsilon_n' A_n' \hat{P}_n^* B_n' M_n C_n \epsilon_n = O_p(1)$, and $\epsilon_n' C_n' M_n A_n \hat{P}_n^* B_n' M_n C_n \epsilon_n = O_p(1)$, where $\hat{P}_n^* = \hat{G}_n^d - \frac{\hat{\eta}_4 - 3}{\hat{\eta}_4 - 1} D(\hat{G}_n^d)$ and $M_n = X_n \left(X_n' X_n\right)^{-1} X_n'$.*

Proof. As A_n' and B_n' are uniformly bounded in row sums in absolute value and elements of X_n are uniformly bounded, elements of $A_n' X_n$ and $B_n' X_n$ are uniformly bounded. Hence, by Lemmas 3.4 and 3.13,

$$\epsilon_n' A_n' \hat{P}_n^* B_n' M_n C_n \epsilon_n = (\frac{1}{\sqrt{n}} \epsilon_n' A_n' \hat{P}_n^* B_n' X_n)(\frac{1}{n} X_n' X_n)^{-1} (\frac{1}{\sqrt{n}} X_n' C_n \epsilon_n) = O_p(1).$$

On the other hand,

$$\begin{aligned} & \epsilon_n' C_n' M_n A_n \hat{P}_n^* B_n' M_n C_n \epsilon_n \\ = \; & (\frac{1}{\sqrt{n}} \epsilon_n' C_n' X_n)(\frac{1}{n} X_n' X_n)^{-1} (\frac{1}{n} X_n' A_n \hat{P}_n^* B_n' X_n)(\frac{1}{n} X_n' X_n)^{-1} (\frac{1}{\sqrt{n}} X_n' C_n \epsilon_n). \end{aligned}$$

Under Assumption 3, because S_n^{-1} is uniformly bounded in both row and column sums in absolute value, $S_n^{-1}(\lambda)$ and, hence, $G_n(\lambda)$ must be uniformly bounded in both row and column sums in absolute value, uniformly in λ in a small neighborhood of λ_0, by Lemma 3.9. As $\hat{\lambda}_n$ and $\hat{\eta}_4$ are consistent, it follow that $\hat{G}_n$ and, hence, $\hat{P}_n^*$ are uniformly bounded in both row and column sums in absolute value with probability one. Therefore, Lemma 3.1 implies that $\frac{1}{n} X_n' A_n \hat{P}_n^* B_n' X_n = O_p(1)$. The desired result follows from Lemma 3.4. ∎

To show the proposed moment conditions are optimal, we show adding additional moment conditions to the optimal moment conditions does not increase the asymptotic efficiency of the GMME using the conditions for redundancy in Breusch et al. (1999). The definition of redundancy is given as follows. "Let $\hat{\theta}$ be the optimal GMME based on a set of (unconditional) moment conditions $E\left[g_1(y,\theta)\right] = 0$. Now add some extra moment conditions $E\left[g_2(y,\theta)\right] = 0$ and let $\tilde{\theta}$ be the optimal GMME based on the whole set of moment conditions $E\left[g(y,\theta)\right] \equiv E\left[g_1'(y,\theta), g_2'(y,\theta)\right]' = 0$. We say that the moment conditions $E\left[g_2(y,\theta)\right] = 0$ are redundant given the moment conditions $E\left[g_1(y,\theta)\right] = 0$, or simply that g_2 is redundant given g_1, if the asymptotic variances of $\hat{\theta}$ and $\tilde{\theta}$ are the same" (Breusch et al., 1999, p. 90). For moment conditions $E\left[g(y,\theta)\right] \equiv E\left[g_1'(y,\theta), g_2'(y,\theta)\right]' = 0$, let

$$\Omega \equiv E\left[g(y,\theta)\, g'(y,\theta)\right] = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix},$$

with $\Omega_{jl} = E\left[g_j(y,\theta)\, g_l'(y,\theta)\right]$ for $j, l = 1, 2$. And define $D_j = E\left[\partial g_j(y,\theta)/\partial\theta'\right]$ for $j = 1, 2$. Let the dimensions of $g_1(y,\theta)$, $g_2(y,\theta)$ and θ be k_1, k_2 and p.

Lemma 3.15 *The following statements are equivalent.*

(a) g_2 *is redundant given* g_1.

(b) $D_2 = \Omega_{21}\Omega_{11}^{-1}D_1$.

(c) There exists a $k_1 \times p$ *matrix* A *such that* $D_1 = \Omega_{11}A$ *and* $D_2 = \Omega_{21}A$.

Proof. Breusch et al. (1999) Theorem 1 (A), (C), and (D), respectively. ■

Lemma 3.16 *Let the set of moment conditions to be considered be*

$$E\left[g(\theta)\right] \equiv E\left[g_1'(\theta), g_2'(\theta), g_3'(\theta)\right]' = 0,$$

or simply $g = \left(g_1', g_2', g_3'\right)'$. *Then* $\left(g_2', g_3'\right)'$ *is redundant given* g_1 *if and only if* g_2 *is redundant given* g_1 *and* g_3 *is redundant given* g_1.

Proof. Breusch et al. (1999) Theorem 2. ■

3.5.3 Proofs

Proof of Proposition 3.1. There are two possible approaches to establish the result. The first approach extends the optimization of variances of the GMME. The second one is a constructive argument based on Breusch et al. (1999). Here we present both approaches.

(1) The first approach derives the best moment function P_n^* analytically. With m quadratic moments in $g_n(\lambda)$, $var(g_n(\lambda_0)) = \sigma_0^4 \Omega_n$, where

$$\Omega_n = (\eta_4 - 3)\omega_m' \omega_m + V_n,$$

with $\omega_m = [vec_D(P_{1n}), \cdots, vec_D(P_{mn})]$ and

$$\begin{aligned} V_n &= \frac{1}{2}(vec(P_{1n}^s), \cdots, vec(P_{mn}^s))'(vec(P_{1n}^s), \cdots, vec(P_{mn}^s)) \\ &= \begin{pmatrix} tr(P_{1n}^s P_{1n}) & \cdots & tr(P_{1n}^s P_{mn}) \\ \vdots & \ddots & \vdots \\ tr(P_{mn}^s P_{1n}) & \cdots & tr(P_{mn}^s P_{mn}) \end{pmatrix}. \end{aligned} \tag{3.23}$$

The two terms in Ω_n can be combined into a unified one as follows. First, because

$$\begin{aligned} & tr(P_{jn}^s P_{ln}) - vec(P_{jn} - D(P_{jn}))^s vec(P_{jn} - D(P_{jn})) \\ = \;& tr(P_{jn}^s P_{ln}) - tr[(P_{jn} - D(P_{jn}))^s (P_{jn} - D(P_{jn}))] \\ = \;& tr(P_{jn}^s P_{ln}) - tr[(P_{jn} - D(P_{jn}))^s P_{ln}] \\ = \;& 2tr[D(P_{jn})P_{ln}] = 2tr[D(P_{jn})D(P_{ln})] = 2vec_D'(P_{jn})vec_D(P_{ln}), \end{aligned}$$

for any j and l, we have

$$\begin{pmatrix} tr(P_{1n}^s P_{1n}) & \cdots & tr(P_{1n}^s P_{mn}) \\ \vdots & \ddots & \vdots \\ tr(P_{mn}^s P_{1n}) & \cdots & tr(P_{mn}^s P_{mn}) \end{pmatrix} - 2\omega_m'\omega_m = \frac{1}{2}\varpi_m'\varpi_m,$$

where $\varpi_m = [vec(P_{1n} - D(P_{1n}))^s, \cdots, vec(P_{mn} - D(P_{mn}))^s]$. Therefore,

$$\Omega_n = \frac{1}{2}\left[2(\eta_4 - 1)\omega_m'\omega_m + \varpi_m'\varpi_m\right].$$

Define the modified matrices $P_{jn}^{+} = P_{jn} - D(P_{jn}) + \sqrt{\frac{\eta_4 - 1}{2}} D(P_{jn})$ for $j = 1, \cdots, m$. As

$$
\begin{aligned}
& vec'(P_{jn}^{+s}) vec(P_{kn}^{+s}) \\
= \ & tr(P_{jn}^{+s} P_{kn}^{+s}) \\
= \ & tr\{[P_{jn}^{s} - D(P_{jn}^{s})][P_{kn}^{s} - D(P_{kn}^{s})]\} + 2(\eta_4 - 1) tr[D(P_{jn}) D(P_{kn})] \\
= \ & vec'[(P_{jn} - D(P_{jn}))^{s}] vec[(P_{kn} - D(P_{kn}))^{s}] + 2(\eta_4 - 1) vec'_D(P_{jn}) vec_D(P_{kn}),
\end{aligned}
$$

it follows that $\Omega_n = \frac{1}{2}(vec(P_{1n}^{+s}), \cdots, vec(P_{kn}^{+s}))'(vec(P_{1n}^{+s}), \cdots, vec(P_{mn}^{+s}))$.

Consider now $tr(P_{jn}^{s} G_n) = tr(P_{jn}^{s}(G_n - \frac{tr(G_n)}{n} I_n))$. We would like to find a matrix A_n such that $tr(P_{jn}^{s}(G_n - \frac{tr(G_n)}{n} I_n)) = tr(P_{jn}^{+s}(G_n - \frac{tr(G_n)}{n} I_n + A_n))$ holds for all j. By taking A_n to be a diagonal matrix, we see that the solution is

$$
A_n = (\sqrt{\frac{2}{\eta_4 - 1}} - 1)(D(G_n) - \frac{tr(G_n)}{n} I_n),
$$

which is invariant with any P_{nj}. Denote

$$
\begin{aligned}
G_n^{-} &= G_n - \frac{tr(G_n)}{n} I_n + A_n \\
&= G_n - \frac{tr(G_n)}{n} I_n + (\sqrt{\frac{2}{\eta_4 - 1}} - 1)(D(G_n) - \frac{tr(G_n)}{n} I_n),
\end{aligned}
$$

which has zero trace. Therefore, $tr(P_{jn}^{s} G_n) = tr(P_{jn}^{+s} G_n^{-})$.

Following Lee (2001*a*), the limit variance of the GMME with P_{jn}, $j = 1, \cdots, m$, is $\Sigma_P^{-1} = (\lim_{n \to \infty} \frac{1}{n} \Sigma_{P,n})^{-1}$, where

$$
\Sigma_{P,n} = (tr(P_{1n}^{s} G_n), \cdots, tr(P_{mn}^{s} G_n)) \Omega_n^{-1} (tr(P_{1n}^{s} G_n), \cdots, tr(P_{mn}^{s} G_n))'.
$$

With the above manipulation, $\Sigma_{P,n}$ can be rewritten as

$$
\begin{aligned}
\Sigma_{P,n} = \ & 2 vec'(G_n^{-})(vec(P_{1n}^{+s}), \cdots, vec(P_{mn}^{+s})) \\
& \cdot [(vec(P_{1n}^{+s}), \cdots, vec(P_{mn}^{+s}))'(vec(P_{1n}^{+s}), \cdots, vec(P_{mn}^{+s}))]^{-1} \\
& \cdot (vec(P_{1n}^{+s}), \cdots, vec(P_{mn}^{+s}))' vec(G_n^{-}).
\end{aligned}
$$

By the generalized Schwartz inequality, $\Sigma_{P,n} \leq 2vec'(G_n^-)vec(G_n^-)$, which provides a bound for the precision matrix $\Sigma_{P,n}$ for any GMME with a finite number of quadratic moments. This bound can be obtained with a corresponding optimum

$$P_n^{+*} = (G_n - \frac{tr(G_n)}{n} I_n) + (\sqrt{\frac{2}{\eta_4 - 1}} - 1)(D(G_n) - \frac{tr(G_n)}{n} I_n).$$

With P_n^+ transformed back to the P_n, the best P_n^* is

$$\begin{aligned} P_n^* &= P_n^{+*} - D(P_n^{+*}) + \sqrt{\frac{2}{\eta_4 - 1}} D(P_n^{+*}) \\ &= (G_n - \frac{tr(G_n)}{n} I_n) - \frac{\eta_4 - 3}{\eta_4 - 1}(D(G_n) - \frac{tr(G_n)}{n} I_n). \end{aligned}$$

(2) In the second approach, we show that $g_n^*(\lambda) = u_n' S_n'(\lambda) P_n^* S_n(\lambda) u_n$ is the best moment function in the sense that any other moment functions are redundant given $g_n^*(\lambda)$. Following Lemma 3.16, to show that any other finite number of moment functions are redundant given $g_n^*(\lambda)$, it is equivalent to show that an arbitrary single moment function is redundant given $g_n^*(\lambda)$. Let P_n be an arbitrary $n \times n$ constant matrix satisfying Assumption 4, and $g_n(\lambda) = u_n' S_n'(\lambda) P_n S_n(\lambda) u_n$. Consider the moment conditions

$$E(\zeta_n(\lambda_0)) = E\begin{pmatrix} g_n^*(\lambda_0) \\ g_n(\lambda_0) \end{pmatrix} = 0.$$

Let

$$E\left[\zeta_n(\lambda_0)\zeta_n'(\lambda_0)\right] = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

$$E(\frac{\partial \zeta_n(\lambda_0)}{\partial \lambda}) = E\begin{pmatrix} \partial g_n^*(\lambda_0)/\partial\lambda \\ \partial g_n(\lambda_0)/\partial\lambda \end{pmatrix} = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix}.$$

According to Lemma 3.15 (b), to show that $g_n(\lambda)$ is redundant given $g_n^*(\lambda)$ it's sufficient to show that $D_2^{-1}\Omega_{21} = D_1^{-1}\Omega_{11}$. And because P_n^* is a special case of P_n, it's sufficient to show $D_2^{-1}\Omega_{21}$ is invariant with P_n. Following Lemma 3.2, we have

$$\Omega_{21} = \sigma_0^4[tr\left(P_n^s P_n^*\right) + \left(\eta_4 - 3\right) vec_D'\left(P_n\right) vec_D\left(P_n^*\right)],$$

where

$$
\begin{aligned}
tr\left(P_n^s P_n^*\right) &= tr\left(P_n^s G_n\right) - \frac{\eta_4-3}{\eta_4-1} tr\left(P_n^s D(G_n)\right) - \frac{1}{n}\left(\frac{2}{\eta_4-1}\right) tr\left(G_n\right) tr\left(P_n^s\right) \\
&= tr\left(P_n^s G_n\right) - \frac{\eta_4-3}{\eta_4-1} tr\left(P_n^s D(G_n)\right),
\end{aligned}
$$

and

$$
vec_D\left(P_n^*\right) = \frac{2}{\eta_4-1}\left(vec_D\left(G_n\right) - \frac{1}{n} tr\left(G_n\right) l_n\right).
$$

As $vec_D'(A)\, vec_D(B) = tr\,(A \cdot D(B))$, and $vec_D'(P_n)\, l_n = 0$, it follows that

$$
\begin{aligned}
vec_D'\left(P_n\right) vec_D\left(P_n^*\right) &= \frac{2}{\eta_4-1}\left(vec_D'\left(P_n\right) vec_D\left(G_n\right) - \frac{1}{n} tr\left(G_n\right) vec_D'\left(P_n\right) l_n\right) \\
&= \frac{2}{\eta_4-1} tr\left(P_n D(G_n)\right) = \frac{1}{\eta_4-1} tr\left(P_n^s D(G_n)\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\Omega_{21} &= \sigma_0^4[tr\left(P_n^s G_n\right) - \frac{\eta_4-3}{\eta_4-1} tr\left(P_n^s D(G_n)\right) + \left(\eta_4-3\right) \frac{1}{\eta_4-1} tr\left(P_n^s D(G_n)\right)] \\
&= \sigma_0^4 tr\left(P_n^s G_n\right).
\end{aligned}
$$

And since $D_2 = -\sigma_0^2 tr\left(P_n^s G_n\right)$ (Lee, 2001*a*), we have $D_2^{-1}\Omega_{21} = -\sigma_0^2$, which is invariant with P_n.

Furthermore, let the asymptotic variance of the consistent root derived from $\min g_n^{*2}(\lambda)$ be Σ_B^{-1}. As $\Sigma_B = \lim_{n\to\infty} \frac{1}{n} D_1' \Omega_{11}^{-1} D_1$ (Lee, 2001*a*), where $D_1 = -\sigma_0^2 tr\left(P_n^{*s} G_n\right)$ and $\Omega_{11}^{-1} D_1 = \Omega_{21}^{-1} D_2 = -\sigma_0^{-2}$, it follows that $\Sigma_B = \lim_{n\to\infty} \frac{1}{n} tr\left(P_n^{*s} G_n\right)$. ■

Proof of Proposition 3.2. The proof is divided into two steps. In the first step, we show that if u_n is observable, $g_n^*(\lambda) = u_n' S_n'(\lambda)\, P_n^* S_n(\lambda)\, u_n$ and $\tilde{g}_n^*(\lambda) = u_n' S_n'(\lambda)\, \hat{P}_n^* S_n(\lambda)\, u_n$ are asymptotic equivalent in the sense that their consistent roots have the same limiting distribution. In the second step, we show that $\tilde{g}_n^*(\lambda) = u_n' S_n'(\lambda)\, \hat{P}_n^* S_n(\lambda)\, u_n$ and $\hat{g}_{F,n}^*(\lambda) = \hat{u}_n' S_n'(\lambda)\, \hat{P}_n^* S_n(\lambda)\, \hat{u}_n$ are asymptotic equivalent.

(1) For consistency of the root of $\tilde{g}_n^*(\lambda) = 0$, it is sufficient to show that $\frac{1}{n}\tilde{g}_n^*(\lambda) - \frac{1}{n}g_n^*(\lambda) = o_p(1)$ uniformly in $\lambda \in \Lambda$. Explicitly, $\frac{1}{n}(\tilde{g}_n^*(\lambda) - g_n^*(\lambda)) = T_{n1} - \lambda T_{n2} + \lambda^2 T_{n3}$ where $T_{n1} = \frac{1}{n}\epsilon_n' S_n'^{-1}(\hat{P}_n^* - P_n^*)S_n^{-1}\epsilon_n$, $T_{n2} = \frac{1}{n}\epsilon_n' G_n'(\hat{P}_n^* - P_n^*)S_n^{-1}\epsilon_n$, and $T_{n3} = \frac{1}{n}\epsilon_n' G_n'(\hat{P}_n^* - P_n^*)G_n\epsilon_n$. The terms T_{nj}, $j = 1, 2, 3$, are all of order $o_p(1)$ by Lemma 3.11. Hence $\frac{1}{n}\tilde{g}_n^*(\lambda) - \frac{1}{n}g_n^*(\lambda) = o_p(1)$ uniformly in $\lambda \in \Lambda$. The consistency of the root of $\tilde{g}_n^*(\lambda) = 0$ follows from the first part of Lemma 3.6.

For the asymptotic distribution of the root of $\tilde{g}_n^*(\lambda) = 0$, consider $\frac{1}{n}(\frac{\partial \tilde{g}_n^*(\lambda)}{\partial \lambda} - \frac{\partial g_n^*(\lambda)}{\partial \lambda})$ and $\frac{1}{\sqrt{n}}(\tilde{g}_n^*(\lambda_0) - g_n^*(\lambda_0))$. As $S_n(\lambda) = S_n - (\lambda - \lambda_0)W_n$,

$$\begin{aligned}
\frac{1}{n}u_n' S_n'(\lambda)\hat{P}_n^{*s} W_n u_n &= \frac{1}{n}\epsilon_n' \hat{P}_n^{*s} G_n \epsilon_n - (\lambda - \lambda_0)\frac{1}{n}\epsilon_n' G_n' \hat{P}_n^{*s} G_n \epsilon_n \\
&= \frac{1}{n}\epsilon_n' P_n^{*s} G_n \epsilon_n - (\lambda - \lambda_0)\frac{1}{n}\epsilon_n' G_n' P_n^{*s} G_n \epsilon_n + R_{n1} + R_{n2} \\
&= \frac{1}{n}u_n' S_n'(\lambda) P_n^{*s} W_n u_n + R_{n1} + R_{n2},
\end{aligned}$$

where $R_{n1} = \frac{1}{n}\epsilon_n'(\hat{P}_n^* - P_n^*)^s G_n \epsilon_n$ and $R_{n2} = \frac{1}{n}\epsilon_n' G_n'(\hat{P}_n^* - P_n^*)^s G_n \epsilon_n$. It follows from Lemma 3.11 that $R_{n1} = o_p(1)$ and $R_{n2} = o_p(1)$. Hence,

$$\frac{1}{n}u_n' S_n'(\lambda)\hat{P}_n^{*s} W_n u_n = \frac{1}{n}u_n' S_n'(\lambda) P_n^{*s} W_n u_n + o_p(1),$$

uniformly in $\lambda \in \Lambda$, i.e., $\frac{1}{n}(\frac{\partial \tilde{g}_n^*(\lambda)}{\partial \lambda} - \frac{\partial g_n^*(\lambda)}{\partial \lambda}) = o_p(1)$ uniformly in $\lambda \in \Lambda$. For the other term,

$$\frac{1}{\sqrt{n}}u_n' S_n' \hat{P}_n^* S_n u_n = \frac{1}{\sqrt{n}}\epsilon_n' P_n^* \epsilon_n + \frac{1}{\sqrt{n}}\epsilon_n'(\hat{P}_n^* - P_n^*)\epsilon_n = \frac{1}{\sqrt{n}}u_n' S_n' P_n^* S_n u_n + o_p(1),$$

by Lemma 3.11, i.e., $\frac{1}{\sqrt{n}}(\tilde{g}_n^*(\lambda_0) - g_n^*(\lambda_0)) = o_p(1)$. Hence, by Lemma 3.6, the feasible GMME derived from $\min_{\lambda \in \Lambda}[u_n' S_n'(\lambda)\hat{P}_n^* S_n(\lambda)u_n]^2$ has the same limiting distribution as that derived from $\min_{\lambda \in \Lambda}[u_n' S_n'(\lambda) P_n^* S_n(\lambda) u_n]^2$.

(2) It is sufficient to show that the moment function $\hat{g}_{F,n}^*(\lambda)$ and its derivative are close enough to those of $\tilde{g}_n^*(\lambda)$ so that Lemma 3.6 is applicable. Specifically, it shall be shown

that $\hat{g}^*_{F,n}(\lambda) - \tilde{g}^*_n(\lambda) = O_p(1)$ and $\frac{\partial \hat{g}^*_{F,n}(\lambda)}{\partial\lambda} - \frac{\partial \tilde{g}^*_n(\lambda)}{\partial\lambda} = O_p(1)$ uniformly in $\lambda \in \Lambda$. These properties are stronger than those sufficient conditions in Lemma 3.6.

Because $\hat{u}_n = (I_n - M_n)u_n$, $\hat{g}^*_{F,n}(\lambda) = \tilde{g}^*_n(\lambda) + E_n(\lambda)$ where

$$E_n(\lambda) = -u_n' S_n'(\lambda)\hat{P}_n^{*s} S_n(\lambda) M_n u_n + u_n' M_n S_n'(\lambda)\hat{P}_n^{*} S_n(\lambda) M_n u_n.$$

Substitute $u_n = S_n^{-1}\epsilon_n$ in the terms of $E_n(\lambda)$. Lemma 3.14 is applicable and all the terms of $E_n(\lambda)$ are of order $O_p(1)$ uniformly in $\lambda \in \Lambda$. The uniform order holds because λ is linear in $S_n(\lambda)$. Hence, $\hat{g}^*_{F,n}(\lambda) = \tilde{g}^*_n(\lambda) + O_p(1)$ uniformly in $\lambda \in \Lambda$. Consequently, one has, in particular, that $\frac{1}{n}\hat{g}^*_{F,n}(\lambda) = \frac{1}{n}\tilde{g}^*_n(\lambda) + o_p(1)$ and $\frac{1}{\sqrt{n}}\hat{g}^*_{F,n}(\lambda_0) = \frac{1}{\sqrt{n}}\tilde{g}^*_n(\lambda_0) + o_p(1)$.

The first order derivative of $\hat{g}^*_{F,n}(\lambda)$ is

$$\begin{aligned} &\frac{\partial \hat{g}^*_{F,n}(\lambda)}{\partial\lambda} \\ = \; & -\hat{u}_n' W_n' \hat{P}_n^{*s} S_n(\lambda)\hat{u}_n = -u_n'(I_n - M_n)W_n'\hat{P}_n^{*s} S_n(\lambda)(I_n - M_n)u_n = \frac{\partial \tilde{g}^*_n(\lambda)}{\partial\lambda} + R_n(\lambda), \end{aligned}$$

where $R_n(\lambda) = u_n' M_n W_n' \hat{P}_n^{*s} S_n(\lambda) u_n + u_n' W_n' \hat{P}_n^{*s} S_n(\lambda) M_n u_n - u_n' M_n W_n' \hat{P}_n^{*s} S_n(\lambda) M_n u_n$. By a similar argument, $R_n(\lambda) = O_p(1)$ uniformly in $\lambda \in \Lambda$ by Lemma 3.14. This implies, in turn, that $\frac{1}{n}\frac{\partial \hat{g}^*_{F,n}(\lambda)}{\partial\lambda} = \frac{1}{n}\frac{\partial \tilde{g}^*_n(\lambda)}{\partial\lambda} + o_p(1)$ uniformly in $\lambda \in \Lambda$. Hence, by Lemma 3.6, the feasible BGMME derived from $\min_{\lambda\in\Lambda}[\hat{u}_n' S_n'(\lambda)\hat{P}_n^* S_n(\lambda)\hat{u}_n]^2$ has the same limiting distribution as that derived from $\min_{\lambda\in\Lambda}[u_n' S_n'(\lambda)\hat{P}_n^* S_n(\lambda)u_n]^2$.

In summary, the above two steps show that the feasible BGMME derived from

$$\min_{\lambda\in\Lambda}[\hat{u}_n' S_n'(\lambda)\hat{P}_n^* S_n(\lambda)\hat{u}_n]^2$$

has the same limiting distribution as the BGMME derived from $\min_{\lambda\in\Lambda}[u_n' S_n'(\lambda) P_n^* S_n(\lambda)u_n]^2$.

■

Proof of Proposition 3.3. Consider the moment conditions

$$E\begin{pmatrix} g_n^*(\theta_0) \\ g_n(\theta_0) \end{pmatrix} = 0,$$

where $g_n(\theta)$ is a vector of arbitrary moment functions taken the form of (3.13). To show the desired results, it is sufficient to show that g_n is redundant given g_n^*, or equivalently that there exists an A_n invariant with P_{jn} $(j = 1, \cdots, m)$ and Q_n st. $D_2 = \Omega_{21} A_n$ according to Lemma 3.15 (c), where

$$D_2 = E\left(\frac{\partial g_n(\theta_0)}{\partial \theta}\right) = -\begin{pmatrix} Q_n' X_n & Q_n' G_n X_n \beta_0 \\ 0 & \sigma_0^2 tr(P_{1n}^s G_n) \\ \vdots & \vdots \\ 0 & \sigma_0^2 tr(P_{mn}^s G_n) \end{pmatrix},$$

and

$$\begin{aligned} & \Omega_{21} \\ = \; & E\left[g_n(\theta_0) g_n^{*\prime}(\theta_0)\right] \\ = \; & \begin{pmatrix} \sigma_0^2 Q_n' Q_{1n}^* & \sigma_0^2 Q_n' Q_{2n}^* & \mu_3 Q_n' vec_D(P_{1n}^*) & \cdots & \mu_3 Q_n' vec_D(P_{k^*+1,n}^*) \\ \mu_3 vec_D'(P_{1n}) Q_{1n}^* & \mu_3 vec_D'(P_{1n}) Q_{2n}^* & \sigma_0^4 tr(P_{1n}^s P_{1n}^*) & \cdots & \sigma_0^4 tr(P_{1n}^s P_{k^*+1,n}^*) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_3 vec_D'(P_{mn}) Q_{1n}^* & \mu_3 vec_D'(P_{mn}) Q_{2n}^* & \sigma_0^4 tr(P_{mn}^s P_{1n}^*) & \cdots & \sigma_0^4 tr(P_{mn}^s P_{k^*+1,n}^*) \end{pmatrix} \\ & + \left(\mu_4 - 3\sigma_0^4\right) \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & vec_D'(P_{1n}) vec_D(P_{1n}^*) & \cdots & vec_D'(P_{1n}) vec_D(P_{k^*+1,n}^*) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & vec_D'(P_{mn}) vec_D(P_{1n}^*) & \cdots & vec_D'(P_{mn}) vec_D(P_{k^*+1,n}^*) \end{pmatrix}. \end{aligned}$$

To simplify notations, denote $\kappa = \sigma_0^6\left[(\eta_4 - 1) - \eta_3^2\right] = \sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2$. Let

$$A_n = -\begin{pmatrix} \sigma_0^{-2} I_k & 0_{k\times 1} \\ 0_{1\times k} & \sigma_0^{-2} \\ 0_{1\times k} & \sigma_0^{-2} \\ b_1 & 0 \\ \vdots & \vdots \\ b_{k^*} & 0 \end{pmatrix},$$

where $b_j = -\frac{\mu_3}{\kappa} e_{kj}'$ for $j = 1, \cdots, k^*$. To check $D_2 = \Omega_{21} A_n$, the following identities are helpful:

(1) $vec_D(P_{j+1,n}^*) = X_{nj}^* - \frac{1}{n} l_n l_n' X_{nj}^*$, for $j = 1, \cdots, k^*$,

(2) $\sum_{j=1}^{k^*} vec_D(P_{j+1,n}^*) e_{kj}' = X_n - \frac{1}{n} l_n l_n' X_n$,

(3) $vec_D(P^*_{1n}) = \frac{2\sigma_0^6}{\kappa} vec_D(G_n - \frac{tr(G_n)}{n} I_n) - \frac{\sigma_0^2 \mu_3}{\kappa}(G_n X_n \beta_0 - \frac{1}{n} l_n l_n' G_n X_n \beta_0)$.

It follows from identity (2) that

(4) $Q^*_{1n} - \frac{\mu_3^2}{\kappa} \sum_{j=1}^{k^*} vec_D(P^*_{j+1,n}) e'_{kj} = X_n$,

and it follows from identity (3) that

(5) $\sigma_0^2 Q^*_{2n} + \mu_3 vec_D(P^*_{1n}) = \sigma_0^2 G_n X_n \beta_0$.

For an arbitrary $n \times n$ matrix P_n with $tr\,(P_n) = 0$, we have:

(6) $vec'_D(P_n) Q^*_{1n} = (\sigma_0^2(\mu_4 - \sigma_0^4)/\kappa) vec'_D(P_n) X_n$,

(7) $\sigma_0^4 tr(P_n^s P^*_{j+1,n}) + (\mu_4 - 3\sigma_0^4) vec'_D(P_n) vec_D(P^*_{j+1,n}) = (\mu_4 - \sigma_0^4) vec'_D(P_n) vec_D(P^*_{j+1,n})$,

for $j = 1, \cdots, k^*$, and

(8) $\mu_3 vec'_D(P_n) Q^*_{2n} + \sigma_0^4 tr(P_n^s P^*_{1n}) + (\mu_4 - 3\sigma_0^4) vec'_D(P_n) vec_D(P^*_{1n}) = \sigma_0^4 tr(P_n^s G_n)$.

It follows from identity (4) that the $(1,1)$ block of $\Omega_{21} A_n$ is $-Q'_n X_n$, and it follows from identity (5) that the $(1,2)$ block of $\Omega_{21} A_n$ is $-Q'_n G_n X_n \beta_0$. Identities (2), (6) and (7) imply that the $(j+1, 1)$ blocks of $\Omega_{21} A_n$ are zeros for $j = 1, \cdots, m$, and (8) implies that the remaining $(j+1, 2)$ blocks of $\Omega_{21} A_n$ are $-\sigma_0^2 tr(P^s_{jn} G_n)$ for $j = 1, \cdots, m$. Therefore, $\Omega_{21} A_n = D_2$.

Furthermore, as $g^*_n(\theta)$ is a special case of $g_n(\theta)$, and A_n is invariant with P_n's and Q_n, it follows that $D_1 = \Omega_{11} A_n$, and hence $\Omega_{11}^{-1} D_1 = A_n$, where $\Omega_{11} = \Omega^*_n = var\left(g^*_n(\theta_0)\right)$ and

$$D_1 = E\left(\frac{\partial g^*_n(\theta_0)}{\partial \theta}\right) = -\begin{pmatrix} Q^{*\prime}_{1n} X_n & Q^{*\prime}_{1n} G_n X_n \beta_0 \\ Q^{*\prime}_{2n} X_n & Q^{*\prime}_{2n} G_n X_n \beta_0 \\ 0 & \sigma_0^2 tr(P^{*s}_{1n} G_n) \\ \vdots & \vdots \\ 0 & \sigma_0^2 tr(P^{*s}_{k^*+1,n} G_n) \end{pmatrix}.$$

Following Lee (2006), $\Sigma_B = \lim_{n\to\infty} \frac{1}{n} D'_1 \Omega_{11}^{-1} D_1 = \lim_{n\to\infty} \frac{1}{n} D'_1 A_n$, where

$$D'_1 A_n = \begin{pmatrix} \sigma_0^{-2} X'_n Q^*_{1n} & \sigma_0^{-2} X'_n Q^*_{2n} \\ \sigma_0^{-2} Q^{*\prime}_{2n} X_n & \sigma_0^{-2} (G_n X_n \beta_0)' Q^*_{2n} + tr(P^{*s}_{1n} G_n) \end{pmatrix}.$$

The desired result follows as $\kappa = \sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2$, $\mu_3 = \eta_3 \sigma_0^3$, and $\mu_4 = \eta_4 \sigma_0^4$. ■

Proof of Proposition 3.4. We shall show that the objective functions $F_n^*(\theta) = \hat{g}_n^{*\prime}(\theta)\hat{\Omega}_n^{*-1}\hat{g}_n^*(\theta)$ and $F_n(\theta) = g_n^{*\prime}(\theta)\Omega_n^{*-1}g_n^*(\theta)$ will satisfy the conditions in Lemma 3.7. If so, the GMME from the minimization of $F_n^*(\theta)$ will have the same limiting distribution as that of the minimization of $F_n(\theta)$. The difference of $F_n^*(\theta)$ and $F_n(\theta)$ and its derivatives involve the difference of $\hat{g}_n^*(\theta)$ and $g_n^*(\theta)$ and their derivatives. Furthermore, one has to consider the difference of $\hat{\Omega}_n^*$ and Ω_n^*.

First, consider $\frac{1}{n}(\hat{g}_n^*(\theta) - g_n^*(\theta))$. Explicitly,

$$\begin{aligned}
&\frac{1}{n}(\hat{g}_n^*(\theta) - g_n^*(\theta))' \\
= \ &[\frac{1}{n}(\hat{Q}_{1n}^* - Q_{1n}^*)'\epsilon_n(\theta), \frac{1}{n}(\hat{Q}_{2n}^* - Q_{2n}^*)'\epsilon_n(\theta), \frac{1}{n}\epsilon_n'(\theta)(\hat{G}_n^* - G_n^*)^d\epsilon_n(\theta), 0_{k^*\times 1}].
\end{aligned}$$

The $\epsilon_n(\theta)$ is related to ϵ_n as $\epsilon_n(\theta) = \epsilon_n + (\lambda_0 - \lambda)G_n\epsilon_n + d_n(\theta)$ where $d_n(\theta) = (\lambda_0 - \lambda)G_nX_n\beta_0 + X_n(\beta_0 - \beta)$. It follows that $\frac{1}{n}(\hat{Q}_{1n}^* - Q_{1n}^*)'\epsilon_n(\theta) = \frac{1}{n}(\hat{Q}_{1n}^* - Q_{1n}^*)'\epsilon_n + (\lambda_0 - \lambda)\frac{1}{n}(\hat{Q}_{1n}^* - Q_{1n}^*)'G_n\epsilon_n + \frac{1}{n}(\hat{Q}_{1n}^* - Q_{1n}^*)'d_n(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$ by Lemma 3.12. The uniformity follows because $d_n(\theta)$ is linear in λ and β. Similarly, it follows that $\frac{1}{n}(\hat{Q}_{2n}^* - Q_{2n}^*)'\epsilon_n(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$ by Lemma 3.12, and $\frac{1}{n}\epsilon_n'(\theta)(\hat{G}_n^* - G_n^*)^d\epsilon_n(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$ by Lemma 3.11. Hence, we conclude that $\frac{1}{n}(\hat{g}_n^*(\theta) - g_n^*(\theta)) = o_p(1)$ uniformly in $\theta \in \Theta$.

Consider the derivatives of $\hat{g}_n^*(\theta)$ and $g_n^*(\theta)$. As the second derivatives of $\epsilon_n(\theta)$ with respect to θ are zero because $\epsilon_n(\theta)$ is linear in θ, it follows that

$$\frac{\partial g_n^*(\theta)}{\partial \theta'} = \begin{pmatrix} Q_n^{*\prime}\frac{\partial \epsilon_n(\theta)}{\partial \theta'} \\ \epsilon_n'(\theta)P_{1n}^{*s}\frac{\partial \epsilon_n(\theta)}{\partial \theta'} \\ \vdots \\ \epsilon_n'(\theta)P_{k^*+1,n}^{*s}\frac{\partial \epsilon_n(\theta)}{\partial \theta'} \end{pmatrix}, \text{ and } \frac{\partial^2 g_n^*(\theta)}{\partial \theta \partial \theta'} = \begin{pmatrix} 0 \\ \frac{\partial \epsilon_n'(\theta)}{\partial \theta}P_{1n}^{*s}\frac{\partial \epsilon_n(\theta)}{\partial \theta'} \\ \vdots \\ \frac{\partial \epsilon_n'(\theta)}{\partial \theta}P_{k^*+1,n}^{*s}\frac{\partial \epsilon_n(\theta)}{\partial \theta'} \end{pmatrix}.$$

The first order derivatives of $\epsilon_n(\theta)$ is $\frac{\partial \epsilon_n(\theta)}{\partial \theta'} = -(X_n, W_n Y_n)$. Because $W_n Y_n = G_n X_n \beta_0 + G_n \epsilon_n$,

$$\begin{aligned}
& \frac{1}{n}(W_n Y_n)'(\hat{P}_{1n}^{*s} - P_{1n}^{*s})\epsilon_n(\theta) \\
= \; & \frac{1}{n}(G_n X_n \beta_0)'(\hat{G}_n^* - G_n^*)^{ds} d_n(\theta) + \frac{1}{n}(G_n X_n \beta_0)'(\hat{G}_n^* - G_n^*)^{ds}(\epsilon_n + (\lambda_0 - \lambda) G_n \epsilon_n) \\
& + \frac{1}{n}\epsilon_n' G_n'(\hat{G}_n^* - G_n^*)^{ds} d_n(\theta) + \frac{1}{n}\epsilon_n' G_n'(\hat{G}_n^* - G_n^*)^{ds}(\epsilon_n + (\lambda_0 - \lambda) G_n \epsilon_n) \\
= \; & o_p(1),
\end{aligned}$$

uniformly in $\theta \in \Theta$, and

$$\begin{aligned}
& \frac{1}{n}(W_n Y_n)'(\hat{P}_{1n}^{*s} - P_{1n}^{*s}) W_n Y_n \\
= \; & \frac{1}{n}(X_n \beta_0)' G_n' (\hat{G}_n^* - G_n^*)^{ds} G_n X_n \beta_0 + \frac{2}{n}(X_n \beta_0)' G_n' (\hat{G}_n^* - G_n^*)^{ds} G_n \epsilon_n \\
& + \frac{1}{n}\epsilon_n' G_n' (\hat{G}_n^* - G_n^*)^{ds} G_n \epsilon_n \\
= \; & o_p(1),
\end{aligned}$$

by Lemma 3.11. Similarly, Lemma 3.11 implies that $\frac{1}{n} X_n' (\hat{G}_n^* - G_n^*)^{ds} \epsilon_n(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$, and $\frac{1}{n} X_n' (\hat{G}_n^* - G_n^*)^{ds} W_n Y_n = o_p(1)$, $\frac{1}{n} X_n' (\hat{G}_n^* - G_n^*)^{ds} X_n = o_p(1)$. Therefore, $\frac{1}{n}\epsilon_n'(\theta)(\hat{P}_{1n}^{*s} - P_{1n}^{*s})\frac{\partial \epsilon_n(\theta)}{\partial \theta'} = o_p(1)$ and $\frac{1}{n}\frac{\partial \epsilon_n'(\theta)}{\partial \theta}(\hat{P}_{1n}^{*s} - P_{1n}^{*s})\frac{\partial \epsilon_n(\theta)}{\partial \theta'} = o_p(1)$ uniformly in $\theta \in \Theta$. Similarly, it follows from Lemma 3.12 that $\frac{1}{n}(\hat{Q}_{1n}^{*\prime} - Q_{1n}^{*\prime})\frac{\partial \epsilon_n(\theta)}{\partial \theta'} = o_p(1)$ and $\frac{1}{n}(\hat{Q}_{2n}^{*\prime} - Q_{2n}^{*\prime})\frac{\partial \epsilon_n(\theta)}{\partial \theta'} = o_p(1)$ uniformly in $\theta \in \Theta$. Hence, we conclude that $\frac{1}{n}(\frac{\partial \hat{g}_n^*(\theta)}{\partial \theta} - \frac{\partial g_n^*(\theta)}{\partial \theta}) = o_p(1)$ and $\frac{1}{n}(\frac{\partial^2 \hat{g}_n^*(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 g_n^*(\theta)}{\partial \theta \partial \theta'}) = o_p(1)$ uniformly in $\theta \in \Theta$.

Consider $\frac{1}{n}(\hat{\Omega}_n^* - \Omega_n^*)$, where

$$\begin{aligned}
\Omega_n^* &= E\left[g_n^*(\theta_0)\, g_n^{*\prime}(\theta_0)\right] \\
&= \begin{pmatrix} \sigma_0^2 Q_n^{*\prime} Q_n^* & \mu_3 Q_n^{*\prime}\omega_{k^*+1}^* \\ \mu_3 \omega_{k^*+1}^{*\prime} Q_n^* & \sigma_0^4 \Delta_{k^*+1}^* + (\mu_4 - 3\sigma_0^4)\omega_{k^*+1}^{*\prime}\omega_{k^*+1}^* \end{pmatrix},
\end{aligned}$$

with $\omega^*_{k^*+1} = [vec_D(P^*_{1n}), \cdots, vec_D(P^*_{k^*+1,n})]$ and

$$\Delta^*_{k^*+1} = \begin{pmatrix} tr\left(P^{*s}_{1n}P^*_{1n}\right) & \cdots & tr\left(P^{*s}_{1n}P^*_{k^*+1,n}\right) \\ \vdots & \ddots & \vdots \\ tr\left(P^{*s}_{k^*+1,n}P^*_{1n}\right) & \cdots & tr\left(P^{*s}_{k^*+1,n}P^*_{k^*+1,n}\right) \end{pmatrix}.$$

First, consider the block matrix $\sigma_0^4\Delta^*_{k^*+1} + (\mu_4 - 3\sigma_0^4)\omega^{*\prime}_{k^*+1}\omega^*_{k^*+1}$. As $\hat{G}^*_n$ is uniformly bounded in column sums in absolute value with probability one, it follows from Lemma 3.11 that $\frac{1}{n}tr(\hat{P}^{*s}_{1n}\hat{P}^*_{1n}) - \frac{1}{n}tr(P^{*s}_{1n}P^*_{1n}) = \frac{1}{n}tr(\hat{G}^{*ds}_n\hat{G}^*_n) - \frac{1}{n}tr(G^{*ds}_nG^*_n) = \frac{1}{n}tr[(\hat{G}^*_n - G^*_n)^{ds}\hat{G}^*_n + G^{*ds}_n(\hat{G}^*_n - G^*_n)] = o_p(1)$, and

$$\begin{aligned} & \frac{1}{n}vec'_D(\hat{P}^*_{1n})vec_D(\hat{P}^*_{1n}) - \frac{1}{n}vec'_D\left(P^*_{1n}\right)vec_D\left(P^*_{1n}\right) \\ = \; & \frac{1}{n}vec'_D(\hat{G}^{*d}_n)vec_D(\hat{G}^{*d}_n - G^{*d}_n) + \frac{1}{n}vec'_D(\hat{G}^{*d}_n - G^{*d}_n)vec_D\left(G^{*d}_n\right) = o_p(1). \end{aligned}$$

Similarly, as $P^*_{j+1,n} = D(X^{*d}_{nj})$ for $j = 1, \cdots, k^*$, Lemma 3.11 implies that $\frac{1}{n}tr[(\hat{P}^{*s}_{1n} - P^{*s}_{1n})D(X^{*d}_{nj})] = \frac{1}{n}vec_D(\hat{G}^*_n - G^*_n)^{ds}X^{*d}_{nj} = o_p(1)$, and $\frac{1}{n}vec'_D(\hat{P}^*_{1n} - P^*_{1n})X^{*d}_{nj} = \frac{1}{n}vec'_D(\hat{G}^*_n - G^*_n)^dX^{*d}_{nj} = o_p(1)$ for $j = 1, \cdots, k^*$. Hence, we conclude that

$$\begin{aligned} & \frac{1}{n}(\hat{\sigma}^2_n)^2tr(\hat{P}^{*s}_{in}\hat{P}^*_{jn}) - \frac{1}{n}\sigma_0^4tr(P^{*s}_{in}P^*_{jn}) \\ = \; & (\hat{\sigma}^2_n)^2\frac{1}{n}(tr(\hat{P}^{*s}_{in}\hat{P}^*_{jn}) - tr(P^{*s}_{in}P^*_{jn})) + ((\hat{\sigma}^2_n)^2 - \sigma_0^4)\frac{1}{n}tr(P^{*s}_{in}P^*_{jn}) = o_p(1), \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{n}(\hat{\mu}_4 - 3(\hat{\sigma}^2_n)^2)vec'_D(\hat{P}^*_{in})vec_D(\hat{P}^*_{jn}) - \frac{1}{n}(\mu_4 - 3\sigma_0^4)vec'_D(P^*_{in})vec_D(P^*_{jn}) \\ = \; & (\hat{\mu}_4 - 3(\hat{\sigma}^2_n)^2)\frac{1}{n}[vec'_D(\hat{P}^*_{in})vec_D(\hat{P}^*_{jn}) - vec'_D(P^*_{in})vec_D(P^*_{jn})] \\ & +[(\hat{\mu}_4 - 3(\hat{\sigma}^2_n)^2) - (\mu_4 - 3\sigma_0^4)]\frac{1}{n}vec'_D(P^*_{in})vec_D(P^*_{jn}) \\ = \; & o_p(1) \end{aligned}$$

for $i, j = 1, \cdots, k^* + 1$.

Next consider the block matrix $\mu_3 Q_n^{*\prime}\omega_{k^*+1}^*$. As elements of $\hat{Q}_{1n}^*$ and $\hat{Q}_{2n}^*$ are uniformly bounded with probability one, it follows from Lemmas 3.11 and 3.12 that $\frac{1}{n}\hat{Q}_{1n}^{*\prime}vec_D(\hat{P}_{1n}^*) - \frac{1}{n}Q_{1n}^{*\prime}vec_D\left(P_{1n}^*\right) = \frac{1}{n}\hat{Q}_{1n}^{*\prime}vec_D(\hat{G}_n^{*d} - G_n^{*d}) + \frac{1}{n}(\hat{Q}_{1n}^* - Q_{1n}^*)'vec_D\left(G_n^{*d}\right) = o_p(1)$ and $\frac{1}{n}\hat{Q}_{2n}^{*\prime}vec_D(\hat{P}_{1n}^*) - \frac{1}{n}Q_{2n}^{*\prime}vec_D\left(P_{1n}^*\right) = \frac{1}{n}\hat{Q}_{2n}^{*\prime}vec_D(\hat{G}_n^{*d} - G_n^{*d}) + \frac{1}{n}(\hat{Q}_{2n}^{*\prime} - Q_{2n}^{*\prime})vec_D\left(G_n^{*d}\right) = o_p(1)$. Similarly, $\frac{1}{n}(\hat{Q}_n^* - Q_n^*)'X_{nj}^{*d} = o_p(1)$ for $j = 1, \cdots, k^*$ by Lemma 3.12. Hence, we conclude that

$$\begin{aligned} & \frac{1}{n}(\hat{\mu}_3\hat{Q}_n^{*\prime}vec_D(\hat{P}_{jn}^*) - \mu_3 Q_n^{*\prime}vec_D(P_{jn}^*)) \\ = \; & \hat{\mu}_3\frac{1}{n}(\hat{Q}_n^{*\prime}vec_D(\hat{P}_{jn}^*) - Q_n^{*\prime}vec_D(P_{jn}^*)) + (\hat{\mu}_3 - \mu_3)\frac{1}{n}Q_n^{*\prime}vec_D(P_{jn}^*) = o_p(1) \end{aligned}$$

for $j = 1, \cdots, k^* + 1$.

Lastly, consider the remaining block matrix $\sigma_0^2 Q_n^{*\prime}Q_n^*$. As elements of $\hat{Q}_{1n}^*$ and $\hat{Q}_{2n}^*$ are uniformly bounded with probability one, Lemma 3.12 implies that $\frac{1}{n}(\hat{Q}_{1n}^{*\prime}\hat{Q}_{1n}^* - Q_{1n}^{*\prime}Q_{1n}^*) = \frac{1}{n}[\hat{Q}_{1n}^{*\prime}(\hat{Q}_{1n}^* - Q_{1n}^*) + (\hat{Q}_{1n}^* - Q_{1n}^*)'Q_{1n}^*] = o_p(1)$. Similarly, by Lemma 3.12, $\frac{1}{n}(\hat{Q}_{2n}^{*\prime}\hat{Q}_{2n}^* - Q_{2n}^{*\prime}Q_{2n}^*) = o_p(1)$ and $\frac{1}{n}(\hat{Q}_{2n}^{*\prime}\hat{Q}_{1n}^* - Q_{2n}^{*\prime}Q_{1n}^*) = o_p(1)$. Therefore, it follows that $\frac{1}{n}(\hat{\sigma}_n^2\hat{Q}_n^{*\prime}\hat{Q}_n^* - \sigma_0^2 Q_n^{*\prime}Q_n^*) = \hat{\sigma}_n^2\frac{1}{n}(\hat{Q}_n^{*\prime}\hat{Q}_n^* - Q_n^{*\prime}Q_n^*) + (\hat{\sigma}_n^2 - \sigma_0^2)\frac{1}{n}Q_n^{*\prime}Q_n^* = o_p(1)$. In conclusion, $\frac{1}{n}\hat{\Omega}_n^* - \frac{1}{n}\Omega_n^* = o_p(1)$. As the limit of $\frac{1}{n}\Omega_n^*$ exists and is a nonsingular matrix, it follows that $(\frac{1}{n}\hat{\Omega}_n^*)^{-1} - (\frac{1}{n}\Omega_n^*)^{-1} = o_p(1)$ by the continuous mapping theorem.

Furthermore, because $\frac{1}{n}(\hat{g}_n^*(\theta) - g_n^*(\theta)) = o_p(1)$, and $\frac{1}{n}[g_n^*(\theta) - E(g_n^*(\theta))] = o_p(1)$ uniformly in $\theta \in \Theta$, and $\sup_{\theta\in\Theta}\frac{1}{n}|E(g_n^*(\theta))| = O(1)$ (Lee, 2006, p. 21), $\frac{1}{n}g_n^*(\theta)$ and $\frac{1}{n}\hat{g}_n^*(\theta)$ are stochastically bounded, uniformly in $\theta \in \Theta$. Similarly, $\frac{1}{n}\frac{\partial g_n^*(\theta)}{\partial\theta}$, $\frac{1}{n}\frac{\partial \hat{g}_n^*(\theta)}{\partial\theta}$, $\frac{1}{n}\frac{\partial^2 g_n^*(\theta)}{\partial\theta\partial\theta}$ and $\frac{1}{n}\frac{\partial^2 \hat{g}_n^*(\theta)}{\partial\theta\partial\theta}$ are stochastically bounded, uniformly in $\theta \in \Theta$.

With the uniform convergence in probability and uniformly stochastic boundedness properties, the difference of $F_n^*(\theta)$ and $F_n(\theta)$ can be investigated. By expansion,

$$\begin{aligned}
&\frac{1}{n}(F_n^*(\theta) - F_n(\theta)) \\
= \;& \frac{1}{n}\hat{g}_n^{*\prime}(\theta)\hat{\Omega}_n^{*-1}(\hat{g}_n^*(\theta) - g_n^*(\theta)) + \frac{1}{n}g_n^{*\prime}(\theta)(\hat{\Omega}_n^{*-1} - \Omega_n^{*-1})\hat{g}_n^*(\theta) + \frac{1}{n}g_n^{*\prime}(\theta)\Omega_n^{*-1}(\hat{g}_n^*(\theta) - g_n^*(\theta)) \\
= \;& o_p(1),
\end{aligned}$$

uniformly in $\theta \in \Theta$. Similarly, for each component θ_l of θ,

$$\begin{aligned}
&\frac{1}{n}\frac{\partial^2 F_n^*(\theta)}{\partial\theta_l\partial\theta'} - \frac{1}{n}\frac{\partial^2 F_n(\theta)}{\partial\theta_l\partial\theta'} \\
= \;& \frac{2}{n}[\frac{\partial\hat{g}_n^{*\prime}(\theta)}{\partial\theta_l}\hat{\Omega}_n^{*-1}\frac{\partial\hat{g}_n^*(\theta)}{\partial\theta'} + \hat{g}_n^{*\prime}(\theta)\hat{\Omega}_n^{*-1}\frac{\partial^2\hat{g}_n^*(\theta)}{\partial\theta_l\partial\theta'} - (\frac{\partial g_n^{*\prime}(\theta)}{\partial\theta_l}\Omega_n^{*-1}\frac{\partial g_n^*(\theta)}{\partial\theta'} + g_n^{*\prime}(\theta)\Omega_n^{*-1}\frac{\partial^2 g_n^*(\theta)}{\partial\theta_l\partial\theta'})] \\
= \;& o_p(1).
\end{aligned}$$

Finally, because $(\frac{\partial\hat{g}_n^{*\prime}(\theta_0)}{\partial\theta}\hat{\Omega}_n^{*-1} - \frac{\partial g_n^{*\prime}(\theta_0)}{\partial\theta}\Omega_n^{*-1}) = o_p(1)$ as above, and $\frac{1}{\sqrt{n}}g_n^*(\theta_0) = O_p(1)$ by the central limit theorems in Lemmas 3.4 and 3.5,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}(\frac{\partial F_n^*(\theta_0)}{\partial\theta} - \frac{\partial F_n(\theta_0)}{\partial\theta}) \\
= \;& 2\{\frac{\partial\hat{g}_n^{*\prime}(\theta_0)}{\partial\theta}\hat{\Omega}_n^{*-1}\frac{1}{\sqrt{n}}(\hat{g}_n^*(\theta_0) - g_n^*(\theta_0)) + (\frac{\partial\hat{g}_n^{*\prime}(\theta_0)}{\partial\theta}\hat{\Omega}_n^{*-1} - \frac{\partial g_n^{*\prime}(\theta_0)}{\partial\theta}\Omega_n^{*-1})\frac{1}{\sqrt{n}}g_n^*(\theta_0)\} \\
= \;& 2\frac{\partial\hat{g}_n^{*\prime}(\theta_0)}{\partial\theta}\hat{\Omega}_n^{*-1}\frac{1}{\sqrt{n}}(\hat{g}_n^*(\theta_0) - g_n^*(\theta_0)) + o_p(1).
\end{aligned}$$

As $\frac{1}{\sqrt{n}}(\hat{g}_n^*(\theta_0) - g_n^*(\theta_0)) = o_p(1)$ by Lemmas 3.11 and 3.12, $\frac{1}{\sqrt{n}}(\frac{\partial F_n^*(\theta_0)}{\partial\theta} - \frac{\partial F_n(\theta_0)}{\partial\theta}) = o_p(1)$. The desired result follows from Lemma 3.7. ■

3.5.4 Joint GMM Estimation of the MRSAR Model

When the disturbances in the MRSAR model are normally distributed, the asymptotic variance matrix of the MLE $\hat{\delta}_{ML} = (\hat{\beta}'_{ML}, \hat{\lambda}_{ML}, \hat{\sigma}^2_{ML})'$ is

$$Avar\left(\hat{\delta}_{ML}\right) = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' X_n & \frac{1}{\sigma_0^2} X_n' \left(G_n X_n \beta_0\right) & 0 \\ \frac{1}{\sigma_0^2}\left(G_n X_n \beta_0\right)' X_n & \frac{1}{\sigma_0^2}\left(G_n X_n \beta_0\right)' \left(G_n X_n \beta_0\right) + tr\left(G_n^s G_n\right) & \frac{1}{\sigma_0^2} tr\left(G_n\right) \\ 0 & \frac{1}{\sigma_0^2} tr\left(G_n\right) & \frac{n}{2\sigma_0^4} \end{pmatrix}^{-1}.$$

As the asymptotic covariance between $\hat{\lambda}_{ML}$ and $\hat{\sigma}^2_{ML}$ are not zero in general, it is not trivial to determine whether the efficiency property of $\hat{\theta}_P = (\hat{\beta}'_P, \hat{\lambda}_P)$ will be affected by concentrating σ^2 out in the GMM estimation. And when the disturbances are not normally distributed, the problem may be more complicated. Here we consider the joint estimation of $\delta_0 = (\beta_0', \lambda_0, \sigma_0^2)'$ in the GMM framework. By comparing the asymptotic variance matrix of the BGMME derived from the joint GMM estimation approach with that of the BGMME described in Proposition 3.3, we conclude that there is no efficiency loss in the estimation of $\theta_0 = (\beta_0', \lambda_0)'$ by concentrating σ^2 out.

For simplicity, we assume an intercept appears in X_n so that the last column of X_n is l_n. Define $\bar{P}^*_{1n} = G^*_n$, $\bar{P}^*_{j+1,n} = D\left(X_{nj}\right)$ for $j = 1, \cdots, k$,[30] and $\bar{Q}^*_n = \left[\bar{Q}^*_{1n}, \bar{Q}^*_{2n}\right]$ with $\bar{Q}^*_{1n} = X_n$ and

$$\bar{Q}^*_{2n} = G_n X_n \beta_0 + \frac{\eta_3^2}{(\eta_4 - 1) - \eta_3^2} G_n X_n \beta_0 - \frac{2\sigma_0 \eta_3}{(\eta_4 - 1) - \eta_3^2} vec_D\left(G_n\right).$$

Let $\bar{g}^*_n(\delta) = \left[\epsilon_n'(\theta)\bar{Q}^*_n, \epsilon_n'(\theta)\bar{P}^*_{jn}\epsilon_n(\theta) - \sigma^2 tr\left(\bar{P}^*_{jn}\right)\right]'$ $(j = 1, \cdots, k+1)$. The consistent root $\hat{\delta}_{BJ}$ derived from $\min_\delta \bar{g}^{*\prime}_n(\delta)\bar{\Omega}^{*-1}_n \bar{g}^*_n(\delta)$ with $\bar{\Omega}^*_n = var\left(\bar{g}^*_n(\delta_0)\right)$ is the BGMME within the class of optimal GMMEs derived from $\min_\delta \bar{g}'_n(\delta)\bar{\Omega}^{-1}_n \bar{g}_n(\delta)$, where

[30] As we assume that $X_{nk} = l_n$, $\bar{P}^*_{k+1,n} = D\left(X_{nk}\right) = I_n$ can be of use as the simple second moment for the estimation of σ_0^2. If there is no intercept in X_n, we need to add a moment associated with I_n to estimate σ_0^2.

$\bar{\Omega}_n = var\left(\bar{g}_n\left(\delta_0\right)\right)$ and

$$\bar{g}_n(\delta) = (\epsilon_n'(\theta)\bar{Q}_n, \epsilon_n'(\theta)\bar{P}_{1n}\epsilon_n(\theta) - \sigma^2 tr\left(\bar{P}_{1n}\right), \cdots, \epsilon_n'(\theta)\bar{P}_{mn}\epsilon_n(\theta) - \sigma^2 tr\left(\bar{P}_{mn}\right))',$$

with $\bar{Q}_n$ being an arbitrary $n \times k'$ matrix of IVs, and $\bar{P}_n$'s being arbitrary $n \times n$ matrices, not necessarily with zero traces. At δ_0, $\bar{g}_n(\delta_0) = [\epsilon_n'\bar{Q}_n, \epsilon_n'\bar{P}_{jn}\epsilon_n - \sigma_0^2 tr\left(\bar{P}_{jn}\right)]'$, which has a zero mean because $E(\bar{Q}_n'\epsilon_n) = \bar{Q}_n'E(\epsilon_n) = 0$ and $E(\epsilon_n'\bar{P}_{jn}'\epsilon_n) = \sigma_0^2 tr(\bar{P}_{jn})$ for $j = 1, \cdots, m$.

Analogous to the proof of Proposition 3.3, the above statement is confirmed by showing that $\bar{g}_n$ is redundant given $\bar{g}_n^*$, or equivalently that there exists a matrix $\bar{A}_n$ invariant with $\bar{P}_{jn}$ $(j = 1, \cdots, m)$ and $\bar{Q}_n$ st. $\bar{D}_2 = \bar{\Omega}_{21}\bar{A}_n$, according to Lemma 3.15 (c), where

$$\bar{D}_2 = E\left(\frac{\partial \bar{g}_n\left(\delta_0\right)}{\partial \delta}\right) = -\begin{bmatrix} \bar{Q}_n'X_n & \bar{Q}_n'G_nX_n\beta_0 & 0 \\ 0 & \sigma_0^2 tr\left(\bar{P}_{1n}^s G_n\right) & tr\left(\bar{P}_{1n}\right) \\ \vdots & \vdots & \vdots \\ 0 & \sigma_0^2 tr\left(\bar{P}_{mn}^s G_n\right) & tr\left(\bar{P}_{mn}\right) \end{bmatrix},$$

and

$$\begin{aligned} & \bar{\Omega}_{21} \\ = \; & E\left[\bar{g}_n\left(\delta_0\right)\bar{g}_n^{*\prime}\left(\delta_0\right)\right] \\ = \; & \begin{bmatrix} \sigma_0^2\bar{Q}_n'\bar{Q}_{1n}^* & \sigma_0^2\bar{Q}_n'\bar{Q}_{2n}^* & \mu_3\bar{Q}_n' vec_D\left(\bar{P}_{1n}^*\right) & \cdots & \mu_3\bar{Q}_n' vec_D\left(\bar{P}_{k+1,n}^*\right) \\ \mu_3 vec_D'\left(\bar{P}_{1n}\right)\bar{Q}_{1n}^* & \mu_3 vec_D'\left(\bar{P}_{1n}\right)\bar{Q}_{2n}^* & \sigma_0^4 tr\left(\bar{P}_{1n}^s\bar{P}_{1n}^*\right) & \cdots & \sigma_0^4 tr\left(\bar{P}_{1n}^s\bar{P}_{k+1,n}^*\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_3 vec_D'\left(\bar{P}_{mn}\right)\bar{Q}_{1n}^* & \mu_3 vec_D'\left(\bar{P}_{mn}\right)\bar{Q}_{2n}^* & \sigma_0^4 tr\left(\bar{P}_{mn}^s\bar{P}_{1n}^*\right) & \cdots & \sigma_0^4 tr\left(\bar{P}_{mn}^s\bar{P}_{k+1,n}^*\right) \end{bmatrix} \\ & +\left(\mu_4 - 3\sigma_0^4\right)\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & vec_D'\left(\bar{P}_{1n}\right)vec_D\left(\bar{P}_{1n}^*\right) & \cdots & vec_D'\left(\bar{P}_{1n}\right)vec_D\left(\bar{P}_{k+1,n}^*\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & vec_D'\left(\bar{P}_{mn}\right)vec_D\left(\bar{P}_{1n}^*\right) & \cdots & vec_D'\left(\bar{P}_{mn}\right)vec_D\left(\bar{P}_{k+1,n}^*\right) \end{bmatrix}. \end{aligned}$$

To simplify notations, denote $\kappa = \sigma_0^6[(\eta_4 - 1) - \eta_3^2] = \sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2$. Let

$$\bar{A}_n = -\begin{bmatrix} \frac{\mu_4-\sigma_0^4}{\kappa} I_k & 0_{k\times 1} & \xi \\ 0_{1\times k} & \sigma_0^{-2} & 0 \\ 0_{1\times k} & \sigma_0^{-2} & 0 \\ b_1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ b_{k-1} & 0 & 0 \\ b_k & 0 & \sigma_0^2/\kappa \end{bmatrix},$$

where the k-dimensional vector $\xi = [0, \cdots, 0, -\mu_3/\kappa]'$, and $b_j = -\left(\mu_3/\kappa\right) e'_{kj}$. With identities analogous to those provided in the proof of Proposition 3.3, straightforward but tedious algebra leads to $\bar{D}_2 = \bar{\Omega}_{21}\bar{A}_n$.

Furthermore, as $\bar{g}_n^*(\delta)$ is a special case of $\bar{g}_n(\delta)$, and $\bar{A}_n$ is invariant with $\bar{P}_n$'s and $\bar{Q}_n$, it follows $\bar{\Omega}_{11}^{-1}\bar{D}_1 = \bar{\Omega}_{21}^{-1}\bar{D}_2 = \bar{A}_n$, where $\bar{\Omega}_{11} = var\left(\bar{g}_n^*(\delta_0)\right)$ and

$$\bar{D}_1 = E\left(\frac{\partial \bar{g}_n^*(\delta_0)}{\partial \delta}\right) = -\begin{bmatrix} \bar{Q}_{1n}^{*\prime} X_n & \bar{Q}_{1n}^{*\prime} G_n X_n \beta_0 & 0 \\ \bar{Q}_{2n}^{*\prime} X_n & \bar{Q}_{2n}^{*\prime} G_n X_n \beta_0 & 0 \\ 0 & \sigma_0^2 tr\left(\bar{P}_{1n}^{*s} G_n\right) & tr\left(\bar{P}_{1n}^*\right) \\ \vdots & \vdots & \vdots \\ 0 & \sigma_0^2 tr\left(\bar{P}_{k+1,n}^{*s} G_n\right) & tr\left(\bar{P}_{k+1,n}^*\right) \end{bmatrix}.$$

The asymptotic precision matrix of $\hat{\delta}_{BJ}$ is $\Sigma_{BJ} = \lim_{n\to\infty} \frac{1}{n}\bar{D}_1'\bar{\Omega}_{11}^{-1}\bar{D}_1 = \lim_{n\to\infty} \frac{1}{n}\bar{D}_1'\bar{A}_n$, where

$$\begin{aligned} &\bar{D}_1'\bar{A}_n \\ = &\begin{bmatrix} \left(\mu_4 - \sigma_0^4\right) X_n' \bar{Q}_{1n}^* / \kappa & \sigma_0^{-2} X_n' \bar{Q}_{2n}^* & -\left(\mu_3/\kappa\right) X_n' l_n \\ \sigma_0^{-2} \bar{Q}_{2n}^{*\prime} X_n & \sigma_0^{-2}\left(G_n X_n \beta_0\right)' \bar{Q}_{2n}^* + tr\left(\bar{P}_{1n}^{*s} G_n\right) & \sigma_0^{-2} tr\left(G_n^*\right) \\ -\left(\mu_3/\kappa\right) l_n' X_n & \sigma_0^{-2} tr\left(G_n^*\right) & n\sigma_0^2/\kappa \end{bmatrix}. \end{aligned}$$

From the inverse of a partitioned matrix, we have $Avar\left(\hat{\theta}_{BJ}\right) = \Sigma_{BJ}^{-1}/n$, where

$$\begin{aligned} n\Sigma_{BJ} &= \begin{bmatrix} \left(\mu_4 - \sigma_0^4\right) X_n' \bar{Q}_{1n}^* / \kappa & \sigma_0^{-2} X_n' \bar{Q}_{2n}^* \\ \sigma_0^{-2} \bar{Q}_{2n}^{*\prime} X_n & \sigma_0^{-2} \left(G_n X_n \beta_0\right)' \bar{Q}_{2n}^* + tr\left(\bar{P}_{1n}^{*s} G_n\right) \end{bmatrix} \\ &\quad - \left(n\sigma_0^2/\kappa\right)^{-1} \begin{bmatrix} -\left(\mu_3/\kappa\right) X_n' l_n \\ \sigma_0^{-2} tr\left(G_n^*\right) \end{bmatrix} \begin{bmatrix} -\left(\mu_3/\kappa\right) l_n' X_n & \sigma_0^{-2} tr\left(G_n^*\right) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_0^{-2} X_n' Q_{1n}^* & \sigma_0^{-2} X_n' Q_{2n}^* \\ \sigma_0^{-2} Q_{2n}^{*\prime} X_n & \sigma_0^{-2} \left(G_n X_n \beta_0\right)' Q_{2n}^* + tr\left(P_{1n}^{*s} G_n\right) \end{bmatrix} \\ &= n\Sigma_B, \end{aligned}$$

since $\kappa = \sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2$, $\mu_3 = \eta_3\sigma_0^3$, and $\mu_4 = \eta_4\sigma_0^4$. Hence the efficiency property of the BGMME of θ_0 is not affected by concentrating σ^2 out in the GMM estimation.

3.5.5 Monte Carlo Results

True parameters: $\lambda_0 = 0.6$

	$n = 49$	$n = 245$	$n = 490$
method	Mean(SD)[RMSE]	Mean(SD)[RMSE]	Mean(SD)[RMSE]
QMLE	.562 (.094) [.102]	.590 (.040) [.041]	.593 (.028) [.029]
OGMME	.568 (.098) [.103]	.593 (.041) [.041]	.596 (.028) [.029]
BGMME	.563 (.095) [.102]	.592 (.039) [.039]	.596 (.027) [.028]

Table 3.1: QMLE and GMME of the SAR disturbance process (normal)

True parameters: $\lambda_0 = 0.6$

	$n = 49$	$n = 245$	$n = 490$
method	Mean(SD)[RMSE]	Mean(SD)[RMSE]	Mean(SD)[RMSE]
QMLE	.559 (.098) [.106]	.589 (.044) [.045]	.592 (.031) [.032]
OGMME	.568 (.100) [.105]	.593 (.041) [.042]	.596 (.029) [.029]
BGMME	.563 (.097) [.103]	.592 (.040) [.041]	.595 (.028) [.028]

Table 3.2: QMLE and GMME of the SAR disturbance process (student t)

True parameters: $\lambda_0 = 0.6$

	$n = 49$	$n = 245$	$n = 490$
method	Mean(SD)[RMSE]	Mean(SD)[RMSE]	Mean(SD)[RMSE]
QMLE	.567 (.098) [.103]	.592 (.041) [.042]	.595 (.028) [.029]
OGMME	.570 (.103) [.107]	.594 (.043) [.044]	.597 (.029) [.029]
BGMME	.569 (.099) [.104]	.594 (.036) [.036]	.597 (.025) [.025]

Table 3.3: QMLE and GMME of the SAR disturbance process (symmetric mixture normal)

True parameters: $\lambda_0 = 0.6$, $\beta_{10} = 1.0$, $\beta_{20} = 0$, $\beta_{30} = -1.0$

method	λ Mean(SD)[RMSE]	β_1 Mean(SD)[RMSE]	β_2 Mean(SD)[RMSE]	β_3 Mean(SD)[RMSE]
	$n = 49$			
QMLE	.573 (.117) [.120]	1.006 (.217) [.217]	−.000 (.206) [.206]	−0.995 (.211) [.211]
OGMME	.590 (.123) [.124]	0.993 (.217) [.217]	−.001 (.206) [.206]	−0.982 (.212) [.213]
BGMME	.603 (.136) [.136]	0.992 (.232) [.233]	.001 (.217) [.217]	−0.986 (.222) [.222]
	$n = 245$			
QMLE	.596 (.046) [.046]	1.001 (.091) [.091]	−.002 (.090) [.090]	−1.003 (.093) [.093]
OGMME	.600 (.046) [.046]	0.998 (.091) [.091]	−.002 (.090) [.090]	−0.999 (.093) [.093]
BGMME	.602 (.046) [.046]	0.997 (.093) [.093]	−.002 (.091) [.091]	−0.999 (.095) [.095]
	$n = 490$			
QMLE	.597 (.034) [.034]	1.003 (.062) [.062]	−.001 (.066) [.066]	−1.001 (.066) [.066]
OGMME	.599 (.034) [.034]	1.001 (.063) [.063]	−.001 (.066) [.066]	−0.999 (.066) [.066]
BGMME	.600 (.034) [.034]	1.001 (.063) [.063]	−.001 (.066) [.066]	−0.999 (.066) [.066]

Table 3.4: QMLE and GMME of the MRSAR model (normal)

True parameters: $\lambda_0 = 0.6$, $\beta_{10} = 1.0$, $\beta_{20} = 0$, $\beta_{30} = -1.0$

method	λ Mean(SD)[RMSE]	β_1 Mean(SD)[RMSE]	β_2 Mean(SD)[RMSE]	β_3 Mean(SD)[RMSE]
	$n = 49$			
QMLE	.575 (.114) [.116]	1.005 (.219) [.219]	−.002 (.207) [.207]	−0.992 (.213) [.213]
OGMME	.590 (.120) [.121]	0.993 (.220) [.220]	−.002 (.206) [.206]	−0.980 (.213) [.214]
BGMME	.605 (.134) [.135]	0.994 (.225) [.226]	.001 (.213) [.213]	−0.985 (.220) [.220]
	$n = 245$			
QMLE	.596 (.047) [.047]	1.000 (.092) [.092]	−.001 (.089) [.089]	−1.001 (.094) [.094]
OGMME	.600 (.047) [.047]	0.996 (.093) [.093]	−.001 (.089) [.089]	−0.998 (.093) [.093]
BGMME	.602 (.047) [.047]	0.995 (.092) [.092]	−.001 (.089) [.089]	−0.999 (.094) [.094]
	$n = 490$			
QMLE	.597 (.034) [.034]	1.002 (.065) [.065]	−.001 (.066) [.066]	−0.999 (.065) [.065]
OGMME	.599 (.033) [.033]	1.000 (.065) [.065]	−.001 (.065) [.065]	−0.997 (.065) [.065]
BGMME	.600 (.033) [.033]	1.000 (.064) [.064]	−.000 (.065) [.065]	−0.997 (.065) [.065]

Table 3.5: QMLE and GMME of the MRSAR model (student t)

True parameters: $\lambda_0 = 0.6$, $\beta_{10} = 1.0$, $\beta_{20} = 0$, $\beta_{30} = -1.0$

method	λ Mean(SD)[RMSE]	β_1 Mean(SD)[RMSE]	β_2 Mean(SD)[RMSE]	β_3 Mean(SD)[RMSE]
	$n = 49$			
QMLE	.567 (.120) [.124]	1.001 (.215) [.215]	−.000 (.208) [.208]	−0.999 (.214) [.214]
OGMME	.585 (.132) [.132]	0.989 (.216) [.217]	.000 (.208) [.208]	−0.986 (.214) [.214]
BGMME	.594 (.136) [.136]	0.985 (.206) [.206]	.001 (.197) [.197]	−0.988 (.209) [.210]
	$n = 245$			
QMLE	.594 (.048) [.048]	0.998 (.091) [.091]	−.002 (.090) [.090]	−0.999 (.090) [.091]
OGMME	.597 (.047) [.048]	0.995 (.091) [.091]	−.002 (.090) [.090]	−0.996 (.090) [.091]
BGMME	.598 (.047) [.047]	0.994 (.089) [.089]	−.003 (.087) [.087]	−0.995 (.088) [.088]
	$n = 490$			
QMLE	.595 (.033) [.033]	1.003 (.065) [.066]	.001 (.061) [.061]	−1.002 (.066) [.066]
OGMME	.598 (.033) [.033]	1.002 (.065) [.065]	.000 (.061) [.061]	−1.000 (.066) [.066]
BGMME	.599 (.032) [.032]	1.001 (.064) [.064]	.000 (.060) [.060]	−1.000 (.064) [.064]

Table 3.6: QMLE and GMME of the MRSAR model (symmetric mixture normal)

True parameters: $\lambda_0 = 0.6$, $\beta_{10} = 1.0$, $\beta_{20} = 0$, $\beta_{30} = -1.0$

method	λ Mean(SD)[RMSE]	β_1 Mean(SD)[RMSE]	β_2 Mean(SD)[RMSE]	β_3 Mean(SD)[RMSE]
	$n = 49$			
QMLE	.569 (.118) [.122]	0.997 (.214) [.214]	−.001 (.207) [.207]	−0.999 (.213) [.213]
OGMME	.586 (.124) [.125]	0.984 (.215) [.215]	.000 (.206) [.206]	−0.986 (.212) [.213]
BGMME	.600 (.138) [.138]	0.986 (.191) [.191]	.002 (.193) [.193]	−0.995 (.193) [.193]
	$n = 245$			
QMLE	.594 (.048) [.048]	0.998 (.091) [.091]	−.001 (.091) [.091]	−0.998 (.091) [.091]
OGMME	.597 (.048) [.048]	0.995 (.091) [.091]	−.001 (.091) [.091]	−0.995 (.091) [.091]
BGMME	.600 (.045) [.045]	0.995 (.073) [.073]	−.001 (.072) [.072]	−0.997 (.073) [.073]
	$n = 490$			
QMLE	.595 (.033) [.033]	1.003 (.065) [.065]	.002 (.063) [.063]	−1.001 (.066) [.066]
OGMME	.597 (.032) [.032]	1.001 (.065) [.065]	.002 (.063) [.063]	−0.999 (.066) [.066]
BGMME	.598 (.029) [.029]	1.001 (.051) [.051]	−.001 (.049) [.049]	−0.999 (.051) [.051]

Table 3.7: QMLE and GMME of the MRSAR model (asymmetric mixture normal)

True parameters: $\lambda_0 = 0.6$, $\beta_{10} = 1.0$, $\beta_{20} = 0$, $\beta_{30} = -1.0$

method	λ Mean(SD)[RMSE]	β_1 Mean(SD)[RMSE]	β_2 Mean(SD)[RMSE]	β_3 Mean(SD)[RMSE]
	$n = 49$			
QMLE	.572 (.122) [.125]	1.002 (.207) [.207]	.007 (.212) [.212]	−0.994 (.211) [.211]
OGMME	.589 (.131) [.131]	0.988 (.207) [.208]	.007 (.209) [.209]	−0.981 (.211) [.211]
BGMME	.610 (.143) [.143]	0.993 (.184) [.184]	.001 (.182) [.182]	−0.984 (.189) [.190]
	$n = 245$			
QMLE	.595 (.046) [.046]	0.999 (.092) [.092]	−.003 (.092) [.092]	−1.001 (.092) [.092]
OGMME	.598 (.046) [.046]	0.996 (.093) [.093]	−.003 (.092) [.092]	−0.998 (.091) [.091]
BGMME	.601 (.043) [.043]	0.996 (.070) [.070]	−.003 (.070) [.070]	−0.999 (.070) [.070]
	$n = 490$			
QMLE	.597 (.034) [.034]	0.997 (.065) [.065]	.000 (.064) [.064]	−0.999 (.066) [.066]
OGMME	.599 (.033) [.033]	0.996 (.065) [.065]	.000 (.064) [.064]	−0.997 (.066) [.066]
BGMME	.600 (.030) [.030]	0.997 (.050) [.051]	.001 (.050) [.050]	−0.997 (.052) [.052]

Table 3.8: QMLE and GMME of the MRSAR model (gamma)

BIBLIOGRAPHY

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L. 1990. “Some robust approaches to testing and estimation in spatial econometrics.” *Regional Science and Urban Economics* 20:141–163.

Athey, S. 2001. “Single crossing properties and the existence of pure strategy equilibria in games of incomplete information.” *Econometrica* 69:861–889.

Bajari, P. and A Hortacsu. 2005. “Are structural estimates of auction models reasonable? evidence from experimental data.” *Journal of Political Economy* 113:703–741.

Bierens, H. J. 1987. Kernel estimators of regression functions. In *Advances in Economic Theory: Fifth World Congress*, ed. T. F. Bewley. Cambridge: Cambridge University Press.

Bollinger, C. R. 2001. “Efficient Two Step GMM Estimation of Spatial Autoregressive Models.” Working paper, Department of Economics, University of Kentucky.

Borsch-Supan, A. and V. Hajivassiliou. 1993. “Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models.” *Journal of Econometrics* 58:347–368.

Breusch, T., H. Qian, P. Schmidt and D. Wyhowski. 1999. “Redundancy of moment conditions.” *Journal of Econometrics* 91:89–111.

Butler, J. S. and R. Moffitt. 1982. “A computationally efficient quadrature procedure for the one-factor multinomial probit model.” *Econometrica* 50:761–764.

Campo, S. 2004. “Asymmetry and risk aversion within the independent private paradigm: the case of the construction procurements.” Working Paper, UNC at Chapel Hill.

Campo, S., E. Guerre, I. Perrigne and Q. Vuong. 2006. “Semiparametric estimation of first-price auctions with risk averse bidders.” Working Paper, Pennsylvania State University.

Cho, I. and D. Kreps. 1987. “Signaling games and stable equilibria.” *Quarterly Journal of Economics* 102:179–221.

Cooper, D. J. and J. H. Kagel. 2003*a*. “The impact of meaningful context on strategic play in signaling games.” *Journal of Economic Behavior and Organization* 50:311–337.

Cooper, D. J. and J. H. Kagel. 2003*b*. “Lessons learned: generalizing learning across games.” *American Economic Review* 93:202–207.

Cooper, D. J. and J. H. Kagel. 2004. “Learning and transfer in signaling games.” Working Paper.

Cooper, D. J., S. Garvin and J. H. Kagel. 1997*a*. “Adaptive learning vs. equilibrium refinements in an entry limit pricing game.” *The Economic Journal* 107:553–575.

Cooper, D. J., S. Garvin and J. H. Kagel. 1997*b*. “Signaling and adaptive learning in an entry limit pricing game.” *RAND Journal of Economics* 28:662–683.

Copson, E. T. 1965. *Asymptotic Expansions*. Cambridge, UK: Cambridge University Press.

Devroye, L. 1987. *A Course in Density Estimation*. Boston, Basel, Stuttgart: Birkhauser.

Fan, J. and J. S. Marron. 1992. “Comparison of Data-Driven Bandwidth Selectors.” *Journal of American Statistical Association* 85:66–72.

Fibich, G. and A. Gavious. 2003. “Asymmetric first-price auction - a perturbation approach.” *Mathematics of Operations Research* 28:836–852.

Fibich, G., A. Gavious and A. Sela. 2004. “Asymptotic analysis of large auctions.” Working Paper.

Fudenberg, D. and D. K. Levine. 1998. *The Theory of Learning in Games*. Cambridge: MIT Press.

Geweke, J. 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In *Computer science and statistics: Proceedings of the Twenty-Third Symposium on the Inference*. American Statistical Association Alexandria, VA: pp. 571–578.

Geweke, J. and M. Keane. 2001. Computationally intensive methods for integration in econometrics. In *Handbook of Econometrics*, ed. J. J. Heckman and E. Leamer. Vol. 5 Amsterdam: North-Holland.

Gourieroux, C. and A. Monfort. 1993. “Simulation-based inference: a survey with special reference to panel data models.” *Journal of Econometrics* 59:5–33.

Groeneboom, P. 1996. Lectures on inverse problems. In *Lectures on Probability Theory and Statistics*, ed. P. Bernard. Number 1648 New York: Springer Verlag.

Guerre, E., I. Perrigne and Q. Vuong. 2000. "Optimal Nonparametric Estimation of First-Price Auction." *Econometrica* 68:525–574.

Hajivassiliou, V. and D. McFadden. 1990. "The method of simulated scores for the estimation of LDV models: with an application to the problem of external debt crises." Cowles foundation discussion paper No. 967. Yale University, New Haven, CT.

Hansen, L. P. 1982. "Large sample properties of generalized method of moments estimators." *Econometrica* 50:1029–1054.

Hardle, W. 1991. *Applied Nonparametric Regression (Econometric Society Monographs)*. Cambridge: Cambridge University Press.

Heckman, J. J. 1981. Statistical models for discrete panel data. In *Structural Analysis of Discrete Data with Econometric Applications*, ed. C. F. Manski and D. McFadden. Cambridge: MIT Press.

Horn, R. and C. Johnson. 1985. *Matrix Analysis*. Cambridge: Cambridge Univsersity Press.

Hsiao, C. 2003. *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Ibragimov, I. A. and R. Z. Has'minskii. 1981. *Statistical Estimation: Asymptotic Theory*. New York: Springer Verlag.

Keane, M. P. 1994. "A computationally practical simulation estimator for panel data." *Econometrica* 62:95–116.

Kelejian, H. H. and D. P. Robinson. 1993. "A suggested method of estimation for spatial interdependence models with autocorrelated errors, and an application to a county expenditure model." *Papers in Regional Science* 72:297–312.

Kelejian, H. H. and I. R. Prucha. 1997. "Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem." *International Regional Science Review* 20:103–111.

Kelejian, H. H. and I. R. Prucha. 1998. "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance." *Journal of Real Estate Finance and Economics* 17:99–121.

Kelejian, H. H. and I. R. Prucha. 1999. "A generalized moments estimator for the autoregressive parameter in a spatial model." *International Economic Review* 40:509–533.

Kelejian, H. H. and I. R. Prucha. 2001. “On the asymptotic distribution of the Moran I test statistic with applications.” *Journal of Econometrics* 104:219–257.

Land, K. and G. Deane. 1992. On the large-sample estimation of regression models with spatial-or networkeffects terms: a two stage least squares approach. In *Sociological Methodology*, ed. P. Marsden. San Francisco: Jossey Bass.

Lebrun, B. 1999. “First price auctions and the asymmetric n bidder case.” *International Economic Review* 40:125–142.

Lee, L. F. 1992. “On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models.” *Econometric Theory* 8:518–552.

Lee, L. F. 1995. “Asymptotic bias in maximum simulated likelihood estimation of discrete choice models.” *Econometric Theory* 11:437–483.

Lee, L. F. 1997. “Simulated maximum likelihood estimation of dynamic discrete choice statistical models: some Monte Carlo results.” *Journal of Econometrics* 82:1–35.

Lee, L. F. 2000. “A numerically stable quadrature procedure for the one-factor random-component discrete choice model.” *Journal of Econometrics* 95:117–129.

Lee, L. F. 2001*a*. “Generalized method of moments estimation of spatial autoregressive processes.” Manuscript, Department of Economics, Ohio State University.

Lee, L. F. 2001*b*. “Interpolation, quadrature, and stochastic integration.” *Econometric Theory* 17:933–961.

Lee, L. F. 2003. “Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances.” *Econometric Reviews* 22:307–335.

Lee, L. F. 2004. “Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models.” *Econometrica* 72:1899–1926.

Lee, L. F. 2006. “GMM and 2SLS estimation of mixed regressive, spatial autoregressive models.” Working paper, Department of Economics, Ohio State University, forthcoming in *Journal of Econometrics*.

Lin, X. and L. F. Lee. 2006. “GMM estimation of spatial autoregressive models with unknown heteroskedasticity.” Working paper, Department of Economics, Ohio State University.

Manski, C. F. 1993. “Identification of endogenous social effects: the reflection problem.” *The Review of Economic Studies* 60:531–542.

Maskin, E. and J. Riley. 1984. “Optimal auctions with risk averse buyers.” *Econometrica* 52:1473–1518.

Maskin, E. and J. Riley. 2000. "Equilibrium in sealed high bid auctions." *Review of Economic Studies* 67:439–454.

Maskin, E. and J. Riley. 2003. "Uniqueness in sealed high bid auctions." *Games and Economic Bahavior* 45:395–409.

Matthews, S. 1987. "Comparing auctions for risk averse buyers: a buyer's point of view." *Econometrica* 55:633–646.

Milgrom, P. and J. Roberts. 1982. "Limit pricing and entry under incomplete information: an equilibrium analysis." *Econometrica* 50:443–459.

Murray, J. D. 1984. *Asymptotic Analysis*. New York: Springer-Verlag.

Ord, J. 1975. "Estimation methods for models of spatial interaction." *Journal of the American Statistical Association* 70:120–126.

Riley, J. G. and W. F. Samuelson. 1981. "Optimal auctions." *American Economic Review* 71:381–392.

Rustichini, A., M. A. Satterthwaite and S. Williams. 1994. "Convergence to efficiency in a simple market with incomplete information." *Econometrica* 62:1041–1063.

Satterthwaite, M. A. and S. Williams. 1989. "The rate of convergence to efficiency in the buyer's bid double auction as the market becomes large." *The Review of Economic Studies* 56:477–498.

Stone, C. J. 1980. "Optimal rates of convergence for nonparametric estimators." *The Annals of Statistics* 8:1348–1360.

Stone, C. J. 1982. "Optimal global rates of convergence for nonparametric regression." *The Annals of Statistics* 10:1040–1053.