ABSTRACT

Title of Dissertation: PANEL DATA MODELS WITH SPATIAL CORRELATION: ESTIMATION THEORY AND EMPIRICAL INVESTIGATION OF THE US WHOLESALE GASOLINE INDUSTRY

Mudit Kapoor, Doctor of Philosophy, 2003

Dissertation Directed by: Professor Ingmar R. Prucha, Professor Roger Betancourt, Professor Harry H. Kelejian, Department of Economics

The first part of my dissertation considers the estimation of a panel data model with error components that are both spatially and time-wise correlated. The dissertation combines widely used model for spatial correlation (Cliff and Ord (1973, 1981)) with the classical error component panel data model. I introduce generalizations of the generalized moments (GM) procedure suggested in Kelejian and Prucha (1999) for estimating the spatial autoregressive parameter in case of a single cross section. I then use those estimators to define feasible generalized least squares (GLS) procedures for the regression parameters. I give formal large sample results concerning the consistency of the proposed GM procedures, as well as the consistency and asymptotic normality of the proposed feasible GLS procedures. The new estimators remain computationally feasible even in large samples.

The second part of my dissertation employs a Cliff-Ord-type model to empirically estimate the nature and extent of price competition in the US

wholesale gasoline industry. I use data on average weekly wholesale gasoline price for 289 terminals (distribution facilities) in the US. Data on demand factors, cost factors and market structure that affect price are also used. I consider two time periods, a high demand period (August 1999) and a low demand period (January 2000).

I find a high level of competition in prices between neighboring terminals. In particular, price in one terminal is significantly and positively correlated to the price of its neighboring terminal. Moreover, I find this to be much higher during the low demand period, as compared to the high demand period. In contrast to previous work, I include for each terminal the characteristics of the marginal customer by controlling for demand factors in the neighboring location. I find these demand factors to be important during period of high demand and insignificant during the low demand period. Furthermore, I have also considered spatial correlation in unobserved factors that affect price. I find it to be high and significant only during the low demand period. Not correcting for it leads to incorrect inferences regarding exogenous explanatory variables.

PANEL DATA MODELS WITH SPATIAL CORRELATION: ESTIMATION THEORY AND EMPIRICAL INVESTIGATION OF THE US WHOLESALE GASOLINE INDUSTRY

By

Mudit Kapoor

Thesis Submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy,
2003

Advisory Committee:

Professor Roger Betancourt, Chair
Professor Ingmar R. Prucha, co-Chair
Professor William Evans
Professor Harry H. Kelejian
Professor Dilip Madan

Contents

[1]This map has been taken from "How Pipelines make the Oil Market Work- Their Networks, Operation and Regulation," Allegro Energy Group.

[2]This map has been taken from "How Pipelines make the Oil Market Work- Their Networks, Operation and Regulation," Allegro Energy Group.

[3]This chart has been taken from "How Pipelines make the Oil Market Work- Their Networks, Operation and Regulation," Allegro Energy Group.

1 Introduction

There has been a rapid increase in the use of models that account for spatial interactions in economics. The first part of my dissertation considers the estimation of a panel data model with error components that are both spatially and time-wise correlated. For the case of a single cross section, a widely used model for spatial correlation is that of Cliff and Ord (1973, 1981). The dissertation combines this model with the classical error component panel data model. I introduce generalizations of the generalized moments (GM) procedure suggested in Kelejian and Prucha (1999) for estimating the spatial autoregressive parameter in case of a single cross section. I then use those estimators to define feasible generalized least squares (GLS) procedures for the regression parameters. I give formal large sample results concerning the consistency of the proposed GM procedures, as well as the consistency and asymptotic normality of the proposed feasible GLS procedures. The new estimators remain computationally feasible even in large samples.

The second part of my dissertation explores the nature and extent of price competition in the US wholesale gasoline industry. Competing models of product differentiation produce contrasting predictions regarding the nature of price competition among firms. In particular, spatial price competition models predict that a firm interacts locally with neighboring firms. In contrast, monopolistic competition models predict low level of strategic interaction among neighbors. In this paper I employ a Cliff-Ord-type model to empirically estimate the nature and extent of price competition in the US wholesale gasoline industry, and to test the competing theories of price competition. I use data on average weekly wholesale gasoline price for 289 terminals (distribution facilities) in the US. Data on demand factors, cost factors and market structure that affect price are also used. Geographic Information System (GIS) software is used to compute the actual road distance between terminals. For each terminal I select the nearest neighbor based on the shortest road distance. I consider two time periods, a high demand period (August 1999) and a low demand period (January 2000).

I find a high level of competition in prices between neighboring terminals. In particular, price in one terminal is significantly and positively correlated to the price of its neighboring terminal. This finding supports the prediction of spatial price competition models. Moreover, I find the extent of spatial correlation to be much higher during the low demand period, as compared to the high demand period. Clearly, given the strategic relevance of gasoline for the U.S. economy, it is important to understand what factors determine and influence its price. Among other determining factors such as the price of crude oil and seasonal supply and demand factors, this paper highlights the significance of competition among distribution facilities as a major factor. In contrast to previous work, I include for each terminal the characteristics of the marginal customer by controlling for demand factors in the neighboring location. The marginal customer plays an important role in spatial price competition models. I find these demand factors to be important during period of high demand and insignificant during the low demand period. Furthermore, I have also considered spatial correlation in unobserved factors that affect price. I find it to be high and significant only during the low demand period. Not correcting for it leads to incorrect inferences regarding exogenous explanatory variables. I have also estimated my model using a Euclidean distance measure between neighbors, as was done in previous work. I find that a Euclidean distance measure as compared to measuring distance in actual road miles underestimates the extent of correlation in prices between neighboring terminals.

The organization of the dissertation is as follows. Chapter 2 considers the estimation of a panel data model with error components that are both spatially and time wise correlated. Chapter 3 explores the nature and extent of price competition in the US wholesale gasoline industry. Chapter 4 is the appendix to chapter 2. Chapter 5 is the appendix to chapter 3. Tables and Figures are at the end of appendix.

2 A Model for Spatially Correlated Panel Data

2.1 Introduction

In recent years there has been a growing interest in spatial issues in empirical economics.[4] These spatial issues typically relate to interaction of various sorts between cross-sectional units. These interactions could reflect economic competition forces, externalities, shocks which affects various cross-sectional units, etc. On a somewhat more formal level, in spatial econometrics these interactions could relate to the models' dependent variable, to the exogenous variables, to the disturbance term, or to various combinations of these three. The most widely used model to estimate spatial interactions are variations of the models considered by Whittle (1954) and Cliff and Ord (1973, 1981). Typically, these models are linear and consider either a spatially correlated disturbance term or a spatial lag in the dependent variable, or both.[5]

In the following we specify a panel data model. We assume that the time dimension is small relative to the number of cross-sectional units. Our specification may be viewed as a generalization of the models considered by

[4] Theoretical and empirical issues have been addressed in papers by Case (1991), Conley (1996), Delong and Summers (1991), Dubin (1988), Kelejian and Robinson (1993), Kelejian and Prucha (1998, 1999, 2001a,b,c), Moulton (1990), Pinske and Slade (1998, 2002), Quah (1992),and Topa (1996) among others.

[5] Anselin (1988) provides a survey of these types of models, as well as estimation and testing procedures.

Whittle (1954) and Cliff and Ord (1973, 1981). Furthermore, we generalize a moments estimator given in Kelejian and Prucha (1999) to our panel data framework and prove its consistency.

Since Marschaks' (1939) original suggestion, the use of panel data sets has become reasonably common in empirical economics.[6] This has led to extensive research in the econometrics of panel data. In the literature, traditional models such as the Seemingly Unrelated Regressions (SUR), originally suggested by Zellner (1962), and Error-component models, have been used to estimate the cross-sectional correlations in the disturbance term via the time dimension. However, there are many panel data sets that have a large cross-section but a short time dimension.[7] This feature of panel data sets makes traditional models "less useful" because of severe difficulties in estimating cross-sectional correlations via a short time dimension. This limitation of panel data sets has led to the adoption of spatial models which compensate for a short time dimension by imposing "reasonable" structural restrictions.

Typically the literature has considered a quasi maximum likelihood estimator for models of the Cliff and Ord variety which contain a spatially

[6]Recent overviews include book length surveys by Hsiao (1986), Dielman (1989), Matyas and Sevestre (1996), and Baltagi (1995) among others.

[7]For example, the National Longitudinal Survey of Labor Market Experience (NLS) and the Michigan Panel Study of Income Dynamics.

correlated disturbance term.[8] However, the limitations of the resulting estimator have been discussed in a paper by Kelejian and Prucha (1999), who show that in many cases involving moderate or large sample sizes in a single cross-section, the estimators may not be computationally feasible. This limitation will prevail when we consider more than one time dimension in addition to cross-sections.

In our framework the disturbances will be assumed to follow a spatially autoregressive process. Motivated by the error-component literature the innovations entering the process will be modelled as a sum of two error components, reflecting unit specific effects and some overall innovation. The implications of this is that the disturbances will be both spatially and time correlated. The time correlation is due to the unit specific effects. Our proposed estimator accounts for both the spatial and time correlation of the disturbance term, and, therefore, it is an important extension of the general moments estimator introduced in Kelejian and Prucha (1999) where the disturbances were only assumed to be spatially correlated. In deriving the large sample properties of our estimator we consider the case in which the number of cross-sectional units increase beyond limit, while the number of time pe-

[8] Kelejian and Prucha (1999) use the term (quasi) ML estimator rather than ML estimator to cover cases in which the true specification of the disturbance term is not that specified by the likelihood function which is typically the normal distribution.

riods is fixed at a finite level. As discussed before, this feature is consistent with many panel data sets. Among other things we provide an application to the generalized least squares (GLS) model. We show the asymptotic equivalence between a feasible GLS estimator, which is formulated based on our consistent general moments estimator, and the true GLS estimator.

It proves helpful to introduce the following notational conventions: Let A_N with $N \geq 1$ be some matrix; we then denote the $(i,j)-th$ element of a matrix A_N as $a_{ij,N}$. Correspondingly, we denote the $i-th$ row and $j-th$ column of A_N as $a_{i.,N}$ and $a_{.j,N}$, respectively. Let D be some vector or matrix; then we will use the norm $||D|| = [tr(D'D)]^{1/2}$. Note that this norm is submultiplicative, that is, $||DB|| \leq ||D|| \; ||B||$, where B is a conformably defined matrix or a vector. In this study we will also define $|D|$ as vector or matrix of absolute values. We will say that the elements of sequence of matrices A_N are uniformly bounded in absolute value if

$$|a_{ij,N}| \leq k < \infty$$

for all $1 \leq i, j \leq N; N \geq 1$, where constant k does not depend on any of the indices.

2.2 Model Specification

Consider the linear regression model

$$y_{it,N} = \mathbf{x}'_{it,N}\beta + u_{it,N}\ , \qquad i = 1, \ldots,\ N;\ t = 1, \ldots,\ T \tag{1}$$

where $y_{it,N}$ is the observation on the dependent variable relating to the $i-$th cross-sectional unit at time t, $\mathbf{x}_{it,N} = [x^1_{it,N}, \ldots,\ x^K_{it,N}]'$ is a corresponding $K \times 1$ vector of observations on exogenous regressors which may contain the constant term, and $u_{it,N}$ is the corresponding disturbance term. We conditionalize our model on the realized value of the regressors and so will view $\mathbf{x}_{it,N}$ as a vector of constants.

Stacking observations over the N cross-sections we have for each time period the following regression model

$$y_N(t) = X_N(t)\beta + u_N(t), \qquad t = 1, \ldots, T \tag{2}$$

where $y_N(t) = [y_{1t,N}, \ldots,\ y_{Nt,N}]'$, $X_N(t) = [\mathbf{x}'_{1t,N}, \ldots,\ \mathbf{x}'_{Nt,N}]'$, and $u_N(t) = [u_{1t,N}, \ldots,\ u_{Nt,N}]'$.

We now model the disturbance process in each time period $t = 1, \ldots,\ T$ as the following spatial autoregressive process of order one:

$$u_N(t) = \rho W_N u_N(t) + \varepsilon_N(t) \tag{3}$$

where W_N is an $N \times N$ matrix of known constants often referred to as a *spatial weighting matrix*, ρ is a scalar parameter, which is typically referred to as a *spatial autoregressive parameter,* and $\varepsilon_N(t) = [\varepsilon_{1t,N}, \ldots, \varepsilon_{Nt,N}]'$ is an $N \times 1$ vector of innovations in period t. For reasons of generality, we permit the elements of W_N and $\varepsilon_N(t)$ to depend on N, that is, to form triangular arrays.[9] In the analysis to follow we maintain, however, that the weighting matrix W_N does not change over time.

As remarked earlier, we consider the case where T is fixed and small; therefore, our asymptotic results are based on the condition, $N \to \infty$. It should be clear that the small time dimension makes it impossible to consistently estimate the general correlation structure via the SUR model.

Stacking the observations over both the cross-section and the time dimensions we have via (2) and (3)

$$y_N = X_N\beta + u_N \tag{4}$$

[9]For a discussion on triangular arrays see Prucha (2002). In this analysis we will allow for the elements of the weighting matrix W_N and the innovation vector $\varepsilon_N(t)$ to depend on the sample size, N. For example, consider $w_{ij,N}$, which is the $(i,j)-th$ element of the weighting matrix, W_N whose dimensions are $N \times N$, where N is the sample size. Triangular array implies that if the sample size changes from N to $\widetilde{N}$, then the corresponding $(i,j)-th$ element of the weighting matrix, $W_{\widetilde{N}}$ will be different from that of W_N, that is, $w_{ij,N}$ is different from $w_{ij,\widetilde{N}}$. Same is true for the innovation vector $\varepsilon_N(t)$. This in turn implies that the elements of the weighting matrix, W and the innovation vector ε should be indexed by the sample size, N.

and

$$u_N = \rho(I_T \otimes W_N)u_N + \varepsilon_N \tag{5}$$

where $y_N = [y_N(1)', \ldots, y_N(T)']'$, $X_N = [X_N(1)', \ldots, X_N(T)']'$, $u_N = [u_N(1)', \ldots, u_N(T)']'$, and $\varepsilon_N = [\varepsilon_N(1)', \ldots, \varepsilon_N(T)']'$.

Finally we assume an error component structure for the innovation vector ε_N.[10] In particular

$$\varepsilon_N = (e_T \otimes I_N)\mu_N + v_N, \tag{6}$$

where e_T is a $T \times 1$ vector of unit elements, I_N is an identity matrix of order N, $\mu_N = [\mu_{1,N}, \ldots, \mu_{N,N}]'$ represents the vector of unit specific error components, and $v_N = [v_N(1)', \ldots, v_N(T)']'$ where $v_N(t) = [v_{1t,N}, \ldots, v_{Nt,N}]'$ contains the error components that vary both over units and time periods. In scalar notation the specification in (6) is

$$\varepsilon_{it,N} = \mu_{i,N} + v_{it,N}, \qquad i = 1, \ldots, N;\ t = 1, \ldots, T$$

In what follows we maintain the following assumptions:

Assumption 1 *For all $1 \leq t \leq T$, $1 \leq i \leq N$, where T is a fixed positive integer and $N \geq 1$, the errors $v_{it,N}$, are identically distributed with mean zero and finite variance σ_v^2, where $0 < \sigma_v^2 < b_v$, and where b_v is*

[10] See Balestra and Nerlove (1966), Nerlove (1971), Maddala (1971), Hsiao (1986) and Baltagi (1995) among others.

a known finite constant. In addition for each $N \geq 1$ *the error terms,* $v_{11,N}, \ldots, v_{N1,N}, \ldots, v_{1T,N}, \ldots, v_{NT,N}$ *are identically and independently distributed. Also for all* $1 \leq i \leq N$ *and* $N \geq 1$*, the errors* $\mu_{i,N}$*, are identically distributed with mean zero and finite variance* σ_μ^2*, where* $0 < \sigma_\mu^2 < b_\mu$*, and where* b_μ *is a known finite constant. In addition for each* $N \geq 1$ *the error terms* $\mu_{1,N}, \ldots, \mu_{N,N}$ *are identically and independently distributed. Furthermore, the two processes* $(v_{it,N})$ *and* $(\mu_{i,N})$ *are independent of each other.*

Assumption 2 *(a) All diagonal elements of* W_N *are zero. (b)* $|\rho| < 1$*. (c) The matrix* $I_N - \rho W_N$ *is non-singular for all* $|\rho| < 1$*.*

In scalar notation the specification in (3) is

$$u_{it,N} = \rho \Sigma_{j=1}^{N} w_{ij,N} u_{jt,N} + \varepsilon_{it,N}, \quad i = 1, \ldots, N; \; t = 1, \ldots, T$$

where $w_{ij,N}$ is the $(i,j) - th$ element of the weighting matrix W_N. The nonzero weights $w_{ij,N}$ are often specified to be those which correspond to units which are related in a meaningful way. Such units are often said to be *neighbors*. As one example, if the cross-sectional units are geographic regions, one might make $w_{ij,N} \neq 0$ if the $i-$th and $j-$th regions are contiguous, and $w_{ij,N} = 0$ otherwise. For reasonable time-periods it is fair to assume that this relationship does not change- i.e., $w_{ij,N}$ is constant through time.

In the above setting, each disturbance term consists of a weighted sum of disturbances in related regions in each time period and an innovation term that contains two stochastic "error components"; one "error component" is unit specific and the other varies both over time and units. Clearly Assumption 2(a) is the normalization of the model. Assumption 2(b) is a stability condition for certain specifications of the weighting matrix, W_N, and Assumption 2(c) ensures that the disturbance vector u_N is uniquely defined in terms of the innovation vector ε_N.

Given the above assumptions it then follows from (6) that $E\varepsilon_N = 0$ and the covariance vector matrix of the innovation vector ε_N is given by

$$\begin{aligned} E\varepsilon_N\varepsilon_N' &= \Omega_{\varepsilon,N} = \sigma_\mu^2(J_T \otimes I_N) + \sigma_v^2 I_{NT} \\ &= \sigma_v^2 Q_{0,N} + \sigma_1^2 Q_{1,N} \end{aligned} \tag{7}$$

where

$$\begin{aligned} \sigma_1^2 &= T\sigma_\mu^2 + \sigma_v^2, \\ Q_{1,N} &= \frac{J_T}{T} \otimes I_N, \\ Q_{0,N} &= I_{NT} - Q_{1,N} = (I_T - \frac{J_T}{T}) \otimes I_N, \end{aligned} \tag{8}$$

and where J_T is a $T \times T$ matrix of unit elements, and in general, I_K is an identity matrix of order K. Observing that $J_T = e_T e_T'$ where e_T is a $T \times 1$

vector of unit elements, it is readily seen that $Q_{0,N}$ and $Q_{1,N}$ are idempotent, orthogonal and sum to the identity matrix[11]. Specifically

$$
\begin{aligned}
Q_{1,N}Q_{1,N} &= (\frac{J_T}{T} \otimes I_N)(\frac{J_T}{T} \otimes I_N) = (\frac{J_T J_T}{T^2} \otimes I_N) = (\frac{e_T e_T' e_T e_T'}{T^2} \otimes I_N) \\
&= (\frac{e_T e_T'}{T} \otimes I_N) = (\frac{J_T}{T} \otimes I_N) = Q_{1,N}, \\
Q_{0,N}Q_{0,N} &= (I_{NT} - Q_{1,N})(I_{NT} - Q_{1,N}) \qquad (9) \\
&= I_{NT} - Q_{1,N} - Q_{1,N} + Q_{1,N}Q_{1,N} \\
&= I_{NT} - Q_{1,N} - Q_{1,N} + Q_{1,N} = I_{NT} - Q_{1,N} = Q_{0,N}, \\
Q_{0,N}Q_{1,N} &= (I_{NT} - Q_{1,N})Q_{1,N} = Q_{1,N} - Q_{1,N} = 0, \\
Q_{0,N} + Q_{1,N} &= I_{NT}.
\end{aligned}
$$

In addition it is readily seen that

$$
\begin{aligned}
tr(Q_{0,N}) &= N(T-1), \\
tr(Q_{1,N}) &= N. \qquad (10)
\end{aligned}
$$

Note that the elements of $Q_{0,N}$ and $Q_{1,N}$ are uniformly bounded by 1.

[11]Variations of these results are available in Baltagi (1995, Pg. 10).

It will prove useful to note that for any $N \times N$ matrix A_N we have[12]

$$
\begin{aligned}
(I_T \otimes A_N)Q_{0,N} &= Q_{0,N}(I_T \otimes A_N), \\
(I_T \otimes A_N)Q_{1,N} &= Q_{1,N}(I_T \otimes A_N). \qquad (11)
\end{aligned}
$$

From (5) it follows that

$$
u_N = [I_T \otimes (I_N - \rho W_N)^{-1}]\varepsilon_N. \qquad (12)
$$

Thus $Eu_N = 0$ and, recalling (7) and (11)

$$
\begin{aligned}
Eu_N u_N^{'} &= \Omega_{u,N}(\rho) = [I_T \otimes (I_N - \rho W_N)^{-1}]\Omega_{\varepsilon,N}[I_T \otimes (I_N - \rho W_N')^{-1}] \\
&= \Omega_{\varepsilon,N}[I_T \otimes (I_N - \rho W_N)^{-1}(I_N - \rho W_N')^{-1}] \qquad (13)
\end{aligned}
$$

We note that in general, the elements of $(I_N - \rho W_N)^{-1}$ will depend on the sample size of the cross-sectional units N. Subsequently, the elements of u_N will depend on N and thus form a triangular array. Also, in general, the elements of $\Omega_{u,N}(\rho)$ will depend on N. Furthermore, the elements of u_N are heteroskedastic, and spatially correlated, as well as correlated over time. In the following sections we explore the estimation strategies for the parameters of the model considered in (4), (5) and (6).

[12] Observing that $Q_{0,N} = (I_T - \frac{J_T}{T}) \otimes I_N$ and $Q_{1,N} = \frac{J_T}{T} \otimes I_N$ we have $(I_T \otimes A_N)Q_{0,N} = (I_T - \frac{J_T}{T}) \otimes A_N = ((I_T - \frac{J_T}{T}) \otimes I_N)(I_T \otimes A_N) = Q_{0,N}(I_T \otimes A_N)$. Similarly, $(I_T \otimes A_N)Q_{1,N} = \frac{J_T}{T} \otimes A_N = (\frac{J_T}{T} \otimes I_N)(I_T \otimes A_N) = Q_{1,N}(I_T \otimes A_N)$.

2.3 Maximum Likelihood Estimation

Recall our model in stacked form, from (4)-(6)

$$\begin{aligned} y_N &= X_N\beta + u_N, \\ u_N &= \rho(I_T \otimes W_N)u_N + \varepsilon_N, \\ \varepsilon_N &= (e_T \otimes I_N)\mu_N + v_N. \end{aligned}$$

Assuming $\varepsilon_N \backsim N\ (0, \Omega_{\varepsilon,N})$ we have

$$u_N \backsim N\ (0, \Omega_{u,N}(\rho)).$$

Therefore,

$$y_N \backsim N\ (X\beta, \Omega_{u,N}(\rho)). \tag{14}$$

Substituting (7) and (8) into (13) we get

$$\begin{aligned} \Omega_{u,N}(\rho) &= \sigma_v^2[(I_T - \frac{J_T}{T}) \otimes (I_N - \rho W_N)^{-1}(I_N - \rho W_N^{'})^{-1}] + \\ &\quad \sigma_1^2[\frac{J_T}{T} \otimes (I_N - \rho W_N)^{-1}(I_N - \rho W_N^{'})^{-1}] \qquad (15) \\ &= [(\sigma_v^2(I_T - \frac{J_T}{T}) + \sigma_1^2\frac{J_T}{T}) \otimes (I_N - \rho W_N)^{-1}(I_N - \rho W_N^{'})^{-1}] \end{aligned}$$

and thus[13]

$$\begin{aligned}\det(\Omega_{u,N}(\rho)) &= \det[(\sigma_v^2(I_T - \frac{J_T}{T}) + \sigma_1^2\frac{J_T}{T}) \otimes (I_N - \rho W_N)^{-1}(I_N - \rho W_N')^{-1}] \\ &= [\det(\sigma_v^2(I_T - \frac{J_T}{T}) + \sigma_1^2\frac{J_T}{T})]^N[\det(I_N - \rho W_N)]^{-2T} \qquad (16)\end{aligned}$$

Given (7)-(9) it is not difficult to show that

$$\Omega_{\varepsilon,N}^{-1} = \sigma_v^{-2}Q_{0,N} + \sigma_1^{-2}Q_{1,N} \qquad (17)$$

and, therefore, from (13)

$$\begin{aligned}\Omega_{u,N}^{-1}(\rho) &= [I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)]\Omega_{\varepsilon,N}^{-1} \\ &= [(\sigma_v^{-2}(I_T - \frac{J_T}{T}) + \sigma_1^{-2}\frac{J_T}{T}) \otimes (I_N - \rho W_N')(I_N - \rho W_N)] \qquad (18)\end{aligned}$$

Given (16) the likelihood function for the model in (4)-(6) is given by

$$\begin{aligned}L &= (2\pi)^{-NT/2}\left|\det(\Omega_{u,N}^{-1}(\rho))\right|^{1/2}\exp(-\frac{1}{2}[y_N - X\beta]'\Omega_{u,N}^{-1}(\rho)[y_N - X\beta]) \\ &= (2\pi)^{-NT/2}|\det(\sigma_v^2(I_T - \frac{J_T}{T}) + \sigma_1^2\frac{J_T}{T})|^{-N/2} * \qquad (19) \\ &\quad |\det(I_N - \rho W_N)|^T\exp(-\frac{1}{2}[y_N - X_N\beta]'\Omega_{u,N}^{-1}(\rho)[y_N - X_N\beta])\end{aligned}$$

Substituting (16) and (18) into (19) and then taking the logs we have the log

[13] Recall that if A is an $N \times N$ matrix and B is an $M \times M$ matrix then $\det(A \otimes B) = [\det(A)]^M[\det(B)]^N$. Furthermore, $\det(A') = \det(A)$ and $\det(A^{-1}) = [\det(A)]^{-1}$.
If A is an $N \times N$ matrix and C is an $N \times N$ matrix then $\det(AC) = \det(A)\det(C)$.

likelihood function

$$\begin{aligned}
\ln(L) \quad = \quad & -\frac{NT}{2}\ln(2\pi) - \\
& \frac{N}{2}\ln\left|\det(\sigma_v^2(I_T - \frac{J_T}{T}) + \sigma_1^2\frac{J_T}{T})\right| + \\
& T\ln|\det(I_N - \rho W_N)| - \qquad (20) \\
& \frac{1}{2}[y_N - X_N\beta]'[(\sigma_v^{-2}(I_T - \frac{J_T}{T}) + \sigma_1^{-2}\frac{J_T}{T}) \otimes \\
& (I_N - \rho W_N')(I_N - \rho W_N)][y_N - X_N\beta].
\end{aligned}$$

As remarked earlier, normality of ε_N is not one of our maintained assumptions, and hence we refer to the maximizers of the above log likelihood as quasi ML estimators. As is evident from (20), the computation of the quasi ML estimators involves among other things, the repeated evaluation of the determinant of the $N \times N$ matrix $I_N - \rho W_N$. To minimize the computational burden, Ord (1975) suggested that $\ln|\det(I_N - \rho W_N)|$ in (20) be determined as $\ln|\det(I_N - \rho W_N)| = \Sigma_{i=1}^N \ln(|1 - \rho\lambda_i|)$, where λ_i denotes the ith eigenvalue of W_N. Given that W_N is a known matrix its eigenvalues have to be computed only once at the outset of the numerical optimization procedure employed in finding the quasi ML estimates and not repeatedly at each of the necessary numerical iterations. However, this still leaves the researcher with the task of finding the eigenvalues of the $N \times N$ matrix W_N. It has been pointed out by Kelejian and Prucha (1999) that this task is typically "challenging" particularly if N is very large and W_N is not properly structured. They considered the case of an "idealized" symmetric weighting matrix, and a sample of size 400. For such matrices, all eigenvalues are real. They then employed a subroutine for computing the eigenvalues of the weighting matrix from the IMSL program library without imposing symmetry. The routine reported eigenvalues with imaginary parts that differed substantially from zero by more than 0.5 in absolute value. Only when they employed a subroutine which utilized the symmetric nature of the weighting matrix were they able to calculate the eigenvalues accurately. In practice, weighting matrices are typically not symmetric. The implication is that an accurate computation of the quasi ML estimator may not be feasible in many cases, even for moderate sample sizes, say 400 or larger. Given the computational problems of the quasi ML estimator, and the small size of the time series which rules out an SUR or an error-component model, it is clearly important to have an alternative estimator of the model parameters which is computationally feasible for general weighting matrices W_N, large cross-sectional units N, and a reasonably small but fixed time series T.

2.4 GLS Estimation

Recall our model in stacked form, from (4)-(6)

$$\begin{aligned} y_N &= X_N\beta + u_N, \\ u_N &= \rho(I_T \otimes W_N)u_N + \varepsilon_N, \\ \varepsilon_N &= (e_T \otimes I_N)\mu_N + v_N. \end{aligned}$$

As discussed earlier the elements of u_N will generally depend on the sample size and hence those of y_N will depend on N. For reasons of generality we will permit the elements of X_N to also depend on N. We maintain the following assumptions for the regressor matrix X_N.

Assumption 3 *The elements of X_N are nonstochastic and uniformly bounded in absolute value by k_x, $0 < k_x < \infty$. Also X_N has full column rank. We also* *assume*

$$
\begin{aligned}
Q_{xx} &= \lim_{N\to\infty}(NT)^{-1}X_N'X_N,\\
\overline{Q}_{xx}(\rho) &= \lim_{N\to\infty}(NT)^{-1}X_N'[I_T\otimes(I_N-\rho W_N)^{-1}][I_T\otimes(I_N-\rho W_N')^{-1}]X_N,\\
\underline{Q}_{xQ_0x}(\rho) &= \lim_{N\to\infty}(NT)^{-1}X_N'[I_T\otimes(I_N-\rho W_N')]Q_{0,N}[I_T\otimes(I_N-\rho W_N)]X_N,\\
\overline{Q}_{xQ_0x}(\rho) &= \lim_{N\to\infty}(NT)^{-1}X_N'[I_T\otimes(I_N-\rho W_N)^{-1}]Q_{0,N}[I_T\otimes(I_N-\rho W_N')^{-1}]X_N,\\
\underline{Q}_{xQ_1x}(\rho) &= \lim_{N\to\infty}(NT)^{-1}X_N'[I_T\otimes(I_N-\rho W_N')]Q_{1,N}[I_T\otimes(I_N-\rho W_N)]X_N,\\
\overline{Q}_{xQ_1x}(\rho) &= \lim_{N\to\infty}(NT)^{-1}X_N'[I_T\otimes(I_N-\rho W_N)^{-1}]Q_{1,N}[I_T\otimes(I_N-\rho W_N')^{-1}]X_N,
\end{aligned}
$$

where the matrices Q_{xx}, $\overline{Q}_{xx}(\rho)$, $[\sigma_v^{-2}\overline{Q}_{xQ_0x}(\rho)+\sigma_1^{-2}\overline{Q}_{xQ_1x}(\rho)]$, $[\sigma_v^{-2}\underline{Q}_{xQ_0x}(\rho)+\sigma_1^{-2}\underline{Q}_{xQ_1x}(\rho)]$ *are finite and nonsingular for all* $|\rho|<1$.

We can rewrite the disturbance term in (5) as

$$
\begin{aligned}
[I_T\otimes(I_N-\rho W_N)]u_N &= \varepsilon_N, \qquad (21)\\
u_N &= [I_T\otimes(I_N-\rho W_N)]^{-1}\varepsilon_N.
\end{aligned}
$$

Premultiplying (4) by $[I_T\otimes(I_N-\rho W_N)]$ we get

$$
y_N^*(\rho)=X_N^*(\rho)\beta+\varepsilon_N, \qquad (22)
$$

where

$$
\begin{aligned}
y_N^*(\rho) &= [I_T\otimes(I_N-\rho W_N)]y_N,\\
X_N^*(\rho) &= [I_T\otimes(I_N-\rho W_N)]X_N.
\end{aligned}
$$

Given the form of ε_N, we have what is often called an "error-component model".

The GLS estimator for such model is

$$\widehat{\beta}_{GLS,N} = [X_N^*(\rho)'\Omega_{\varepsilon,N}^{-1}X_N^*(\rho)]^{-1}[X_N^*(\rho)'\Omega_{\varepsilon,N}^{-1}y_N^*(\rho)]. \tag{23}$$

Using $\Omega_{\varepsilon,N}^{-1} = \sigma_v^{-2}Q_0 + \sigma_1^{-2}Q_{1,N}$, we can rewrite (23) as

$$\begin{aligned}
\widehat{\beta}_{GLS,N} &= \left(\frac{X_N^*(\rho)'Q_{0,N}X_N^*(\rho)}{\sigma_v^2} + \frac{X_N^*(\rho)'Q_{1,N}X_N^*(\rho)}{\sigma_1^2}\right)^{-1} * \\
&\quad \left(\frac{X_N^*(\rho)'Q_{0,N}y_N^*(\rho)}{\sigma_v^2} + \frac{X_N^*(\rho)'Q_{1,N}y_N^*(\rho)}{\sigma_1^2}\right), \\
&= (X_N^*(\rho)'Q_{0,N}X_N^*(\rho) + \delta X_N^*(\rho)'Q_{1,N}X_N^*(\rho))^{-1} * \qquad (24) \\
&\quad (X_N^*(\rho)'Q_{0,N}y_N^*(\rho) + \delta X_N^*(\rho)'Q_{1,N}y_N^*(\rho)),
\end{aligned}$$

where

$$\delta = \frac{\sigma_v^2}{\sigma_1^2} = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\mu^2}.$$

Given that $Q_{1,N} = I_{NT} - Q_{0,N}$ we can rewrite (24) as

$$\begin{aligned}
\widehat{\beta}_{GLS,N} &= [(1-\delta)X_N^*(\rho)'Q_{0,N}X_N^*(\rho) + \delta X_N^*(\rho)'X_N^*(\rho)]^{-1} * \\
&\quad [(1-\delta)X_N^*(\rho)'Q_{0,N}y_N^*(\rho) + \delta X_N^*(\rho)'y_N^*(\rho)]. \qquad (25)
\end{aligned}$$

Let

$$
\begin{aligned}
H &= [(1-\delta)X_N^*(\rho)'Q_{0,N}X_N^*(\rho) + \delta X_N^*(\rho)'X_N^*(\rho)], \\
F_w &= H^{-1}[(1-\delta)X_N^*(\rho)'Q_{0,N}X_N^*(\rho)], \\
\widehat{\beta}_{Q_{0,N}} &= [X_N^*(\rho)'Q_{0,N}X_N^*(\rho)]^{-1}[X_N^*(\rho)'Q_{0,N}y_N^*(\rho)], \qquad (26) \\
\widehat{\beta}_{OLS,N} &= [X_N^*(\rho)'X_N^*(\rho)]^{-1}[X_N^*(\rho)'y_N^*(\rho)].
\end{aligned}
$$

From the above expressions in (26) it is very clear that

$$
I - F_w = I - H^{-1}[(1-\delta)X_N^*(\rho)'Q_{0,N}X_N^*(\rho)], \qquad (27)
$$

where I is an identity matrix.

Premultipying (27) by H we get

$$
\begin{aligned}
H(I-F_w) &= H - HF_w \\
&= H - (1-\delta)X_N^*(\rho)'Q_{0,N}X_N^*(\rho) \qquad (28) \\
&= \delta X_N^*(\rho)'X_N^*(\rho).
\end{aligned}
$$

The result in (28) implies

$$
I - F_w = H^{-1}[\delta X_N^*(\rho)'X_N^*(\rho)]. \qquad (29)
$$

Using the above definitions for H, F_w, $I-F_w$, and by noting that

$$
\begin{aligned}
[X_N^*(\rho)'Q_{0,N}X_N^*(\rho)]\widehat{\beta}_{Q_{0,N}} &= [X_N^*(\rho)'Q_{0,N}y_N^*(\rho)], \\
[X_N^*(\rho)'X_N^*(\rho)]\widehat{\beta}_{OLS,N} &= [X_N^*(\rho)'y_N^*(\rho)],
\end{aligned}
$$

we can express (25) as

$$\begin{aligned}
\widehat{\beta}_{GLS,N} &= H^{-1}[(1-\delta)(X_N^*(\rho)'Q_{0,N}X_N^*(\rho))\widehat{\beta}_{Q_0,N} + \delta(X_N^*(\rho)'X_N^*(\rho))\widehat{\beta}_{OLS,N}] \\
&= H^{-1}[(1-\delta)X_N^*(\rho)'Q_{0,N}X_N^*(\rho)]\widehat{\beta}_{Q_0,N} + H^{-1}[\delta X_N^*(\rho)'X_N^*(\rho)]\widehat{\beta}_{OLS,N} \\
&= F_w\widehat{\beta}_{Q_0,N} + (I - F_w)\widehat{\beta}_{OLS,N}.
\end{aligned}$$

We note that $\widehat{\beta}_{Q_0,N}$ corresponds to the OLS estimator of the transformed model

$$Q_{0,N}y_N^*(\rho) = Q_{0,N}X_N^*(\rho)\beta + Q_{0,N}\varepsilon_N.$$

Furthermore, $\widehat{\beta}_{OLS,N}$ corresponds to the OLS estimator of the model

$$y_N^*(\rho) = X_N^*(\rho)\beta + \varepsilon_N.$$

The typical element of $Q_{0,N}y_N^*(\rho)$ is $y_{it,N}^* - \overline{y}_{i.,N}^*$. The estimator, $\widehat{\beta}_{Q_0,N}$ is based on within group (unit) variation of the data and corresponds to the within group estimator defined in the literature.[14] Clearly if $\delta = 1$ so that $\sigma_\mu^2 = 0$, the generalized least squares estimator $\widehat{\beta}_{GLS,N}$ reduces to the OLS estimator $\widehat{\beta}_{OLS,N}$. Furthermore, if $\delta = 1$ then $\Omega_{\varepsilon,N} = \sigma_v^2 I_{NT}$, in which case $\widehat{\beta}_{OLS,N}$ is the best linear unbiased estimator (BLUE). If normality is assumed then $\widehat{\beta}_{OLS,N}$ would be efficient. However, if $\delta \neq 1$ in which case $\sigma_\mu^2 \neq 0$, then $\widehat{\beta}_{OLS,N}$ is not BLUE.

[14] See Greene (2000).

Clearly

$$\begin{aligned} E\widehat{\beta}_{GLS,N} &= \beta, \\ VC(\widehat{\beta}_{GLS,N}) &= [\frac{X_N^*(\rho)'Q_{0,N}X_N^*(\rho)}{\sigma_v^2} + \frac{X_N^*(\rho)'Q_{1,N}X_N^*(\rho)}{\sigma_1^2}]^{-1}. \end{aligned} \tag{30}$$

It is evident that the GLS estimator depends on unknown parameters, in particular ρ, σ_v^2 and σ_1^2. Therefore, such an estimator is not feasible. In order to compute the feasible GLS estimator of β we need consistent estimators of ρ, σ_v^2 and σ_1^2. For the moment assume that consistent estimators of ρ, σ_v^2 and σ_1^2 exist. Define the feasible GLS estimator of β as identical to $\widehat{\beta}_{GLS,N}$ except that ρ, σ_v^2 and σ_1^2 are replaced by any consistent estimators. Given this, we put forth the following theorem.

Theorem 1 *Given that Assumptions 1 to 3 hold:*

(a) The true GLS estimator $\widehat{\beta}_{GLS,N}$ is a consistent estimator of β, and

$$(NT)^{1/2}[\widehat{\beta}_{GLS,N} - \beta] \xrightarrow{D} N\ \{0, [\sigma_{v,N}^{-2}\underline{Q}_{xQ_0x}(\rho) + \sigma_{1,N}^{-2}\underline{Q}_{xQ_1x}(\rho)]^{-1}\}$$

(b) Let $\widehat{\rho}_N$, $\widehat{\sigma}_{v,N}^2$, $\widehat{\sigma}_{1,N}^2$ be consistent estimators of ρ, σ_v^2, σ_1^2. Then the true GLS estimator $\widehat{\beta}_{GLS,N}$ and the feasible GLS estimator $\widehat{\beta}_{FGLS,N}$ have the same asymptotic distribution. More specifically,

$$(NT)^{1/2}[\widehat{\beta}_{GLS,N} - \widehat{\beta}_{FGLS,N}] \xrightarrow{p} 0 \text{ as } N \to \infty.$$

(c) Furthermore,

$$(NT)^{-1}X'\{[I_T \otimes (I_N - \widehat{\rho}_N W_N')]\widehat{\Omega}_{\varepsilon,N}^{-1}[I_T \otimes (I_N - \widehat{\rho}_N W_N)]\}X$$

where

$$\widehat{\Omega}_{\varepsilon,N}^{-1} = \widehat{\sigma}_{v,N}^{-2} Q_{0,N} + \widehat{\sigma}_{1,N}^{-2} Q_{1,N},$$

is a consistent estimator of

$$\sigma_{v,N}^{-2} \underline{Q}_{xQ_0x}(\rho) + \sigma_{1,N}^{-2} \underline{Q}_{xQ_1x}(\rho).$$

In the spatial model we have considered, a rigorous proof of the asymptotic distribution of the GLS estimator $\widehat{\beta}_{GLS,N}$ requires the use of a central limit theorem for triangular arrays. We will consider such a theorem in the appendix.

Note that Theorem 1 will hold for any consistent estimators of ρ, σ_v^2, σ_1^2. Therefore, these parameters can be viewed as nuisance parameters. In our next section we provide a simple estimation strategy which produces consistent estimators of ρ, σ_v^2, σ_1^2.

2.5 A Generalized Moment Estimator of the Spatial Autoregressive Parameter for Panel Data

Recall our model in stacked form, from (4)-(6)

$$
\begin{aligned}
y_N &= X_N\beta + u_N, \\
u_N &= \rho(I_T \otimes W_N)u_N + \varepsilon_N, \\
\varepsilon_N &= (e_T \otimes I_N)\mu_N + v_N.
\end{aligned}
$$

As discussed earlier the variance covariance matrix of u_N, $\Omega_{u,N}$, depends on ρ, σ_v^2 and σ_1^2. Therefore, we need consistent estimators of ρ, σ_v^2 and σ_1^2 to be able to formulate the feasible GLS estimator.

In the following we define generalized moments (GM) estimators of ρ, σ_v^2 and σ_1^2. These GM estimators generalize, in essence, the GM estimators given in Kelejian and Prucha (1999) for the case of a single cross section. The estimation procedure involves two steps. In the first step we obtain a predictor of u_N, say $\widetilde{u}_N$. In the second step the predictor $\widetilde{u}_N$ is used in the GM approach to consistently estimate ρ, σ_v^2 and σ_1^2.

For notational convenience, let

$$
\begin{aligned}
\overline{u}_N &= (I_T \otimes W_N)u_N, \\
\overline{\overline{u}}_N &= (I_T \otimes W_N)\overline{u}_N, \qquad (31)
\end{aligned}
$$

and correspondingly, let

$$\begin{aligned}
\widetilde{\overline{u}}_N &= (I_T \otimes W_N)\widetilde{u}_N, \\
\widetilde{\overline{\overline{u}}}_N &= (I_T \otimes W_N)\widetilde{\overline{u}}_N. \qquad (32)
\end{aligned}$$

Furthermore, let

$$\overline{\varepsilon}_N = (I_T \otimes W_N)\varepsilon_N. \qquad (33)$$

Given Assumptions 1 and 2 we demonstrate in the appendix the following moments:[15]

$$\begin{aligned}
E\frac{1}{N(T-1)}\varepsilon_N^{'} Q_{0,N}\varepsilon_N &= \sigma_v^2, \\
E\frac{1}{N(T-1)}\overline{\varepsilon}_N^{'} Q_{0,N}\overline{\varepsilon}_N &= \sigma_v^2 N^{-1} tr(W_N^{'} W_N), \\
E\frac{1}{N(T-1)}\overline{\varepsilon}_N^{'} Q_{0,N}\varepsilon_N &= 0, \qquad (34) \\
E\frac{1}{N}\varepsilon_N^{'} Q_{1,N}\varepsilon_N &= \sigma_1^2, \\
E\frac{1}{N}\overline{\varepsilon}_N^{'} Q_{1,N}\overline{\varepsilon}_N &= \sigma_1^2 N^{-1} tr(W_N^{'} W_N), \\
E\frac{1}{N}\overline{\varepsilon}_N^{'} Q_{1,N}\varepsilon_N &= 0.
\end{aligned}$$

Our GM estimators of ρ, σ_v^2, and σ_1^2 are based on the six moments described in (34). Note that in light of (5), (31) and (33) we have

$$\varepsilon_N = u_N - \rho\overline{u}_N \text{ and } \overline{\varepsilon}_N = \overline{u}_N - \rho\overline{\overline{u}}_N. \qquad (35)$$

[15] See Kapoor, Kelejian and Prucha (2002).

By substituting the expressions for ε_N and $\overline{\varepsilon}_N$ in (35) into the six moments described in (34) we get the following six-equation systems:

$$\underbrace{\Gamma_N}_{6\times 4}\underbrace{[\rho, \rho^2, \sigma_v^2, \sigma_1^2]'}_{4\times 1} - \underbrace{\gamma_N}_{6\times 1} = 0 \tag{36}$$

where

$$\Gamma_N = \frac{1}{N}\begin{bmatrix} \frac{2Eu_N' Q_{0,N}\overline{u}_N}{T-1} & \frac{-E\overline{u}_N' Q_{0,N}\overline{u}_N}{T-1} & N & 0 \\ \frac{2E\overline{\overline{u}}_N' Q_{0,N}\overline{u}_N}{T-1} & \frac{-E\overline{\overline{u}}_N' Q_{0,N}\overline{\overline{u}}_N}{T-1} & tr(W_N' W_N) & 0 \\ \frac{\substack{E(u_N' Q_{0,N}\overline{\overline{u}}_N \\ +\overline{u}_N' Q_{0,N}\overline{u}_N)}}{T-1} & \frac{-E\overline{u}_N' Q_{0,N}\overline{\overline{u}}_N}{T-1} & 0 & 0 \\ 2Eu_N' Q_{1,N}\overline{u}_N & -E\overline{u}_N' Q_{1,N}\overline{u}_N & 0 & N \\ 2E\overline{\overline{u}}_N' Q_{1,N}\overline{u}_N & -E\overline{\overline{u}}_N' Q_{1,N}\overline{\overline{u}}_N & 0 & tr(W_N' W_N) \\ \substack{E(u_N' Q_{1,N}\overline{\overline{u}}_N \\ +\overline{u}_N' Q_{1,N}\overline{u}_N)} & -E\overline{u}_N' Q_{1,N}\overline{\overline{u}}_N & 0 & 0 \end{bmatrix}_{6\times 4},$$

$$\gamma_N = \frac{1}{N}\begin{bmatrix} \frac{Eu_N' Q_{0,N}u_N}{T-1} \\ \frac{E\overline{u}_N' Q_{0,N}\overline{u}_N}{T-1} \\ \frac{Eu_N' Q_{0,N}\overline{u}_N}{T-1} \\ Eu_N' Q_{1,N}u_N \\ E\overline{u}_N' Q_{1,N}\overline{u}_N \\ Eu_N' Q_{1,N}\overline{u}_N \end{bmatrix}_{6\times 1},$$

$$\Gamma_{0,N} \quad = \quad \frac{1}{N}\begin{bmatrix} \frac{2Eu_N'Q_{0,N}\overline{u}_N}{T-1} & \frac{-E\overline{u}_N'Q_{0,N}\overline{u}_N}{T-1} & N \\ \frac{2E\overline{\overline{u}}_N'Q_{0,N}\overline{u}_N}{T-1} & \frac{-E\overline{\overline{u}}_N'Q_{0,N}\overline{\overline{u}}_N}{T-1} & tr(W_N'W_N) \\ \frac{\substack{E(u_N'Q_{0,N}\overline{\overline{u}}_N \\ +\overline{u}_N'Q_{0,N}\overline{u}_N)}}{T-1} & \frac{-E\overline{u}_N'Q_{0,N}\overline{\overline{u}}_N}{T-1} & 0 \end{bmatrix}_{3\times 3} \tag{37}$$

$$\Gamma_{1,N} \quad = \quad \frac{1}{N}\begin{bmatrix} 2Eu_N'Q_{1,N}\overline{u}_N & -E\overline{u}_N'Q_{1,N}\overline{u}_N & N \\ 2E\overline{\overline{u}}_N'Q_{1,N}\overline{u}_N & -E\overline{\overline{u}}_N'Q_{1,N}\overline{\overline{u}}_N & tr(W_N'W_N) \\ \substack{E(u_N'Q_{1,N}\overline{\overline{u}}_N \\ +\overline{u}_N'Q_{1,N}\overline{u}_N)} & -E\overline{u}_N'Q_{1,N}\overline{\overline{u}}_N & 0 \end{bmatrix}_{3\times 3} . \tag{38}$$

Note that the elements of upper three rows and the first three columns of the matrix, Γ_N, correspond to the elements of the matrix $\Gamma_{0,N}$. Furthermore, the elements of the lower three rows and the first, second and fourth column correspond to the elements of the matrix $\Gamma_{1,N}$. The usefulness of the matrices $\Gamma_{0,N}$ and $\Gamma_{1,N}$ will be evident later.

Now consider the following analogue to (36) in terms of sample moments based on $\widetilde{u}_N$, $\widetilde{\overline{u}}_N$ and $\widetilde{\overline{\overline{u}}}_N$ which are described in (32):

$$\underbrace{G_N}_{6\times 4}\underbrace{[\rho, \rho^2, \sigma_v^2, \sigma_1^2]'}_{4\times 1} - \underbrace{g_N}_{6\times 1} = \underbrace{\xi_N(\rho, \sigma_v^2, \sigma_1^2)}_{6\times 1}, \tag{39}$$

where

$$G_N = \frac{1}{N}\begin{bmatrix} \frac{2\widetilde{u}'_N Q_{0,N}\widetilde{\overline{u}}_N}{T-1} & \frac{-\widetilde{\overline{u}}'_N Q_{0,N}\widetilde{\overline{u}}_N}{T-1} & N & 0 \\ \frac{2\widetilde{\overline{\overline{u}}}'_N Q_{0,N}\widetilde{\overline{u}}_N}{T-1} & \frac{-\widetilde{\overline{\overline{u}}}'_N Q_{0,N}\widetilde{\overline{\overline{u}}}_N}{T-1} & tr(W'_N W_N) & 0 \\ \frac{\begin{matrix}(\widetilde{u}'_N Q_{0,N}\widetilde{\overline{\overline{u}}}_N \\ +\widetilde{\overline{u}}'_N Q_{0,N}\widetilde{\overline{u}}_N)\end{matrix}}{T-1} & \frac{-\widetilde{\overline{u}}'_N Q_{0,N}\widetilde{\overline{\overline{u}}}_N}{T-1} & 0 & 0 \\ 2\widetilde{u}'_N Q_{1,N}\widetilde{\overline{u}}_N & -\widetilde{\overline{u}}'_N Q_{1,N}\widetilde{\overline{u}}_N & 0 & N \\ 2\widetilde{\overline{\overline{u}}}'_N Q_{1,N}\widetilde{\overline{u}}_N & -\widetilde{\overline{\overline{u}}}'_N Q_{1,N}\widetilde{\overline{\overline{u}}}_N & 0 & tr(W'_N W_N) \\ \begin{matrix}(\widetilde{u}'_N Q_{1,N}\widetilde{\overline{\overline{u}}}_N \\ +\widetilde{\overline{u}}'_N Q_{1,N}\widetilde{\overline{u}}_N)\end{matrix} & -\widetilde{\overline{u}}'_N Q_{1,N}\widetilde{\overline{\overline{u}}}_N & 0 & 0 \end{bmatrix}_{6\times 4},$$

$$g_N = \frac{1}{N}\begin{bmatrix} \frac{\widetilde{u}'_N Q_{0,N}\widetilde{u}_N}{T-1} \\ \frac{\widetilde{\overline{u}}'_N Q_{0,N}\widetilde{\overline{u}}_N}{T-1} \\ \frac{\widetilde{u}'_N Q_{0,N}\widetilde{\overline{u}}_N}{T-1} \\ \widetilde{u}'_N Q_{1,N}\widetilde{u}_N \\ \widetilde{\overline{u}}'_N Q_{1,N}\widetilde{\overline{u}}_N \\ \widetilde{u}'_N Q_{1,N}\widetilde{\overline{u}}_N \end{bmatrix}_{6\times 1},$$

Since G_N and g_N are observable and $\alpha' = [\rho, \rho^2, \sigma_v^2, \sigma_1^2]$ is the parameter vector to be estimated, we can view the 6×1 vector $\xi_N(\rho,\, \sigma_v^2,\, \sigma_1^2)$ as a vector of residuals.

We now define our generalized moments estimators of ρ, σ_v^2 and σ_1^2 as a two step procedure which is similar to the FGLS procedure in the literature. In the first step we get the nonlinear least squares estimators, say $\widetilde{\rho}_{NLS,N}$, $\widetilde{\sigma}^2_{v\ NLS,N}$ and $\widetilde{\sigma}^2_{1\ NLS,N}$ corresponding to (39). More specifically,

$$\begin{aligned}(\widetilde{\rho}_{NLS,N}, \widetilde{\sigma}^2_{v\ NLS,N}, \widetilde{\sigma}^2_{1\ NLS,N}) &= \arg\min\{\xi_N(\rho,\sigma_v^2,\sigma_1^2)'\xi_N(\rho,\sigma_v^2,\sigma_1^2) : \\ \underline{\rho} &\in [-a,a],\ \underline{\sigma}_v^2 \in [0,b_v],\ \underline{\sigma}_1^2 \in [0,b_1]\} \qquad (40)\end{aligned}$$

where $a \geq 1$.

For the second step we define a weighting matrix $\widehat{\Theta}_N$. More specifically,

$$\widehat{\Theta}_N = \begin{bmatrix} \widehat{\Theta}^1_{3\times 3,N} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \widehat{\Theta}^2_{3\times 3,N} \end{bmatrix}_{6\times 6},$$

where

$$\widehat{\Theta}^1_{3\times 3,N} = \frac{\widetilde{\sigma}^4_{v\ NLS,N}}{(T-1)} \begin{bmatrix} 2 & 2tr(\frac{W_N'W_N}{N}) & 0 \\ 2tr(\frac{W_N'W_N}{N}) & 2tr(\frac{W_N'W_NW_N'W_N}{N}) & tr(\frac{W_N'W_N(W_N'+W_N)}{N}) \\ 0 & tr(\frac{W_N'W_N(W_N'+W_N)}{N}) & tr(\frac{W_NW_N+W_N'W_N}{N}) \end{bmatrix}_{3\times 3},$$

$$\widehat{\Theta}^2_{3\times 3,N} = \widetilde{\sigma}^4_{1\ NLS,N} \begin{bmatrix} 2 & 2tr(\frac{W_N'W_N}{N}) & 0 \\ 2tr(\frac{W_N'W_N}{N}) & 2tr(\frac{W_N'W_NW_N'W_N}{N}) & tr(\frac{W_N'W_N(W_N'+W_N)}{N}) \\ 0 & tr(\frac{W_N'W_N(W_N'+W_N)}{N}) & tr(\frac{W_NW_N+W_N'W_N}{N}) \end{bmatrix}_{3\times 3},$$

W_N is a spatial weighting matrix, and $\widetilde{\sigma}^4_{v\ NLS,N}$ and $\widetilde{\sigma}^4_{1\ NLS,N}$ are consistent estimators of σ^4_v and σ^4_1, which are defined in (40).[16] Furthermore, let Θ_N be the weighting matrix which is similar to $\widehat{\Theta}_N$ except that $\widetilde{\sigma}^4_{v\ NLS,N}$ and $\widetilde{\sigma}^4_{1\ NLS,N}$ are replaced by their true parameters σ^4_v and σ^4_1, respectively.

In the second step we define our generalized moments estimators of ρ, σ^2_v and σ^2_1 as a weighted nonlinear least squares estimators, say $\widehat{\rho}_{NLS,N}$, $\widehat{\sigma}^2_{v\ NLS,N}$ and $\widehat{\sigma}^2_{1\ NLS,N}$. More specifically,

$$\begin{aligned}(\widehat{\rho}_{NLS,N}, \widehat{\sigma}^2_{v\ NLS,N}, \widehat{\sigma}^2_{1\ NLS,N}) &= \arg\min\{\xi_N(\rho, \sigma^2_v, \sigma^2_1)'\widehat{\Theta}_N^{-1}\xi_N(\rho, \sigma^2_v, \sigma^2_1) : \\ \underline{\rho} &\in [-a, a],\ \underline{\sigma}^2_v \in [0, b_v],\ \underline{\sigma}^2_1 \in [0, b_1]\} \qquad (41)\end{aligned}$$

where $a \geq 1$.

Remark 1. If the innovations ε_N, were normally distributed the matrix, $N^{-1}\widehat{\Theta}_N$, would correspond to the estimated variance-covariance matrix of the six moment conditions described in (34). Recall that normality of the error term is not one of our maintained assumptions. However, our Monte Carlo results show that even in those cases in which the innovation ε_N is not normally distributed, using the weighting matrix, $\widehat{\Theta}_N$ instead of an identity matrix, improves the efficiency of our general moments estimators defined in

[16] Consistency of $\tilde{\sigma}^4_v$ and $\tilde{\sigma}^4_1$ is shown in the appendix.

(41).[17]

Remark 2. Note that (41) implies that $\left|\widehat{\rho}_{NLS,N}\right| \leq a$ with $a \geq 1$. Since $|\rho| \leq 1$, if the bound a is sufficiently large, $\left|\widehat{\rho}_{NLS,N}\right|$ is essentially the unconstrained nonlinear least squares estimator of ρ. The existence and measurability of $\widehat{\rho}_{NLS,N}$, $\widehat{\sigma}^2_{v\ NLS,N}$, and $\widehat{\sigma}^2_{1\ NLS,N}$ are ensured by, for example Lemma 2 in Jennrich (1969).

In the following let $P_N(\rho) = [I_N - \rho W_N]^{-1}$. We now specify three additional assumptions:

Assumption 4 *(a) The row and column sums of W_N, more specifically $\Sigma_{j=1}^N |w_{ij,N}|$ and $\Sigma_{i=1}^N |w_{ij,N}|$, are uniformly bounded by, say, $k_w < \infty$ for all $i \geq 1$, $j \leq N$, $N \geq 1$. (b) The row and column sums of $P_N(\rho)$, more specifically $\Sigma_{j=1}^N |p_{ij,N}(\rho)|$ and $\Sigma_{i=1}^N |p_{ij,N}(\rho)|$, are uniformly bounded by, say, $k_p < \infty$ for all $i \geq 1$, $j \leq N$, $N \geq 1$, $|\rho| < 1$, where k_p may depend on ρ.*

Assumption 5 *Let $\widetilde{u}_{it,\ N}$ denote the $(i,t)-th$ element of $\widetilde{u}_N$. We then assume that there exists (finite dimensional) random vectors $d_{it,N}$ and Δ_N such that $|\widetilde{u}_{it,N} - u_{it,N}| \leq ||d_{it,N}||\,\|\Delta_N\|$ where $(NT)^{-1}\Sigma_{t=1}^T\Sigma_{i=1}^N \|d_{it,N}\|^{2+\delta} = O_p(1)$ for some $\delta > 0$ and $N^{1/2}\|\Delta_N\| = O_p(1)$.*

[17] Monte Carlo results are available in a paper, Kapoor, Kelejian and Prucha (2002).

Assumption 6 *The smallest eigen value of* $\Gamma'_{0,N}\Gamma_{0,N}$, $\Gamma'_{1,N}\Gamma_{1,N}$, $\Gamma'_{N}\Gamma_{N}$, *where* $\Gamma_{0,N}$, $\Gamma_{1,N}$ *and* Γ_N *are defined in (37), (38) and (36) respectively, are bounded away from zero. This implies that,* $\lambda_{\min}(\Gamma'_{0,N}\Gamma_{0,N}) \geq \lambda_* > 0$, $\lambda_{\min}(\Gamma'_{1,N}\Gamma_{1,N}) \geq \lambda_* > 0$, $\lambda_{\min}(\Gamma'_{N}\Gamma_{N}) \geq \lambda_* > 0$, *where* λ_* *may depend on* ρ, σ_v^2 *and* σ_1^2.

Remark 3. (a) In practice, spatial models are often formulated in such a way that each cross-sectional unit has a limited number of "neighbors" regardless of the sample size (see, for example, Case 1991 and Kelejian and Robinson 1995). In such cases the weighting matrix W_N is sparse for large N, and so Assumption 4(a) would be satisfied. There are many cases in which the elements of W_N are taken to be non-negative and row normalized such that $\Sigma_{i=1}^{N} w_{ij,N} = 1$. Still in other cases the weighting matrix does not contain zeros, but its elements are assumed to decline rapidly in certain directions because they are defined in terms of variables such as distance (see for example, Dubin 1988 and DeLong and Summers 1991). Therefore, under reasonable conditions, Assumption 4(a) is typically satisfied.

(b) Recall from (15) and the definition of $P_N(\rho)$, that $\Omega_{u,N}(\rho) = [(\sigma_v^2(I_T - \frac{J_T}{T}) + \sigma_1^2\frac{J_T}{T}) \otimes P_N(\rho)P_N(\rho)']$. Assumption 4(b), together with the specification that T is fixed, implies that $(NT)^{-1}\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT}\left|\omega_{ij,N}(\rho)\right|$ is uniformly bounded[18], where $\omega_{ij,N}(\rho)$ is the $(i,j)-$th element of $\Omega_{u,N}(\rho)$. Therefore, Assumption 4(b) restricts the degree of correlation of the elements of u_N.

Remark 4. Assumption 5 should be satisfied for most cases in which $\widetilde{u}_N$ is based on $N^{1/2}$-consistent estimators of regression coefficients. We demonstrate in the appendix that under our maintained assumptions and the model specification in (4), (5) and (6), the OLS estimator $\widehat{\beta}_{OLS,N} = (X_N' X_N)^{-1} X_N' y_N$ is $N^{1/2}$-consistent. Given this we can compute the corresponding residuals $\widetilde{u}_{it,N} = y_{it,N} - \mathbf{x}_{it,N}' \widehat{\beta}_{OLS,N}$. Furthermore, we show in the appendix that these residuals satisfy Assumption 5 with $d_{it,N} = \mathbf{x}_{it,N}$ and $\Delta_N = \widehat{\beta}_{OLS,N} - \beta$.

Remark 5. Assumption 6 is an identifiability condition.

Our basic result is Theorem 2, whose proof is given in the appendix.

Theorem 2 *Let $\widehat{\rho}_{NLS,N}$, $\widehat{\sigma}^2_{v\ NLS,N}$, $\widehat{\sigma}^2_{1\ NLS,N}$ be the nonlinear least squares estimators defined by (41). Suppose Assumptions 1 to 6 and the smallest and largest eigenvalues of the matrices Θ_N^{-1} satisfy $0 < \overline{\lambda}_* \leq \lambda_{\min}(\Theta_N^{-1}) \leq \lambda_{\max}(\Theta_N^{-1})$. Suppose furthermore that $\widehat{\beta}_{OLS,N}$ and $\widehat{\Theta}_N$ are consistent estimators of β and Θ_N, respectively. Then, the GM estimators $\widehat{\rho}_{NLS,N}$, $\widehat{\sigma}^2_{v\ NLS,N}$,*

[18] This is demonstrated in the appendix.

$\widehat{\sigma}^2_{1\ NLS,N}$ *defined in (41) are consistent for* ρ, σ_v^2, σ_1^2, *i.e.,*

$$(\widehat{\rho}_{NLS,N}, \widehat{\sigma}^2_{v\ NLS,N}, \widehat{\sigma}^2_{1\ NLS,N}) \xrightarrow{p} (\rho, \sigma_v^2, \sigma_1^2) \text{ as } N \to \infty.$$

The assumptions relating to eigenvalues of Θ_N^{-1} together with Assumption 6 ensure identifiably uniqueness of the parameters $\rho, \sigma_v^2, \sigma_1^2$. They also ensure that the elements of Θ_N^{-1} are $O(1)$.

2.6 Conclusion

This paper considers the estimation of a panel data model with error components that are both spatially and time-wise correlated. The dissertation combines the model for spatial correlation (that of Cliff and Ord (1973, 1981)) with the classical error component panel data model. I introduce generalizations of the generalized moments (GM) procedure suggested in Kelejian and Prucha (1999) for estimating the spatial autoregressive parameter in case of a single cross section. I then use those estimators to define feasible generalized least squares (GLS) procedures for the regression parameters. I give formal large sample results concerning the consistency of the proposed GM procedures, as well as the consistency and the asymptotic normality of the proposed feasible GLS procedures. The new estimators remain computationally feasible even in large samples.

3 Estimation of Price Competition in a Spatial Model: An Investigation of the US Wholesale Gasoline Industry

3.1 Introduction

Significance of gasoline emerges from the fact that transportation costs accounts for 19% of the average annual expenditure of US households and gasoline price is an important factor that influence these costs. Having acknowledged this, it is important to understand the factors that determine gasoline prices. Among several factors such as price of crude oil, seasonal supply and demand and weather conditions, I highlight the role that competition among distribution facilities play in determining price of gasoline. More specifically, this paper analyzes the nature and extent of price competition in the US wholesale gasoline industry.

The nature of competition among wholesale gasoline distributors can be studied using insights from theoretical models of product differentiation. While product differentiation, within the gasoline industry arises from different sources such as brand names, quality (regular, premium etc.), the location of distribution facilities is also an important dimension of differentiation. In theoretical literature, classic models of product differentiation include, among others, the spatial models such as linear city model by Hotelling (1929) and circular city model by Salop (1979) and the Monopolistic Competition models such as Chamberlin (1933), representative consumer models by Spence (1976), and Dixit and Stiglitz (1977). These competing models, however, produce contrasting predictions regarding the nature of price competition among firms.

In spatial models of product differentiation, each firm is identified with an "address" in product space.[19] In general it can be imagined that the firms' products are located in some N-dimensional characteristic space and the consumers' optimum points of consumption are distributed over this characteristic space. The unique feature of these models is that firms compete only for the local customers.[20] Therefore, firms compete locally and there is high level of strategic interaction between them.[21] In contrast, monopolistic competition models predict that firms compete with all other firms for customers. This implies that there is low level of interaction between firms and strategies of one firm do not affect or have negligible effect on payoffs of other firms.

[19]Product differentiation between firms selling homogenous products might exist due to different geographical locations of the firms.

[20]This result is robust even in the presence of a continuum of firms.

[21]A crucial assumption that generates this result is that consumers have a sufficiently high valuation for the products.

In this paper, while analyzing price competition in the US wholesale gasoline industry, I empirically distinguish between spatial model and monopolistic competition model of product differentiation. This work is most closely related to Pinske, Slade, and Brett (2002) who propose an "instrumental variables series estimator" to investigate the nature of price competition among firms in a differentiated product market.[22] My work, however, makes crucial innovations along conceptual and methodological grounds. Specifically there are three main innovations in this paper. Firstly, at a conceptual level, I account for the marginal customer who plays a significant role in price determination in spatial models of product differentiation. And while limited data availability makes empirical identification of the marginal customer a very challenging task, the proposed variable has the desirable feature in that it captures the characteristics of the marginal customer and is also empirically simple to compute. Secondly, this work is based on a very comprehensive dataset comprising of two time periods - a high demand period (August 1999) and a low demand period (January 2000). Thirdly, my innovation is at a methodological level where I introduce an estimation strategy that allows for spatial correlation in the explanatory variables and in the unobserved fac-

[22] For "instrumental variable series estimator" see Pinske, Slade and Brett (2002).

tors, in addition to spatial correlation in the dependent variable that has been considered in Pinske, Slade, and Brett (2002). From my results, it becomes clear that this new strategy has important implications for the estimation of parameters of the model as well as in making accurate inferences.

The organization of the paper is as follows. In the next section I describe the US wholesale gasoline industry. Section 3.3 describes the theoretical model and the empirical specification and section 3.4 provides a brief description of the estimation strategy. Section 3.5 discusses the data and section 3.6 describes the empirical model. Section 3.7 describes the results and section 3.8 provides the robustness test. Conclusion and further extensions are in Section 3.9.

3.2 US Wholesale Gasoline Industry

In the US, gasoline is either imported (from Saudi Arabia, Venezuela, etc.) or produced domestically (refineries in Texas, East Coast, etc.).[23] The US petroleum industry is divided into five regions called Petroleum Administration for Defense Districts (PADDs).[24] These are PADD1 (East coast), PADD2 (Midwest), PADD3 (Gulf coast), PADD4 (Rocky Mountain), and PADD5

[23] Local production here refers to the processing of crude oil which is either imported or drilled within the US.

[24] See Figure 1 at the end of the paper for map of the PADDs.

(West coast). Each of these regions are different in terms of production and consumption of gasoline, for example, PADD1 has limited refining capacity and has the highest non-feedstock demand. To meet demand in this region, output is augmented by imports from the Middle East and shipments from the Gulf Coast. PADD2s' regional demand is met by local refineries which is also supplemented by imports from Canada and Gulf Coast. PADD3 is the largest supplier of refined products accounting for 47% of the entire supply in the US. It also accounts for 80% of interregional trade of refined products among the PADDs. PADD4 has the lowest demand for refined products and PADD5 is logistically separate from the other regions and all of its production comes from California.[25]

On a functional basis, the petroleum industry can be divided into two main sectors: the upstream market that includes exploration and production and the downstream market that includes refining, transportation and marketing. These markets are highly integrated and there are firms that

[25]For details on regional differences in production and consumption see "How Pipelines make the Oil Market Work- Their Networks, Operation and Regulation," *Association of Oil Pipelines and the American Petroleums Institute's Pipeline Committee. December (2001)*

References

[1]

operate in both the sectors. Large companies known as "majors" are fully integrated and may own and operate establishments in all of these sectors.[26] Smaller non integrated companies are referred to as "independents" generally specializing in one aspect of the industry.

A very complex and an efficient infrastructure exists to transport gasoline from regions of supply to regions of demand. Four modes of transportation are available to suppliers of gasoline. They are pipelines, waterways, trucks and railroads.[27] Suppliers, on their part, select the shipping modes that minimizes the costs of transportation. Pipelines are the most cost effective and safest way of shipping refined products over long distances, in particular from refineries and coastal areas to distribution and storage facilities or terminals which are typically located near large cities. From these terminals, then, gasoline is trucked by the wholesalers to the retail outlets.[28]

Two types of gasoline, "branded" and "unbranded", are sold and while

[26] See Figure 3 at the end of the paper for the list of "majors".

[27] See Figure 2 at the end of the paper for map of pipelines for refined products in the US.

[28] The US has the largest network of oil pipelines of any nation. All of Europe, for instance, has a pipeline network that is only 1/10 the size of the US network. Suppliers of gasoline select transportation modes on basis of costs and economics favors pipelines. Trucking is generally limited to short haul movements where alternatives are often not available; between distribution facilities or terminals and retail outlets. Railroad is very expensive compared to pipelines and is far from being universally available in US. Waterborne shipments can be priced competitively with pipelines, their use is, however, limited by geography.

branded gasoline bears the name of a major supplier like Exxon, Shell, etc., unbranded gasoline does not bear any brand name. A "major" supplier can sell unbranded gasoline in addition to branded gasoline, however, this cannot be resold bearing a brand name. An "independent" supplier, on the other hand can sell only unbranded gasoline.

Distributors purchase gasoline from suppliers at the terminals and resell it to retail outlets. There can be two types of distributors, "integrated" and "independent". Integrated distributors are owned by the "major" company and only supply to own brand retail outlets. On the other hand, the independent distributors buy from any supplier, "major" or "independent". If they buy branded gasoline, it can only be resold to retail outlets of the same brand, however, they can sell unbranded gasoline to any independent retailers. Market power of a distributor depends on the total number of distributors in the market as well as on the ability of outside distributors to enter this market. Entry into the independent distribution market is easier due to the low costs involved while entry into the integrated market is restricted. This characteristic of the independent distribution sector makes it very competitive.

Independent distributors play a very important role in price competition among suppliers at a terminal and across terminals, particularly in case of unbranded gasoline. At each terminal they purchase unbranded gasoline from the supplier with the lowest price. To a large extent this ensures competition among suppliers within the terminal. Furthermore, they also take advantage of arbitrage opportunities that might exist across terminals, for example, if the price differential across terminal is larger than the transportation cost then they will buy from the terminal with the lowest price. Price competition across terminals, however, is limited due to the transportation costs involved. Given that unbranded gasoline is a very homogenous product and competition across terminals is limited, product differentiation within this industry arises based on the location of the terminal. This feature of the unbranded gasoline industry makes it interesting and appropriate for my empirical analysis.[29]

From policy perspective, the relevance of this industry was highlighted by Hastings in the hearing before the Committee on Government Affairs, US Senate, May 2002. He emphasized that the unbranded gasoline market is necessary to ensure sufficient unbranded gasoline supply at competitive

[29]Besides this, purchase decision of gasoline depends on dynamic issues like brand loyalty, switching costs, long term contracts etc. These issues are difficult to address with limited data availability. Unbranded gasoline price is not discounted which makes it a true transaction price: Pinske, Slade and Brett (2002)

prices, which in turn is crucial for the entry and survival of independent retailers including new chains such as Kmart, Walmart, Costco and RaceTrac. These independent retailers are important because they increase competition at the retail level. For this reason it is important to understand the nature and extent of competition within the unbranded gasoline industry.

3.3 Theoretical Model

The theoretical model used is the same as in Pinske et. al. (2002). The advantage of this model is that it nests models of spatial competition and monopolistic competition. This nesting allows me to assess the nature of competition.

Suppose there are N firms, where $N \geq 1$, which produce a differentiated product in each time period t. Each firm is indexed by a subscript i, where $i = 1, \ldots, N$. Let q_{it} be the product firm i produces in time period t. Each product is associated with a unique characteristic y_{it} and is sold at a nominal price $\widetilde{p}_{it}$ in each time period t.

There are K buyers, where $K \geq 1$. The model allows for the possibility that buyers can purchase more than one variety of product at a time. A buyer is indexed by a subscript k, where $k = 1, \ldots, K$. Each buyer is located at a point in a geographical space. Depending on their geographical location, buyers will choose an optimal location for the purchase of the product. Then buyers will resell this product in competitive markets that are indexed by j, where $j = 1, \ldots, J$. In each of these markets buyers will face a parametric nominal price $\widetilde{v}_{jt}$. Typically buyers resell in one of these markets, however, the model can accommodate for the possibility where buyers can resell in multiple markets.

Let competitive profit of buyer k in time period t be denoted by $\widetilde{\pi}_{kt}(\widetilde{v}_t, \widetilde{p}_t, y_t)$, where $\widetilde{v}_t = (\widetilde{v}_{1t}, \ldots, \widetilde{v}_{Jt})'$, $\widetilde{p}_t = (\widetilde{p}_{1t}, \ldots, \widetilde{p}_{Nt})'$, $y_t = (y_{1t}, \ldots, y_{Nt})'$. This profit function implies that buyers do not hold inventories. The justification for this assumption is that there are huge costs of holding inventories in the gasoline industry. The aggregate profit function for the entire buying industry in each time period t is given by

$$\widetilde{\pi}_t(\widetilde{v}_t, \widetilde{p}_t, y_t) = \sum_{k=1}^{K} \widetilde{\pi}_{kt}(\widetilde{v}_t, \widetilde{p}_t, y_t). \tag{42}$$

It has been shown that the aggregate profit that is obtained when each firm maximizes profit separately, taking prices as given, is the same as that which would be obtained if firms were to jointly maximize profits.[30] In brief, there is no loss of generality in treating the entire buying industry as a single firm.

In order to approximate the profit function of the buying industry, a flexi-

[30] This has been shown in Koopmans (1957), Mas-Collel et. al. (1995).

ble functional form is considered. This is a second order approximation of any arbitrary profit function. Similar to previous work, I simplify the empirical analysis by using a normalized quadratic functional form.[31] Specifically,

$$\begin{aligned}
\widetilde{\pi}_t(\widetilde{v}_t, \widetilde{p}_t, y_t) &= \sum_{j=1}^{J} \alpha_j^{(1)} \widetilde{v}_{jt} + \sum_{i=1}^{N} \alpha_i^{(2)} \widetilde{p}_{it} + \sum_{i=1}^{N} \alpha_i^{(3)} y_{it} \\
&\quad + \frac{1}{2} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}^{(1)} \widetilde{p}_{it} \widetilde{p}_{jt} + \sum_{i=1}^{J} \sum_{j=1}^{J} w_{ij}^{(2)} \widetilde{v}_{it} \widetilde{v}_{jt} \right. \\
&\quad + \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}^{(3)} y_{it} y_{jt} + \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}^{(4)} \widetilde{p}_{it} y_{jt} \\
&\quad \left. + \sum_{i=1}^{N} \sum_{j=1}^{J} w_{ij}^{(5)} \widetilde{p}_{it} \widetilde{v}_{jt} + \sum_{i=1}^{N} \sum_{j=1}^{J} w_{ij}^{(6)} y_{it} \widetilde{v}_{jt} \right].
\end{aligned} \tag{43}$$

After normalizing by an index of output prices, say V_t, we get[32]

$$\begin{aligned}
\pi_t(v_t, p_t, y_t) &= \sum_{j=1}^{J} \alpha_j^{(1)} v_{jt} + \sum_{i=1}^{N} \alpha_i^{(2)} p_{it} + V_t^{-1} \sum_{i=1}^{N} \alpha_i^{(3)} y_{it} \\
&\quad + \frac{1}{2} V_t \left[\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ijt}^{(1)} p_{it} p_{jt} + \sum_{i=1}^{J} \sum_{j=1}^{J} w_{ijt}^{(2)} v_{it} v_{jt} \right. \\
&\quad + V_t^{-2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ijt}^{(3)} y_{it} y_{jt} + V_t^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ijt}^{(4)} p_{it} y_{jt} \\
&\quad \left. + \sum_{i=1}^{N} \sum_{j=1}^{J} w_{ijt}^{(5)} p_{it} v_{jt} + V_t^{-1} \sum_{i=1}^{N} \sum_{j=1}^{J} w_{ijt}^{(6)} y_{it} v_{jt} \right],
\end{aligned} \tag{44}$$

where $\pi_t = V_t^{-1} \widetilde{\pi}_t$, $v_{jt} = V_t^{-1} \widetilde{v}_{jt}$, and $p_{it} = V_t^{-1} \widetilde{p}_{it}$. Furthermore, $v_{jt} s'$ are

[31] Berndt, Fuss, Waverman (1977), McFadden (1978). Also see Jorgenson (1983) and Diewert (1974) for useful surveys on this topic.

[32] Similar to previous work I also assume that this index is exogenous to the prices of the selling industry, p_t.

normalized output prices in competitive markets and can be treated as constants. This will reduce the expression in (44) to the following

$$\begin{aligned}\pi_t(v_t, p_t, y_t) &= a_{1t} + \sum_{i=1}^{N} a_{2it} p_{it} + \sum_{i=1}^{N} a_{3it} y_{it} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ijt}^{(1)} p_{it} p_{jt} \\ &\quad + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ijt}^{(2)} y_{it} y_{jt} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ijt}^{(3)} p_{it} y_{jt}, \qquad (45)\end{aligned}$$

where $a_{1t} = \sum_{j=1}^{J} \alpha_j^{(1)} v_{jt} + \frac{V_t}{2} \sum_{i=1}^{J} \sum_{j=1}^{J} w_{ijt}^{(2)} v_{it} v_{jt}$, $a_{2it} = \alpha_i^{(2)} + \frac{V_t}{2} \sum_{j=1}^{J} w_{ijt}^{(5)} v_{jt}$, $a_{3it} = V_t^{-1} \alpha_i^{(3)} + \frac{1}{2} \sum_{j=1}^{J} w_{ijt}^{(6)} v_{jt}$, $b_{ijt}^{(1)} = V_t w_{ijt}^{(1)}$, $b_{ijt}^{(2)} = V_t^{-1} w_{ijt}^{(3)}$, $b_{ijt}^{(3)} = w_{ijt}^{(4)}$. Moreover, $b_{ijt}^{(1)} = b_{jit}^{(1)}$, $b_{ijt}^{(2)} = b_{jit}^{(2)}$ and $b_{ijt}^{(3)} = b_{jit}^{(3)}$.

Now we can derive quantity demanded for each product by using the Hotelling Lemma. Taking the derivative of (45) with respect to p_{it}, for $i = 1, \ldots, N$, we get

$$\begin{aligned}q_{it} &= \frac{\partial \pi_t(v_t, p_t, y_t)}{\partial p_{it}} \\ &= a_{2it} + \sum_{j=1}^{N} b_{ijt}^{(1)} p_{jt} + \frac{1}{2}\sum_{j=1}^{N} b_{ijt}^{(3)} y_{jt}, \qquad (46)\end{aligned}$$

where q_{it} is the quantity demanded from seller i at time t.

Next we turn to the sellers. As stated earlier, imperfect competition is assumed in the seller side of the market. Furthermore, it is assumed that sellers, indexed by subscript i, face a constant marginal cost C_{it}, in each time period t. This marginal cost is a linear function of various cost factors, therefore, $C_{it} = \sum_{h=1}^{H} \gamma_{ht} c_{it,h}$, where $c_{it,h}$ is the marginal cost associated with h cost factor, and $h = 1, \ldots, H$. In each time period t, seller $i's$ profit function is denoted by

$$\varphi_{it} = (p_{it} - C_{it}) q_{it} - F_i, \tag{47}$$

where F_i is the fixed cost.

By substituting for q_{it} from (46) into (47) we get

$$\varphi_{it} = (p_{it} - C_{it}) \left[a_{2it} + \sum_{j=1}^{N} b_{ijt}^{(1)} p_{jt} + \frac{1}{2} \sum_{j=1}^{N} b_{ijt}^{(3)} y_{jt} \right] - F_i, \tag{48}$$

for $i = 1, \ldots, N$.

In time period t seller i will maximize profits with respect to its own prices, given the prices of other sellers. Therefore, the first order conditions (foc) can be solved to yield seller $i's$ best reply function with respect to the prices of other sellers in each time period t. Solving the foc yields

$$\begin{aligned} p_{it} &= \frac{-1}{2b_{iit}^{(1)}} \left[a_{2it} + \sum_{\substack{j=1 \\ j \neq i}}^{N} b_{ijt}^{(1)} p_{jt} + \frac{1}{2} \sum_{j=1}^{N} b_{ijt}^{(3)} y_{jt} - b_{iit}^{(1)} C_{it} \right] \\ &= \frac{-1}{2b_{iit}^{(1)}} \left[a_{2it} + \sum_{\substack{j=1 \\ j \neq i}}^{N} b_{ijt}^{(1)} p_{jt} + \frac{1}{2} \sum_{j=1}^{N} b_{ijt}^{(3)} y_{jt} \right] + \frac{1}{2} \sum_{h=1}^{H} \gamma_{ht} c_{it,h}. \end{aligned} \tag{49}$$

Equation (49) is the basis of empirical specification. It is evident from the above equation that it will not be possible to estimate all the parameters of the model from a single equation or a short panel data. Therefore, it is necessary to put restrictions on the parameters of the model. In the empirical analysis of gasoline industry, the constant term can have several interpretations. It could capture some overall cost factors which affect all sellers, for example, price of crude oil. In order to model these phenomena I use a random effects model where,

$$\frac{-1}{2b_{iit}^{(1)}} a_{2it} = a_t + \mu_{it},$$

and where a_t is some finite constant in time period t and μ_{it} is independently and identically distributed with a zero mean and a finite variance.

Seller $i's$ reaction curve with respect to seller $j's$ price, p_{jt}, has a slope of $(-\frac{b_{ijt}^{(1)}}{2b_{iit}^{(1)}})$. In the literature this has also been referred to as the "short run market vulnerability".[33] This measures the damage that can be done to a firm by short-run market action of an opponent. Recall that product differentiation in this analysis arises from location in geographical space. Therefore, it is assumed that these ratios will depend on some measure of distance. In the analysis to follow I assume that

$$-\frac{b_{ijt}^{(1)}}{2b_{iit}^{(1)}} = \lambda_t d_{ij},$$

[33] See Shubik (1959).

where λ_t is the parameter to be estimated and d_{ij} is a dummy variable which takes a positive value if i and $j-th$ cross-sectional units are neighbors by some measure of closeness and otherwise it is zero.

Similarly $-\frac{b^{(3)}_{ijt}}{4b^{(1)}_{iit}}$ measures the slope of the reaction curve with respect to seller $j's$ product characteristic, y_{jt}. As before

$$\begin{aligned} -\frac{b^{(3)}_{ijt}}{4b^{(1)}_{iit}} &= \eta_t d^{(1)}_{ij}, \qquad i \neq j \\ -\frac{b^{(3)}_{iit}}{4b^{(1)}_{iit}} &= \delta_t, \end{aligned}$$

where η_t and δ_t are the parameters to be estimated and $d^{(1)}_{ij}$ is a dummy variable which takes a positive value if i and $j-th$ cross-sectional units are neighbors by some measure of closeness and otherwise it is zero. In the model, variable y_{jt}, for $j = 1, \ldots, N$, define the product characteristics. In our analysis of gasoline industry, product of each terminal is uniquely characterized by the location of the terminal. The buyers will have an optimal choice of location to purchase the product based on price and location of each seller. In this analysis the distributors purchase gasoline from suppliers at terminals and resell it to retail outlets. Therefore, I assume that y_{jt} reflect factors that affect demand for retail gasoline in the region where terminal j is located. Furthermore, I assume that $d_{ij} = d^{(1)}_{ij}$.

After imposing these restrictions, the model in (49) reduces to

$$p_{it} = a_t + \lambda_t \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{ij} p_{jt} + \delta_t y_{it} + \eta_t \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{ij} y_{jt} + \sum_{h=1}^{H} \beta_{ht} c_{it,h} + u_{it},\ i = 1, \ldots, N, \tag{50}$$

where $\beta_{ht} = \frac{1}{2}\gamma_{ht}$ and u_{it} is the disturbance term that captures the unexplained factors that affect the prices. Next we stack the model and get

$$p_N(t) = X_N(t)\beta_t + \lambda_t D_N p_N(t) + u_N(t), \tag{51}$$

where $p_N(t) = (p_{1t}, \ldots, p_{Nt})'$, $X_N(t) = [\mathbf{k}, y_N(t), D_N y_N(t), C_N(t)]_{N\times(3+H)}$ is the $N \times (3 + H)$ matrix of observations on $(3 + H)$ exogenous variables, $\mathbf{k}$ is the vector of constants which is identical for each cross sectional unit, $\beta_t = [a_t, \delta_t, \eta_t, \beta_{1t}, \ldots, \beta_{Ht}]'$ is $(3+H)\times 1$ vector of regression parameters, $y_N(t) = (y_{1t}, \ldots, y_{Nt})'$, $C_N(t) = [c_{N,1}(t), \ldots, c_{N,H}(t)]_{N\times H}$, $c_{N,h}(t) = (c_{1t,h}, \ldots, c_{Nt,h})$ for $h = 1, \ldots, H$, D_N is an $N \times N$ weighting matrix whose $(i, j) - th$ element is d_{ij}, u_t is the $N \times 1$ vector of disturbances. Note that the parameters of the model are time dependent. Therefore, this allows for the possibility of different parameters in different time periods. However, I assume that the model specification remains the same for every time period. In other words the agents play the same game in each time period.

The disturbance term captures the effect of unobserved demand and cost factors. I allow these to be spatially correlated. However, it is assumed that the disturbance term has a zero mean, finite variance and finite fourth moments. Furthermore, I also assume that the disturbance term is independent of the explanatory variables, $X_N(t)$. More specifically, $E(u_N(t)|X_N(t)) = 0$.

3.4 Estimation

The model is a first order autoregressive spatial model which is a variation of the model considered by Whittle (1954) and Cliff and Ord (1973, 1981). Recall the model in (51)

$$p_N(t) = X_N(t)\beta_t + \lambda_t D_N p_N(t) + u_N(t). \tag{52}$$

I also consider spatial correlation in the disturbance term. More specifically,

$$u_N(t) = \rho_t D_N u_N(t) + \varepsilon_N(t). \tag{53}$$

where ρ_t is the spatially autoregressive parameter in the disturbance term and $\varepsilon_N(t)$ is an $N \times 1$ vector of innovations. In order to estimate the parameters of the model for each cross-section, I use a generalized spatial two-stage least squares (GS2SLS) procedure suggested in Kelejian and Prucha (1998). The advantages of using this procedure over the conventional estimation procedure, maximum likelihood estimator, are (a) computationally feasible for large samples, (b) the results are not based on the assumption that the disturbance term is normally distributed. Note that for each time period, t, I estimate the parameters of the model separately. I am assuming that there is a structural change in parameters in different time periods. It will become evident later from our results that the vector of coefficients is different for different time periods.[34]

In the following it will prove useful to rewrite the model in (52) and (53) in a more compact form as

$$\begin{aligned} p_N(t) &= Z_N(t)\theta_t + u_N(t), \\ u_N(t) &= \rho_t D_N u_N(t) + \varepsilon_N(t), \end{aligned} \tag{54}$$

where $Z_N(t) = [X_N(t), D_N p_N(t)]$ and $\theta_t = [\beta_t', \lambda_t]'$. Furthermore, ρ_t is unknown and is, therefore, estimated.

Kelejian and Prucha (1998) suggest a three step procedure for estimation of unknown parameters in the model in (54).[35] In the first step, the regression model in (52) is estimated by two-stage least squares (2SLS) using the instruments $H_N(t)$. For instruments they suggest a subset of linearly independent columns of $(X_N(t), D_N X_N(t), D_N^2 X_N(t), \ldots)$ where the subset at

[34] If there was no structural change in the parameters for different time periods then we could estimate the parameters of the model by using an estimation startegy suggested in Kapoor, Kelejian and Prucha (2002).

[35] For rigorous proof of consistenecy, large sample properties refer to Kelejian and Prucha (1998, 2001).

least contains the linearly independent columns of $(X_N(t), D_N X_N(t))$. The resulting 2SLS estimator is as follows:

$$\widetilde{\theta}_{t,N} = [\widehat{Z}_N(t)'\widehat{Z}_N(t)]^{-1}\widehat{Z}_N(t)'p_N(t), \tag{55}$$

where $\widehat{Z}_N(t) = P_{H_N}(t)Z_N(t) = (X_N(t), \widehat{D_N p_N(t)})$, where $\widehat{D_N p_N(t)} = P_{H_N}(t)D_N p_N(t)$ and $P_{H_N}(t) = H_N(t)[H_N(t)'H_N(t)]^{-1}H_N(t)'$.

Before proceeding with estimation of spatial autoregressive parameter, ρ_t, I take the residuals from the first step of the estimation to test whether the disturbance term is spatially correlated. I use the Moran I statistic suggested in Kelejian and Prucha (2001). In order to test the null hypothesis of zero spatial correlation in the disturbance, the following Moran I statistic is constructed:

$$I_N(t) = \frac{Q_N^*(t)}{\widetilde{\sigma}_{Q_N^*}(t)} \xrightarrow{D} N(0,1), \tag{56}$$

where $Q_N^*(t) = \widehat{u}_N(t)'D_N\widehat{u}_N(t)$, with $\widehat{u}_N(t) = p_N(t) - Z_N(t)\widetilde{\theta}_t$ and $\widetilde{\sigma}_{Q_N^*}(t)$ is a normalizing factor. Kelejian and Prucha (2001) specify the normalizing factor as

$$\widetilde{\sigma}^2_{Q_N^*}(t) = \widehat{\sigma}^4_N(t)tr(D'_N D_N + D_N D_N) + \widehat{\sigma}^2_N(t)\widehat{b}'_N(t)\widehat{b}_N(t),$$

where $\widehat{\sigma}^2_N(t) = N^{-1}\widehat{u}_N(t)'\widehat{u}_N(t)$, $\widehat{b}_N(t) = -H_N(t)P_N(t)'\widehat{d}_N(t)$ with $\widehat{d}_N(t)' = N^{-1}\widehat{u}_N(t)'(D'_N + D_N)Z_N(t)$ and where

$$P_N(t) = [N^{-1}\widehat{Z}_N(t)'\widehat{Z}_N(t)]^{-1}N^{-1}Z_N(t)'H_N(t)[N^{-1}H_N(t)'H_N(t)]^{-1}.$$

If the null hypothesis is rejected then I proceed to the second step of the estimation procedure. In the second step ρ_t and $\sigma^2_{\varepsilon,t}$ are estimated, where ρ_t is the spatial autoregressive parameter in the disturbances and $\sigma^2_{\varepsilon,t}$ is the variance of the innovation term $\varepsilon_N(t)$. The second step estimators of ρ_t and $\sigma^2_{\varepsilon,t}$, say $\widetilde{\rho}_t$ and $\widetilde{\sigma}^2_{\varepsilon,t}$, are nonlinear least squares estimator defined as minimizers of

$$\left[g_N(t) - G_N(t)\begin{bmatrix}\rho_t \\ \rho_t^2 \\ \sigma^2_{\varepsilon,t}\end{bmatrix}\right]'\left[g_N(t) - G_N(t)\begin{bmatrix}\rho_t \\ \rho_t^2 \\ \sigma^2_{\varepsilon,t}\end{bmatrix}\right], \qquad (57)$$

where

$$G_N(t) = \frac{1}{N}\begin{bmatrix} 2\widehat{u}_N(t)'\widehat{v}_N(t) & -\widehat{v}_N(t)'\widehat{v}_N(t) & N \\ 2\widehat{w}_N(t)'\widehat{v}_N(t) & -\widehat{w}_N(t)'\widehat{w}_N(t) & tr(D_N'D_N) \\ \begin{matrix}(\widehat{u}_N(t)'\widehat{w}_N(t)+ \\ \widehat{v}_N(t)'\widehat{v}_N(t))\end{matrix} & -\widehat{v}_N(t)'\widehat{w}_N(t) & 0 \end{bmatrix}_{3\times 3},$$

$$g_N(t) = \frac{1}{N}\begin{bmatrix} \widehat{u}_N(t)'\widehat{u}_N(t) \\ \widehat{v}_N(t)'\widehat{v}_N(t) \\ \widehat{u}_N(t)'\widehat{v}_N(t) \end{bmatrix}_{3\times 1},$$

where $\widehat{v}_N(t) = D_N\widehat{u}_N(t)$ and $\widehat{w}_N(t) = D_N\widehat{v}_N(t)$.

In the third step of the procedure a Cochrane-Orcutt type transformation is applied to the model in (54). More specifically,

$$p_{N*}(t) = Z_{N*}(t)\theta_t + \varepsilon_N(t), \tag{58}$$

where $p_{N*}(t) = p_N(t) - \rho_t D_N p_N(t)$, $Z_{N*}(t) = Z_N(t) - \rho_t D_N Z_N(t)$ and $\varepsilon_N(t) = u_N(t) - \rho_t D_N u_N(t)$. Since ρ_t is unknown we replace it with its estimate $\widetilde{\rho}_t$ defined in (57) and estimate the model in (58) using 2SLS. The resulting estimator is termed as the feasible GS2SLS and is given by

$$\widetilde{\theta}_{tF,N} = [\widehat{Z}_{N*}(t)'\widehat{Z}_{N*}(t)]^{-1}\widehat{Z}_{N*}(t)'\widehat{p}_{N*}(t), \tag{59}$$

where $\widehat{Z}_{N*}(t) = P_{H_N}(t)[Z_N(t) - \widetilde{\rho}_t D_N Z_N(t)]$, $\widehat{p}_{N*}(t) = p_N(t) - \widetilde{\rho}_t D_N p_N(t)$.

3.5 Data

I use data on weekly average unbranded gasoline prices, cost factors (wages, average net earnings, price of crude oil), demand factors (population, per capita personal income) as well as data on market structure (spot markets for gasoline and percentage change in stocks of gasoline). The data used is for two time periods, a high demand period (third week of August 1999) and a low demand period (third week of January 2000).

There are 289 wholesale rack locations or terminals in the US which sold unbranded gasoline for third week of August 1999 and third week of January 2000. I have included only those terminals which are located in the mainland

and have excluded those that did not sell unbranded gasoline in the two time periods under consideration. There are 238 terminals located in metropolitan statistical areas (MSAs). For the terminals which are not in MSAs, I identify the zipcode and locate the county in which these terminals are situated using Geographic Information System (GIS) software.

Data on terminal prices was obtained from Oil Price Information Service (OPIS), a private data collection agency. Price data includes regular unbranded gasoline prices charged to the distributors at the terminal. The prices are denoted by $PR99$ for August 1999 and $PR00$ for January 2000. The prices are in cents/gallon.

Population data for the regions where terminals are located was obtained from two sources. For terminals located in MSAs the data was obtained from Census and for the remaining terminals it was obtained from Regional Economic Information System (REIS). In the analysis I use the log of population and denote this by $POP99$ for August 1999 and $POP00$ for January 2000.

Data on per capita personal income and per capita net earnings for regions where terminals are located was also obtained from two sources. For terminals located in MSAs data from Bureau of Labor Statistics (BLS) is used and for the remaining terminals I use data from REIS. Per capita personal income is denoted by $INC99$ for August 1999 and $INC00$ for January 2000. Per capita net earnings is denoted by $EA99$ for August 1999 and $EA00$ for January 2000. Both income and earnings are in 10^3 dollars.

Price of crude oil is for the entire US and was obtained from Energy Information Administration (EIA). This is the same for all terminals and therefore is treated as a constant. It is denoted by $CRPR99$ for August 1999 and $CRPR00$ for January 2000. These are in cents/gallon.

In order to capture the effects of market structure I include the spot markets for gasoline. There are seven spot markets in the US. They are located in New York, Gulf Coast, Midwest, Chicago, Los Angeles, San Francisco and Northwest. The data on spot prices was obtained from OPIS and EIA. The spot price for terminal i is the price that prevailed in the spot market closest to terminal i. The spot prices are denoted by $SPOT99$ for August 1999 and $SPOT00$ for January 2000. They are in cents/gallon.

Changes in stocks of gasoline are a measure of imbalances in demand and supply. The data on stock is available from EIA for each of the PADDs. Moreover, PADD1 (East coast) is further subdivided into three broad regions and data on stocks is available for them as well. I compute the percentage change in stocks as a difference in stock between the third and the second week divided by the stock in second week times 100. I do this for both August 1999 and January 2000. This variable is denoted by $PERST99$ for August 1999 and $PERST00$ for January 2000.

In order to capture intra-terminal competition I include the number of suppliers at each terminal. Data on this was obtained from OPIS. This is denoted by $CO99$ for August 1999 and $CO00$ for January 2000.

In order to capture broad regional differences, I introduce dummy variable for each PADDs. There are 5 PADDs and I denote the dummy variables as $PADD_i$, where $i = 1, \ldots, 5$. These dummy variables capture the broad regional differences in demand and supply that have been discussed before.

Most importantly, I discuss the construction of the weighting matrix, D_N. The weighting matrix is the measure of closeness between terminals. Theoretical models in this area have suggested measures of closeness like: terminals that are nearest to each other, terminal that share a common market boundary, that share a market boundary with a third competitor etc. These measures could be endogenously or exogenously determined. In this analysis I will focus only on the first measure of closeness, that is, terminals that are nearest to each other, geographically.

I use the GIS software to construct a weighting matrix of nearest neighbor.

I identify each terminal with a zipcode and then use the software to locate the nearest neighbor for each terminal. There are two ways to identify the nearest neighbor for each terminal. One way is to use a Euclidean distance and the second way is to look at actual road distance between terminals. I find that for some, the nearest terminal is different when I use a Euclidean distance measure rather than actual road distance. For example, consider a terminal in Artesia in New Mexico, the nearest terminal by a Euclidean distance measure is in El Paso, Texas whereas using the actual road distance the nearest terminal is in Midland, Texas. This has important implications in the model specification and the consistency of estimates. Misspecified neighbor for any terminal will make the estimation inconsistent. In this paper I consider both measures to compute the nearest terminal.

More specifically, the $(i, j)-th$ element of the weighting matrix is denoted by d_{ij} and is a dummy variable which takes a positive value if terminal j is nearest to terminal to i and otherwise it is 0. Weighting matrices are not symmetric, that is, if terminal j is the nearest to terminal i then it need not be the case that terminal i is the nearest to terminal j.

3.6 Empirical Model

The econometric model to be estimated is specified separately for third week of August 1999 and third week of January 2000. The model is the following:

$$\begin{aligned} PR(t) &= \lambda_{(t)} D_N PR(t) + \alpha_{1,(t)} CRPR(t) + \sum_{i=2}^{5} \alpha_{i,(t)} PADD_{i-1} + \alpha_{6,(t)} CO(t) \\ &\quad + \alpha_{7,(t)} SPOT(t) + \alpha_{8,(t)} PERST(t) + \alpha_{9,(t)} EA(t) + \alpha_{10,(t)} POP(t) \\ &\quad + \alpha_{11,(t)} D_N POP(t) + \alpha_{12,(t)} INC(t) + \alpha_{13,(t)} D_N INC(t) \qquad (60) \\ &\quad + u_{(t)}, \end{aligned}$$

where $(t) = 99$ for third week of August 1999 and $(t) = 00$ for third week of January 2000, D_N is the weighting matrix, hence $D_N PR(t)$, $D_N POP(t)$, $D_N INC(t)$ are the nearest neighbor's price, population and income, respectively, $u_{(t)}$ is the disturbance term and $\lambda_{(t)}$, $\alpha_{s,(t)}$, for $s = 1, \ldots, 13$, are the parameters to be estimated. A number of issues must be addressed in order to develop consistent estimates of the model in (60). Firstly, I need to instrument for the nearest neighbor's price, $D_N PR(t)$ which is an endogenous variable.[36] As instruments for this variable, I use exogenous explanatory variables for nearest terminal. I also use the exogenous explanatory variables of terminal nearest to the nearest terminal. I then estimate the model using

[36] I assume that all explanatory variables other than $D_N PR(t)$ are exogenous, that is, they are independent of the disturbances, $u_{(t)}$.

two-stage least squares (2SLS) procedure.[37] Secondly, I correct for spatial correlation in the disturbances. More specifically,

$$u_{(t)} = \rho_{(t)} D_N u_{(t)} + \varepsilon_{(t)}, \tag{61}$$

where, $\rho_{(t)}$ is the spatial autoregressive parameter. Using the estimation strategy described in previous section, I estimate the spatial autoregressive parameter, $\rho_{(t)}$, in (61). Then I use the estimate of $\rho_{(t)}$, say $\widetilde{\rho}_{(t)}$, and apply

[37]In choosing instruments for estimation I follow Kelejian and Prucha (1999). In presence of spatial lag in the dependent variable, the instruments they suggest are linearly independent columns of the own explanatory variables, nearest neighbors explanatory variables and nearest to nearest neighbor's explanatory variables. In our model, for each terminal we have constructed the nearest neighbor. However, this is not symmentric, for example, consider terminal A whose nearest neighbor is terminal B, then for terminal B the nearest neighbor could be terminal C and not terminal A. In our analysis terminal A competes directly for customer with terminal B, while terminal B competes directly for customers with terminal C. Clearly, the prices set in terminal A will depend on terminal B, therefore, as instruments for prices of terminal B we could choose exogenous characteristics of terminal C which have direct effect on the prices at terminal B but not on prices of terminal A.

the Cochrane-Orcutt type transformation on (60) and get

$$
\begin{aligned}
(I_N - \widetilde{\rho}_{(t)} D_N) PR(t) \;=\;& \lambda_{(t)} (I_N - \widetilde{\rho}_{(t)} D_N) D_N PR(t) \\
& + \alpha_{1,(t)} (I_N - \widetilde{\rho}_{(t)} D_N) CRPR(t) \\
& + \sum_{i=2}^{5} \alpha_{i,(t)} (I_N - \widetilde{\rho}_{(t)} D_N) PADD_{i-1} \\
& + \alpha_{6,(t)} (I_N - \widetilde{\rho}_{(t)} D_N) CO(t) \\
& + \alpha_{7,(t)} (I_N - \widetilde{\rho}_{(t)} D_N) SPOT(t) \\
& + \alpha_{8,(t)} (I_N - \widetilde{\rho}_{(t)} D_N) PERST(t) \\
& + \alpha_{9,(t)} (I_N - \widetilde{\rho}_{(t)} D_N) EA(t) \\
& + \alpha_{10,(t)} (I_N - \widetilde{\rho}_{(t)} D_N) POP(t) \\
& + \alpha_{11,(t)} D_N (I_N - \widetilde{\rho}_{(t)} D_N) POP(t) \\
& + \alpha_{12,(t)} (I_N - \widetilde{\rho}_{(t)} D_N) INC(t) \\
& + \alpha_{13,(t)} (I_N - \widetilde{\rho}_{(t)} D_N) D_N INC(t) + innovation\ term,
\end{aligned} \tag{62}
$$

where, I_N is an identity matrix. I again instrument for the nearest neighbor's price, $D_N PR(t)$ which is an endogenous variable. I then estimate the transformed model in (62) using two-stage least squares (2SLS) procedure.

One of the main innovations of this paper is to look at the role of marginal customer in price determination at a terminal. Spatial models of product differentiation crucially rest on the characteristics of the marginal customer.

In models with homogenous customers, where each customer has the same demand, pricing decision of a terminal is independent of the level of demand. This however would change when we look at heterogenous customers. In the appendix, I formally develop a theoretical model to show how pricing decision of a terminal depends on the level of demand of the marginal customer. In particular I show that price of a terminal is negatively related to the level of demand of the marginal customer. Empirically, the marginal customer is extremely difficult to identify. In order to address this issue, I introduce a variable that captures the characteristics of the marginal customer and is easy to compute. More specifically, I incorporate the per capita income of the neighboring region. This is a good proxy for the marginal customer under the assumption that terminals are competing for customers.

Another innovation of this paper is to look at two time periods with different demand intensities. In periods of high demand, one expects regional variations in demand to be higher. This would in turn imply that customers in different regions can be treated as being heterogenous in terms of demand. The spatial competition model that I have developed with heterogenous customers is the relevant model for periods of high demand. Whereas in a low demand period, the regional variation in demand is lower. In this situation, the appropriate model is the existing spatial competition model with homogenous customers. In our empirical analysis we nest both these models by incorporating the characteristics of the marginal customer in both time periods. The model predicts that in periods of high demand, price in a given terminal is positively related to price of neighboring terminal and is negatively related to the level of demand of the marginal customer. Whereas, in periods of low demand, prices are positively related to prices of the neighboring terminal and are not related to level of demand of the marginal customer.

3.7 Results

I report the results in four tables, given that the analysis is for two time periods and two weighting matrices[38]

	Actual Road Distance	*Euclidean Distance*
High Demand (August 1999)	Table 1	Table 3
Low Demand (January 2000)	Table 2	Table 4

Each table contains, first, the estimates from ordinary least squares (OLS) and 2SLS in columns 1 and 2, respectively. Columns 3 and 4 report the OLS and 2SLS estimates after correcting for spatial correlation in disturbances.[39]

[38] Tables are at the end of the appendix.

[39] Wald test rejects the hypothesis that the same coefficient vector applies in the two time periods.

Looking at these results, we notice several interesting features. One general fact being that extent of spatial correlation in prices between terminals is significant and positive for both high and low demand periods. This result confirms the prediction of spatial price competition model. More specifically, the coefficient on price of the nearest terminal is positive and significant for both specifications of the weighting matrix as well as both time periods. As noted earlier this coefficient is the measure of price competition. Comparing the two periods we notice that the extent of competition is much higher in period of low demand (0.84) as compared to high demand (0.58).[40]

One must observe some caution in interpreting the above result. Positive correlation in prices is a strong indicator of price competition, however, it is plausible to think of situations where prices move together even in the absence of competition.[41] This fear can be put to rest as the presence of price competition between terminals is confirmed by the finding that prices at a terminal are inversely affected by income in the neighboring region during periods of high demand. This result is predicted by my model of spatial prod-

[40] This observation is based on comparing column 2 of Table 1 with column 4 of Table 2. The reason for this comparison being that in the high demand period, there is no spatial correlation in disturbances and therefore the meaningful estimate is the uncorrected 2SLS, as against the low demand period, when there is spatial correlation in the disturbances, therefore we look at the corrected 2SLS.

[41] We could observe positive correlation in prices even when firms collude in their pricing decision.

uct differentiation with heterogenous customers (appendix). The intuition for this result is easy to grasp. During periods of high demand, we observe greater regional variation in demand. If a terminal competes for customers in neighboring region then higher income in that region implies greater demand, which then means that the terminal should lower its price to attract some of the high demand customers from the neighboring region. During low demand periods, however, it is reasonable to assume low regional variation in demand therefore customers can be treated as homogenous. Spatial price competition model with homogenous customers predicts that pricing decision of a terminal is independent of demand in the neighboring region. These predictions are driven by the central role played by the marginal customer in price determination.

In the estimation strategy, I account for the characteristics of the marginal customer by incorporating per capita income in the neighboring region of each terminal. The estimation results confirm the predictions stated above as we observe that during the high demand period, the coefficient on neighbor's income ($D_N INC99$) is negative and significant (-0.11) and during the low demand period the coefficient on ($D_N INC00$) is insignificant (-0.02).

Next we compare the two specifications of the weighting matrix. On comparing results, we notice that using a Euclidean measure of distance underestimates the extent of correlation in prices between terminals in both periods. This is as expected because actual road distance captures the true neighbor for a terminal while a Euclidean measure leads to misspecifications.

My estimation strategy allows for spatial correlation in disturbances. From the results, it is clear that not correcting for spatial correlation in disturbances leads to inaccurate inferences. In particular, we compare column 2 (uncorrected 2SLS) and 4 (corrected 2SLS) of Table 2. which analyses the data for the low demand period with actual road distance specification of the weighting matrix. The Moran I statistic which tests for spatial correlation in disturbances, rejects the hypothesis of zero spatial correlation. Not correcting for this leads to biased estimation of standard errors which in turn leads to faulty conclusions. The results of Table 2 confirm this.

3.8 Robustness Test

In our analysis the $(i, j)-th$ element of weighting matrix D_N, d_{ij} is a dummy variable which is positive if the $j-th$ terminal is the nearest to the $i-th$ terminal and zero otherwise. However, one limitation of this weighting matrix is that it does not account for the actual road distance between nearest terminals. For example, the nearest terminal to a terminal in Rapid City, South Dakota is in Sidney, Nebraska which is 240 miles, and for terminal in Aberdeen, South Dakota the nearest terminal is 77 miles away in Wolsey, South Dakota. Irrespective of the distance the weighting matrix in both cases assigns an equal weight of 1 to the nearest terminal. I overcome this limitation by constructing different weighting matrices which are dependent on the distance between the nearest terminals. In particular, I construct five weighting matrices, $D_N^{(i)}$ where $i = 1, \ldots, 5$. Let $d_{ij}^{(i)}$ be the $(i, j) - th$ element of matrix $D_N^{(i)}$, where $i = 1, \ldots, 5$. More specifically:

$$\begin{aligned} d_{ij}^{(1)} &= 1/n, \quad \text{if } j \text{ is the nearest terminal to } i \text{ and } 0 \leq x < 30, \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

$$\begin{aligned} d_{ij}^{(2)} &= 1/n, \quad \text{if } j \text{ is the nearest terminal to } i \text{ and } 30 \leq x < 70, \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

$$\begin{aligned} d_{ij}^{(3)} &= 1/n, \quad \text{if } j \text{ is the nearest terminal to } i \text{ and } 70 \leq x < 110, \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

$$\begin{aligned} d_{ij}^{(4)} &= 1/n, \quad \text{if } j \text{ is the nearest terminal to } i \text{ and } 110 \leq x < 150, \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

$$
\begin{aligned}
d_{ij}^{(5)} &= 1/n, \quad \text{if } j \text{ is the nearest terminal to } i \text{ and } 150 \leq x, \\
&= 0 \quad \text{otherwise.}
\end{aligned}
$$

where n is the number of terminals which are nearest to terminal i and x is the actual road distance between terminals.

In order to study the effect of distance on the nature and extent price competition I construct the following empirical model

$$
\begin{aligned}
PR(t) &= \sum_{i=1}^{5} \lambda_{(t)}^{(i)} D_N^{(i)} PR(t) + \alpha_{1,(t)} CRPR(t) + \sum_{i=2}^{5} \alpha_{i,(t)} PADD_{i-1} \\
&+ \alpha_{6,(t)} CO(t) + \alpha_{7,(t)} SPOT(t) + \alpha_{8,(t)} PERST(t) \\
&+ \alpha_{9,(t)} EA(t) + \alpha_{10,(t)} POP(t) + \sum_{i=1}^{5} \alpha_{11,(t)}^{(i)} D_N^{(i)} POP(t) \qquad (63) \\
&+ \alpha_{12,(t)} INC(t) + \sum_{i=1}^{5} \alpha_{13,(t)}^{(i)} D_N^{(i)} INC(t) + u_{(t)},
\end{aligned}
$$

where $(t) = 99$ for third week of August 1999 and $(t) = 00$ for third week of January 2000.

The results reported in Table 5 in the appendix are striking. The results indicate that for third week of August 1999, $\lambda_{99}^{(i)}$, for $i = 1, \ldots, 5$ are almost identical. I also find similar results for third week of January 2000. In other words, it appears from the result that the nature and extent of competition between terminals is not influenced by the actual road distance between the terminals. This exercise proves the robustness of the result, that is, accounting or not accounting for the actual road distance between the nearest terminal has little impact on the nature and extent of competition between terminals.

We have also provided another test of robustness of our results. In addition to the variables considered above we have included taxes on wholesale gasoline charged by each state as one of the explanatory variables. The results (reported in Tables 6 and 7) indicate that inclusion of this explanatory variable does not change the main conclusion of the above analysis. In particular, we find that nature and extent of competition is less during high demand period as compared to the low demand period, these results are similar to the analysis without taxes. Furthermore, in both periods taxes have a positive and a significant impact on the prices. Moreover, we still find that characteristics of the marginal customer plays an important role during the high demand period when there is a high regional variation in demand and is insignificant during the low demand period. One possible explanation for why the results are not sensitive to inclusion of this variable is that in the analysis we are using transaction prices. These transaction prices are inclusive of the tax rates, hence, tax rates are significant in affecting the prices but do not affect the results on the nature and extent of price competition between terminals.

3.9 Conclusion and Further Extensions

In this paper, while analyzing price competition in the US wholesale gasoline industry, I empirically distinguish between spatial model and monopolistic competition model of product differentiation. While product differentiation, within the industry arises from different sources such as brand names, quality (regular, premium etc.), the location of distribution facilities is also an important dimension of differentiation. There are three main innovations in this paper. Firstly, I look at the role of marginal customer in price determination in spatial models of product differentiation. These models crucially rest on the characteristics of the marginal customer. In models with homogenous customers, pricing decision of a terminal is independent of the level of demand. This however changes when one considers heterogenous customers. I formally develop a theoretical model in the appendix to show that in the presence of heterogenous customers, pricing decision of a terminal depends on the level of demand of the marginal customer. In particular I show that price of a terminal is negatively related to the level of demand of the marginal customer. I introduce a variable that captures the characteristics of the marginal customer and is easy to compute. Secondly, I consider two time periods with different demand intensities. In a period of high demand, one observes regional variations in demand to be higher whereas in a low demand period, the regional variation in demand is lower. The difference in demand intensities in different time periods has important empirical implications. In particular during the high demand period the relevant spatial model of product differentiation is the one in which the customers are heterogenous in their demand levels. In such a scenario the model in the appendix shows that prices across terminals are correlated and are negatively related to the level of demand of the marginal customer. Whereas, in periods of low demand, the regional variation in demand is lower. Therefore, the relevant spatial model of product differentiation is the one in which the customers are homogenous in their demand levels. In such a scenario prices are correlated, but are not affected by the level of demand of the marginal customer. Thirdly, I use an estimation strategy that allows for spatial correlation in the explanatory variables and in the unobserved factors that affect prices, in addition to spatial correlation in the dependent variable that has been considered in previous work.

The main results from the estimation confirm all the above predictions of spatial models of product differentiation. More specifically, I find the extent of price competition between terminals to be significant and positive for both high and low demand periods. On comparing the two periods, I observe the extent of competition to be much higher in period of low demand as compared to high demand period. Secondly, the results also confirm the predictions of spatial model of product differentiation with heterogenous customers and also with homogenous customers. In particular, the results show that during the period of high demand (August 1999) when the regional variation is high, prices are significantly and negatively affected by neighbor's income (which is a measure of level of demand of the marginal customer). Whereas during the period of low demand (January 2000) when the regional variation in demand is lower, prices are independent of neighbor's income. Thirdly, the results reveal a high and significant spatial correlation in the unobserved factors that affect prices during the low demand period. Not correcting for which leads to inaccurate inferences. Lastly, I have estimated my model using two measures of distance between neighboring terminals, actual road distance and a Euclidean distance. I find that using a Euclidean distance measure as compared to actual road distance underestimates the extent of competition between terminals.

This research is relevant from policy perspective. Given the strategic relevance of gasoline for the US economy, it is important to understand what factors determine and influence its price. This paper highlights the significance of competition among distribution facilities as a major factor and also that the wholesale gasoline markets are geographically segmented and small. Furthermore, in studying the unbranded gasoline market we have addressed an important issue which is being considered by policy makers, as was highlighted by Hastings in the hearing before the Committee on Government Affairs, US Senate, May 2002. He emphasized that the unbranded gasoline market is necessary to ensure sufficient unbranded gasoline supply at competitive prices, which in turn is crucial for the entry and survival of independent retailers including new chains such as Kmart, Walmart, Costco and RaceTrac. These independent retailers are important because they increase competition at the retail level. For this reason it is important to understand the nature and extent of competition within the unbranded gasoline industry.

There is an interesting extension to this work which is underway. Throughout this paper I have assumed each terminal as a single firm. This assumption puts a restriction on the market structure, as typically at a terminal, more than one suppliers compete for consumers.[42] Furthermore, in many cases a

[42] On an average there are six suppliers per terminal.

supplier is present at both the neighboring terminals. In extending this paper I plan to formally incorporate this market structure and develop a model that can account for both inter as well as intra terminal competition. Availability of data on suppliers at each terminal has made it feasible to address these more interesting issues. Using this data I also plan to study the effect of mergers and acquisitions within the gasoline industry on prices which is very relevant for policy issues related to antitrust.

4 Appendix to Chapter 2

In this appendix we will make use of the following definitions.[43]

Definition 1. A sequence $\{a_N\}_{N=1}^{\infty}$ is at most of order N^{δ}, and is written as $a_N = O(N^{\delta})$, if there exists a real number $M > 0$, such that

$$N^{-\delta} |a_N| \leq M.$$

Definition 2. A sequence $\{a_N\}_{N=1}^{\infty}$ is of smaller order than N^{δ}, and is written as $a_N = o(N^{\delta})$, if

$$\lim_{N\to\infty} N^{-\delta} |a_N| = 0.$$

Definition 3. A sequence of random variables $\{c_N\}_{N=1}^{\infty}$ is at most of order N^{δ} in probability, and is written as $c_N = Op(N^{\delta})$, if for every $\epsilon > 0$ there exists a real number $M > 0$, such that

$$P\{N^{-\delta} |c_N| \geq M\} \leq \epsilon.$$

Definition 4. A sequence of random variables $\{c_N\}_{N=1}^{\infty}$ is smaller order than N^{δ} in probability, and is written as $c_N = o_p(N^{\delta})$, if

$$p \lim N^{-\delta} c_N = 0.$$

[43] These definitions are from Judge and et.al., Pages 145-148.

These definitions extend to vectors and matrices if the conditions hold for every element in the vector or matrix.

In the following we derive the six moment conditions described in (34). In light of (6) and the definition of $Q_{0,N}$ in (8) we note that $Q_{0,N}\varepsilon_N = Q_{0,N}v_N$. Specifically,

$$\begin{aligned} Q_{0,N}\varepsilon_N &= Q_{0,N}(e_T \otimes I_N)\mu_N + Q_{0,N}v_N \\ &= ((I_T - \frac{J_T}{T})e_T \otimes I_N)\mu_N + Q_{0,N}v_N \\ &= ((e_T - \frac{J_T e_T}{T}) \otimes I_N)\mu_N + Q_{0,N}v_N \\ &= ((e_T - \frac{e_T e_T' e_T}{T}) \otimes I_N)\mu_N + Q_{0,N}v_N \\ &= ((e_T - e_T) \otimes I_N)\mu_N + Q_{0,N}v_N \\ &= Q_{0,N}v_N. \end{aligned} \tag{A.1}$$

Furthermore, by the definition of $\overline{\varepsilon}_N$ in (33) and from (11) we note that

$$\begin{aligned} Q_{0,N}\overline{\varepsilon}_N &= Q_{0,N}(I_T \otimes W_N)\varepsilon_N \\ &= (I_T \otimes W_N)Q_{0,N}\varepsilon_N \\ &= (I_T \otimes W_N)Q_{0,N}v_N. \end{aligned} \tag{A.2}$$

Given Assumptions 1 and 2, (A.1), (A.2), and by using (9), (10)and (11) we have

$$
\begin{aligned}
E\frac{1}{N(T-1)}\varepsilon_N^{'}Q_{0,N}\varepsilon_N &= E\frac{1}{N(T-1)}v_N^{'}Q_{0,N}v_N \\
&= \frac{1}{N(T-1)}\sigma_v^2 tr(Q_{0,N}) \qquad\qquad (A.3)\\
&= \frac{1}{N(T-1)}\sigma_v^2 N(T-1) \\
&= \sigma_v^2.
\end{aligned}
$$

$$
\begin{aligned}
E\frac{1}{N(T-1)}\overline{\varepsilon}_N^{'}Q_{0,N}\overline{\varepsilon}_N &= E\frac{1}{N(T-1)}v_N^{'}Q_{0,N}(I_T\otimes W_N^{'}W_N)Q_{0,N}v_N \\
&= \sigma_v^2\frac{1}{N(T-1)}tr\left[Q_{0,N}(I_T\otimes W_N^{'}W_N)\right] \\
&= \sigma_v^2\frac{1}{N(T-1)}tr\left[(I_T-\frac{J_T}{T})\otimes W_N^{'}W_N\right] \qquad (A.4)\\
&= \sigma_v^2\frac{1}{N(T-1)}tr(I_T-\frac{J_T}{T})tr(W_N^{'}W_N) \\
&= \sigma_v^2 N^{-1}tr(W_N^{'}W_N).
\end{aligned}
$$

$$
\begin{aligned}
E\frac{1}{N(T-1)}\overline{\varepsilon}_N^{'}Q_{0,N}\varepsilon_N &= E\frac{1}{N(T-1)}v_N^{'}Q_{0,N}(I_T\otimes W_N^{'})Q_{0,N}v_N \\
&= \sigma_v^2\frac{1}{N(T-1)}tr\left[Q_{0,N}(I_T\otimes W_N^{'})\right] \\
&= \sigma_v^2\frac{1}{N(T-1)}tr\left[(I_T-\frac{J_T}{T})\otimes W_N^{'}\right] \qquad (A.5)\\
&= \sigma_v^2\frac{1}{N(T-1)}tr(I_T-\frac{J_T}{T})tr(W_N^{'}) \\
&= \sigma_v^2 N^{-1}tr(W_N^{'}) \\
&= 0.
\end{aligned}
$$

In light of (6) and the definition of $Q_{1,N}$ in (8) we note that $Q_{1,N}\varepsilon_N = [e_T \otimes I_N]\mu_N + Q_{1,N}\upsilon_N$. Specifically,

$$\begin{aligned} Q_{1,N}\varepsilon_N &= Q_{1,N}[e_T \otimes I_N]\mu_N + Q_{1,N}\upsilon_N \\ &= (\frac{J_T e_T}{T} \otimes I_N)\mu_N + Q_{1,N}\upsilon_N \\ &= (\frac{e_T e_T' e_T}{T} \otimes I_N)\mu_N + Q_{1,N}\upsilon_N \\ &= (e_T \otimes I_N)\mu_N + Q_{1,N}\upsilon_N. \end{aligned} \tag{A.6}$$

Furthermore, by the definition of $\overline{\varepsilon}_N$ in (33) and from (11) we note that

$$\begin{aligned} Q_{1,N}\overline{\varepsilon}_N &= Q_{1,N}(I_T \otimes W_N)\varepsilon_N \\ &= (I_T \otimes W_N)Q_{1,N}\varepsilon_N \\ &= (I_T \otimes W_N)[(e_T \otimes I_N)\mu_N + Q_{1,N}\upsilon_N] \\ &= (e_T \otimes W_N)\mu_N + (I_T \otimes W_N)Q_{1,N}\upsilon_N. \end{aligned} \tag{A.7}$$

Given Assumptions 1 and 2, (A.6), (A.7), and by using (9), (10), (11) and

$T\sigma_\mu^2 + \sigma_v^2 = \sigma_1^2$, we have

$$\begin{aligned} E\frac{1}{N}\varepsilon_N^{'} Q_{1,N}\varepsilon_N &= E\frac{1}{N}\mu_N^{'}(e_T^{\prime}e_T \otimes I_N)\mu_N + E\frac{1}{N}v_N^{'}Q_{1,N}v_N \\ &= \frac{T}{N}\sigma_\mu^2 tr(I_N) + \frac{1}{N}\sigma_v^2 tr(\frac{J_T}{T})tr(I_N) \\ &= T\sigma_\mu^2 + \sigma_v^2 = \sigma_1^2. \end{aligned} \tag{A.8}$$

$$\begin{aligned} E\frac{1}{N}\overline{\varepsilon}_N^{'} Q_{1,N}\overline{\varepsilon}_N &= E\frac{1}{N}\mu_N^{'}(e_T^{\prime}e_T \otimes W_N^{'}W_N)\mu_N + \\ &\quad E\frac{1}{N}v_N^{'}Q_{1,N}(I_T \otimes W_N^{'}W_N)Q_{1,N}v_N \\ &= \frac{T}{N}\sigma_\mu^2 tr(W_N^{'}W_N) + \sigma_v^2\frac{1}{N}tr\left[Q_{1,N}(I_T \otimes W_N^{'}W_N)\right] \\ &= \frac{T}{N}\sigma_\mu^2 tr(W_N^{'}W_N) + \sigma_v^2\frac{1}{N}tr\left[\frac{J_T}{T} \otimes W_N^{'}W_N\right] \\ &= \frac{T}{N}\sigma_\mu^2 tr(W_N^{'}W_N) + \sigma_v^2\frac{1}{N}tr(\frac{J_T}{T})tr(W_N^{'}W_N) \\ &= \frac{T}{N}\sigma_\mu^2 tr(W_N^{'}W_N) + \sigma_v^2\frac{1}{N}tr(W_N^{'}W_N) \\ &= (T\sigma_\mu^2 + \sigma_v^2)N^{-1}tr(W_N^{'}W_N) \\ &= \sigma_1^2 N^{-1}tr(W_N^{'}W_N). \end{aligned} \tag{A.9}$$

$$\begin{aligned} E\frac{1}{N}\overline{\varepsilon}_N^{'} Q_{1,N}\varepsilon_N &= E\frac{1}{N}\mu_N^{'}(e_T^{\prime}e_T \otimes W_N^{\prime})\mu_N + \\ &\quad E\left[\frac{1}{N}v_N^{'}Q_{1,N}(I_T \otimes W_N^{'})Q_{1,N}v_N\right] \\ &= \frac{T}{N}\sigma_\mu^2 tr(W_N^{'}) + \sigma_v^2\frac{1}{N}tr\left[Q_{1,N}(I_T \otimes W_N^{'})\right] \\ &= \frac{T}{N}\sigma_\mu^2 tr(W_N^{'}) + \sigma_v^2\frac{1}{N}tr\left[\frac{J_T}{T} \otimes W_N^{'}\right] \\ &= \frac{T}{N}\sigma_\mu^2 tr(W_N^{'}) + \sigma_v^2\frac{1}{N}tr(\frac{J_T}{T})tr(W_N^{'}) \\ &= \frac{T}{N}\sigma_\mu^2 tr(W_N^{'}) + \sigma_v^2\frac{1}{N}tr(W_N^{'}) \\ &= 0. \end{aligned} \tag{A.10}$$

Recall from (15) and the definition of $P_N(\rho) = [I_N - \rho W_N]^{-1}$, that

$$\Omega_{u,N}(\rho) = [(\sigma_v^2(I_T - \frac{J_T}{T}) + \sigma_1^2 \frac{J_T}{T}) \otimes P_N(\rho)P_N(\rho)'].$$

Given Assumption 4 we observe that

$$\begin{aligned}(NT)^{-1}\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT} |\omega_{ij,N}(\rho)| &\leq T\sigma_1^2[(NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\Sigma_{k=1}^{N} |p_{ik,N}(\rho)| \, |p_{jk,N}(\rho)|] \\ &= \sigma_1^2[N^{-1}\Sigma_{k=1}^{N}\Sigma_{i=1}^{N} |p_{ik,N}(\rho)| \, \Sigma_{j=1}^{N} |p_{jk,N}(\rho)|] \\ &\leq \sigma_1^2 k_p^2 < \infty, \qquad \text{(A.11)}\end{aligned}$$

where $\omega_{ij,N}(\rho)$ is the $(i,j)-th$ element of $\Omega_{u,N}(\rho)$, and $p_{ik,N}(\rho)$ and $p_{jk,N}(\rho)$ are $(i,k)-th$ and $(j,k)-th$ elements of $P_N(\rho)$, respectively. This proves that $(NT)^{-1}\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT} |\omega_{ij,N}(\rho)|$ is uniformly bounded, thus limiting the degree of correlation of the elements of u_N.

It proves helpful to introduce the following expressions.

Recall from (12), (31), and the definition of $P_N(\rho) = [I_N - \rho W_N]^{-1}$, that

$$\begin{aligned}u_N &= [I_T \otimes (I_N - \rho W_N)^{-1}]\varepsilon_N = [I_T \otimes P_N]\varepsilon_N, \\ \overline{u}_N &= (I_T \otimes W_N)u_N = (I_T \otimes W_N P_N)\varepsilon_N, \qquad \text{(A.12)} \\ \overline{\overline{u}}_N &= (I_T \otimes W_N)\overline{u}_N = (I_T \otimes W_N^2 P_N)\varepsilon_N.\end{aligned}$$

Furthermore, from (33), (A.1), (A.2), (A.6), (A.7) we have

$$
\begin{aligned}
\overline{\varepsilon}_N &= (I_T \otimes W_N)\varepsilon_N, \\
Q_{0,N}\varepsilon_N &= Q_{0,N}v_N, \\
Q_{0,N}\overline{\varepsilon}_N &= (I_T \otimes W_N)Q_{0,N}v_N, \qquad \text{(A.13)} \\
Q_{1,N}\varepsilon_N &= (e_T \otimes I_N)\mu_N + Q_{1,N}v_N, \\
Q_{1,N}\overline{\varepsilon}_N &= (e_T \otimes W_N)\mu_N + (I_T \otimes W_N)Q_{1,N}v_N.
\end{aligned}
$$

Using the expressions in (A.12), (A.13), (8) and (11) we have

$$
\begin{aligned}
Q_{0,N}u_N &= Q_{0,N}[I_T \otimes P_N]\varepsilon_N = [I_T \otimes P_N]Q_{0,N}\varepsilon_N = [I_T \otimes P_N]Q_{0,N}v_N \\
&= \left[(I_T - \frac{J_T}{T}) \otimes P_N\right] v_N, \\
Q_{0,N}\overline{u}_N &= Q_{0,N}(I_T \otimes W_N P_N)\varepsilon_N = (I_T \otimes W_N P_N)Q_{0,N}\varepsilon_N \\
&= (I_T \otimes W_N P_N)Q_{0,N}v_N \\
&= \left[(I_T - \frac{J_T}{T}) \otimes W_N P_N\right] v_N, \\
Q_{0,N}\overline{\overline{u}}_N &= Q_{0,N}(I_T \otimes W_N^2 P_N)\varepsilon_N = (I_T \otimes W_N^2 P_N)Q_{0,N}\varepsilon_N \\
&= (I_T \otimes W_N^2 P_N)Q_{0,N}v_N \\
&= \left[(I_T - \frac{J_T}{T}) \otimes W_N^2 P_N\right] v_N, \qquad (A.14) \\
Q_{1,N}u_N &= Q_{1,N}[I_T \otimes P_N]\varepsilon_N = [I_T \otimes P_N]Q_{1,N}\varepsilon_N \\
&= [e_T \otimes P_N]\mu_N + [I_T \otimes P_N]Q_{1,N}v_N \\
&= [e_T \otimes P_N]\mu_N + \left[\frac{J_T}{T} \otimes P_N\right] v_N, \\
Q_{1,N}\overline{u}_N &= Q_{1,N}(I_T \otimes W_N P_N)\varepsilon_N = (I_T \otimes W_N P_N)Q_{1,N}\varepsilon_N \\
&= [e_T \otimes W_N P_N]\mu_N + [I_T \otimes W_N P_N]Q_{1,N}v_N \\
&= [e_T \otimes W_N P_N]\mu_N + \left[\frac{J_T}{T} \otimes W_N P_N\right] v_N, \\
Q_{1,N}\overline{\overline{u}}_N &= Q_{1,N}(I_T \otimes W_N^2 P_N)\varepsilon_N = (I_T \otimes W_N^2 P_N)Q_{1,N}\varepsilon_N \\
&= [e_T \otimes W_N^2 P_N]\mu_N + [I_T \otimes W_N^2 P_N]Q_{1,N}v_N \\
&= [e_T \otimes W_N^2 P_N]\mu_N + \left[\frac{J_T}{T} \otimes W_N^2 P_N\right] v_N,
\end{aligned}
$$

Using the expressions in (A.13), (A.14), (8) and (9) we can write the following moments. These moments will be used later to prove Theorem 2.

$$
\begin{aligned}
\Psi_{1,N} &= \frac{1}{N(T-1)} u_N^{'} Q_{0,N} u_N = \frac{1}{N(T-1)} v_N^{'} C_{1,N} v_N, \\
C_{1,N} &= (I_T - \frac{J_T}{T}) \otimes P_N^{'} P_N, \\
\\
\Psi_{2,N} &= \frac{1}{N(T-1)} u_N^{'} Q_{0,N} \overline{u}_N = \frac{1}{N(T-1)} v_N^{'} C_{2,N} v_N, \\
C_{2,N} &= (I_T - \frac{J_T}{T}) \otimes P_N^{'} W_N P_N, \\
\\
\Psi_{3,N} &= \frac{1}{N(T-1)} \overline{u}_N^{'} Q_{0,N} \overline{u}_N = \frac{1}{N(T-1)} v_N^{'} C_{3,N} v_N, \\
C_{3,N} &= (I_T - \frac{J_T}{T}) \otimes P_N^{'} W_N^{'} W_N P_N, \qquad \qquad \text{(A.15a)} \\
\\
\Psi_{4,N} &= \frac{1}{N(T-1)} \overline{\overline{u}}_N^{'} Q_{0,N} \overline{u}_N = \frac{1}{N(T-1)} v_N^{'} C_{4,N} v_N, \\
C_{4,N} &= (I_T - \frac{J_T}{T}) \otimes P_N^{'} (W_N^{'})^2 W_N P_N, \\
\\
\Psi_{5,N} &= \frac{1}{N(T-1)} \overline{\overline{u}}_N^{'} Q_{0,N} \overline{\overline{u}}_N = \frac{1}{N(T-1)} v_N^{'} C_{5,N} v_N, \\
C_{5,N} &= (I_T - \frac{J_T}{T}) \otimes P_N^{'} (W_N^{'})^2 W_N^2 P_N, \\
\\
\Psi_{6,N} &= \frac{1}{N(T-1)} u_N^{'} Q_{0,N} \overline{\overline{u}}_N = \frac{1}{N(T-1)} v_N^{'} C_{6,N} v_N, \\
C_{6,N} &= (I_T - \frac{J_T}{T}) \otimes P_N^{'} W_N^2 P_N,
\end{aligned}
$$

$$\begin{aligned}
\Psi_{7,N} &= \frac{1}{N}u_N^{'}Q_{1,N}u_N = \frac{T}{N}\mu_N^{'}\widetilde{C}_{7,N}\mu_N + \frac{1}{N}v_N^{'}C_{7,N}v_N + \frac{2}{N}\mu_N^{'}\widehat{C}_{7,N}v_N,\\
\widetilde{C}_{7,N} &= P_N^{'}P_N,\ C_{7,N} = \frac{J_T}{T}\otimes P_N^{'}P_N,\ \widehat{C}_{7,N} = e_T^{'}\otimes P_N^{'}P_N,
\end{aligned}$$

$$\begin{aligned}
\Psi_{8,N} &= \frac{1}{N}u_N^{'}Q_{1,N}\overline{u}_N = \frac{T}{N}\mu_N^{'}\widetilde{C}_{8,N}\mu_N + \frac{1}{N}v_N^{'}C_{8,N}v_N + \frac{2}{N}\mu_N^{'}\widehat{C}_{8,N}v_N,\\
\widetilde{C}_{8,N} &= P_N^{'}W_NP_N,\ C_{8,N} = \frac{J_T}{T}\otimes P_N^{'}W_NP_N,\ \widehat{C}_{8,N} = e_T^{'}\otimes P_N^{'}W_NP_N,
\end{aligned}$$

$$\begin{aligned}
\Psi_{9,\,N} &= \frac{1}{N}\overline{u}_N^{'}Q_{1,N}\overline{u}_N = \frac{T}{N}\mu_N^{'}\widetilde{C}_{9,N}\mu_N + \frac{1}{N}v_N^{'}C_{9,N}v_N + \frac{2}{N}\mu_N^{'}\widehat{C}_{9,N}v_N,\\
\widetilde{C}_{9,N} &= P_N^{'}W_N^{'}W_NP_N,\ C_{9,N} = \frac{J_T}{T}\otimes P_N^{'}W_N^{'}W_NP_N, \qquad \text{(A.15b)}\\
\widehat{C}_{9,N} &= e_T^{'}\otimes P_N^{'}W_N^{'}W_NP_N,
\end{aligned}$$

$$\begin{aligned}
\Psi_{10,N} &= \frac{1}{N}\overline{\overline{u}}_N^{'}Q_{1,N}\overline{u}_N = \frac{T}{N}\mu_N^{'}\widetilde{C}_{10,N}\mu_N + \frac{1}{N}v_N^{'}C_{10,N}v_N + \frac{2}{N}\mu_N^{'}\widehat{C}_{10,N}v_N,\\
\widetilde{C}_{10,N} &= P_N^{'}(W_N^{'})^2W_NP_N,\ C_{10,N} = \frac{J_T}{T}\otimes P_N^{'}(W_N^{'})^2W_NP_N,\\
\widehat{C}_{10,N} &= e_T^{'}\otimes P_N^{'}(W_N^{'})^2W_NP_N,
\end{aligned}$$

$$\begin{aligned}
\Psi_{11,N} &= \frac{1}{N}\overline{\overline{u}}_N^{'}Q_{1,N}\overline{\overline{u}}_N = \frac{T}{N}\mu_N^{'}\widetilde{C}_{11,N}\mu_N + \frac{1}{N}v_N^{'}C_{11,N}v_N + \frac{2}{N}\mu_N^{'}\widehat{C}_{11,N}v_N,\\
\widetilde{C}_{11,N} &= P_N^{'}(W_N^{'})^2W_N^2P_N,\ C_{11,N} = \frac{J_T}{T}\otimes P_N^{'}(W_N^{'})^2W_N^2P_N,\\
\widehat{C}_{11,N} &= e_T^{'}\otimes P_N^{'}(W_N^{'})^2W_N^2P_N,
\end{aligned}$$

$$\begin{aligned}
\Psi_{12,N} &= \frac{1}{N}u_N^{'}Q_{1,N}\overline{\overline{u}}_N = \frac{T}{N}\mu_N^{'}\widetilde{C}_{12,N}\mu_N + \frac{1}{N}v_N^{'}C_{12,N}v_N + \frac{2}{N}\mu_N^{'}\widehat{C}_{12,N}v_N,\\
\widetilde{C}_{12,N} &= P_N^{'}W_N^2P_N,\ C_{12,N} = \frac{J_T}{T}\otimes P_N^{'}W_N^2P_N,\ \widehat{C}_{12,N} = e_T^{'}\otimes P_N^{'}W_N^2P_N,
\end{aligned}$$

The corresponding quadratic forms are based on predictors of u_N, $\overline{u}_N$ and $\overline{\overline{u}}_N$, say $\widetilde{u}_N$, $\widetilde{\overline{u}}_N$ and $\widetilde{\overline{\overline{u}}}_N$ respectively, where u_N, $\overline{u}_N$ and $\overline{\overline{u}}_N$ are defined in (5) and (31), respectively, and where $\widetilde{\overline{u}}_N$ and $\widetilde{\overline{\overline{u}}}_N$ are defined in (32). The $(i)-th$ element of $u_N(t)$, $\overline{u}_N(t)$, $\overline{\overline{u}}_N(t)$, $\widetilde{u}_N(t)$, $\widetilde{\overline{u}}_N(t)$, $\widetilde{\overline{\overline{u}}}_N(t)$ are $u_{it,N}$, $\overline{u}_{it,N}$, $\overline{\overline{u}}_{it,N}$, $\widetilde{u}_{it,N}$, $\widetilde{\overline{u}}_{it,N}$, $\widetilde{\overline{\overline{u}}}_{it,N}$, respectively. We define

$$
\begin{aligned}
\overline{u}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} u_{jt,N}, \\
\overline{\overline{u}}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} \overline{u}_{jt,N} = \Sigma_{j=1}^{N} w_{ij,N} \Sigma_{l=1}^{N} w_{jl,N} u_{lt,N}, \\
\widetilde{\overline{u}}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} \widetilde{u}_{jt,N}, \\
\widetilde{\overline{\overline{u}}}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} \widetilde{\overline{u}}_{jt,N} = \Sigma_{j=1}^{N} w_{ij,N} \Sigma_{l=1}^{N} w_{jl,N} \widetilde{u}_{lt,N},
\end{aligned}
$$

where $w_{ij,N}$, $w_{jl,N}$ are $(i,j)-th$ and $(j,l)-th$ element of W_N, respectively, and where $\widetilde{u}_{it,N}$, $\widetilde{u}_{lt,N}$, $\widetilde{u}_{jt,N}$ are predictors for $u_{it,N}$, $u_{lt,N}$, $u_{jt,N}$, respectively, which satisfy Assumption 5. The sample quadratic forms will be denoted by $\widetilde{\Psi}_{h,N}$, for $h = 1, \ldots, 12$.

Lemma 1 *Let A_N and B_N be two square matrices of dimension $kN \times kN$ whose row and column sums are uniformly bounded in absolute value by a finite constant, say k_A and k_B, respectively, and where k is some finite positive integer and $N \geq 1$. Define $C_N = A_N B_N$, then the row and column sums of*

C_N *are uniformly bounded in absolute value* $k_A k_B$.

Furthermore, let E_N *be a matrix of dimension* $N \times N$, *where* $N \geq 1$, *whose row and column sums are uniformly bounded in absolute value by a finite constant, say* k_E. *In addition consider a finite matrix* D *of dimension* $k \times l$, *whose row and column sums are bounded in absolute value by a finite constant, say* $k_{D,}$ *where* k *and* l *are some finite positive integers. Define* $F_N = D \otimes E_N$, *then the row and column sums of* F_N *are uniformly bounded in absolute value* $k_D k_E$.

Proof: Consider matrices $A_N = (a_{ij,kN})$, $B_N = (b_{ij,kN})$, and $C_N = (c_{ij,kN}) = A_N B_N$, where $a_{ij,kN}$, $b_{ij,kN}$ and $c_{ij,kN}$ are $(i,j) - th$ element of matrices A_N, B_N and C_N, respectively. Then

$$\begin{aligned}
\Sigma_{i=1}^{kN} |a_{ij,kN}| &\leq k_A, \\
\Sigma_{i=1}^{kN} |b_{ij,kN}| &\leq k_B, \\
\Sigma_{i=1}^{kN} |c_{ij,kN}| &= \Sigma_{i=1}^{kN} \Sigma_{l=1}^{kN} |a_{il,kN}| |b_{lj,kN}| \\
&= \Sigma_{l=1}^{kN} |b_{lj,kN}| \Sigma_{i=1}^{kN} |a_{il,kN}| \leq k_A k_B.
\end{aligned}$$

Similarly

$$\begin{aligned}
\Sigma_{j=1}^{kN} |a_{ij,kN}| &\leq k_A, \\
\Sigma_{j=1}^{kN} |b_{ij,kN}| &\leq k_B, \\
\Sigma_{j=1}^{kN} |c_{ij,kN}| &= \Sigma_{j=1}^{kN} \Sigma_{l=1}^{kN} |a_{il,kN}| \, |b_{lj,kN}| \\
&= \Sigma_{l=1}^{kN} |a_{il,kN}| \, \Sigma_{j=1}^{kN} |b_{lj,kN}| \leq k_B k_A.
\end{aligned}$$

For second part of the Lemma consider the matrices $D = (d_{ij})$, $E_N = (e_{ij,N})$, and $F_N = (f_{ij,N}) = D \otimes E_N$, where d_{ij}, $e_{ij,N}$ and $f_{ij,N}$ are $(i,j) - th$ element of matrices D, E_N and F_N, respectively. Then

$$\begin{aligned}
\Sigma_{j=1}^{l} |d_{ij}| &\leq k_D, \\
\Sigma_{i=1}^{k} |d_{ij}| &\leq k_D, \\
\Sigma_{i=1}^{N} |e_{ij,N}| &\leq k_E, \\
\Sigma_{j=1}^{N} |e_{ij,N}| &\leq k_E.
\end{aligned}$$

Furthermore, let $i = (r-1)N + h$, and $j = (p-1)N + x$, where $r = 1, \ldots, k$, $h = 1, \ldots, N$, $p = 1, \ldots, l$ and $x = 1, \ldots, N$. Then

$$\begin{aligned}
\Sigma_{i=1}^{kN} |f_{ij,N}| &= \Sigma_{r=1}^{k} |d_{rp}| \, \Sigma_{h=1}^{N} |e_{hx,N}| \leq k_D k_E, \\
\Sigma_{j=1}^{lN} |f_{ij,N}| &= \Sigma_{p=1}^{l} |d_{rp}| \, \Sigma_{j=1}^{N} |e_{hx,N}| \leq k_D k_E.
\end{aligned}$$

Lemma 2 *Under Assumption 4 the elements of the matrix* $C_{h,N}$*, for* $h = 1, \ldots, 12$*, defined above have the following properties:*

$$\begin{aligned} \Sigma_{i=1}^{NT} |c_{h,ij}| &\leq k_c, \\ \Sigma_{j=1}^{NT} |c_{h,ij}| &\leq k_c, \end{aligned}$$

for all $N \geq 1$ *and where* T *is a fixed positive integer, and* $1 \leq i,\ j \leq NT$ *for some* $0 < k_c < \infty$*, where* $c_{h,ij}$ *is the* $(i,j)-th$ *element of matrix* $C_{h,N}$*.*

Similarly the elements of the matrix $\widetilde{C}_{h,N}$*, for* $h = 7, \ldots, 12$*, defined above have the following properties:*

$$\begin{aligned} \Sigma_{i=1}^{N} |\widetilde{c}_{h,ij}| &\leq k_{\widetilde{c}}, \\ \Sigma_{j=1}^{N} |\widetilde{c}_{h,ij}| &\leq k_{\widetilde{c}}, \end{aligned}$$

for all $N \geq 1$ *and* $1 \leq i,\ j \leq N$ *for some* $0 < k_{\widetilde{c}} < \infty$*, where* $\widetilde{c}_{h,ij}$ *is the* $(i,j)-th$ *element of matrix* $\widetilde{C}_{h,N}$*.*

Similarly the elements of the matrix $\widehat{C}_{h,N}$*, for* $h = 7, \ldots, 12$*, defined above have the following properties:*

$$\begin{aligned} \Sigma_{i=1}^{N} |\widehat{c}_{h,ij}| &\leq k_{\widehat{c}}, \\ \Sigma_{j=1}^{NT} |\widehat{c}_{h,ij}| &\leq k_{\widehat{c}}, \end{aligned}$$

for all $N \geq 1$ *and where* T *is a fixed positive integer, and* $1 \leq i,\ j \leq NT$ *for some* $0 < k_{\widehat{c}} < \infty$*, where* $\widehat{c}_{h,ij}$ *is the* $(i,j)-th$ *element of matrix* $\widehat{C}_{h,N}$

Furthermore,

$$
\begin{aligned}
(NT)^{-2}\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT}(c_{h,ij}+c_{h,ji})^2 &= o(1), \\
N^{-2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}(\widetilde{c}_{h,ij}+\widetilde{c}_{h,ji})^2 &= o(1), \\
N^{-2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{NT}\widehat{c}_{h,ij}^2 &= o(1),
\end{aligned}
$$

as $N \to \infty$.

Proof: By Lemma 1 the row and column sums of the matrices $C_{h,N}$, $\widetilde{C}_{h,N}$ and $\widehat{C}_{h,N}$ are uniformly bounded. Next observe that the row and column sums of the matrices $C_{h,N}+C'_{h,N}$, $[C_{h,N}+C'_{h,N}][C_{h,N}+C'_{h,N}]$, $\widetilde{C}_{h,N}+\widetilde{C}'_{h,N}$, $[\widetilde{C}_{h,N}+\widetilde{C}'_{h,N}][\widetilde{C}_{h,N}+\widetilde{C}'_{h,N}]$, $\widehat{C}_{h,N}\widehat{C}'_{h,N}$ are uniformly bounded by $2k_c$, $4k_c^2$, $2k_{\widetilde{c}}$, $4k_{\widetilde{c}}^2$ and $k_{\widehat{c}}^2$, respectively. The second claim of the lemma now follows because

$$
\begin{aligned}
(NT)^{-2}\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT}(c_{h,ij}+c_{h,ji})^2 &= (NT)^{-2}Tr\{[C_{h,N}+C'_{h,N}][C_{h,N}+C'_{h,N}]\} \\
&\leq \frac{4k_c^2}{NT} \to 0 \ as \ N \to \infty.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
N^{-2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}(\widetilde{c}_{h,ij}+\widetilde{c}_{h,ji})^2 &= N^{-2}Tr\{[\widetilde{C}_{h,N}+\widetilde{C}'_{h,N}][\widetilde{C}_{h,N}+\widetilde{C}'_{h,N}]\} \\
&\leq \frac{4k_{\widetilde{c}}^2}{N} \to 0 \ as \ N \to \infty.
\end{aligned}
$$

Furthermore,

$$
N^{-2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{NT}\widehat{c}_{h,ij}^2 \leq \frac{k_{\widehat{c}}^2}{N} \to 0 \ as \ N \to \infty.
$$

Lemma 3 [44] *Let* $\pi_N = (\pi_{1,N}, \ldots, \pi_{NT,N})'$, *where* $\pi_{i,N}$ *is a real valued random variable such that for all* $1 \leq i \leq NT$, *where* T *is a fixed positive integer and* $N \geq 1$, *the real valued random variables* $\pi_{i,N}$, *are identically distributed with mean zero, finite variance* σ_π^2 *and* $\varkappa_4 = E(\pi_{i,N}^4) < \infty$, *where* $0 < \sigma_\pi^2 < b_\pi$, *and where* b_π *is a known finite constant. In addition for each* $N \geq 1$ *the real valued random variables* $\pi_{1,N}, \ldots, \pi_{NT,N}$ *are identically and independently distributed.*

Define a quadratic form $Q_N = \pi'_N A_N \pi_N$, *where* A_N *is a square matrix of dimension* $NT \times NT$. *Then*

$$\begin{aligned} EQ_N &= \sigma_\pi^2 \Sigma_{i=1}^{NT} a_{ii,N}, \\ var(Q_N) &= (\varkappa_4 - \sigma_\pi^4)\Sigma_{i=1}^{NT} a_{ii,N}^2 + \sigma_\pi^4 \Sigma_{i=1}^{NT} \Sigma_{j=1}^{i-1} (a_{ij,N} + a_{ji,N})^2, \end{aligned}$$

where $a_{ij,N}$ *is the* $(i,j) - th$ *element of* A_N.

Proof: Observe that

$$\begin{aligned} Q_N &= \pi'_N A_N \pi_N \\ &= \Sigma_{i=1}^{NT} \Sigma_{j=1}^{NT} a_{ij,N} \pi_{i,N} \pi_{j,N}, \\ &= \Sigma_{i=1}^{NT} a_{ii,N} \pi_{i,N}^2 + \Sigma_{i=1}^{NT} \Sigma_{j=1}^{i-1} (a_{ij,N} + a_{ji,N}) \pi_{i,N} \pi_{j,N}. \end{aligned}$$

[44] This Lemma has been proved in Kelejian and Prucha (2001).

In light of the i.i.d. assumption relating to $\pi_{i,N}$ and also, given that $E(\pi_{i,N}) = 0$, we have

$$\begin{aligned} EQ_N &= \Sigma_{i=1}^{NT} a_{ii,N} E(\pi_{i,N}^2) + \Sigma_{i=1}^{NT} \Sigma_{j=1}^{i-1} (a_{ij,N} + a_{ji,N}) E(\pi_{i,N}\pi_{j,N}) \\ &= \sigma_\pi^2 \Sigma_{i=1}^{NT} a_{ii,N}^2. \end{aligned}$$

Next observe that the random variables $\Sigma_{i=1}^{NT} a_{ii,N}\pi_{i,N}^2$ and $\Sigma_{i=1}^{NT}\Sigma_{j=1}^{i-1}(a_{ij,N} + a_{ji,N})\pi_{i,N}\pi_{j,N}$ have a zero covariance since $E(\pi_{i,N}^2\pi_{r,N}\pi_{j,N}) = 0$ unless $i = r = j$, which is ruled out. Furthermore, $cov(\pi_{i,N}^2, \pi_{j,N}^2) = 0$, unless $i = j$, and $cov(\pi_{i,N}\pi_{j,N}, \pi_{r,N}\pi_{k,N}) = 0$, unless $i = r$ and $j = k$, or $i = k$ and $j = r$, which is ruled out by the indices of the summation. Therefore,

$$var(Q_N) = \Sigma_{i=1}^{NT} a_{ii,N}^2 var(\pi_{i,N}^2) + \Sigma_{i=1}^{NT}\Sigma_{j=1}^{i-1}(a_{ij,N} + a_{ji,N})^2 var(\pi_{i,N}\pi_{j,N}).$$

Note that

$$\begin{aligned} var(\pi_{i,N}^2) &= E(\pi_{i,N}^4) - (E(\pi_{i,N}^2))^2 = \varkappa_4 - \sigma_\pi^4, \\ var(\pi_{i,N}\pi_{j,N}) &= var(\pi_{i,N}) var(\pi_{j,N}) = \sigma_\pi^4,\ when\ i \neq j. \end{aligned}$$

Therefore,

$$var(Q_N) = (\varkappa_4 - \sigma_\pi^4)\Sigma_{i=1}^{NT} a_{ii,N}^2 + \sigma_\pi^4 \Sigma_{i=1}^{NT}\Sigma_{j=1}^{i-1}(a_{ij,N} + a_{ji,N})^2.$$

Lemma 4 [45] *Let* $\eta_N = (\eta_{1,N}, \ldots, \eta_{NT,N})'$, *where* $\eta_{i,N}$ *is a real valued random variable such that for all* $1 \leq i \leq NT$, *where* T *is a fixed positive integer and* $N \geq 1$, *the real valued random variables* $\eta_{i,N}$, *are identically distributed with mean zero, finite variance* σ_η^2 *and* $E\eta_{i,N}^4 < \infty$, *where* $0 < \sigma_\eta^2 < b_\eta$, *and where* b_η *is a known finite constant. In addition for each* $N \geq 1$ *the real valued random variables* $\eta_{1,N}, \ldots, \eta_{NT,N}$ *are identically and independently distributed. Furthermore, let* $\xi_N = (\xi_{1,N}, \ldots, \xi_{N,N})'$, *where* $\xi_{i,N}$ *is a real valued random variable such that for all* $1 \leq i \leq N$ *and* $N \geq 1$, *the real valued random variables* $\xi_{i,N}$, *are identically distributed with mean zero, finite variance* σ_ξ^2 *and* $E\xi_{i,N}^4 < \infty$, *where* $0 < \sigma_\xi^2 < b_\xi$, *and where* b_ξ *is a known finite constant. In addition for each* $N \geq 1$ *the real valued random variables* $\xi_{1,N}, \ldots, \xi_{N,N}$ *are identically and independently distributed. In addition* η_N *and* ξ_N *are independent of each other.*

Define $H_N = \xi_N' B_N \eta_N$, *where* B_N *is a matrix of dimension* $N \times NT$, *also where* T *is a fixed positive integer and* $N \geq 1$. *Then*

$$\begin{aligned} EH_N &= 0, \\ var(H_N) &= \sigma_\eta^2 \sigma_\xi^2 \Sigma_{i=1}^{N} \Sigma_{j=1}^{NT} b_{ij,N}^2. \end{aligned}$$

where $b_{ij,N}$ *is the* $(i,j) - th$ *element of* B_N.

[45] This Lemma has been proved in Kelejian and Prucha (2001).

Proof: Note that

$$\begin{aligned} H_N &= \xi_N' B_N \eta_N = \Sigma_{j=1}^{NT} \Sigma_{i=1}^{N} \eta_{j,N} \xi_{i,N} b_{ij,N}, \\ &= \Sigma_{j=1}^{NT} \eta_{j,N} \Sigma_{i=1}^{N} \xi_{i,N} b_{ij,N} \end{aligned}$$

Given our assumptions relating to the independence of η_N and ξ_N and in light of the i.i.d. assumption relating to $\eta_{j,N}$ and $\xi_{i,N}$, we observe that $cov(\eta_{j,N}\Sigma_{i=1}^{N}\xi_{i,N}b_{ij,N}, \eta_{k,N}\Sigma_{i=1}^{N}\xi_{i,N}b_{ik,N},) = 0$ unless $j = k$. Furthermore, by noting that $E\eta_{j,N} = E\xi_{i,N} = 0$, we have

$$\begin{aligned} EH_N &= \Sigma_{j=1}^{NT} E(\eta_{j,N}) \Sigma_{i=1}^{N} E(\xi_{i,N}) b_{ij,N} = 0, \\ var(H_N) &= \Sigma_{j=1}^{NT} var(\eta_{j,N} \Sigma_{i=1}^{N} \xi_{i,N} b_{ij,N}) \\ &= \Sigma_{j=1}^{NT} var(\eta_{j,N}) var(\Sigma_{i=1}^{N} \xi_{i,N} b_{ij,N}) \\ &= \Sigma_{j=1}^{NT} \sigma_\eta^2 \Sigma_{i=1}^{N} \sigma_\xi^2 b_{ij,N}^2 \\ &= \sigma_\eta^2 \sigma_\xi^2 \Sigma_{j=1}^{NT} \Sigma_{i=1}^{N} b_{ij,N}^2. \end{aligned}$$

Lemma 5 *Under Assumptions 1 to 4, the quadratic forms* $\Psi_{h,N}$*, for* $h = 1, \ldots, 12$*, have the following properties:*

$$\begin{aligned} E\Psi_{h,N} &= O(1), \\ var(\Psi_{h,N}) &= o(1). \end{aligned}$$

Therefore,

$$\Psi_{h,N} - E\Psi_{h,N} \xrightarrow{p} 0, \qquad \Psi_{h,N} = O_p(1).$$

Proof: In light of Lemma 3, observe that for $h = 1, \ldots, 6$,

$$\begin{aligned} E\Psi_{h,N} &= \frac{1}{N(T-1)} E v_N' C_{h,N} v_N \\ &= \frac{1}{N(T-1)} \sigma_v^2 \Sigma_{i=1}^{NT} c_{h,ii}, \end{aligned}$$

and

$$\begin{aligned} var(\Psi_{h,N}) &= \frac{1}{(N(T-1))^2} var(v_N' C_{h,N} v_N) \\ &= \frac{1}{(N(T-1))^2} [(\vartheta_{v,4} - \sigma_v^4) \Sigma_{i=1}^{NT} c_{h,ii}^2 + \sigma_v^4 \Sigma_{i=1}^{NT} \Sigma_{j=1}^{i-1} (c_{h,ij} + c_{h,ji})^2], \end{aligned}$$

where $c_{h,ij}$ is the $(i,j) - th$ element of $C_{h,N}$, $\vartheta_{v,4} = E v_{it}^4$, since $cov(v_{ik} v_{jl}, v_{rm} v_{sn}) = 0$ unless $i = r$, $k = m$ and $j = s$, $l = n$, or $i = s$, $k = n$ and $j = r$, $l = m$; compare, e.g., Kelejian and Prucha (2001). By Lemma 2, the row and column sums of $C_{h,N}$ are uniformly bounded in absolute value by some finite constant, say k_c. Hence

$$\begin{aligned} |E\Psi_{h,N}| &\leq \frac{1}{N(T-1)} \sigma_v^2 \Sigma_{i=1}^{NT} |c_{h,ii}| \\ &\leq \frac{NT}{N(T-1)} \sigma_v^2 k_c \\ &\leq \frac{T}{(T-1)} \sigma_v^2 k_c \leq 2\sigma_v^2 k_c < \infty, \end{aligned}$$

since $T > 1$ but finite, this proves that $E\Psi_{h,N} = O(1)$.

Next we observe that

$$\begin{aligned}
var(\Psi_{h,N}) &\leq \frac{1}{(N(T-1))^2}[(\vartheta_{v,4}-\sigma_v^4)\Sigma_{i=1}^{NT}c_{h,ii}^2+\sigma_v^4\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT}(c_{h,ij}+c_{h,ji})^2] \\
&\leq (\vartheta_{v,4}-\sigma_v^4)\frac{NT}{(N(T-1))^2}k_c^2+\frac{\sigma_v^4}{(N(T-1))^2}\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT}(c_{h,ij}+c_{h,ji})^2 \\
&\leq (\vartheta_{v,4}-\sigma_v^4)\frac{T}{(T-1)^2}\frac{k_c^2}{N}+\frac{T^2}{(T-1)^2}\frac{\sigma_v^4}{N^2T^2}\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT}(c_{h,ij}+c_{h,ji})^2.
\end{aligned}$$

Given that $T>1$ and finite, a sufficient condition for $var(\Psi_{h,N})=o(1)$ is that the terms $\frac{k_c^2}{N}$ and $\frac{1}{N^2T^2}\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT}(c_{h,ij}+c_{h,ji})^2$ on the r.h.s. be $o(1)$ as $N\rightarrow\infty$, which holds in light of Lemma 2. The last two claims follow from Chebychev's inequality and, for example, corollary 5.1.1.2 in Fuller(1976, p.186), respectively.

In light of Lemma 3 and 4 given our assumption that μ_N and v_N are independent, we observe that for $h=7,\ldots,12$,

$$\begin{aligned}
E\Psi_{h,N} &= \frac{T}{N}E\mu_N^{'}\widetilde{C}_{h,N}\mu_N+\frac{1}{N}Ev_N^{'}C_{h,N}v_N+\frac{2}{N}E\mu_N^{'}\widehat{C}_{h,N}v_N \\
&= T\sigma_\mu^2N^{-1}\Sigma_{i=1}^{N}\widetilde{c}_{h,ii}+\sigma_v^2N^{-1}\Sigma_{i=1}^{NT}c_{h,ii},
\end{aligned}$$

and

$$\begin{aligned}
var(\Psi_{h,N}) &= \frac{T^2}{N^2}var(\mu_N^{'}\widetilde{C}_{h,N}\mu_N)+\frac{1}{N^2}var(v_N^{'}C_{h,N}v_N)+\frac{4}{N^2}var(\mu_N^{'}\widehat{C}_{h,N}v_N) \\
&= T^2N^{-2}[(\vartheta_{\mu,4}-\sigma_\mu^4)\Sigma_{i=1}^{N}\widetilde{c}_{h,ii}^2+\sigma_\mu^4\Sigma_{i=1}^{N}\Sigma_{j=1}^{i-1}(\widetilde{c}_{h,ij}+\widetilde{c}_{h,ji})^2] \\
&\quad +N^{-2}[(\vartheta_{v,4}-\sigma_v^4)\Sigma_{i=1}^{NT}c_{h,ii}^2+\sigma_v^4\Sigma_{i=1}^{NT}\Sigma_{j=1}^{i-1}(c_{h,ij}+c_{h,ji})^2] \\
&\quad +N^{-2}[4\sigma_\mu^2\sigma_v^2\Sigma_{i=1}^{N}\Sigma_{j=1}^{NT}\widehat{c}_{h,ij}^2],
\end{aligned}$$

where $c_{h,ij}$, $\widetilde{c}_{h,ij}$, $\widehat{c}_{h,ij}$ are the $(i,j)-th$ element of $C_{h,N}$, $\widetilde{C}_{h,N}$, $\widehat{C}_{h,N}$, respectively, $\vartheta_{\mu,4} = E\mu_i^4$ and $\vartheta_{\nu,4} = Ev_{it}^4$, since $cov(\mu_i\mu_j,\, \mu_r\mu_s) = 0$ unless $i = r$, and $j = s$, or $i = s$, and $j = r$ and $cov(v_{ik}v_{jl},\, v_{rm}v_{sn}) = 0$ unless $i = r$, $k = m$ and $j = s$, $l = n$, or $i = s$, $k = n$ and $j = r$, $l = m$; compare, e.g., Kelejian and Prucha (2001). By Lemma 2, the row and column sums of $\widetilde{C}_{h,N}$, $C_{h,N}$, $\widehat{C}_{h,N}$ are uniformly bounded in absolute value by some finite constant, $k_{\widetilde{c}}$, k_c and $k_{\widehat{c}}$, respectively. Hence, in light of our assumption that μ and v are independent we observe,

$$\begin{aligned} |E\Psi_{h,N}| &\leq T\sigma_\mu^2 N^{-1}\Sigma_{i=1}^N |\widetilde{c}_{h,ii}| + \sigma_v^2 N^{-1}\Sigma_{i=1}^{NT} |c_{h,ii}| \\ &\leq T\sigma_\mu^2 k_{\widetilde{c}} + T\sigma_v^2 k_c < \infty, \end{aligned}$$

given that $T > 1$ and finite, this proves that $E\Psi_{h,N} = O(1)$.

Next we observe that

$$\begin{aligned} var(\Psi_{h,N}) &\leq T^2N^{-2}[(\vartheta_{\mu,4} - \sigma_\mu^4)Nk_{\widetilde{c}}^2 + \sigma_\mu^4\Sigma_{i=1}^N\Sigma_{j=1}^N(\widetilde{c}_{h,ij} + \widetilde{c}_{h,ji})^2] \\ &\quad + N^{-2}[(\vartheta_{v,4} - \sigma_v^4)NTk_c^2 + \sigma_v^4\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT}(c_{h,ij} + c_{h,ji})^2] \\ &\quad + N^{-2}[4\sigma_\mu^2\sigma_v^2Nk_{\widehat{c}}^2] \\ &= [T^2(\vartheta_{\mu,4} - \sigma_\mu^4)\frac{k_{\widetilde{c}}^2}{N} + T^2\sigma_\mu^4N^{-2}\Sigma_{i=1}^N\Sigma_{j=1}^N(\widetilde{c}_{h,ij} + \widetilde{c}_{h,ji})^2] \\ &\quad + [T(\vartheta_{v,4} - \sigma_v^4)\frac{k_c^2}{N} + T^2\sigma_v^4(NT)^{-2}\Sigma_{i=1}^{NT}\Sigma_{j=1}^{NT}(c_{h,ij} + c_{h,ji})^2] \\ &\quad + [4\sigma_\mu^2\sigma_v^2\frac{k_{\widehat{c}}^2}{N}] \end{aligned}$$

Given that $T > 1$ and finite, a sufficient condition for $var(\Psi_{h,N}) = o(1)$ is that the terms $\frac{k_{\tilde{c}}^2}{N}$, $\frac{k_c^2}{N}$, $\frac{k_{\hat{c}}^2}{N}$, $N^{-2}\Sigma_{i=1}^N \Sigma_{j=1}^N (\widetilde{c}_{h,ij} + \widetilde{c}_{h,ji})^2$ and $(NT)^{-2}\Sigma_{i=1}^{NT} \Sigma_{j=1}^{NT} (c_{h,ij} + c_{h,ji})^2$ be $o(1)$ as $N \to \infty$, which holds in light of Lemma 2. The last two claims follow from Chebychev's inequality and, for example, corollary 5.1.1.2 in Fuller(1976, p.186), respectively.

Lemma 6 *Consider random variables* $\upsilon_{it,N}$*,* $\omega_{it,N}$ *,* $\widetilde{\upsilon}_{it,N}$*, and* $\widetilde{\omega}_{it,N}$ *and assume that*

$$|\widetilde{\upsilon}_{it,N} - \upsilon_{it,N}| \leq D_{it,N}^{\upsilon} \tau_N^{\upsilon}, \qquad |\widetilde{\omega}_{it,N} - \omega_{it,N}| \leq D_{it,N}^{\omega} \tau_N^{\omega},$$

where $D_{it,N}^{\upsilon}$*,* $D_{it,N}^{\omega}$*,* τ_N^{υ} *and* τ_N^{ω} *are,respectively, nonnegative random variables with*

$$\begin{aligned} (NT)^{-1}\Sigma_{t=1}^T \Sigma_{i=1}^N (D_{it,N}^{\upsilon})^2 &= O_p(1), \\ (NT)^{-1}\Sigma_{t=1}^T \Sigma_{i=1}^N (D_{it,N}^{\omega})^2 &= O_p(1), \end{aligned}$$

and

$$\begin{aligned} \tau_N^{\upsilon} &= o_p(1), \\ \tau_N^{\omega} &= o_p(1). \end{aligned}$$

Suppose furthermore that

$$
\begin{aligned}
(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\upsilon_{it,N}^{2} &= O_p(1), \\
(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\omega_{it,N}^{2} &= O_p(1).
\end{aligned}
$$

Then

$$
(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\widetilde{\upsilon}_{it,N}\widetilde{\omega}_{it,N} - (NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\upsilon_{it,N}\omega_{it,N} \xrightarrow{p} 0, \ \ as\ N \to \infty.
$$

Proof: Observe that

$$
\begin{aligned}
& \left|(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\widetilde{\upsilon}_{it,N}\widetilde{\omega}_{it,N} - (NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\upsilon_{it,N}\omega_{it,N}\right| \\
\leq\ & (NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\left|\widetilde{\upsilon}_{it,\,N} - \upsilon_{it,\,N}\right|\left|\omega_{it,N}\right| \\
& +(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\left|\widetilde{\omega}_{it,N} - \omega_{it,N}\right|\left|\upsilon_{it,N}\right| \\
& +(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\left|\widetilde{\upsilon}_{it,N} - \upsilon_{it,N}\right|\left|\widetilde{\omega}_{it,N} - \omega_{it,N}\right| \\
\leq\ & [(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}D_{it,N}^{\upsilon}\left|\omega_{it,N}\right|]\tau_{N}^{\upsilon} \\
& +[(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}D_{it,N}^{\omega}\left|\upsilon_{it,N}\right|]\tau_{N}^{\omega} \\
& +[(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}D_{it,N}^{\upsilon}D_{it,N}^{\omega}]\tau_{N}^{\upsilon}\tau_{N}^{\omega} \\
\leq\ & [(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}(D_{it,N}^{\upsilon})^{2}]^{1/2}[(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\omega_{it,N}^{2}]^{1/2}\tau_{N}^{\upsilon} \\
& +[(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}(D_{it,N}^{\omega})^{2}]^{1/2}[(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\upsilon_{it,N}^{2}]^{1/2}\tau_{N}^{\omega} \\
& +[(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}(D_{it,N}^{\upsilon})^{2}]^{1/2}[(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}(D_{it,N}^{\omega})^{2}]^{1/2}\tau_{N}^{\upsilon}\tau_{N}^{\omega}.
\end{aligned}
$$

The last inequality follows from the above equation and Hölder's inequality. Since $\tau_N^{\upsilon} = o_p(1)$ and $\tau_N^{\omega} = o_p(1)$, the claim in the lemma follows by observing that all other terms are bounded in probability.

Lemma 7 *Under Assumptions 1 to 5 we have for* $h = 1, \ldots, 12$*:*

$$\widetilde{\Psi}_{h,N} - \Psi_{h,N} \xrightarrow{p} 0, \ as \ N \to \infty,$$

where $\Psi_{h,N}$ *is expressed in (A.15a) and (A.15b).* $\widetilde{\Psi}_{h,N}$ *is a sample quadratic form of* $\Psi_{h,N}$ *which is based on predictors of* u_N*,* $\overline{u}_N$ *and* $\overline{\overline{u}}_N$*, say* $\widetilde{u}_N$*,* $\widetilde{\overline{u}}_N$ *and* $\widetilde{\overline{\overline{u}}}_N$ *respectively, where* u_N*,* $\overline{u}_N$ *and* $\overline{\overline{u}}_N$ *are defined in (5) and (31), respectively, and where* $\widetilde{\overline{u}}_N$ *and* $\widetilde{\overline{\overline{u}}}_N$ *are defined in (32). The* $i-th$ *element of* $u_N(t)$*,* $\overline{u}_N(t)$*,* $\overline{\overline{u}}_N(t)$*,* $\widetilde{u}_N(t)$*,* $\widetilde{\overline{u}}_N(t)$*,* $\widetilde{\overline{\overline{u}}}_N(t)$ *are* $u_{it,N}$*,* $\overline{u}_{it,N}$*,* $\overline{\overline{u}}_{it,N}$*,* $\widetilde{u}_{it,N}$*,* $\widetilde{\overline{u}}_{it,N}$*,* $\widetilde{\overline{\overline{u}}}_{it,N}$*, respectively. We define*

$$\begin{aligned}
\overline{u}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} u_{jt,N}, \\
\overline{\overline{u}}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} \overline{u}_{jt,N} = \Sigma_{j=1}^{N} w_{ij,N} \Sigma_{l=1}^{N} w_{jl,N} u_{lt,N}, \\
\widetilde{\overline{u}}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} \widetilde{u}_{jt,N}, \\
\widetilde{\overline{\overline{u}}}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} \widetilde{\overline{u}}_{jt,N} = \Sigma_{j=1}^{N} w_{ij,N} \Sigma_{l=1}^{N} w_{jl,N} \widetilde{u}_{lt,N},
\end{aligned}$$

where $w_{ij,N}$*,* $w_{jl,N}$ *are* $(i,j)-th$ *and* $(j,l)-th$ *element of* W_N*, respectively, and where* $\widetilde{u}_{it,N}$*,* $\widetilde{u}_{lt,N}$*,* $\widetilde{u}_{jt,N}$ *are predictors for* $u_{it,N}$*,* $u_{lt,N}$*,* $u_{jt,N}$*, respectively, which satisfy Assumption 5.*

Proof: It will prove useful to introduce the following expressions:

$$
\begin{aligned}
\varphi_N &= Q_{0,N} u_N, \\
\overline{\varphi}_N &= (I_T \otimes W_N)\varphi_N = (I_T \otimes W_N) Q_{0,N} u_N, \\
\overline{\overline{\varphi}}_N &= (I_T \otimes W_N)\overline{\varphi}_N = (I_T \otimes W_N^2)\varphi_N = (I_T \otimes W_N^2) Q_{0,N} u_N, \\
\psi_N &= Q_{1,N} u_N, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.16)} \\
\overline{\psi}_N &= (I_T \otimes W_N)\psi_N = (I_T \otimes W_N) Q_{1,N} u_N, \\
\overline{\overline{\psi}}_N &= (I_T \otimes W_N)\overline{\psi}_N = (I_T \otimes W_N^2)\psi_N = (I_T \otimes W_N^2) Q_{1,N} u_N.
\end{aligned}
$$

In light of (A.16), (11) and (31) we have

$$
\begin{aligned}
\overline{\varphi}_N &= (I_T \otimes W_N) Q_{0,N} u_N = Q_{0,N}(I_T \otimes W_N) u_N = Q_{0,N}\overline{u}_N, \\
\overline{\overline{\varphi}}_N &= (I_T \otimes W_N^2) Q_{0,N} u_N = Q_{0,N}(I_T \otimes W_N^2) u_N = Q_{0,N}\overline{\overline{u}}_N, \\
\overline{\psi}_N &= (I_T \otimes W_N) Q_{1,N} u_N = Q_{1,N}(I_T \otimes W_N) u_N = Q_{1,N}\overline{u}_N, \qquad \text{(A.17)} \\
\overline{\overline{\psi}}_N &= (I_T \otimes W_N^2) Q_{1,N} u_N = Q_{1,N}(I_T \otimes W_N^2) u_N = Q_{1,N}\overline{\overline{u}}_N.
\end{aligned}
$$

The $i-th$ element of $\varphi_N(t)$, $\overline{\varphi}_N(t)$, $\overline{\overline{\varphi}}_N(t)$, $\psi_N(t)$, $\overline{\psi}_N(t)$, $\overline{\overline{\psi}}_N(t)$, respectively, in expressions (A.16) and (A.17) are,

$$\begin{aligned}
\varphi_{it,N} &= u_{it,N} - \frac{1}{T}\Sigma_{s=1}^{T} u_{is,N}, \\
\overline{\varphi}_{it,N} &= \overline{u}_{it,N} - \frac{1}{T}\Sigma_{s=1}^{T} \overline{u}_{is,N}, \\
\overline{\overline{\varphi}}_{it,N} &= \overline{\overline{u}}_{it,N} - \frac{1}{T}\Sigma_{s=1}^{T} \overline{\overline{u}}_{is,N}, \\
\psi_{it,N} &= \frac{1}{T}\Sigma_{s=1}^{T} u_{is,N}, \qquad\qquad (A.18) \\
\overline{\psi}_{it,N} &= \frac{1}{T}\Sigma_{s=1}^{T} \overline{u}_{is,N}, \\
\overline{\overline{\psi}}_{it,N} &= \frac{1}{T}\Sigma_{s=1}^{T} \overline{\overline{u}}_{is,N}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\overline{u}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} u_{jt,N}, \qquad\qquad (A.19) \\
\overline{\overline{u}}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N} \overline{u}_{jt,N} = \Sigma_{j=1}^{N} w_{ij,N} \Sigma_{l=1}^{N} w_{jl,N} u_{lt,N},
\end{aligned}$$

where $u_{jt,N}$ is the $j-th$ element of $u_N(t)$ and $w_{ij,N}$ is the $(i,j)-th$ element of W_N.

Premultiplying the quadratic forms in (A.15a) and (A.15b) by $\frac{T}{T}$ and rewriting the expressions using (A.16) and (A.17) into scalar notation yields

$$
\begin{aligned}
\Psi_{1,N} &= \frac{T}{T-1}(NT)^{-1}\varphi_N'\varphi_N = \frac{T}{T-1}(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\varphi_{it,N}^2, \\
\Psi_{2,N} &= \frac{T}{T-1}(NT)^{-1}\varphi_N'\overline{\varphi}_N = \frac{T}{T-1}(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\varphi_{it,N}\overline{\varphi}_{it,N}, \\
\Psi_{3,N} &= \frac{T}{T-1}(NT)^{-1}\overline{\varphi}_N'\overline{\varphi}_N = \frac{T}{T-1}(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\overline{\varphi}_{it,N}^2, \\
\Psi_{4,N} &= \frac{T}{T-1}(NT)^{-1}\overline{\overline{\varphi}}_N'\overline{\varphi}_N = \frac{T}{T-1}(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\overline{\overline{\varphi}}_{it,N}\overline{\varphi}_{it,N}, \\
\Psi_{5,N} &= \frac{T}{T-1}(NT)^{-1}\overline{\overline{\varphi}}_N'\overline{\overline{\varphi}}_N = \frac{T}{T-1}(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\overline{\overline{\varphi}}_{it,N}^2, \\
\Psi_{6,N} &= \frac{T}{T-1}(NT)^{-1}\varphi_N'\overline{\overline{\varphi}}_N = \frac{T}{T-1}(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\varphi_{it,N}\overline{\overline{\varphi}}_{it,N}, \\
\Psi_{7,\,N} &= T(NT)^{-1}\psi_N'\psi_N = T(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\psi_{it,N}^2, \qquad \text{(A.20)} \\
\Psi_{8,\,N} &= T(NT)^{-1}\psi_N'\overline{\psi}_N = T(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\psi_{it,N}\overline{\psi}_{it,N}, \\
\Psi_{9,\,N} &= T(NT)^{-1}\overline{\psi}_N'\overline{\psi}_N = T(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\overline{\psi}_{it,N}^2, \\
\Psi_{10,\,N} &= T(NT)^{-1}\overline{\overline{\psi}}_N'\overline{\psi}_N = T(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\overline{\overline{\psi}}_{it,N}\overline{\psi}_{it,N}, \\
\Psi_{11,\,N} &= T(NT)^{-1}\overline{\overline{\psi}}_N'\overline{\overline{\psi}}_N = T(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\overline{\overline{\psi}}_{it,N}^2, \\
\Psi_{12,\,N} &= T(NT)^{-1}\psi_N'\overline{\overline{\psi}}_N = T(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\psi_{it,N}\overline{\overline{\psi}}_{it,N},
\end{aligned}
$$

where $\varphi_{it,N}$, $\overline{\varphi}_{it,N}$, $\overline{\overline{\varphi}}_{it,N}$, $\psi_{it,N}$, $\overline{\psi}_{it,N}$, $\overline{\overline{\psi}}_N$ are the $i-th$ element of $\varphi_N(t)$, $\overline{\varphi}_N(t)$, $\overline{\overline{\varphi}}_N(t)$, $\psi_N(t)$, $\overline{\psi}_N(t)$, $\overline{\overline{\psi}}_N(t)$, respectively

In the following let $\widetilde{\varphi}_N$, $\widetilde{\overline{\varphi}}_N$, $\widetilde{\overline{\overline{\varphi}}}_N$, $\widetilde{\psi}_N$, $\widetilde{\overline{\psi}}_N$, $\widetilde{\overline{\overline{\psi}}}_N$ be predictors of φ_N, $\overline{\varphi}_N$, $\overline{\overline{\varphi}}_N$, ψ_N, $\overline{\psi}_N$, $\overline{\overline{\psi}}_N$, which are based on $\widetilde{u}_N$, $\widetilde{\overline{u}}_N$, $\widetilde{\overline{\overline{u}}}_N$, where $\widetilde{u}_N$, $\widetilde{\overline{u}}_N$, $\widetilde{\overline{\overline{u}}}_N$ are predictors of u_N, $\overline{u}_N$, $\overline{\overline{u}}_N$, respectively, which are defined above. Following our convention, the $i-th$ element of $\widetilde{\varphi}_N(t)$, $\widetilde{\overline{\varphi}}_N(t)$, $\widetilde{\overline{\overline{\varphi}}}_N(t)$, $\widetilde{\psi}_N(t)$, $\widetilde{\overline{\psi}}_N(t)$, $\widetilde{\overline{\overline{\psi}}}_N(t)$ will be denoted by $\widetilde{\varphi}_{it,N}$, $\widetilde{\overline{\varphi}}_{it,N}$, $\widetilde{\overline{\overline{\varphi}}}_{it,N}$, $\widetilde{\psi}_{it,N}$, $\widetilde{\overline{\psi}}_{it,N}$, $\widetilde{\overline{\overline{\psi}}}_{it,N}$, respectively.

To prove the lemma we verify that for each of the quadratic forms in (A.15a) and (A.15b) and reexpressed in scalar notation in (A.20), the assumptions maintained in Lemma 6 w.r.t. the respective variables are satisfied. Since $T > 1$ and finite, the terms $T/(T-1)$ and T are finite constants and can be ignored in our arguments.

We first verify that $\varphi_{it,N}$, $\overline{\varphi}_{it,N}$, $\overline{\overline{\varphi}}_{it,N}$, $\psi_{it,N}$, $\overline{\psi}_{it,N}$, and $\overline{\overline{\psi}}_{it,N}$ satisfy conditions maintained for $\upsilon_{it,N}$ and $\omega_{it,N}$ in Lemma 6. Since $\Psi_{h,N} = O_p(1)$ by Lemma 5, for $h = 1, \ldots, 12$, therefore, it follows from (A.20) that

$$\begin{aligned}
(NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\varphi_{it,N}^{2} &= O_p(1), \\
(NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\overline{\varphi}_{it,N}^{2} &= O_p(1), \\
(NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\overline{\overline{\varphi}}_{it,N}^{2}, &= O_p(1), \qquad \text{(A.21)} \\
(NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\psi_{it,N}^{2} &= O_p(1), \\
(NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\overline{\psi}_{it,N}^{2} &= O_p(1), \\
(NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\overline{\overline{\psi}}_{it,N}^{2}, &= O_p(1).
\end{aligned}$$

We next show that $\varphi_{it,N}$, $\overline{\varphi}_{it,N}$, $\overline{\overline{\varphi}}_{it,N}$, $\varphi_{it,N}$, $\overline{\varphi}_{it,N}$, $\overline{\overline{\varphi}}_{it,N}$ and their predictors $\widetilde{\varphi}_{it,N}$, $\widetilde{\overline{\varphi}}_{it,N}$, $\widetilde{\overline{\overline{\psi}}}_{it,N}$, $\widetilde{\psi}_{it,N}$, $\widetilde{\overline{\varphi}}_{it,N}$, $\widetilde{\overline{\overline{\psi}}}_{it,N}$ satisfy the remaining conditions in

Lemma 6 for $\upsilon_{it,N}$ and $\omega_{it,N}$ and their predictors, $\widetilde{\upsilon}_{it,N}$ and $\widetilde{\omega}_{it,N}$. Analogous to (A.18) we have

$$\begin{aligned}
\widetilde{\varphi}_{it,N} &= \widetilde{u}_{it,N} - \frac{1}{T}\Sigma_{s=1}^{T}\widetilde{u}_{is,N}, \\
\widetilde{\overline{\varphi}}_{it,N} &= \widetilde{\overline{u}}_{it,N} - \frac{1}{T}\Sigma_{s=1}^{T}\widetilde{\overline{u}}_{is,N}, \\
\widetilde{\overline{\overline{\varphi}}}_{it,N} &= \widetilde{\overline{\overline{u}}}_{it,N} - \frac{1}{T}\Sigma_{s=1}^{T}\widetilde{\overline{\overline{u}}}_{is,N}, \qquad \text{(A.22)}\\
\widetilde{\psi}_{it,N} &= \frac{1}{T}\Sigma_{s=1}^{T}\widetilde{u}_{is,N}, \\
\widetilde{\overline{\psi}}_{it,N} &= \frac{1}{T}\Sigma_{s=1}^{T}\widetilde{\overline{u}}_{is,N}, \\
\widetilde{\overline{\overline{\psi}}}_{it,N} &= \frac{1}{T}\Sigma_{s=1}^{T}\widetilde{\overline{\overline{u}}}_{is,N}.
\end{aligned}$$

Furthermore, analogous to (A.19), we have

$$\begin{aligned}
\widetilde{\overline{u}}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N}\widetilde{u}_{jt,N}, \qquad \text{(A.23)}\\
\widetilde{\overline{\overline{u}}}_{it,N} &= \Sigma_{j=1}^{N} w_{ij,N}\widetilde{\overline{u}}_{jt,N} = \Sigma_{j=1}^{N} w_{ij,N}\Sigma_{l=1}^{N} w_{jl,N}\widetilde{u}_{lt,N}.
\end{aligned}$$

Recall that by Assumption 5,

$$\left|\widetilde{u}_{it,N} - u_{it,N}\right| \leq ||d_{it,N}|| \, \|\Delta_N\|, \qquad \text{(A.24)}$$

where $(NT)^{-1}\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\|d_{it,N}\|^{2+\delta} = O_p(1)$ for some $\delta > 0$ and $N^{1/2}\|\Delta_N\| = O_p(1)$. Then, by Holder's inequality with $q = 2+\delta$, $\delta > 0$, and $\frac{1}{q} + \frac{1}{p} = 1$ we have

$$\begin{aligned}
\left|\varphi_{it,N}-\widetilde{\varphi}_{it,N}\right| &= \left|u_{it,N}-\widetilde{u}_{it,N}-\frac{1}{T}\Sigma_{s=1}^{T}(u_{is,N}-\widetilde{u}_{is,N})\right| \\
&\leq \left|u_{it,N}-\widetilde{u}_{it,N}\right|+\frac{1}{T}\Sigma_{s=1}^{T}\left|u_{is,N}-\widetilde{u}_{is,N}\right| \\
&\leq 2\Sigma_{s=1}^{T}\left|u_{is,N}-\widetilde{u}_{is,N}\right| \\
&\leq 2\left|\left|\Delta_N\right|\right|\Sigma_{s=1}^{T}\left\|d_{is,N}\right\| \\
&\leq 2\left|\left|\Delta_N\right|\right|[\Sigma_{s=1}^{T}\left|1\right|^{p}]^{1/p}[\Sigma_{s=1}^{T}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q} \qquad \text{(A.25)} \\
&\leq 2\left|\left|\Delta_N\right|\right|T^{1/p}[\Sigma_{s=1}^{T}\Sigma_{i=1}^{N}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q} \\
&= 2\left|\left|\Delta_N\right|\right|T^{1/p}[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{i=1}^{N}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q}(NT)^{1/q} \\
&= 2TN^{1/q}\left|\left|\Delta_N\right|\right|[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{i=1}^{N}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q} \\
&= D_N\tau_N,
\end{aligned}$$

where $D_N=[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{i=1}^{N}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q}$, and $\tau_N=2TN^{1/q}\left|\left|\Delta_N\right|\right|$ $=2TN^{-\delta/[2(2+\delta)]}N^{1/2}\left|\left|\Delta_N\right|\right|$. Given that $T>1$ and finite, and by Assumption 5, $D_N=O_p(1)$, and $\tau_N=o_p(1)$. Therefore, $\varphi_{it,N}$ and $\widetilde{\varphi}_{it,N}$ satisfy the properties maintained for υ_{it} and $\widetilde{\upsilon}_{it}$ in Lemma 6.

Next observe that by Assumption 4 we have

$$\Sigma_{j=1}^{N}\left|w_{ij,N}\right|^{p}=k_w^{p-1}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|[\left|\frac{w_{ij,N}}{k_w}\right|^{p-1}]\leq k_w^{p-1}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|]\leq k_w^{p}. \qquad \text{(A.26)}$$

Recall the expressions for $\overline{\varphi}_{it,N}$, $\overline{u}_{it,N}$, $\widetilde{\overline{\varphi}}_{it,\ N}$ and $\widetilde{\overline{u}}_{it,N}$ given in (A.18), (A.19), (A.22) and (A.23), and the inequality (A.24). Then by the triangle and Holder inequalities with $q = 2 + \delta$, $\delta > 0$, and $\frac{1}{q} + \frac{1}{p} = 1$, as well as (A.26) yields

$$\begin{aligned}
\left|\overline{\widetilde{\varphi}}_{it,N} - \overline{\varphi}_{it,N}\right| &= \left|\overline{\widetilde{u}}_{it,N} - \overline{u}_{it,N} - \frac{1}{T}\Sigma_{s=1}^{T}(\overline{\widetilde{u}}_{is,N} - \overline{u}_{is,N})\right| \\
&\leq \left|\overline{\widetilde{u}}_{it,N} - \overline{u}_{it,N}\right| + \frac{1}{T}\Sigma_{s=1}^{T}\left|\overline{\widetilde{u}}_{is,N} - \overline{u}_{is,N}\right| \\
&\leq 2\Sigma_{s=1}^{T}\left|\overline{\widetilde{u}}_{is,N} - \overline{u}_{is,N}\right| \\
&= 2\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\left|\widetilde{u}_{js,N} - u_{js,N}\right| \\
&\leq 2\left\|\Delta_N\right\|\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\left|\left|d_{js,N}\right|\right| \\
&\leq 2\left|\left|\Delta_N\right|\right|[\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|^{p}]^{1/p} * \qquad\qquad (A.27) \\
&\quad [\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|\left|d_{js,N}\right|\right|^{q}]^{1/q} \\
&= 2\left|\left|\Delta_N\right|\right|[T\Sigma_{j=1}^{N}\left|w_{ij,N}\right|^{p}]^{1/p} * \\
&\quad [(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|\left|d_{js,N}\right|\right|^{q}]^{1/q}(NT)^{1/q} \\
&\leq 2TN^{1/q}\left|\left|\Delta_N\right|\right| k_w * \\
&\quad [(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|\left|d_{js,N}\right|\right|^{q}]^{1/q} \\
&= \overline{D}_N\overline{\tau}_N,
\end{aligned}$$

where $\overline{D}_N = [(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|\left|d_{js,\,N}\right|\right|^{q}]^{1/q}$,and $\overline{\tau}_N = 2Tk_wN^{1/q}\left|\left|\Delta_N\right|\right|$ $= 2Tk_wN^{-\delta/[2(2+\delta)]}N^{1/2}\left|\left|\Delta_N\right|\right|$. It now follows immediately from Assumption

5 and given that $T > 1$ and finite, that $\overline{D}_N = O_p(1)$ and $\overline{\tau}_N = o_p(1)$. Therefore, $\overline{\varphi}_{it,N}$ and $\overline{\widetilde{\varphi}}_{it,N}$ also satisfy the properties maintained for υ_{it} and $\widetilde{\upsilon}_{it}$ in Lemma 6.

Now recall the expressions for $\overline{\overline{\varphi}}_{it,N}, \overline{\overline{u}}_{it,N}, \overline{\overline{\widetilde{\varphi}}}_{it,\ N}$ and $\overline{\overline{\widetilde{u}}}_{it,N}$ given in (A.18), (A.19), (A.22) and (A.23), and the inequality (A.24). Then by the triangle and Holder inequalities with $q = 2 + \delta$, $\delta > 0$, and $\frac{1}{q} + \frac{1}{p} = 1$, as well as

(A.26) yields

$$
\begin{aligned}
\left|\widetilde{\overline{\overline{\varphi}}}_{it,N} - \overline{\overline{\varphi}}_{it,N}\right| &= \left|\widetilde{\overline{\overline{u}}}_{it,N} - \overline{\overline{u}}_{it,N} - \frac{1}{T}\Sigma_{s=1}^{T}(\widetilde{\overline{\overline{u}}}_{is,N} - \overline{\overline{u}}_{is,N})\right| \\
&\leq \left|\widetilde{\overline{\overline{u}}}_{it,N} - \overline{\overline{u}}_{it,N}\right| + \frac{1}{T}\Sigma_{s=1}^{T}\left|\widetilde{\overline{\overline{u}}}_{is,N} - \overline{\overline{u}}_{is,N}\right| \\
&\leq 2\Sigma_{s=1}^{T}\left|\widetilde{\overline{\overline{u}}}_{is,N} - \overline{\overline{u}}_{is,N}\right| \\
&= 2\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\Sigma_{l=1}^{N}\left|w_{jl,N}\right|\left|\widetilde{u}_{ls,N} - u_{ls,N}\right| \\
&\leq 2\left|\left|\Delta_N\right|\right|\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\Sigma_{l=1}^{N}\left|w_{jl,N}\right|\left|\left|d_{ls,N}\right|\right| \\
&= 2\left|\left|\Delta_N\right|\right|\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|w_{jl,N}\right|\left|\left|d_{ls,N}\right|\right| \qquad \text{(A.28)} \\
&\leq 2\left|\left|\Delta_N\right|\right|\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\left[\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|w_{jl,N}\right|^{p}\right]^{1/p} * \\
&\quad \left[\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}\right]^{1/q} \\
&\leq 2\left|\left|\Delta_N\right|\right|\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\left[T\Sigma_{l=1}^{N}\left|w_{jl,N}\right|^{p}\right]^{1/p} * \\
&\quad \left[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}\right]^{1/q}(NT)^{1/q} \\
&\leq 2TN^{1/q}\left|\left|\Delta_N\right|\right|\Sigma_{j=1}^{N}\left|w_{ij,N}\right|k_w * \\
&\quad \left[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}\right]^{1/q} \\
&\leq 2TN^{1/q}\left|\left|\Delta_N\right|\right|k_w^2\left[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}\right]^{1/q} \\
&= \overline{\overline{D}}_N\overline{\overline{\tau}}_N,
\end{aligned}
$$

where $\overline{\overline{D}}_N = [(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}]^{1/q}$ and $\overline{\overline{\tau}}_N = 2Tk_w^2N^{1/q}\left|\left|\Delta_N\right|\right|$ $= 2Tk_w^2N^{-\delta/[2(2+\delta)]}N^{1/2}\left|\left|\Delta_N\right|\right|$. Again, it follows immediately from Assumption 5 and given that $T > 1$ and finite, that $\overline{\overline{\overline{D}}}_N = O_p(1)$ and $\overline{\overline{\overline{\tau}}}_N = o_p(1)$. Therefore, $\overline{\overline{\varphi}}_{it,N}$ and $\widetilde{\overline{\overline{\varphi}}}_{it,N}$ also satisfy the properties maintained for υ_{it} and $\widetilde{\upsilon}_{it}$ in Lemma 6.

Recall the expressions for $\psi_{it,N}, \widetilde{\psi}_{it,N}$, given in (A.18), (A.22), and the inequality (A.24). Then using the triangle and Holder inequalities with $q = 2 + \delta$, $\delta > 0$, and $\frac{1}{q} + \frac{1}{p} = 1$, yields

$$\begin{aligned}
\left|\psi_{it,N} - \widetilde{\psi}_{it,N}\right| &= \left|\frac{1}{T}\Sigma_{s=1}^{T}(u_{is,N} - \widetilde{u}_{is,N})\right| \\
&\leq \frac{1}{T}\Sigma_{s=1}^{T}\left|u_{is\ N} - \widetilde{u}_{is,N}\right| \\
&\leq \frac{1}{T}\left|\left|\Delta_N\right|\right|\Sigma_{s=1}^{T}\left\|d_{is,N}\right\| \\
&\leq \frac{1}{T}\left|\left|\Delta_N\right|\right|[\Sigma_{s=1}^{T}\left|1\right|^{p}]^{1/p}[\Sigma_{s=1}^{T}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q} \qquad \text{(A.29)} \\
&\leq \frac{1}{T}\left|\left|\Delta_N\right|\right|T^{1/p}[\Sigma_{s=1}^{T}\Sigma_{i=1}^{N}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q} \\
&= \frac{1}{T}\left|\left|\Delta_N\right|\right|T^{1/p}[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{i=1}^{N}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q}(NT)^{1/q} \\
&\leq N^{1/q}\left|\left|\Delta_N\right|\right|[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{i=1}^{N}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q} \\
&= D_N\tau_N^{*},
\end{aligned}$$

with $D_N = [(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{i=1}^{N}\left|\left|d_{is,N}\right|\right|^{q}]^{1/q}$, and $\tau_N^{*} = N^{1/q}\left|\left|\Delta_N\right|\right|$ $= N^{-\delta/[2(2+\delta)]}N^{1/2}\left|\left|\Delta_N\right|\right|$. By Assumption 5, $D_N = O_p(1)$, and $\tau_N^{*} = o_p(1)$. Therefore, $\psi_{it,N}$ and $\widetilde{\psi}_{it,N}$ satisfy the properties maintained for υ_{it} and $\widetilde{\upsilon}_{it}$ in Lemma 6.

Recall the expressions for $\overline{\psi}_{it,N}$, $\overline{u}_{it,N}$, $\overline{\widetilde{\psi}}_{it,N}$ and $\widetilde{\overline{\psi}}_{it,N}$ given in (A.18), (A.19), (A.22) and (A.23), and the inequality (A.24). Then by the triangle and Holder inequalities with $q = 2+\delta$, $\delta > 0$, and $\frac{1}{q}+\frac{1}{p} = 1$, as well as (A.26) yields

$$
\begin{aligned}
\left|\widetilde{\overline{\psi}}_{it,N} - \overline{\psi}_{it,N}\right| &= \left|\frac{1}{T}\Sigma_{s=1}^{T}(\widetilde{\overline{u}}_{is,N} - \overline{u}_{is,N})\right| \\
&\leq \frac{1}{T}\Sigma_{s=1}^{T}\left|\widetilde{\overline{u}}_{is,N} - \overline{u}_{is,N}\right| \\
&= \frac{1}{T}\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\left|\widetilde{u}_{js,N} - u_{js,N}\right| \\
&\leq \frac{1}{T}\left\|\Delta_N\right\|\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\left|\left|d_{js,N}\right|\right| \qquad \text{(A.30)} \\
&\leq \frac{1}{T}\left|\left|\Delta_N\right|\right|\left[\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|^{p}\right]^{1/p}\left[\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|\left|d_{js,N}\right|\right|^{q}\right]^{1/q} \\
&= \frac{1}{T}\left|\left|\Delta_N\right|\right|\left[T\Sigma_{j=1}^{N}\left|w_{ij,N}\right|^{p}\right]^{1/p} * \\
&\quad \left[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|\left|d_{js,N}\right|\right|^{q}\right]^{1/q}(NT)^{1/q} \\
&\leq N^{1/q}\left|\left|\Delta_N\right|\right|k_w * \\
&\quad \left[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|\left|d_{js,N}\right|\right|^{q}\right]^{1/q} \\
&= \overline{D}_N\overline{\tau}_N^{*},
\end{aligned}
$$

where $\overline{D}_N = [(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|\left|d_{js,N}\right|\right|^{q}]^{1/q}$,and $\overline{\tau}_N^{*} = k_w N^{1/q}\left|\left|\Delta_N\right|\right|$ $= k_w N^{-\delta/[2(2+\delta)]}N^{1/2}\left|\left|\Delta_N\right|\right|$. It now follows immediately from Assumption 5 that $\overline{D}_N = O_p(1)$ and $\overline{\tau}_N^{*} = o_p(1)$. Therefore, $\overline{\psi}_{it,N}$ and $\widetilde{\overline{\psi}}_{it,N}$ also satisfy the properties maintained for υ_{it} and $\widetilde{\upsilon}_{it}$ in Lemma 6.

Now recall the expressions for $\overline{\overline{\psi}}_{it,N}$, $\overline{\overline{u}}_{it,N}$, $\widetilde{\overline{\overline{\psi}}}_{it,N}$ and $\widetilde{\overline{\overline{u}}}_{it,N}$ given in (A.18), (A.19), (A.22) and (A.23), and the inequality (A.24). Then by the triangle and Holder inequalities with $q = 2 + \delta$, $\delta > 0$, and $\frac{1}{q} + \frac{1}{p} = 1$, as well as

(A.26) yields

$$
\begin{aligned}
\left|\overset{\approx}{\overline{\psi}}_{it,N}-\overline{\overline{\psi}}_{it,N}\right| &= \left|\frac{1}{T}\Sigma_{s=1}^{T}(\overset{\approx}{\overline{u}}_{is,N}-\overline{\overline{u}}_{is,N})\right| \\
&\leq \frac{1}{T}\Sigma_{s=1}^{T}\left|\overset{\approx}{\overline{u}}_{is,N}-\overline{\overline{u}}_{is,N}\right| \\
&= \frac{1}{T}\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\Sigma_{l=1}^{N}\left|w_{jl,N}\right|\left|\widetilde{u}_{ls,N}-u_{ls,N}\right| \\
&\leq \frac{1}{T}\left|\left|\Delta_N\right|\right|\Sigma_{s=1}^{T}\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\Sigma_{l=1}^{N}\left|w_{jl,N}\right|\left|\left|d_{ls,N}\right|\right| \\
&= \frac{1}{T}\left|\left|\Delta_N\right|\right|\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|w_{jl,N}\right|\left|\left|d_{ls,N}\right|\right| \\
&\leq \frac{1}{T}\left|\left|\Delta_N\right|\right|\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\left[\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|w_{jl,N}\right|^{p}\right]^{1/p} * \\
&\quad \left[\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}\right]^{1/q} \qquad \text{(A.31)} \\
&\leq \frac{1}{T}\left|\left|\Delta_N\right|\right|\Sigma_{j=1}^{N}\left|w_{ij,N}\right|\left[T\Sigma_{l=1}^{N}\left|w_{jl,N}\right|^{p}\right]^{1/p} * \\
&\quad \left[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}\right]^{1/q}(NT)^{1/q} \\
&\leq \frac{1}{T}(NT)^{1/q}\left|\left|\Delta_N\right|\right|\Sigma_{j=1}^{N}\left|w_{ij,N}\right|T^{1/p}k_w * \\
&\quad \left[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}\right]^{1/q} \\
&\leq N^{1/q}\left|\left|\Delta_N\right|\right|k_w^2 * \\
&\quad \left[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}\right]^{1/q} \\
&= \overline{\overline{D}}_N\overline{\overline{\tau}}_N^{*},
\end{aligned}
$$

where $\overline{\overline{D}}_N=[(NT)^{-1}\Sigma_{s=1}^{T}\Sigma_{l=1}^{N}\left|\left|d_{ls,N}\right|\right|^{q}]^{1/q}$ and $\overline{\overline{\tau}}_N^{*}=k_w^2N^{1/q}\left|\left|\Delta_N\right|\right|$ $=k_w^2N^{-\delta/[2(2+\delta)]}N^{1/2}\left|\left|\Delta_N\right|\right|$. Again, it follows immediately from Assumption

5 that $\overline{\overline{D}}_N = O_p(1)$ and $\overline{\overline{\tau}}_N^* = o_p(1)$. Therefore, $\overline{\overline{\psi}}_{it,N}$ and $\overline{\overline{\widetilde{\psi}}}_{it,N}$ also satisfy the properties maintained for υ_{it} and $\widetilde{\upsilon}_{it}$ in Lemma 6.

Having verified that all variables involved in the quadratic forms in (A.15) considered in Lemma 7 satisfy the conditions of Lemma 6, then Lemma 7 follows from Lemma 6.

The proof of Theorem 2 requires a central limit theorem (CLT) for triangular arrays. The CLT follows directly from a corollary to the Lindeberg-Feller CLT for triangular arrays using the Cramer-Wold device. That corollary is, for example, given in Billingsley (1995, problem 27.6, pg. 368).

Theorem A[46]**.** Let $\{\nu_{i,N}, 1 \leq i \leq NT, N \geq 1\}$ be a triangular array of random variables, where T is a fixed positive finite integer, that are identically distributed, and for each $N \geq 1$ (jointly) independent, with $E\nu_{i,N} = 0$ and $E\nu_{i,N}^2 = \sigma^2$, $0 < \sigma^2 < \infty$. Let $\{z_{ij,N}, 1 \leq i \leq NT, N \geq 1\}$, $j = 1, \ldots, K$, be a triangular array of real numbers, where T is a fixed positive finite integer, that are uniformly bounded in absolute value, that is, $k_z = \sup_{NT} \sup_{i \leq NT,\, j \leq K} |z_{ij,\, N}| < \infty$. Furthermore, consider $\{V_N : N \geq 1\}$ and $\{Z_N : N \geq 1\}$, where $V_N = (\nu_{i,N})_{i=1,\ldots,NT}$ and $Z_N = (z_{ij,N})_{i=1,\ldots,NT;\, j=1,\ldots,K}$ denote the corresponding sequences of $NT \times 1$ random vectors and $NT \times K$

[46] A similar theorem has been used in Kelejian and Prucha (1999).

real matrices, respectively. Let

$$\lim_{N\to\infty}(NT)^{-1}Z_N'Z_N = Q,$$

be finite and positive definite. Then

$$(NT)^{-1/2}Z_N'V_N \xrightarrow{D} N(0,\sigma^2 Q).$$

Proof of Theorem 1. To prove part (a) of the theorem observe that

$$(NT)^{1/2}[\widehat{\beta}_{GLS,N}-\beta] = [(NT)^{-1}Z_N'\Omega_{\varepsilon,N}^{-1}Z_N]^{-1}(NT)^{-1/2}Z_N'\Omega_{\varepsilon,N}^{-1}\varepsilon_N,$$

where $\widehat{\beta}_{GLS,N}$ is defined in (23) and $Z_N = [I_T \otimes (I_N - \rho W_N)]X_N$. Note that in general the elements of Z_N will depend on the sample size. Furthermore, under our maintained assumptions the elements of Z_N are uniformly bounded in absolute value by $(1+k_w)k_x$.

Recall from (17) that

$$\Omega_{\varepsilon,N}^{-1} = \sigma_1^{-2}Q_{1,N} + \sigma_v^{-2}Q_{0,N}. \tag{A.32}$$

Therefore,

$$\begin{aligned}\lim_{N\to\infty}(NT)^{-1}Z_N'\Omega_{\varepsilon,N}^{-1}Z_N &= \lim_{N\to\infty}(NT)^{-1}[\sigma_v^{-2}Z_N'Q_{0,N}Z_N + \\ &\quad \sigma_1^{-2}Z_N'Q_{1,N}Z_N] \\ &= \sigma_v^{-2}\underline{Q}_{xQ_0x}(\rho) + \sigma_1^{-2}\underline{Q}_{xQ_1x}(\rho),\end{aligned}$$

where $\underline{Q}_{xQ_0x}(\rho)$ and $\underline{Q}_{xQ_1x}(\rho)$ are defined in Assumption 3, and are finite and nonsingular.

Recall from (6) that

$$\varepsilon_N = (e_T \otimes I_N)\mu_N + \upsilon_N.$$

Furthermore, by Assumption 1 the error terms, $\upsilon_{11,N}, \ldots, \upsilon_{N1,N}, \ldots, \upsilon_{1T,N}, \ldots, \upsilon_{NT,N}$ are identically and independently distributed and for each $N \geq 1$ jointly independent with mean zero and finite variance σ_υ^2. Similarly the error terms, $\mu_{1,N}, \ldots, \mu_{N,N}$ are identically and independently distributed and for each $N \geq 1$ jointly independent with mean zero and finite variance σ_μ^2. In addition, the two processes $\{\upsilon_{it,N}\}$ and $\{\mu_{i,N}\}$ are independent of each other.

Now we express

$$(NT)^{-1/2} Z_N' \Omega_{\varepsilon,N}^{-1} \varepsilon_N = (NT)^{-1/2} Z_N' \Omega_{\varepsilon,N}^{-1} (e_T \otimes I_N)\mu_N + (NT)^{-1/2} Z' \Omega_{\varepsilon,N}^{-1} \upsilon_N.$$

Using the definitions of $Q_{1,N}$ and $Q_{0,N}$ in (8) we have

$$\begin{aligned}
Q_{0,N}(e_T \otimes I_N) &= ((I_T - \frac{J_T}{T}) \otimes I_N)(e_T \otimes I_N) \\
&= (I_T - \frac{J_T}{T})e_T \otimes I_N \\
&= (e_T - e_T) \otimes I_N = 0, \qquad\qquad (A.33) \\
Q_{1,N}(e_T \otimes I_N) &= (\frac{J_T}{T} \otimes I_N)(e_T \otimes I_N) \\
&= (\frac{J_T}{T} e_T \otimes I_N) \\
&= (e_T \otimes I_N).
\end{aligned}$$

Using the results in (A.32) and (A.33) we express

$$(NT)^{-1/2} Z_N' \Omega_{\varepsilon,N}^{-1}(e_T \otimes I_N)\mu_N = (NT)^{-1/2} Z_N' \sigma_1^{-2}(e_T \otimes I_N)\mu_N.$$

Next by Assumption 3 and (8) observe that

$$\begin{aligned}
\lim_{N\to\infty}(NT)^{-1}\sigma_1^{-4} Z_N'(e_T e_T' \otimes I_N) Z_N &= \lim_{N\to\infty}(NT)^{-1} T \sigma_1^{-4} Z_N'(\frac{e_T e_T'}{T} \otimes I_N) Z_N \\
&= \lim_{N\to\infty}(NT)^{-1} T \sigma_1^{-4} Z_N' Q_{1,N} Z_N \\
&= T\sigma_1^{-4} \underline{Q}_{xQ_1x}(\rho),
\end{aligned}$$

$$\begin{aligned}
\lim_{N\to\infty}(NT)^{-1} Z_N' \Omega_{\varepsilon,N}^{-2} Z_N &= \lim_{N\to\infty}(NT)^{-1}\sigma_1^{-4} Z_N' Q_{1,N} Z_N + \\
&\quad \lim_{N\to\infty}(NT)^{-1}\sigma_v^{-4} Z_N' Q_{0,N} Z_N \\
&= \sigma_1^{-4}\underline{Q}_{xQ_1x}(\rho) + \sigma_v^{-4}\underline{Q}_{xQ_0x}(\rho).
\end{aligned}$$

Then by Theorem A

$$(NT)^{-1/2} Z_N' \Omega_{\varepsilon,N}^{-1} (e_T \otimes I_N)\mu_N \xrightarrow{D} N\{0, \sigma_\mu^2 [T\sigma_1^{-4} \underline{Q}_{xQ_1x}(\rho)]\},$$

$$(NT)^{-1/2} Z_N' \Omega_{\varepsilon,N}^{-1} \upsilon_N \xrightarrow{D} N\{0, \sigma_v^2 [\sigma_1^{-4} \underline{Q}_{xQ_1x}(\rho) + \sigma_v^{-4} \underline{Q}_{xQ_0x}(\rho)]\}.$$

where $\underline{Q}_{xQ_1x}(\rho)$ and $\underline{Q}_{xQ_0x}(\rho)$ are defined in Assumption 3. Given $\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2$ and by Assumption 1 that the two processes $\{\mu_{i,N}\}$ and $\{\upsilon_{it,N}\}$ are independent of each other, it then follows that

$$\begin{aligned} (NT)^{-1/2} Z_N' \Omega_{\varepsilon,N}^{-1} \varepsilon_N &= (NT)^{-1/2} Z_N' \Omega_{\varepsilon,N}^{-1} (e_T \otimes I_N)\mu_N + (NT)^{-1/2} Z' \Omega_{\varepsilon,N}^{-1} \upsilon_N \\ &\xrightarrow{D} [N\ \{0, [\sigma_\mu^2 (T\sigma_1^{-4} \underline{Q}_{xQ_1x}(\rho))]\} + \qquad \text{(A.34)} \\ &\quad N\{0, [\sigma_v^2 (\sigma_1^{-4} \underline{Q}_{xQ_1x}(\rho) + \sigma_v^{-4} \underline{Q}_{xQ_0x}(\rho))]\}] \\ &= N\ \{0, [(T\sigma_\mu^2 + \sigma_v^2)\sigma_1^{-4} \underline{Q}_{xQ_1x}(\rho) + \sigma_v^{-2} \underline{Q}_{xQ_0x}(\rho)]\} \\ &= N\ \{0, [\sigma_1^{-2} \underline{Q}_{xQ_1x}(\rho) + \sigma_v^{-2} \underline{Q}_{xQ_0x}(\rho)]\}. \end{aligned}$$

By Assumption 3 and (A.32)

$$\begin{aligned} \lim_{N\to\infty} (NT)^{-1} Z_N' \Omega_{\varepsilon,N}^{-1} Z_N &= \lim_{N\to\infty} (NT)^{-1} [\sigma_v^{-2} Z_N' Q_{0,N} Z_N + \\ &\quad \sigma_1^{-2} Z_N' Q_{1,N} Z_N] \qquad \text{(A.35)} \\ &= \sigma_v^{-2} \underline{Q}_{xQ_0x}(\rho) + \sigma_1^{-2} \underline{Q}_{xQ_1x}(\rho), \end{aligned}$$

then it follows from (A.34) and (A.35) that[47]

$$
\begin{aligned}
& [\lim_{N\to\infty}(NT)^{-1}Z_N'\Omega_{\varepsilon,N}^{-1}Z_N]^{-1}(NT)^{-1/2}Z_N'\Omega_{\varepsilon,N}^{-1}\varepsilon_N \\
= \;& [\sigma_v^{-2}\underline{Q}_{xQ_0x}(\rho)+\sigma_1^{-2}\underline{Q}_{xQ_1x}(\rho)]^{-1} * \\
& (NT)^{-1/2}Z_N'\Omega_{\varepsilon,N}^{-1}\varepsilon_N \qquad\qquad (A.36) \\
& \xrightarrow{D} N\ \{0,[\sigma_1^{-2}\underline{Q}_{xQ_1x}(\rho)+ \\
& \sigma_v^{-2}\underline{Q}_{xQ_0x}(\rho)]^{-1}\}.
\end{aligned}
$$

Therefore,

$$(NT)^{1/2}[\widehat{\beta}_{GLS,N}-\beta] \xrightarrow{D} N\ \{0,[\sigma_1^{-2}\underline{Q}_{xQ_1x}(\rho)+\sigma_v^{-2}\underline{Q}_{xQ_0x}(\rho)]^{-1}\},$$

which also implies that $\widehat{\beta}_{GLS,N}$ is a consistent estimator of β.

We prove part (b) of the theorem by showing that

$$(NT)^{1/2}[\widehat{\beta}_{GLS,N}-\widehat{\beta}_{FGLS,N}] \xrightarrow{p} 0 \;\; as\ N\to\infty.$$

To prove this it suffices to show that[48]

$$
\begin{aligned}
& (NT)^{-1}X'\{[I_T\otimes(I_N-\widehat{\rho}_N W_N')]\widehat{\Omega}_{\varepsilon,N}^{-1}[I_T\otimes(I_N-\widehat{\rho}_N W_N)]- \\
& [I_T\otimes(I_N-\rho W_N')]\Omega_{\varepsilon,N}^{-1}[I_T\otimes(I_N-\rho W_N)]\}X \xrightarrow{p} 0, \qquad (A.37)
\end{aligned}
$$

[47] See Greene Ch. 12 Pg. 501-502, 2000.
[48] See Schmidt, pg. 71, 1976.

and

$$(NT)^{-1/2}X_N'\{[I_T \otimes (I_N - \widehat{\rho}_N W_N')]\widehat{\Omega}_{\varepsilon,\,N}^{-1}[I_T \otimes (I_N - \widehat{\rho}_N W_N)] -$$

$$[I_T \otimes (I_N - \rho W_N')]\Omega_{\varepsilon,\,N}^{-1}[I_T \otimes (I_N - \rho W_N)]\}u_N \xrightarrow{p} 0, \tag{A.38}$$

where $\widehat{\Omega}_{\varepsilon,N}^{-1} = \widehat{\sigma}_{1,N}^{-2}Q_{1,N} + \widehat{\sigma}_{v,N}^{-2}Q_{0,N}$ and $\widehat{\sigma}_{1,N}^{-2}$ and $\widehat{\sigma}_{v,N}^{-2}$ are consistent estimators of σ_1^{-2} and σ_v^{-2}, respectively.

Using the above expressions for $\Omega_{\varepsilon,N}^{-1}$ and $\widehat{\Omega}_{\varepsilon,N}^{-1}$ and by (11) we get

$$\begin{aligned}
& [I_T \otimes (I_N - \widehat{\rho}_N W_N')]\widehat{\Omega}_{\varepsilon,N}^{-1}[I_T \otimes (I_N - \widehat{\rho}_N W_N)] - \\
& [I_T \otimes (I_N - \rho W_N')]\Omega_{\varepsilon,N}^{-1}[I_T \otimes (I_N - \rho W_N)] \\
= \; & [I_T \otimes (I_N - \widehat{\rho}_N W_N')(I_N - \widehat{\rho}_N W_N)]\widehat{\Omega}_{\varepsilon,N}^{-1} - \\
& [I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)]\Omega_{\varepsilon,N}^{-1} \qquad \text{(A.39)} \\
= \; & [I_T \otimes (I_N - \widehat{\rho}_N W_N')(I_N - \widehat{\rho}_N W_N)]\widehat{\Omega}_{\varepsilon,N}^{-1} - \\
& [I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)]\widehat{\Omega}_{\varepsilon,N}^{-1} + \\
& [I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)]\widehat{\Omega}_{\varepsilon,N}^{-1} - \\
& [I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)]\Omega_{\varepsilon,N}^{-1} \\
= \; & [I_T \otimes [-(\widehat{\rho}_N - \rho)(W_N' + W_N) + (\widehat{\rho}_N^2 - \rho^2)W_N'W_N]](\widehat{\sigma}_{1,N}^{-2}Q_{1,N} + \widehat{\sigma}_{v,N}^{-2}Q_{0,N}) + \\
& [I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)][(\widehat{\sigma}_{1,N}^{-2} - \sigma_1^{-2})Q_{1,N} + (\widehat{\sigma}_{v,N}^{-2} - \sigma_v^{-2})Q_{0,N}].
\end{aligned}$$

Using (A.39) in (A.37) and (A.38) we have

$$
\begin{aligned}
& (NT)^{-1} X' \{[I_T \otimes (I_N - \widehat{\rho}_N W_N')] \widehat{\Omega}_{\varepsilon,N}^{-1} [I_T \otimes (I_N - \widehat{\rho}_N W_N)] \\
& -[I_T \otimes (I_N - \rho W_N')] \Omega_{\varepsilon,N}^{-1} [I_T \otimes (I_N - \rho W_N)]\} X \qquad \text{(A.40)} \\
= \; & -(\widehat{\rho}_N - \rho)(NT)^{-1} X' \{[I_T \otimes (W_N' + W_N)][\widehat{\sigma}_{1,\,N}^{-2} Q_{1,N} + \widehat{\sigma}_{v,\,N}^{-2} Q_{0,N}]\} X + \\
& (\widehat{\rho}_N^2 - \rho^2)(NT)^{-1} X' \{[I_T \otimes (W_N' W_N)][\widehat{\sigma}_{1,\,N}^{-2} Q_{1,N} + \widehat{\sigma}_{v,\,N}^{-2} Q_{0,N}]\} X + \\
& (\widehat{\sigma}_{1,\,N}^{-2} - \sigma_1^{-2})(NT)^{-1} X' \{[I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)] Q_{1,N}\} X + \\
& (\widehat{\sigma}_{v,\,N}^{-2} - \sigma_v^{-2})(NT)^{-1} X' \{[I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)] Q_{0,N}\} X,
\end{aligned}
$$

and

$$
\begin{aligned}
& (NT)^{-1/2} X_N' \{[I_T \otimes (I_N - \widehat{\rho}_N W_N')] \widehat{\Omega}_{\varepsilon,N}^{-1} [I_T \otimes (I_N - \widehat{\rho}_N W_N)] - \\
& [I_T \otimes (I_N - \rho W_N')] \Omega_{\varepsilon,N}^{-1} [I_T \otimes (I_N - \rho W_N)]\} u_N \qquad \text{(A.41)} \\
= \; & -(\widehat{\rho}_N - \rho)(NT)^{-1/2} X_N' \{[I_T \otimes (W_N' + W_N)][\widehat{\sigma}_{1,N}^{-2} Q_{1,N} + \widehat{\sigma}_{v,N}^{-2} Q_{0,N}]\} u_N + \\
& (\widehat{\rho}_N^2 - \rho^2)(NT)^{-1/2} X_N' \{[I_T \otimes (W_N' W_N)][\widehat{\sigma}_{1,N}^{-2} Q_{1,N} + \widehat{\sigma}_{v,N}^{-2} Q_{0,N}]\} u_N + \\
& (\widehat{\sigma}_{1,N}^{-2} - \sigma_1^{-2})(NT)^{-1/2} X_N' \{[I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)] Q_{1,N}\} u_N + \\
& (\widehat{\sigma}_{v,N}^{-2} - \sigma_v^{-2})(NT)^{-1/2} X_N' \{[I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)] Q_{0,N}\} u_N.
\end{aligned}
$$

Next we note that under our maintained assumptions the elements of X_N, W_N, $Q_{0,N}$ and $Q_{1,N}$ are uniformly bounded in absolute value by k_x, k_w, 1, 1, respectively. It then follows from Lemma 1 that the elements of

$$(NT)^{-1}X_N'\{[I_T \otimes (W_N' + W_N)]Q_{1,N}\}X_N,$$

$$(NT)^{-1}X_N'\{[I_T \otimes (W_N' + W_N)]Q_{0,N}\}X_N,$$

$$(NT)^{-1}X_N'\{[I_T \otimes (W_N'W_N)]Q_{1,N}\}X_N,$$

$$(NT)^{-1}X_N'\{[I_T \otimes (W_N'W_N)]Q_{0,N}\}X_N,$$

$$(NT)^{-1}X_N'\{[I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)]Q_{1,N}\}X_N,$$

$$(NT)^{-1}X_N'\{[I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)]Q_{0,N}\}X_N,$$

are uniformly bounded in absolute value by $2k_x^2k_w$, $2k_x^2k_w$, $k_x^2k_w^2$, $k_x^2k_w^2$, $k_x^2(1+k_w)^2$, $k_x^2(1+k_w)^2$, respectively. Condition (A.37) then follows from (A.40), since $\widehat{\rho}_N$, $\widehat{\sigma}_{v,N}^2$ and $\widehat{\sigma}_{1,N}^2$ are consistent estimators of ρ, σ_v^2 and σ_1^2, respectively.

Next consider the following expressions in (A.41):

$$\begin{aligned}
\delta_{1,N} &= (NT)^{-1/2}X_N'\{[I_T \otimes (W_N' + W_N)]Q_{1,N}\}u_N, \\
\delta_{2,N} &= (NT)^{-1/2}X_N'\{[I_T \otimes (W_N' + W_N)]Q_{0,N}\}u_N, \\
\delta_{3,N} &= (NT)^{-1/2}X_N'\{[I_T \otimes (W_N' + W_N)]Q_{1,N}\}u_N, \\
\delta_{4,N} &= (NT)^{-1/2}X_N'\{[I_T \otimes (W_N' + W_N)]Q_{0,N}\}u_N, \\
\delta_{5,N} &= (NT)^{-1/2}X_N'\{[I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)]Q_{1,N}\}u_N, \\
\delta_{6,N} &= (NT)^{-1/2}X_N'\{[I_T \otimes (I_N - \rho W_N')(I_N - \rho W_N)]Q_{0,N}\}u_N.
\end{aligned} \tag{A.42}$$

It is evident that $E\delta_{i,N} = 0$, for $i = 1, \ldots, 6$. Using (7), (13) and $P_N(\rho) = [I_N - \rho W_N]^{-1}$ the variance-covariance matrix of $\delta_{i,N}$ for $i = 1, \ldots, 6$ is

$$E\delta_{i,N}\delta'_{i,N} = (NT)^{-1}X'_N\Phi_i X_N \qquad i = 1, \ldots, 6$$

where

$$\begin{aligned}
\Phi_1 &= \sigma_1^2\{I_T \otimes [(W'_N + W_N)P_N(\rho)P_N(\rho)'(W_N + W'_N)]\}Q_{1,N}, \\
\Phi_2 &= \sigma_v^2\{I_T \otimes [(W'_N + W_N)P_N(\rho)P_N(\rho)'(W_N + W'_N)]\}Q_{0,N}, \\
\Phi_3 &= \sigma_1^2\{I_T \otimes [(W'_N W_N)P_N(\rho)P_N(\rho)'(W_N W'_N)]\}Q_{1,N}, \\
\Phi_4 &= \sigma_v^2\{I_T \otimes [(W'_N W_N)P_N(\rho)P_N(\rho)'(W_N W'_N)]\}Q_{0,N}, \\
\Phi_5 &= \sigma_1^2(\frac{J_T}{T} \otimes I_N) = \sigma_1^2 Q_{1,N}, \\
\Phi_6 &= \sigma_v^2((I_T - \frac{J_T}{T}) \otimes I_N) = \sigma_1^2 Q_{0,N}.
\end{aligned} \tag{A.43}$$

Under our maintained assumptions the row and column sums of W_N, $P_N(\rho)$, $Q_{1,N}$ and $Q_{0,N}$ are uniformly bounded in absolute value. It then follows from Lemma 1 that row and column sums of matrices Φ_i are also uniformly bounded by some finite constant, say k_i ($i = 1, \ldots, 6$). Since the elements of X_N are uniformly bounded in absolute value by k_x, the elements of the variance-covariance matrices $(NT)^{-1}X'_N\Phi_i X_N$ are bounded in absolute value by $k_x^2 k_i < \infty$. It then follows from, for example, Corollary 5.1.1.2 in Fuller (1976) pg. 186, that the elements of $\delta_{i,N}$ for $i = 1, \ldots, 6$ in (A.42) are $O_p(1)$.

Condition (A.38) is seen to hold from (A.41) because $\widehat{\rho}_N$, $\widehat{\sigma}^2_{v,\,N}$ and $\widehat{\sigma}^2_{1,\,N}$ are consistent estimators of ρ, σ^2_v and σ^2_1, respectively.

Part (c) of the theorem follows immediately from (A.32) and Assumption 3 and also the fact that $\widehat{\rho}_N$, $\widehat{\sigma}^2_{v,N}$ and $\widehat{\sigma}^2_{1,N}$ are consistent estimators of ρ, σ^2_v and σ^2_1, respectively.

Next we prove that under Assumptions 1 to 5, the OLS estimator for the model in (4)

$$y_N = X_N\beta + u_N,$$

$\widehat{\beta}_{OLS,N} = [X_N'X_N]^{-1}X_N'y_N$ is $(NT)^{1/2}$ consistent, where $N \geq 1$ and T is a fixed positive integer. Observe that

$$\begin{aligned} (NT)^{1/2}(\widehat{\beta}_{OLS,N} - \beta) &= [(NT)^{-1}X_N'X_N]^{-1}(NT)^{-1/2}X_N'u_N \\ &= [(NT)^{-1}X_N'X_N]^{-1}(NT)^{-1/2}Z_N'\varepsilon_N, \end{aligned}$$

where u_N is defined in (21) and Z_N is defined here as $Z_N = (I_T \otimes (I_N - \rho W_N')^{-1})X_N$. Note that in general the elements of Z_N will depend on the sample size.

Recall from (6) that

$$\varepsilon_N = (e_T \otimes I_N)\mu_N + v_N,$$

Furthermore, by Assumption 1 the error terms, $v_{11,N}, \ldots, v_{N1,N}, \ldots, v_{1T,N}$,

$\ldots, v_{NT,N}$ are identically and independently distributed and for each $N \geq 1$ jointly independent with mean zero and finite variance σ_v^2. Similarly the error terms, $\mu_{1,N}, \ldots, \mu_{N,N}$ are identically and independently distributed and for each $N \geq 1$ jointly independent with mean zero and finite variance σ_μ^2. In addition, the two processes $\{v_{it,N}\}$ and $\{\mu_{i,N}\}$ are independent of each other.

Now we express

$$(NT)^{-1/2} Z_N' \varepsilon_N = (NT)^{-1/2} Z_N' (e_T \otimes I_N) \mu_N + (NT)^{-1/2} Z_N' v_N.$$

Assumption 3 and (8) implies that

$$\begin{aligned}
\lim_{N\to\infty} (NT)^{-1} Z_N' (e_T e_T' \otimes I_N) Z_N &= \lim_{N\to\infty} (NT)^{-1} T Z_N' (\frac{e_T e_T'}{T} \otimes I_N) Z_N \\
&= \lim_{N\to\infty} (NT)^{-1} T Z_N' Q_{1,N} Z_N \\
&= T \overline{Q}_{xQ_1x}(\rho),
\end{aligned}$$

$$\begin{aligned}
\lim_{N\to\infty} (NT)^{-1} Z_N' Z_N &= \lim_{N\to\infty} (NT)^{-1} X_N' (I_T \otimes (I_N - \rho W_N)^{-1} (I_N - \rho W_N')^{-1}) X_N \\
&= \overline{Q}_{xx}(\rho).
\end{aligned}$$

Then by Theorem A

$$\begin{aligned}
&(NT)^{-1/2} Z_N' (e_T \otimes I_N) \mu_N \xrightarrow{D} N\{0, \sigma_\mu^2 T \overline{Q}_{xQ_1x}(\rho)\}, \\
&(NT)^{-1/2} Z_N' v_N \xrightarrow{D} N\{0, \sigma_v^2 \overline{Q}_{xx}(\rho)\}, \qquad \text{(A.44)}
\end{aligned}$$

where $\overline{Q}_{xQ_1x}(\rho)$ and $\overline{Q}_{xx}(\rho)$, defined in Assumption 3, are finite and nonsingular. Next note that

$$\begin{aligned} \overline{Q}_{xx}(\rho) &= \overline{Q}_{xQ_0x}(\rho) + \overline{Q}_{xQ_1x}(\rho), \\ \sigma_1^2 &= T\sigma_\mu^2 + \sigma_v^2. \end{aligned}$$

Thus

$$(NT)^{-1/2} Z_N' v_N \xrightarrow{D} N\{0, \sigma_v^2(\overline{Q}_{xQ_0x}(\rho) + \overline{Q}_{xQ_1x}(\rho))\}.$$

By Assumption 1 the two processes $\{\mu_{i,N}\}$ and $\{v_{it,N}\}$ are independent of each other and therefore,

$$\begin{aligned} (NT)^{-1/2} Z_N' \varepsilon_N &= (NT)^{-1/2} Z_N'(e_T \otimes I_N)\mu_N + (NT)^{-1/2} Z_N' v_N \\ &\xrightarrow{D} [N\{0, \sigma_\mu^2 T \overline{Q}_{xQ_1x}(\rho)\} + N\{0, \sigma_v^2(\overline{Q}_{xQ_0x}(\rho) + \overline{Q}_{xQ_1x}(\rho))\}] \\ &= N\{0, [(\sigma_\mu^2 T + \sigma_v^2)\overline{Q}_{xQ_1x}(\rho) + \sigma_v^2 \overline{Q}_{xQ_0x}(\rho)]\} \qquad \text{(A.45)} \\ &= N\{0, [\sigma_1^2 \overline{Q}_{xQ_1x}(\rho) + \sigma_v^2 \overline{Q}_{xQ_0x}(\rho)]\}. \end{aligned}$$

From Assumption 3 we have

$$\lim_{N\to\infty} (NT)^{-1} X_N' X_N = Q_{xx}(\rho), \qquad \text{(A.46)}$$

where is $Q_{xx}(\rho)$ finite and nonsingular. It then follows from (A.45) and (A.46) that

$$
\begin{aligned}
& [\lim_{N\to\infty}(NT)^{-1}X_N'X_N]^{-1}(NT)^{-1/2}Z_N'\varepsilon_N \\
= \; & [Q_{xx}(\rho)]^{-1}(NT)^{-1/2}Z_N'\varepsilon_N \\
& \xrightarrow{D} N\{0,[Q_{xx}(\rho)]^{-1}(\sigma_1^2\overline{Q}_{xQ_1x}(\rho)+\sigma_v^2\overline{Q}_{xQ_0x}(\rho))[Q_{xx}(\rho)]^{-1}\}.
\end{aligned}
$$

Therefore,

$$
(NT)^{1/2}(\widehat{\beta}_{OLS,N}-\beta) \xrightarrow{D} N\{0,[Q_{xx}(\rho)]^{-1}(\sigma_1^2\overline{Q}_{xQ_1x}(\rho)+\sigma_v^2\overline{Q}_{xQ_0x}(\rho))[Q_{xx}(\rho)]^{-1}\},
$$

which implies that $\widehat{\beta}_{OLS,N}$ is $(NT)^{1/2}$ consistent.

Proof of Theorem 2. In order to prove this theorem we first need to show that $\widetilde{\sigma}^2_{v\;NLS,N}$ and $\widetilde{\sigma}^2_{1\;NLS,N}$, which are used in the weighting matrix $\widehat{\Theta}_N$ in the definition of the GM estimator in (41), are consistent estimates of σ_v^2 and σ_1^2.

The existence and measurability of $\widetilde{\rho}_{NLS,N}$, $\widetilde{\sigma}^2_{v\;NLS,N}$ and $\widetilde{\sigma}^2_{1\;NLS,N}$ which are defined in (40), are ensured by, for example, Lemma 2 in Jennrich (1969). The objective function of the nonlinear least squares estimator and its corresponding nonstochastic counterpart are given by, respectively,

$$
\begin{aligned}
R_N(\underline{\theta}) &= [G_N[\underline{\rho},\underline{\rho}^2,\underline{\sigma}_v^2,\underline{\sigma}_1^2]'-g_N]'[G_N[\underline{\rho},\underline{\rho}^2,\underline{\sigma}_v^2,\underline{\sigma}_1^2]'-g_N], \\
\overline{R}_N(\underline{\theta}) &= [\Gamma_N[\underline{\rho},\underline{\rho}^2,\underline{\sigma}_v^2,\underline{\sigma}_1^2]'-\gamma_N]'[\Gamma_N[\underline{\rho},\underline{\rho}^2,\underline{\sigma}_v^2,\underline{\sigma}_1^2]'-\gamma_N].
\end{aligned}
$$

where $\underline{\theta} = (\underline{\rho}, \underline{\sigma}_v^2, \underline{\sigma}_1^2)'$. To prove the consistency of $(\widetilde{\rho}_{NLS,N}, \widetilde{\sigma}^2_{v\ NLS,N}, \widetilde{\sigma}^2_{1\ NLS,N})$, we show that conditions of, for example, Lemma 3.1 in Pötscher and Prucha (1991a) are satisfied for the problem at hand. We first show that $\theta = (\rho, \sigma_v^2, \sigma_1^2)'$ is identifiably unique. By (36) we have

$$\gamma_N = \Gamma_N[\rho, \rho^2, \sigma_v^2, \sigma_1^2].$$

Hence $\overline{R}_N(\theta) = 0$ and

$$\begin{aligned}
& \overline{R}_N(\underline{\theta}) - \overline{R}_N(\theta) \\
= \; & [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2]\Gamma_N'\Gamma_N \\
& [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2]' \\
\geq \; & \lambda_{\min}(\Gamma_N'\Gamma_N)[\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2] \\
& [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2]' \\
\geq \; & \lambda_*[\underline{\rho} - \rho, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2][\underline{\rho} - \rho, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2]' = \lambda_* \left|\left|\underline{\theta} - \theta\right|\right|^2
\end{aligned}$$

utilizing Assumption 6. Hence for every $\epsilon > 0$ and any N, we have

$$\inf_{\{\underline{\theta} : ||\underline{\theta} - \theta|| \geq \epsilon} [\overline{R}_N(\underline{\theta}) - \overline{R}_N(\theta)] \geq \inf_{\{\underline{\theta} : ||\underline{\theta} - \theta|| \geq \epsilon} \lambda_* \left|\left|\underline{\theta} - \theta\right|\right|^2 = \lambda_* \epsilon^2 > 0,$$

which proves that θ is identifiably unique. Next, let $F_N = [G_N, -g_N]$ and

$\Phi_N = [\Gamma_N, -\gamma_N]$; then for $\rho \in [-a, a]$, $\sigma_v^2 \in [0, b_\nu]$ and $\sigma_1^2 \in [0, b_1]$

$$\begin{aligned}
\left|R_N(\underline{\theta}) - \overline{R}_N(\underline{\theta})\right| &= \left|[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1][F_N' F_N - \Phi_N' \Phi_N][\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1]\right| \\
&\leq \left|\left|F_N' F_N - \Phi_N' \Phi_N\right|\right| \left|\left|\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1\right|\right|^2 \\
&\leq \left|\left|F_N' F_N - \Phi_N' \Phi_N\right|\right| [1 + a^2 + a^4 + b_\nu^2 + b_1^2].
\end{aligned}$$

Lemmata 5 and 7 imply that $F_N - \Phi_N \xrightarrow{p} 0$ and that the elements of F_N and Φ_N are $O_p(1)$ and $O(1)$, respectively. It then follows that $R_N(\underline{\theta}) - \overline{R}_N(\underline{\theta})$ converge to zero uniformly over the (extended) parameter space, that is,

$$\begin{aligned}
&\sup_{\rho \in [-a,a], \sigma_v^2 \in [0, b_v], \sigma_1^2 \in [0, b_1]} \left|R_N(\underline{\theta}) - \overline{R}_N(\underline{\theta})\right| \\
\leq\ &\left|\left|F_N' F_N - \Phi_N' \Phi_N\right|\right| [1 + a^2 + a^4 + b_\nu^2 + b_1^2] \xrightarrow{p} 0
\end{aligned}$$

as $N \to \infty$. The consistency of $(\widetilde{\rho}_{NLS,N}, \widetilde{\sigma}^2_{v\ NLS,N}, \widetilde{\sigma}^2_{1\ NLS,N})$ now follows directly from Lemma 3.1 in Pötscher and Prucha (1991a).

Given that $\widetilde{\sigma}^2_{v\ NLS,N}$ and $\widetilde{\sigma}^2_{1\ NLS,N}$ are consistent estimates of σ_v^2 and σ_1^2, we now show the consistency of $\widehat{\rho}_{NLS,N}$, $\widehat{\sigma}^2_{v\ NLS,N}$ and $\widehat{\sigma}^2_{1\ NLS,N}$ which are defined in (41).

The existence and measurability of $\widehat{\rho}_{NLS,N}$, $\widehat{\sigma}^2_{v\ NLS,N}$ and $\widehat{\sigma}^2_{1\ NLS,N}$ are ensured by, for example, Lemma 2 in Jennrich (1969). The objective function of the nonlinear least squares estimator and its corresponding nonstochastic counterpart are given by, respectively,

$$\begin{aligned}
R_N(\underline{\theta}) &= [G_N[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2]' - g_N]' \widehat{\Theta}_N^{-1} [G_N[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2]' - g_N], \\
\overline{R}_N(\underline{\theta}) &= [\Gamma_N[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2]' - g_N]' \Theta_N^{-1} [\Gamma_N[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2]' - \gamma_N].
\end{aligned}$$

where $\underline{\theta} = (\underline{\rho}, \underline{\sigma}_v^2, \underline{\sigma}_1^2)'$.

We now establish two preliminary results implied by the assumptions. First we show that $\lambda_{\min}(\Gamma_N' \Theta_N^{-1} \Gamma_N) \geq \lambda_\circ$ for some $\lambda_\circ > 0$. To see this let $A = (a_{ij}) = \Gamma_{0,N}' \Gamma_{0,N}$ and $B = (b_{ij}) = \Gamma_{1,N}' \Gamma_{1,N}$. Then in light of (36)

$$\Gamma_N' \Gamma_N = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & 0 & b_{13} \\ b_{21} & b_{22} & 0 & b_{23} \\ 0 & 0 & 0 & 0 \\ b_{31} & b_{32} & 0 & b_{33} \end{bmatrix}.$$

Hence in light of using Assumption 6,

$$\begin{aligned}
x'\Gamma_N'\Gamma_N x &= [x_1, x_2, x_3]A[x_1, x_2, x_3]' + [x_1, x_2, x_4]B[x_1, x_2, x_4]' \\
&\geq \lambda_{\min}(A)[x_1, x_2, x_3][x_1, x_2, x_3]' + \lambda_{\min}(B)[x_1, x_2, x_4][x_1, x_2, x_4]' \\
&\geq \lambda_* x'x,
\end{aligned}$$

and thus in light of, e.g., Rao (1973, p. 62)

$$\lambda_{\min}(\Gamma_N'\Gamma_N) = \inf_x \frac{x'\Gamma_N'\Gamma_N x}{x'x} \geq \lambda_* > 0.$$

Next observe that

$$\begin{aligned}
\lambda_{\min}(\Gamma_N'\Theta_N^{-1}\Gamma_N) &= \inf_x \frac{x'\Gamma_N'\Theta_N^{-1}\Gamma_N x}{x'x} \geq \lambda_{\min}(\Theta_N^{-1}) \inf_x \frac{x'\Gamma_N'\Gamma_N x}{x'x} \\
&= \lambda_{\min}(\Theta_N^{-1})\lambda_{\min}(\Gamma_N'\Gamma_N) \geq \lambda_\circ > 0,
\end{aligned}$$

with $\lambda_\circ = \overline{\lambda}_* \lambda_*$, since $\lambda_{\min}(\Theta_N^{-1}) \geq \overline{\lambda}_* > 0$ by assumption.

Second we show that $\Theta_N^{-1} = O(1)$. For notational convenience let $S_N = \Theta_N^{-1}$. To verify the claim we need to show that $|s_{ij,N}| \leq k < \infty$ for some constant k that does not depend on N. Again in light of Rao (1973, p. 62)

$$\lambda_{\min}(S_N) = \inf_x \frac{xS_Nx}{x'x}, \qquad \lambda_{\max}(S_N) = \max_x \frac{xS_Nx}{x'x}.$$

Hence it follows from the maintained assumptions concerning the smallest and largest eigenvalues of $S_N = \Theta_N^{-1}$ that

$$0 < \overline{\lambda}_* \leq \frac{xS_Nx}{x'x} \leq \overline{\lambda}_{**} < \infty.$$

Taking x to be a vector that has a one in the $i-th$ and $j-th$ positions and zeros elsewhere we have

$$\begin{aligned} 0 &< \overline{\lambda}_* \leq s_{ii,N} \leq \overline{\lambda}_{**} < \infty, && i = j, \\ 0 &< \overline{\lambda}_* \leq (s_{ii,N} + s_{jj,N} + 2s_{ij,N})/2 \leq \overline{\lambda}_{**} < \infty && i \neq j. \end{aligned}$$

From this it is readily seen that $|s_{ij,N}| \leq \overline{\lambda}_{**}$ for all i, j which proves the claim.

To prove the consistency of $(\widehat{\rho}_{NLS,N}, \widehat{\sigma}^2_{v\ NLS,N}, \widehat{\sigma}^2_{1\ NLS,N})$, we show that conditions of, for example, Lemma 3.1 in Pötscher and Prucha (1991a) are satisfied for the problem at hand. We first show that $\theta = (\rho,\ \sigma_v^2,\ \sigma_1^2)'$ is identifiably unique. By (36) we have

$$\gamma_N = \Gamma_N[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2].$$

Hence $\overline{R}_N(\theta) = 0$ and

$$\begin{aligned}
& \overline{R}_N(\underline{\theta}) - \overline{R}_N(\theta) \\
= \; & [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2]\Gamma_N'\Theta_N^{-1}\Gamma_N \\
& [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2]' \\
\geq \; & \lambda_{\min}(\Gamma_N'\Theta_N^{-1}\Gamma_N)[\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2] \\
& [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2]' \\
\geq \; & \lambda_\circ[\underline{\rho} - \rho, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2][\underline{\rho} - \rho, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2]' = \lambda_\circ \left|\left|\underline{\theta} - \theta\right|\right|^2
\end{aligned}$$

utilizing Assumption 6. Hence for every $\epsilon > 0$ and any N, we have

$$\inf_{\{\underline{\theta} : ||\underline{\theta} - \theta|| \geq \epsilon} [\overline{R}_N(\underline{\theta}) - \overline{R}_N(\theta)] \geq \inf_{\{\underline{\theta} : ||\underline{\theta} - \theta|| \geq \epsilon} \lambda_\circ \left|\left|\underline{\theta} - \theta\right|\right|^2 = \lambda_\circ \epsilon^2 > 0$$

which proves that θ is identifiably unique. Next, let $F_N = [G_N, -g_N]$ and $\Phi_N = [\Gamma_N, -\gamma_N]$; then for $\rho \in [-a, a]$, $\sigma_v^2 \in [0, b_\nu]$ and $\sigma_1^2 \in [0, b_1]$

$$\begin{aligned}
\left|R_N(\underline{\theta}) - \overline{R}_N(\underline{\theta})\right| \; &= \; \left|[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1][F_N'\widehat{\Theta}_N^{-1}F_N - \Phi_N'\Theta_N^{-1}\Phi_N][\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1]\right| \\
&\leq \; \left|\left|F_N'\widehat{\Theta}_N^{-1}F_N - \Phi_N'\Theta^{-1}\Phi_N\right|\right| \left|\left|\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1\right|\right|^2 \\
&\leq \; \left|\left|F_N'\widehat{\Theta}_N^{-1}F_N - \Phi_N'\Theta^{-1}\Phi_N\right|\right| [1 + a^2 + a^4 + b_\nu^2 + b_1^2]
\end{aligned}$$

Lemmata 5 and 7 imply that $F_N - \Phi_N \xrightarrow{p} 0$ and that the elements of F_N and Φ_N are $O_p(1)$ and $O(1)$, respectively. We also note that by the consistency of $\widetilde{\sigma}^2_{v\ NLS,N}$ and $\widetilde{\sigma}^2_{1\ NLS,N}$ and the definition of $\widehat{\Theta}_N$ in (41), $\widehat{\Theta}_N - \Theta_N \xrightarrow{p} 0$, furthermore, the elements of $\widehat{\Theta}_N$ and Θ_N are $O_p(1)$ and $O(1)$, respectively. It then follows that $R_N(\underline{\theta}) - \overline{R}_N(\underline{\theta})$ converge to zero uniformly over the (extended) parameter space, that is,

$$\begin{aligned} & \sup_{\rho \in [-a,a], \sigma_v^2 \in [0,b_v], \sigma_1^2 \in [0,b_1]} \left| R_N(\underline{\theta}) - \overline{R}_N(\underline{\theta}) \right| \\ \leq & \left\| F_N' \widehat{\Theta}_N^{-1} F_N - \Phi_N' \Theta_N^{-1} \Phi_N \right\| [1 + a^2 + a^4 + b_\nu^2 + b_1^2] \xrightarrow{p} 0 \end{aligned}$$

as $N \to \infty$. The consistency of $(\widehat{\rho}_{NLS,N}, \widehat{\sigma}^2_{v\ NLS,N}, \widehat{\sigma}^2_{1\ NLS,N})$ now follows directly from Lemma 3.1 in Pötscher and Prucha (1991a).

5 Appendix to Chapter 3

5.1 A Linear City Model of Product Differentiation with Heterogenous Consumers

Consider a city that can be represented as lying on a line segment of length 1, as shown in the figure

There is a continuum of consumers whose total number is N who are assumed to be uniformly located along this straight line segment. A consumers location is indexed by $z \in [0, 1]$, the distance from the left end of the city. At each end of city is located one supplier of a commodity: Firm 1 is located at the left end of the city and Firm 2 is located at the right. This commodity is produced at constant cost of $c > 0$. The consumers are heterogeneous in terms of the quantity demanded. Without loss of generality, let us assume that consumers located between $[0, z^*]$ demand d_1 units of commodity, consumers located between $[z^*, z^{**}]$ demand d_2 and consumers located between $[z^{**}, 1]$ demand d_3 units of commodity and derive gross benefits v_1, v_2 and v_3, respectively from its consumption. The total cost of buying from firm j

for a consumer located at distance x from firm j is $p_j + tx$, where p_j is the price charged by firm j and $t/2$ is the cost or disutility per unit of distance traveled by the consumer in going to and from the firm $j's$ location. The presence of travel cost induces product differentiation between firms. As a simplifying assumption let $z^* < 1/2 < z^{**}$.

Ignoring the possibility of nonpurchase by consumers, will purchase from the location with the minimum delivered price, that is, price charged at the firm and also the cost of travelling to and from the firm. Given pair of prices, p_1 and p_2 charged by firm 1 and firm 2, respectively, consumers at locations $[0, \widehat{z})$ buy from firm 1. At these locations $p_1 + tz < p_2 + t(1-z)$ (purchasing from firm 1 is better that purchasing from firm 2). Consumers at locations $(\widehat{z}, 1]$ will buy form firm 2. The location of the consumer who is indifferent between the two firms is the point $\widehat{z}$ such that

$$p_1 + t\widehat{z} = p_2 + t(1 - \widehat{z}),$$

or

$$\widehat{z} = \frac{t + p_2 - p_1}{2t}. \tag{B.1}$$

Given p_1 and p_2, let $\widehat{z}$ be defined as in (B.1). Then firm 1's demand given a pair of prices (p_1, p_2), equal $d_1 z^* + d_2[\widehat{z} - z^*]$, when $z^* < \widehat{z} < z^{**}$, $d_1\widehat{z}$ when $0 < \widehat{z} < z^*$, $d_1 z^* + d_2[z^{**} - z^*] + d_3\widehat{z}$ when $z^{**} < \widehat{z} < 1$. Substituting for $\widehat{z}$

from (A1) we have

$$x_1(p_1,p_2) = \begin{cases} 0 & \quad p_1 > p_2 + t \\ d_1(\frac{t+p_2-p_1}{2t})N & \quad p_1 \in [p_2 + t - 2tz^*, p_2 + t] \\ [d_2(\frac{t+p_2-p_1}{2t}) \\ +(d_1 - d_2)z^*]N & \quad p_1 \in [p_2 + t - 2tz^{**}, p_2 + t - 2tz^*] \\ & if \\ [d_3(\frac{t+p_2-p_1}{2t}) \\ +(d_2 - d_3)z^{**} \\ +(d_1 - d_2)z^*]N & \quad p_1 \in [p_2 - t, p_2 + t - 2tz^{**}] \\ [d_3 + (d_2 - d_3)z^{**} \\ +(d_1 - d_2)z^*]N & \quad p_1 < p_2 - t \end{cases} \tag{B.2}$$

and by symmetry of the two firms the demand function of firm 2, $x_2(p_1,p_2)$

$$x_2(p_1,p_2) = \begin{cases} 0 & \quad p_2 > p_1 + t \\ d_3(\frac{t+p_1-p_2}{2t})N & \quad p_2 \in [p_1 - t + 2tz^{**}, p_1 + t] \\ [d_2(\frac{t+p_1-p_2}{2t}) \\ +(d_3 - d_2)(1 - z^{**})]N & \quad p_2 \in [p_1 - t + 2tz^*, p_1 - t + 2tz^{**}] \\ & if \\ [d_1(\frac{t+p_1-p_2}{2t}) \\ +(d_2 - d_1)(1 - z^*) \\ +(d_3 - d_2)(1 - z^{**})]N & \quad p_2 \in [p_1 - t, p_1 - t + 2tz^*] \\ [d_3 + (d_2 - d_3)z^{**} \\ +(d_1 - d_2)z^*]N & \quad p_2 < p_1 - t \end{cases} . \tag{B.3}$$

I assume that $(d_1, d_2, d_3, z^*, z^{**})$ are such that firm 1 in searching for its best response to any price choice of firm 2 will restrict itself prices in the interval $[p_2 + t - 2tz^{**}, p_2 + t - 2tz^*]$ and firm 2 in searching for its best response to any price choice of firm 1 will restrict itself prices in the interval $p_2 \in [p_1 - t + 2tz^*, p_1 - t + 2tz^{**}]$. Thus, firm 1 best response to firm 2 prices, p_2 solves

$$\max_{p_1}(p_1 - c)(d_2(\frac{t + p_2 - p_1}{2t}) + (d_1 - d_2)z^*)N$$

$$s.t. p_1 \in [p_2 + t - 2tz^{**}, p_2 + t - 2tz^*]$$

The necessary and sufficient (Kuhn-Tucker) first order conditions (foc) for this problem (assuming interior solution) is

$$t + p_2 + c - 2p_1 + \frac{(d_1 - d_2)}{d_2} 2tz^* = 0. \tag{B.4}$$

Solving the (foc) in (B.4) yields

$$p_1 = \frac{t + p_2 + c}{2} + \frac{(d_1 - d_2)}{d_2} tz^*. \tag{B.5}$$

Similarly we can solve the (foc) for firm 2 given the prices of firm 1, p_1 which would yield

$$p_2 = \frac{t + p_1 + c}{2} + \frac{(d_3 - d_2)}{d_2} tz^{**}. \tag{B.6}$$

It is evident from (B.5) and (B.6) that p_i for $i = 1, 2$, is inversely related to d_2. This implies that price is determined by the level of demand of the marginal customer. However, if the agents were identical, that is, $d_1 = d_2 = d_3$, then prices are independent of the level of demand.

6 List of Tables

6.1 Table 1: Estimation results, the weighting matrix is based on measures of closeness by actual road distance, for third week of August 1999

	EQUATION (60)		EQUATION (62)	
VARIABLES	**OLS**	**2SLS**	**OLS**	**GS2SLS**
$D_N PR99$	.811 **	.578 **	.900 **	.574 **
	(.032)	(.090)	(.022)	(.091)
CRPR99	.227 **	.501 **	.126 **	.513 **
	(.066)	(.121)	(.047)	(.125)
$PADD_1$	-.004	-2.232 *	-.372	-2.20 *
	(.885)	(1.25)	(.602)	(1.30)
$PADD_2$	-.389	-2.546 **	.040	-2.54 **
	(.701)	(1.084)	(.474)	(1.12)
$PADD_3$	-.707	-.967	-.856	-.934
	(1.20)	(1.44)	(.822)	(1.51)
$PADD_4$	-.155	.563	-.459	.561
	(.71)	(.816)	(.490)	(.85)
SPOT99	.064	.134 **	.030	.132 **
	(.032)	(.043)	(.024)	(.043)
PERST99	.315**	.477 **	.222 **	.487 **
	(.14)	(.164)	(.099)	(.170)
CO99	-.007	-.016	.004	-.017
	(.021)	(.023)	(.017)	(.023)
POP99	-.281**	-.283 **	-.283 **	-.285 **
	(.091)	(.099)	(.089)	(.099)
$D_N POP99$	.190	.160	.233	.152
	(.091)	(.100)	(.089)	(.100)
EA99	.247**	.292 **	.175 *	.302 **
	(.113)	(.126)	(.090)	(.126)
INC99	-.128	-.178	-.077	-.185 *
	(.100)	(.110)	(.081)	(.111)
$D_N INC99$	-.108**	-.107 **	-.095**	-.108 **
	(.033)	(.036)	(.033)	(.036)
ρ_{99}	-.31	.06		
Moran I	-4.67	.40		
Observations	289	289	289	289
R^2	.92		.97	

Dependent variable *PR99*. The weighting matrix is based on measures of closeness by actual road distance. Standard errors are in brackets.
Note: Column 4 and 5 are estimates after correcting for spatial correlation in the disturbances.
(*) indicates significance at 10%. (**) indicates significance at 5%.

6.2 Table 2: Estimation results, the weighting matrix is based on measures of closeness by actual road distance, for third week of January 2000

	EQUATION (60)		EQUATION (62)	
VARIABLES	OLS	2SLS	OLS	GS2SLS
$D_N PR00$	.691**	.689**	.875**	.835**
	(.046)	(.127)	(.032)	(.093)
$CRPR00$	.385**	.388**	.180**	.220**
	(.090)	(.150)	(.063)	(.104)
$PADD_1$	-1.47**	-1.473**	-.833	-.875
	(.738)	(.695)	(.527)	(.536)
$PADD_2$	-.505	-.505	-.246	-.255
	(.454)	(.626)	(.304)	(.307)
$PADD_3$	-1.928	-.930	-.538	-.555
	(.627)	(.782)	(.455)	(.460)
$PADD_4$	-2.43	-2.44**	-1.32**	-1.463**
	(.896)	(.713)	(.591)	(.670)
$SPOT00$	.024	.024	.005	-.003
	(.060)	(.06)	(.042)	(.046)
$PERST00$	.190**	.191**	.108*	.117*
	(.092)	(.097)	(.061)	(.064)
$CO00$	.015	.015	.011	.012
	(.019)	(.020)	(.015)	(.016)
$OXY00$	-.454	-.45	-.205	-.162
	(.538)	(.552)	(.468)	(.482)
$POP00$	-.153*	-.152*	-.137*	-.124
	(.081)	(.087)	(.077)	(.083)
$D_N POP00$	.172**	.172**	.115	.116
	(.082)	(.082)	(.076)	(.077)
$EA00$	-.009	.008	-.007	-.009
	(.092)	(.100)	(.070)	(.079)
$INC00$	.008	-.010	.002	-.009
	(.082)	(.087)	(.064)	(.069)
$D_N INC00$	-.056**	-.055**	-.020	-.021
	(.028)	(.028)	(.026)	(.027)
ρ_{00}	-.32	-.31		
Moran I	-4.36	-2.15		
Observations	289	289	289	289
R^2	.60		.81	

Dependent variable *PR00*. The weighting matrix is based on measures of closeness by actual road distance. Standard errors are in brackets.
Note: Column 4 and 5 are estimates after correcting for spatial correlation in the disturbances.

6.3 Table 3: Estimation results, the weighting matrix is based on measures of closeness by a Euclidean distance, for third week of August 1999

	EQUATION (60)		EQUATION (62)	
VARIABLES	OLS	2SLS	OLS	GS2SLS
$D_N PR99$	.802 ** (.031)	.526 ** (.092)	.893 ** (.022)	.517 ** (.092)
$CRPR99$	.218 ** (.066)	.552 ** (.127)	.114 ** (.047)	.580 ** (.133)
$PADD_1$	-.311 (.876)	-2.86 ** (1.26)	-.112 (.594)	-2.86 ** (1.36)
$PADD_2$	-.556 (.695)	-3.068 ** (1.10)	.096 (.469)	-3.08 ** (1.18)
$PADD_3$	.343 ** (1.19)	-1.54 (1.46)	-.587 (.814)	-1.53 (1.61)
$PADD_4$	.37 (.702)	.815 (.828)	-.298 (.482)	.825 (.90)
$SPOT99$	.078 ** (.032)	.156 ** (.043)	.043* (.024)	.150 ** (.043)
$PERST99$	.269 * (.140)	.481 ** (.171)	.181 * (.098)	.498 ** (.183)
$CO99$	-.010 (.021)	-.019 (.024)	.001 (.017)	-.021 (.024)
$POP99$	-.197 ** (.091)	-.226 ** (.103)	-.180* (.091)	-.226 ** (.100)
$D_N POP99$	.156 * (.089)	.131 (.101)	.158 * (.089)	.121 (.099)
$EA99$	.209 * (.112)	.275 ** (.128)	.146 * (.088)	.287 ** (.129)
$INC99$	-.119 (.099)	-.178 (.113)	-.080 (.080)	-.190 * (.113)
$D_N INC99$	-.100 ** (.032)	-.109 ** (.036)	-.070 ** (.032)	-.113 ** (.035)
ρ_{99}	-.32	.12		
Moran I	-4.67	.83		
Observations	289	289	289	289
R^2	.92		.96	

Dependent variable *PR99*. The weighting matrix is based on measures of closeness by a Euclidean distance. Standard errors are in brackets.
Note: Column 4 and 5 are estimates after correcting for spatial correlation in the disturbances.
(*) indicates significance at 10%. (**) indicates significance at 5%.

6.4 Table 4: Estimation results, the weighting matrix is based on measures of closeness by a Euclidean distance, for third week of January 2000

	EQUATION (60)		EQUATION (62)	
VARIABLES	OLS	2SLS	OLS	GS2SLS
$D_N PR00$	.669**	.602**	.860**	.736**
	(.046)	(.125)	(.033)	(.103)
$CRPR00$	.400**	.466**	.180**	.308**
	(.092)	(149)	(.064)	(.118)
$PADD_1$	-1.522**	-1.587**	-.766	-1.04*
	(.750)	(.761)	(.540)	(.607)
$PADD_2$	-.530	.534	-.223	.308
	(.462)	(.464)	(.311)	(.353)
$PADD_3$	-.930	-.973	-.414	-.583
	(.637)	(.644)	(.464)	(.516)
$PADD_4$	-.257**	-2.785**	-1.44**	-2.003**
	(.914)	(.994)	(.610)	(.769)
$SPOT00$	.037	.048	.013	.033
	(.061)	(.064)	(.043)	(.051)
$PERST00$	.197**	.210**	.114*	.154**
	(.094)	(.097)	(.063)	(.075)
$CO00$	.019	.019	.014	.018
	(.019)	(.198)	(.016)	(.017)
$OXY00$	-.413	-.346	.005	-.028
	(.546)	(.56)	(.476)	(.517)
$POP00$	-.167**	-.147	-.133*	-.127
	(.083)	(.09)	(.079)	(.086)
$D_N POP00$	.208**	.207**	.130*	.161**
	(.082)	(.082)	(.078)	(.080)
$EA00$	.024	.0002	.023	-.009
	(.094)	(.103)	(.071)	(.088)
$INC00$	-.002	.015	-.017	.015
	(.083)	(.089)	(.066)	(.077)
$D_N INC00$	-.068**	-.067**	-.028	-.043
	(.028)	(.028)	(.026)	(.027)
ρ_{00}	-.30	-.28		
Moran I	-4.17	-1.48		
Observations	289	289	289	289
R^2	.58		.80	

Dependent variable *PR00*. The weighting matrix is based on measures of closeness by euclidean distance. Standard errors are in brackets.
Note: Column 4 and 5 are estimates after correcting for spatial correlation in the disturbances.

6.5 Table 5: Estimation results of the robustness test

	EQUATION (63)	
VARIABLES	HIGH DEMAND (t=99)	LOW DEMAND (t=00)
$D^{1}{}_{N}PR(t)$	.811 (.058)	.741 (.086)
$D^{2}{}_{N}PR(t)$	.808 (.053)	.723 (.085)
$D^{3}{}_{N}PR(t)$	.819 (.045)	.750 (.084)
$D^{4}{}_{N}PR(t)$	.791 (.075)	.720 (.096)
$D^{5}{}_{N}PR(t)$	.811 (.049)	.654 (.088)

Dependent variable *PR(t)*. The weighting matrix is based on measures of closeness by actual road distance. Standard errors are in brackets.

6.6 Table 6: Estimation results, after including tax as one of the explanatory variables, for third week of August 1999

	EQUATION (60)	EQUATION (62)
VARIABLES	2SLS	GS2SLS
$D_N PR99$	0.516** (0.097)	0.486** (0.097)
$CRPR99$	0.57** (0.13)	0.622** (0.137)
$PADD_1$	-4.26** (1.644)	-4.577** (1.757)
$PADD_2$	-4.06** (1.355)	-4.336** (1.44)
$PADD_3$	-3.93* (2.08)	-4.313** (2.244)
$PADD_4$	1.186 (0.903)	1.296 (0.993)
$SPOT99$	0.15** (0.045)	0.148** (0.045)
$PERST99$	0.192 (0.212)	0.188 (0.232)
$CO99$	-0.02 (0.024)	-0.026 (0.024)
$POP99$	-0.27** (0.104)	-0.278** (0.101)
$D_N POP99$	0.159 (0.104)	0.138 (0.102)
$EA99$	0.299** (0.13)	0.322** (0.132)
$INC99$	-0.2* (0.115)	-0.217** (0.116)
$D_N INC99$	-0.1** (0.037)	-0.106** (0.037)
TAX	0.091** (0.043)	0.106** (0.046)
ρ_{99}	.14	
Moran I	.93	
Observations	289	289
R^2	.90	.87

Dependent variable *PR99*. The weighting matrix is based on measures of closeness by actual road distance. Standard errors are in brackets.

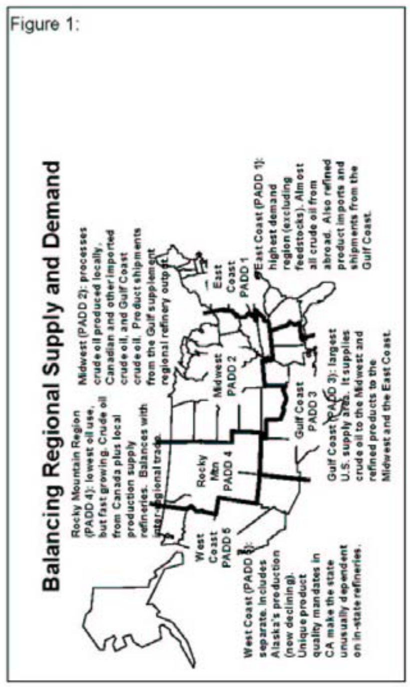
6.7 Table 7: Estimation results, after including tax as one of the explanatory variables, for third week of January 2000

	EQUATION (60)	EQUATION (62)
VARIABLES	2SLS	GS2SLS
$D_N PR00$	0.642** (0.129)	0.796** (0.101)
$CRPR00$	0.439** (0.15)	0.267** 0.113)
$PADD_1$	-1.00 (0.756)	-0.593 (0.571)
$PADD_2$	-0.31 (0.456)	-0.132 (0.328)
$PADD_3$	-1.1* (0.631)	-0.711 (0.482)
$PADD_4$	-2.03** (0.954)	-1.279* (0.688)
$SPOT00$	0.017 (0.062)	-0.002 (0.048)
$PERST00$	0.106 (0.097)	0.066 (0.07)
$CO00$	0.013 (0.019)	0.01 (0.016)
$OXY00$	-0.45 (0.546)	-0.166 (0.492)
$POP00$	-0.15* (0.086)	-0.133 (0.083)
$D_N POP00$	0.168** (0.081)	0.118 (0.077)
$EA00$	0.025 (0.098)	0.016 (0.08)
$INC00$	-0.01 (0.085)	-0.014 (0.07)
$D_N INC00$	-0.05* (0.028)	-0.023 (0.027)
TAX	0.064** (0.026)	0.044** (0.019)
ρ_{00}	-0.26	
Moran I	-1.77	
Observations	289	289
R^2	.60	.80

Dependent variable *PR00*. The weighting matrix is based on measures of closeness by actual road distance. Standard errors are in brackets.

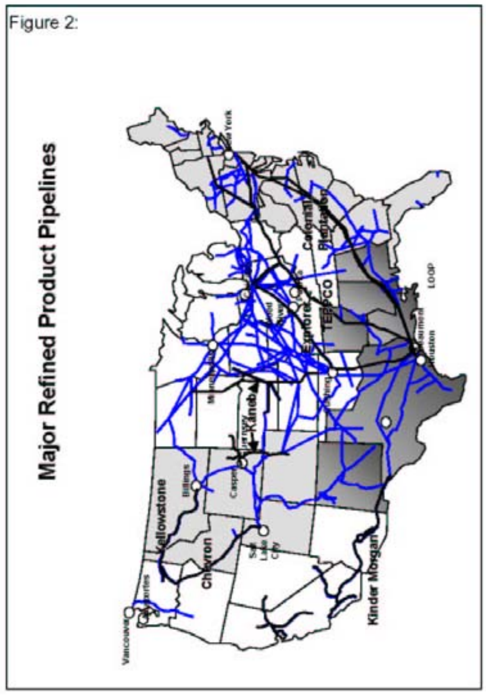
7 List of Figures

7.1 Figure 1: Petroleum Allocation for Defense Districts (PADDs)[49]



[49] This map has been taken from "How Pipelines make the Oil Market Work- Their Networks, Operation and Regulation," Allegro Energy Group.

7.2 Figure 2: Network of Oil pipelines in the US[50]



[50]This map has been taken from "How Pipelines make the Oil Market Work- Their Networks, Operation and Regulation," Allegro Energy Group.

7.3 Figure 3: The FRS companies in 2000[51]

Figure 3:

The FRS Companies in 2000 (* Denotes new survey entrant in 2000)	
Amerada Hess Corporation	Exxon Mobil Corporation
American Petrofina Holding Company	Kerr-McGee Corporation
(formerly known as Fina, Inc.)	Lyondell-CITGO Refining, L.P.
Anadarko Petroleum Corporation	Motiva Enterprises, L.L.C.
Apache Corporation *	Occidental Petroleum Corporation
BP America, Inc.	Phillips Petroleum Company
Burlington Resources, Inc.	Premcor, Inc.
Chevron Corporation	Shell Oil Company
CITGO Petroleum Corporation	Sunoco, Inc.
Coastal Corporation	Tesoro Petroleum Corporation
Conoco, Inc.	Texaco, Inc.
Devon Energy Corporation*	Tosco Corporation
Dominion Resources, Inc.*	Ultramar Diamond Shamrock Corporation
El Paso Energy Corporation	Unocal Corporation
Enron Corporation	USX Corporation
EOG Resources, Inc.*	Valero Energy Corporation
Equilon Enterprises, L.L.C.	The Williams Companies, Inc.

[51] This chart has been taken from "How Pipelines make the Oil Market Work- Their Networks, Operation and Regulation," Allegro Energy Group.

References

[1] Amemiya, T., 1971. The Estimation of the Variances in a Variance-Component Model. *International Economic Review* 12, 1-13.

[2] Amemiya, T., 1985. *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

[3] Anderson, S. P., A. D. Palma and J. Thisse (1989): "Demand for Differentiated Products, Discrete Choice Models, and the Characteristics Approach," *Review of Economic Studies,* 56, 21-35.

[4] Anselin, L., 1988. *Spatial Econometrics: Methods and Models.* Boston: Kluwer Academic Publishers.

[5] Audretsch, D.B. and Feldmann, M.P., 1996. R&D Spillovers and the Geography of Innovation and Production. *American Economic Review* 86, 630-640.

[6] Balestra, P. and Nerlove, M., 1966. Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. *Econometrica 34, 585-612.*

[7] Baltagi, B.H., 1980. On Seemingly Unrelated Regressions with Error Components. *Econometrica* 48, 1547-1551.

[8] Baltagi, B.H., 1981. Simultaneous Equations with Error Components. *Journal of Econometrics* 17, 189-200.

[9] Baltagi, B. H., 1995. *Econometric Analysis of Panel Data*. New York: Wiley.

[10] Baltagi, B.H. and D. Li, 1999. Prediction in Panel Data Models with Spatial Correlation. In L. Anselin and R.J.G.M. Raymond, eds., *New Advances in Spatial Econometrics*. New York: Springer Verlag, forthcoming.

[11] Baltagi, B.H. and D. Li, 2001a. Double Length Artificial Regressions for Testing Spatial Dependence. *Econometric Reviews* 20, 31-40.

[12] Baltagi, B.H. and D. Li, 2001b. LM Test for Functional Form and Spatial Error Correlation. *International Regional Science Review* 24, 194-225.

[13] Baltagi, B.H., S.H. Song, and W. Koh, 2001. Testing Panel Data Regression Models with Spatial Error Correlation, Department of Economics, Texas A&M University, mimeo.

[14] Bernat Jr., G., 1996. Does Manufacturing Matter? A Spatial Econometric View of Kaldor's Laws. *Journal of Regional Science* 36, 463-477.

[15] Bertrand, J. (1883): "Thėorie Mathėmatique de la Richesse Sociale," *Journal des Savants*, 67, 499-508.

[16] Besley, T. and Case, A., 1995. Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition. *American Economic Review* 85, 25-45.

[17] Betancourt, R. and M. Malanoski (1999): "An Estimable Model of Supermarket Behavior: Prices, Distribution Services and Some Effects of Competition," *Empirica,* 26, 55-73.

[18] Bollinger. C. and Ihlanfeldt, K., 1997. The Impact of Rapid Rail Transit on Economic Development: The Case of Atlanta's Marta. *Journal of Urban Economics* 42, 179-204.

[19] Borenstein, S., and A. Shepard (1996): "Dynamic Pricing in Retail Gasoline Markets," *Rand Journal of Economics,* 27, 429-451.

[20] Buettner, T., 1999. The Effect of Unemployment, Aggregate Wages, and Spatial Contiguity on Local Wages: An Investigation with German District Level Data. *Papers in Regional Science* 78, 47-67.

[21] Case, A., 1991. Spatial Patterns in Household Demand. *Econometrica* 59, 953-966.

[22] Case, A., Hines Jr., J., and Rosen, H., 1993. Budget Spillovers and Fiscal Policy Independence: Evidence from the States. *Journal of Public Economics* 52, 285-307.

[23] Chamberlin, E. (1933): "The Theory of Monopolistic Competition," *Cambridge, Mass.: Harvard University Press.*

[24] Cliff, A. and Ord J., 1973. *Spatial Autocorrelation.* London: Pion.

[25] Cliff, A. and Ord, J., 1981. *Spatial Processes, Models and Applications.* London: Pion.

[26] Conley, T., 1999. GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics* 92, 1-45.

[27] Cressie, N.A.C., 1993. *Statistics of Spatial Data.* New York: Wiley.

[28] Das, D., Kelejian, H.H., and Prucha, I.R., 2001. Small Sample Properties of Estimators of Spatial Autoregressive Models with Autoregressive Disturbances. *Papers in Regional Science*, forthcoming.

[29] Deneckere, R. and M. Rothschild (1992): "Monopolistic competition and Preference Diversity," *Review of Economic Studies*, 59, 361-373.

[30] Dixit, A. and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review,* 291-308.

[31] Dowd, M. R. and LeSage, J. P., 1997. Analysis of Spatial Contiguity Influences on State Price Level Formation. *International Journal of Forecasting* 13, 245-253.

[32] ESRI ArcView 3.3 and Network Analyst 1.0a.

[33] Feenstra, R. C., and J. A. Levinsohn (1995): "Estimating Markups and Market Conduct with Multidimensional Product Attributes," *Review of Economic Studies,* 62, 19-52.

[34] Gallant, A.R. and White, H., 1988. *A Unified Theory of Estimation and Inference in Nonlinear Dynamic Models*, New York: Basil Blackwell.

[35] Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Economterica,* 53, 891-951.

[36] Green, W. H. (2000): "Econometric Analysis," 4th edition, Prentice Hall.

[37] Haltiwanger, J. and J. E. Harrington Jr. (1991): "The Impact of Cyclical Demand Movements on Collusive Behavior," *Rand Journal of Economics,* 22, 89-106.

[38] Hanning, R. (1984): "Testing a Spatial Interacting-Markets Hypothesis," *Review of Economics and Statistics,* 576-583.

[39] Holtz-Eakin, D., 1994. Public Sector Capital and the Productivity Puzzle. *Review of Economics and Statistics* 76, 12-21.

[40] Horn, R. and Johnson, C., 1985. *Matrix Analysis.* New York: Cambridge University Press.

[41] Hotelling, H. (1929): "Stabilty in Competition," *Economic Journal*, 10, 41-57.

[42] "How Pipelines make the Oil Market Work- Their Networks, Operation and Regulation," *Association of Oil Pipelines and the American Petroleums Institute's Pipeline Committee.*

[43] Hsiao, C., 1986. *Analysis for Panel Data.* New York: Cambridge Press.

[44] Jennrich, R., 1969. Asymptotic Properties of Non-linear Least Squares Estimators. *The Annals of Mathematical Statistics* 40, 633-643.

[45] Kapoor, M., Kelejian, H. and I. R. Prucha (2002): "Panel Data Models with Spatially Correlated Error Components," *Working Paper,* University of Maryland.

[46] Kelejian, H.H. and Prucha., I.R., 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.

[47] Kelejian, H.H. and Prucha., I.R., 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* 17, 99-121.

[48] Kelejian, H.H. and Prucha., I.R., 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40, 509-533.

[49] Kelejian, H.H. and Prucha., I.R., 2001. On the Asymptotic Distribution of the Moran I Test Statistic with Applications. *Journal of Econometrics 104, 219-257.*

[50] Kelejian, H.H. and Prucha., I.R., 2002. Estimation of Systems of Spatially Interrelated Cross Sectional Equations. University of Maryland, manuscript. (Forthcoming in *Journal of Econometrics*).

[51] Kelejian, H. and Robinson, D., 1997. Infrastructure Productivity Estimation and its Underlying Econometric Specifications. *Papers in Regional Science* 76, 115-131.

[52] Kelejian, H. and Robinson, D., 2000. Returns to Investment in Navigation Infrastructure: An Equilibrium Approach. *Annals of Regional Science 34, 83-108.*

[53] Koopmans, T. J. (1957): "Three Essays on the State of Economic Science," *Essay 1,* New York: McGraw Hill.

[54] Lee, L.F., 1999a. Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. Department of Economics, HKUST, Hong Kong.

[55] Lee, L.F., 1999b. Asymptotic Distributions of Maximum Likelihood Estimators for Spatial Autoregressive Models. Department of Economics, HKUST, Hong Kong.

[56] Lee, L.-F., 2001a, Generalized Method of Moments Estimation of Spatial Autoregressive Processes, Department of Economics, Ohio State University, mimeo.

[57] Lee, L.-F., 2001b, GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models, Department of Economics, Ohio State University, mimeo.

[58] Lee, L.F., 2002. Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregessive Models. *Econometric Theory 18, 252-277.*

[59] LeSage, J. P., 1997. Bayesian Estimation of Spatial Autoregressive Models. *International Regional Science Review 20, 113-129.*

[60] LeSage, J. P., 1999. A Spatial Econometric Analysis of China's Economic Growth. *Journal of Geographic Information Sciences 5, 143-153.*

[61] LeSage, J. P., 2000. Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. *Geographic Analysis 32, 19-35.*

[62] Mas-Colell, A., M. D. Whinston, J. R. Green (1995): "Microeconomic Theory," Oxford University Press, New York.

[63] Maskin, E. and J. Tirole (1988): "A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs," *Econometrica,* 56, 549-569.

[64] Maskin, E. and J. Tirole (1988): "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles," *Econometrica,* 56, 571-599.

[65] McFadden, D. (1978): "The General Linear Profit Function," in *Production Economics: A Dual Approach to Theory and Applications,* Vol. 1, ed. by M. Fuss and D. McFadden. Amsterdam, The Netherlands: North Holland, 269-286.

[66] Nerlove, M., 1971. A Note on Error Component Models. *Econometrica 39, 383-396.*

[67] *Oil Price Information Service (* August 1999, January 2000) Rockville, Maryland.

[68] Pace, R. and Barry, R., 1997. Sparse Spatial Autoregressions. *Statistics and Probability Letters* 33, 291-297.

[69] Pinkse, J., and Slade, M.E., 1998. Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models. *Journal of Econometrics* 85, 125-154.

[70] Pinkse, J., Slade, M. E., and Brett, C., 2002. Spatial Price Competition: A Semiparametric Approach. *Econometrica* 70, 1111-53.

[71] Pötscher, B.M. and Prucha, I.R., 1997. *Dynamic Nonlinear Econometric Models, Asymptotic Theory.* New York: Springer Verlag.

[72] Pötscher, B.M. and Prucha, I.R., 2001. Basic Elements of Asymptotic Theory. In B.H. Baltagi, ed., *A Companion to Theoretical Econometrics.* Oxford: Blackwell.

[73] Prucha, I.R., 1984. On the Asymptotic Efficiency of Feasible Aitken Estimators for Seemingly Unrelated Regression Models with Error Components. *Econometrica* 52, 203-207.

[74] Prucha, I.R., 1985. Maximum Likelihood and Instrumental Variable Estimation in Simultaneous Equation Systems with Error Components. *International Economic Review* 26, 491-506.

[75] Rao, C.R., 1973. *Linear Statistical Inference and Its Applications.* New York: Wiley.

[76] Rey, S.J. and Boarnet, M.G., 1998. A Taxonomy of Spatial Econometric Models for Simultaneous Equation Systems. Department of Geography, San Diego State University.

[77] Rotemberg, J. J. and G. Saloner (1986): "A Supergame-Theoretic Model of Price Wars During Boom," *American Economic Review,* 76, 390-407.

[78] Salop, S. (1979): "Monopolistic Competition with Outside Goods," *Bell Journal of Economics*, 10, 141-156.

[79] Schmidt, P., 1976. *Econometrics.* New York: Marcel Dekker.

[80] Shroder, M., 1995. Games the States Don't Play: Welfare Benefits and the Theory of Fiscal Federalism. *Review of Economics and Statistics* 77, 183-191.

[81] Shubik, M. (1959): "Strategy and Market Structure," John Wiley & Sons, New York.

[82] Slade, M. E. (1986): "Exogeneity Test of Market Boundaries Applied to Petroleum Products," *Journal of Industrial Economics,* 34, 291-303.

[83] Spence, A. M. (1976): "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies*, 43, 217-235.

[84] Spiller, P. T., and C. J. Huang (1986): "On the Extent of Market: Wholesale Gasoline in the Northeastern United States," *Journal of Industrial Economics,* 35, 131-145.

[85] U.S. Department of Commerce, Census Bureau, (1999 and 2000).

[86] U.S. Department of Commerce, Census Bureau, (1999 and 2000): "Regional Economic Information System."

[87] U.S. Department of Labor, Bureau of Labor Statistics (1999 and 2000): "Occupational Employment and Wages."

[88] Vigil, R., 1998. Interactions Among Municipalities in the Provision of Police Services: A Spatial Econometric Approach. University of Maryland, Ph.D. Thesis.