ABSTRACT

Title of dissertation: DYNAMIC PANEL DATA MODELS WITH SPATIALLY CORRELATED DISTURBANCES

Jan Mutl, Doctor of Philosophy, 2006

Dissertation directed by: Professor Ingmar Prucha
Department of Economics

This thesis considers a dynamic panel data model with error components that are correlated both spatially (cross-sectionally) and time-wise. The model extends the literature on dynamic panel data models with cross-sectionally independent error components. The model for spatial dependence is a Cliff-Ord type model.

We introduce a three step estimation procedure and give formal large sample results for the case of a finite time dimension. In particular, we show that a simple first stage instrumental variable (IV) estimator, that ignores the spatial correlation of the errors, is consistent and $\sqrt{N}$-consistent, where N denotes the cross-sectional dimension. We then extend the generalized moments estimator introduced by Kelejian and Prucha (1999) for estimating the spatial autoregressive parameter and show that if it is based on a $\sqrt{N}$-consistently estimated disturbances, it will also be consistent. Finally, we derive a large sample distribution of a second stage generalized method of moments (GMM) estimator based on a consistent estimator of the spatial autoregressive parameter. We also present results from a small Monte Carlo study to illustrate the small sample performance of the proposed estimation procedure.

JEL Classification and Keywords: Cross-Sectional Models; Spatial Models (C21), Models with Panel Data (C23); Dynamic panels, Spatial Autocorrelation

DYNAMIC PANEL DATA MODELS WITH SPATIALLY CORRELATED DISTURBANCES

by

Jan Mutl

Dissertation submited to the Faculty of the Graduate School of the University of Maryland, College Park in partial fullfilment of the requirements for the degree of Doctor of Philosophy
2006

Advisory Committee:

Professor Ingmar Prucha, Chair
Professor John Chao, Co-Chair
Professor Harry Kelejian
Professor John Rust
Professor Francis Alt

Contents

LIST OF TABLES

LIST OF FIGURES

1 Introduction

This thesis considers estimation of panel data models when the dependent variable is allowed to be correlated in both dimensions. Using a natural terminology, I investigate models in which there is correlation both across time and between the cross-sectional units. Although there might be many ways to write down such model, I choose to concentrate on concrete specification that arises as an extension of the existing literature on dynamic panel data models and on spatial modelling. In doing so, I hope to offer a useful synthesis of the two strands of the literature. My model is applicable to situations where the number of time periods over which the data are observed is limited.[1]

In the next chapter, I review the existing literature related to this topic. I first focus on theoretical contributions to dynamic panels estimation methods, then briefly outline the specifications used in spatial econometrics, and close with a review of papers that have used specifications in which time and space are interacting in a nontrivial way.

Chapter 3 will then spell out the specification I chose to concentrate on. It will also provide the general assumptions maintained throughout the thesis and discuss some implication of the model.

In Chapter 4, I provide an outline of several estimation methods and provide a formal statements of their asymptotic properties. I start with an initial instru-

[1] Of course, if the time dimension of the panel is sufficiently large, one can consider, for example, a seemingly unrelated regression model that allows for a fairly general specification of the correlation pattern in the cross-sectional dimension.

mental variable (IV) technique suggested by Anderson and Hsiao (1981) to estimate the slope coefficients of the model. Although this method ignores possible cross-sectional correlation in the data, I show that it is still consistent and asymptotically normal under the specification considered in this thesis. Next, I outline a spatial generalized moments estimation technique that estimates the degree of cross-sectional dependence in the disturbances. The method was suggested by Kapoor et al. (2005) for a static model and is based on Kelejian and Prucha (1999). I extend the proofs in Kapoor et al. (2005) for the dynamic case. The last step of the proposed estimation method consists of a generalized method of moments (GMM) estimation of the slope coefficients. I discuss the optimal choice weighting matrix for a given set of moment conditions. I provide formal large sample results for a generic GMM estimator based on linear moment conditions with stochastic instruments. I also provide formal large sample properties of a feasible GMM estimator and its small sample covariance matrix approximation.

In Chapter 5, I investigate small sample properties of the different estimation method via a Monte Carlo study. I also provide some simulation evidence that supports the formal large sample claims made in the thesis.

2 Review of Literature

The purpose of this review is not to provide a comprehensive treatment of the econometric work that has been done on panel data methods. For such there are excellent book-length works, such as Hsiao (2003) or Baltagi (2002). Instead, I will provide a more in depth review of the theoretical work that has been done on dynamic panel data models on the one hand and then review the literature relaxing the assumption of independently and identically distributed (iid) errors both in panel and purely cross-sectional setting.

It proves to be useful to introduce the following notational conventions: I use bold letters for matrices and vectors, and regular font letters to denote scalars. Furthermore, I use lower case letters for vectors and upper case letters for matrices. In general I will denote the cross-sectional dimension of the panel as N and the time dimension as T.

2.1 Dynamic Panel Data Models

Models with individual effects and limited time dimension face the problem of incidental parameters. Hence these are estimated after a suitable transformation that removes the individual effects. In most cases this would be after first differencing. If the model also includes a lagged endogenous variable, the first difference of the error term will then be correlated with the explanatory variables. It has been long recognized in the literature that in this situation, the ordinary least squares (OLS) estimator will be biased, see, e.g., Trognon (1978) for an analytical treatment, or Nerlove (1967 and 1971) who explores the properties of the bias of the OLS estimation by Monte Carlo work. Trognon (1978), Nickell (1981) and Sevestre and Trognon (1985) derive analytical expressions for the asymptotic biases of the OLS estimator of an autoregressive panel data models with fixed time dimension. Small sample bias correction has also been suggested by Kiviet (1995).

The bias of the OLS estimation also resulted in attention to other estimation methods. Hence Anderson and Hsiao (1981, 1982) discuss maximum likelihood (ML) estimation of various model specifications and provide a comprehensive classification of the different conceptual possibilities of dynamic panel data models. They also suggest a simple instrumental variables (IV) estimator that is consistent. Bhargava and Sargan (1983) provide a framework for maximum likelihood estimation for a panel with lagged dependent variable and individual effects. As an alternative, Chamberlain (1982) proposed a minimum distance (MD) type of estimator for distributed lag models with heterogenous coefficients.

The subsequent developments have shifted attention to generalized method of moments (GMM) estimators that utilize linear moment conditions. The literature has focused on exploiting as many possible moment conditions while keeping the resulting GMM estimator linear. Most of the large sample results are usually backed by a reference to 'standard central limit theorems' or assumed to follow from the general results on the asymptotic properties of GMM estimators in, for example, Hansen (1982). The (non)optimality of utilizing redundant moment conditions has also not been explored in detail. Papers in this line of research include Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995) and Blundell and Bond (1998). The use of all lags as available instruments was suggested by Holtz-Eakin, Newey and Rosen (1988). Keane and Runkle (1992) provide an alternative method of exploiting the moment conditions.[2] Large sample results for the GMM estimators are in Alvarez and Arellano (2003), while Harris and Tzavalis (1999) obtain the limiting distributions of pooled OLS, the within-group (WG) and WG with individual trends estimators, under the null of a unit root and normally distributed errors. Observe that, as noted by Kiviet (1995) and Judson and Owen (1999), the number of possible instruments used by the GMM estimators increases with T^2, the GMM estimators may perform poorly in samples with moderate and large T.

More recently several authors have proposed maximum likelihood and quasi-maximum likelihood (ML and QML) procedures arguing that these are computationally feasible and providing some Monte Carlo evidence of improved small sample performance even for non-normal errors. See the papers by Hsiao, Pesaran and Tahmicsioglu (2002) and Binder, Hsiao and Pesaran (2000) discussed below. Some further Monte Carlo evidence is provided by Binder, Hsiao, Mutl and Pesaran (2002).

Below I will review papers on the GMM, bias corrected OLS, MD and ML estimation mentioned above and compare the various model specifications, assumptions on the disturbance process involved and estimation methods. When required, I modify the original notation to make the comparison feasible.

[2] They propose to transform the model by a Cholesky decomposition of an initial estimate of the variance covariance matrix and use the untransformed instruments in the second step of the estimation. See below for a more detailed review.

2.1.1 GMM Estimation

I will now review the papers proposing GMM type of estimators in more detail. The model under consideration can be written as

$$y_{it} = \phi y_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}, \qquad t = 1, .., T, \quad i = 1, ..., N, \tag{2.1.1}$$

where y_{it} and $\mathbf{x}_{it}$ denote the (scalar) dependent variable and the $1 \times p$ vector of exogenous variables corresponding to cross sectional unit i in period t, ϕ and $\boldsymbol{\beta}$ represent corresponding 1×1 and $p \times 1$ parameters, and $u_{it} = \mu_i + \varepsilon_{it}$ denotes the overall disturbance term consisting of individual effects μ_i and an innovation ε_{it}. Under different assumptions on the disturbance process we obtain different possible moment restrictions that are exploited by the estimator. The proposed estimator also differs under different exogeneity assumptions on the $p \times 1$ vector of explanatory variables.

Arellano and Bond (1991) assume that the error terms are distributed as

$$\mu_i \sim IID\left(0, \sigma_\mu^2\right), \tag{2.1.2}$$

and

$$\varepsilon_{it} \sim IID\left(0, \sigma_\varepsilon^2\right), \tag{2.1.3}$$

independent of each other.[3] Because the disturbances as well as the endogenous variable contain individual effects, they will be correlated when interacted in levels. Therefore, the moment conditions considered involve first differences of the disturbances and in particular they are

$$E\left[\left(u_{it}-u_{i,t-1}\right)y_{i,t-k}\right]=0, \qquad t=2,..,T, \quad k=2,..,t-1 \quad i=1,...,N, \tag{2.1.4}$$

and with strictly exogenous variables also

$$E\left[\mathbf{x}'_{is}\left(u_{it}-u_{i,t-1}\right)\right]=\mathbf{0}_{p\times 1}, \qquad t=2,..,T, \quad s=1,..,T \quad i=1,...,N, \tag{2.1.5}$$

while with the variables being only predetermined these conditions hold only for $s=1,..,t-1$.

Stacking the model by grouping the observation first by time and then by individuals[4] we can write the first differenced model (after dropping the initial observation) as

$$\underset{(T-1)N\times 1}{\Delta\mathbf{y}} = \underset{(T-1)N\times 2}{\Delta\mathbf{Z}\cdot}\underset{2\times 1}{\boldsymbol{\delta}} + \underset{(T-1)N\times 1}{\Delta\varepsilon}, \tag{2.1.6}$$

[3]These assumptions are not formally stated in the paper. However, the asymptotic claims are based on the iid assumptions.

[4]This stacking is commonly used in the literature on dynamic panel. Observe, however, that we will use a different order of stacking in our model presented in later chapters.

where $\Delta\mathbf{Z} = [\Delta\mathbf{y}_{-1}, \mathbf{X}]$ with

$$\Delta\mathbf{y} = \begin{pmatrix} y_{12} - y_{11} \\ \vdots \\ y_{1T} - y_{1,T-1} \\ \vdots \\ y_{N2} - y_{N1} \\ \vdots \\ y_{NT} - y_{N,T-1} \end{pmatrix}, \ \Delta\mathbf{y}_{-1} = \begin{pmatrix} y_{11} - y_{10} \\ \vdots \\ y_{1,T-1} - y_{1,T-2} \\ \vdots \\ y_{N1} - y_{N0} \\ \vdots \\ y_{N,T-1} - y_{N,T-2} \end{pmatrix},$$

$$\Delta\mathbf{X} = \begin{pmatrix} \mathbf{x}_{12} - \mathbf{x}_{11} \\ \vdots \\ \mathbf{x}_{1T} - \mathbf{x}_{1,T-1} \\ \vdots \\ \mathbf{x}_{N2} - \mathbf{x}_{N1} \\ \vdots \\ \mathbf{x}_{NT} - \mathbf{x}_{N,T-1} \end{pmatrix}, \ \Delta\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{12} - \varepsilon_{11} \\ \vdots \\ \varepsilon_{1T} - \varepsilon_{1,T-1} \\ \vdots \\ \varepsilon_{N2} - \varepsilon_{N1} \\ \vdots \\ \varepsilon_{NT} - \varepsilon_{N,T-1} \end{pmatrix}. \qquad (2.1.7)$$

We can define the matrix of instruments as $\mathbf{H} = (\mathbf{H}_1', ..., \mathbf{H}_N')'$ where for the case of strictly exogenous variables we have

$$\mathbf{H}_i' = \begin{bmatrix} \begin{pmatrix} y_{i0} \\ \mathbf{x}_{i1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix} & & & \\ & \begin{pmatrix} y_{i0} \\ y_{i2} \\ \mathbf{x}_{i1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix} & & \\ & & \ddots & \\ & & & \begin{pmatrix} y_{i0} \\ \vdots \\ y_{i,T-2} \\ \mathbf{x}_{i1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix} \end{bmatrix}. \tag{2.1.8}$$

The proposed estimator is of the form

$$\widehat{\boldsymbol{\delta}} = \left(\Delta\mathbf{Z}'\mathbf{H}\mathbf{A}^{-1}\mathbf{H}'\Delta\mathbf{Z}\right)^{-1} \Delta\mathbf{Z}'\mathbf{H}\mathbf{A}^{-1}\mathbf{H}'\Delta\mathbf{y}, \tag{2.1.9}$$

where $\mathbf{A}$ is some weights matrix for the moments. More specifically, the first step of the estimation uses a simple weighting matrix

$$\begin{aligned}\mathbf{A} &= \sum_{i=1}^{N} \mathbf{H}_i' \mathbf{D} \mathbf{D}' \mathbf{H}_i \\ &= \mathbf{H}' \left(\mathbf{I}_N \otimes \mathbf{D}\mathbf{D}' \right) \mathbf{H},\end{aligned} \tag{2.1.10}$$

where $\mathbf{D}$ is a $T-1 \times T$ first difference operator matrix:

$$\mathbf{D} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}_{T-1\times T} . \tag{2.1.11}$$

In the second step the moment conditions are weighted by their estimated variance covariance matrix and the authors propose to use

$$\begin{aligned}\mathbf{A} &= \sum_{i=1}^{N} \mathbf{H}_i' \Delta\widehat{\mathbf{u}}_i \Delta\widehat{\mathbf{u}}_i' \mathbf{H} \\ &= \mathbf{H}' \left[\mathbf{I}_N \otimes \mathbf{D} \left(\sum_{i=1}^{N} \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i' \right) \mathbf{D}' \right] \mathbf{H},\end{aligned} \tag{2.1.12}$$

where $\Delta\widehat{\mathbf{u}}_i = (\Delta\widehat{u}_{i2}, ..., \Delta\widehat{u}_{iT})'$ and $\widehat{\mathbf{u}}_i = (\widehat{u}_{i1}, ..., \widehat{u}_{iT})'$ are the fitted residuals from the first step estimator.

Arellano and Bover (1995) consider a general nonsingular transformation of the model that removes the individual effects. Consider again the model in (2.1.1) and let $\mathbf{K}$ be any $(T-1) \times T$ transformation matrix of rank $(T-1)$ such that $\mathbf{K}\mathbf{e}_T = \mathbf{0}_{T-1}$, where $\mathbf{e}_T$ is a $T \times 1$ vector of ones. That is, the transformation by $\mathbf{K}$ is nonsingular and removes the individual effects. Hence $\mathbf{K}$ can, for example, be the matrix $\mathbf{D}$ considered above, or be equal to the 'Within Group' operator, with

$$\mathbf{K} = \begin{pmatrix} \left[1-\frac{1}{T}\right] & -\frac{1}{T} & \cdots & -\frac{1}{T} & -\frac{1}{T} \\ -\frac{1}{T} & \left[1-\frac{1}{T}\right] & \cdots & -\frac{1}{T} & -\frac{1}{T} \\ \vdots & & \ddots & \vdots & \vdots \\ -\frac{1}{T} & -\frac{1}{T} & \cdots & \left[1-\frac{1}{T}\right] & -\frac{1}{T} \end{pmatrix}. \tag{2.1.13}$$

Arellano and Bover (1995) also suggest the orthogonal deviations operator defined as:

$$\mathbf{K} = \begin{pmatrix} 1 & -\frac{1}{(T-1)} & -\frac{1}{(T-1)} & \cdots & -\frac{1}{(T-1)} & -\frac{1}{(T-1)} & -\frac{1}{(T-1)} \\ 0 & 1 & -\frac{1}{(T-2)} & \cdots & -\frac{1}{(T-2)} & -\frac{1}{(T-2)} & -\frac{1}{(T-2)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}. \tag{2.1.14}$$

This transformation subtracts the mean of future observations available in the sample from the first $T-1$ observations.

The transformed model is then

$$\left(\mathbf{I}_N \otimes \mathbf{K}\right)\mathbf{y} = \left(\mathbf{I}_N \otimes \mathbf{K}\right)\mathbf{Z}\boldsymbol{\delta} + \left(\mathbf{I}_N \otimes \mathbf{K}\right)\boldsymbol{\varepsilon}, \tag{2.1.15}$$

If the transformation matrix is upper triangular and the disturbances ε_{it} are not serially correlated, then the same moment conditions as consider by Arellano and Bond (1991) remain valid for the transformed model. Arellano and Bover (1995) then show that the resulting GMM estimator is in fact invariant to the choice of the transformation matrix.

If the exogenous variables are uncorrelated with the individual effects, Arellano and Bover (1995) also suggest the use of additional moment conditions in the form of

$$E\left[\left(\frac{1}{T}\sum_{t=1}^{T} u_{it}\right)\mathbf{x}_{is}\right] = \mathbf{0}_{p\times 1}. \tag{2.1.16}$$

In this case the transformation matrix is appended with a row consisting of $\mathbf{e}_T/T$ and can be denoted as:

$$\mathbf{C} = \begin{pmatrix} \mathbf{K} \\ \mathbf{e}_T/T \end{pmatrix}. \tag{2.1.17}$$

The instrument matrix $\mathbf{H}_i$ becomes

$$\mathbf{H}_i' = \begin{bmatrix} \begin{pmatrix} y_{i0} \\ \mathbf{x}_{i1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix} & & & & \\ & \begin{pmatrix} y_{i0} \\ y_{i2} \\ \mathbf{x}_{i1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix} & & & \\ & & \ddots & & \\ & & & \begin{pmatrix} y_{i0} \\ \vdots \\ y_{i,T-2} \\ \mathbf{x}_{i1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix} & \\ & & & & \begin{pmatrix} \mathbf{x}_{i1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix} \end{bmatrix}. \quad (2.1.18)$$

The GMM estimator of Arellano and Bover (1995) can then be expressed as

$$\widehat{\boldsymbol{\delta}} = \left[\mathbf{Z}'\left(\mathbf{I}_N \otimes \mathbf{C}'\right)\mathbf{H}\mathbf{A}^{-1}\mathbf{H}'\left(\mathbf{I}_N \otimes \mathbf{C}\right)\mathbf{Z}\right]^{-1}\mathbf{Z}'\left(\mathbf{I}_N \otimes \mathbf{C}'\right)\mathbf{H}\mathbf{A}^{-1}\mathbf{H}'\left(\mathbf{I}_N \otimes \mathbf{C}\right)\mathbf{y}. \tag{2.1.19}$$

The preliminary estimates are obtained with $\mathbf{A} = \mathbf{H}'\left(\mathbf{I}_N \otimes \mathbf{C}\mathbf{C}'\right)\mathbf{H}$ and the second stage estimator uses consistently with (2.1.12):

$$\mathbf{A} = \mathbf{H}'\left[\mathbf{I}_N \otimes \mathbf{C}\left(\sum_{i=1}^{N}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\right)\mathbf{C}'\right]\mathbf{H}, \tag{2.1.20}$$

where $\widehat{\mathbf{u}}_i$ are the fitted residuals from the preliminary estimation. Given that the estimator is invariant to the choice of the transformation matrix, the filtering is in fact irrelevant and the estimator can be obtained by performing three stage least squares (3SLS).

Ahn and Schmidt (1995) show that there are additional moment conditions that can be exploited. Ahn and Schmidt also make weaker assumptions that lead to the set of moment restriction utilized by the Arellano and Bond (1991) and Arellano and Bover (1995) estimators. In particular, Ahn and Schmidt assume that the disturbances satisfy:

$$\begin{aligned} Cov\left(\varepsilon_{it}, y_{i0}\right) &= 0, \qquad t = 1, .., T \qquad (2.1.21) \\ Cov\left(\varepsilon_{it}, \mu_i\right) &= 0, \qquad t = 1, .., T \\ Cov\left(\varepsilon_{it}, \varepsilon_{is}\right) &= 0, \qquad t, s = 1, .., T; t \neq s \end{aligned}$$

The additional moment conditions pointed out by Ahn and Schmidt are

$$E\left[u_{iT}\left(\varepsilon_{it}-\varepsilon_{i,t-1}\right)\right]=0, \qquad t=2,..,T-1. \tag{2.1.22}$$

These restrictions, together with the moment conditions utilized by the Arellano and Bond (1991) estimator, represent all the moment conditions implied by the assumption that the innovations ε_{it} are mutually uncorrelated among themselves and with μ_i and y_{i0}.

Ahn and Schmidt also point out that further restrictions can be derived from homogeneity and stationarity assumptions. The assumption that the innovations ε_{it} have a variance that does not change over time implies the following additional moment restrictions:

$$E\left[y_{i,t-2}\Delta\varepsilon_{i,t-1}-y_{i,t-1}\Delta\varepsilon_{it}\right]=0, \qquad t=4,..,T. \tag{2.1.23}$$

In a model without exogenous variables the homogeneity restrictions can be implemented by utilizing the extended instrument set defined as

$$\mathbf{H}_i^{+}=\begin{pmatrix} \mathbf{H}_i & & & & & \\ & y_{i2} & -y_{i3} & & & \\ & & y_{i3} & -y_{i4} & & \\ & & & \ddots & \ddots & \\ & & & & y_{i,T-2} & -y_{i,T-3} \end{pmatrix}, \tag{2.1.24}$$

where $\mathbf{H}_i$ is the Arellano and Bond instrument matrix for the case without exogenous variables, i.e.

$$\mathbf{H}_i = \begin{pmatrix} y_{i0} & & & & & & \\ & y_{i0} & y_{i1} & & & & \\ & & & \ddots & & & \\ & & & & y_{i0} & \cdots & y_{i,T-2} \end{pmatrix}. \qquad (2.1.25)$$

Ahn and Schmidt show that the GMM estimator based on the full set of moment restrictions is asymptotically equivalent to Chamberlain's (1982, 1984) optimal minimum distance estimator and that it reaches the semiparametric efficiency bound.

Blundell and Bond (1998) document a potential gain in efficiency arising from exploiting restrictions on the initial observations when the time dimension of the panel is small and the degree of autocorrelation is high. The estimation approaches discussed so far usually drop the first observation. With N going to infinity and T fixed this amounts to ignoring information from a fixed proportion of the sample and thus can lead to sizeable inefficiency.

In their simulation study Blundell and Bond consider two types of additional restrictions. The first type of restriction justifies the use of an extended linear GMM estimator that uses lagged differences of y_{it} as instruments for equations in levels (in addition to lagged levels of y_{it} as instruments for equations in first differences). The second type of restriction validates the use of the error components GLS estimator on an extended model that conditions on the observed initial values. This provides a consistent estimator under homoscedasticity which, under normality, is asymptotically equivalent to conditional maximum likelihood (see also Blundell and Smith, 1991).

In a model without exogenous variables, Blundell and Bond show that after removing redundant restrictions the extended GMM estimator they consider utilizes the following instrument matrix:

$$\mathbf{H}_i^{++} = \begin{pmatrix} \mathbf{H}_i^{+} & & & \\ & \Delta y_{i2} & & \\ & & \ddots & \\ & & & \Delta y_{i,T-1} \end{pmatrix}, \tag{2.1.26}$$

where $\mathbf{H}_i^{+}$ is the instrument matrix employed by the Anh and Schmidt estimator and is defined in (2.1.24) above.

Their Monte Carlo simulations and asymptotic variance calculations show that this extended GMM estimator offers considerable efficiency gains in situations where the basic GMM estimator performs poorly. The GLS estimator that conditions on the initial values is also found to have good finite sample properties. However, the conditional GLS estimator requires homoscedasticity, and only extends to a model with regressors if the regressors are strictly exogenous which is not the case for the GMM estimators.

The efficiency gain from incorporating the information in the initial observation is also documented by a simulation study of Hahn (1999).

Alvarez and Arellano (2002) consider the same model (2.1.1) with $|\phi| < 1$ and $E(\varepsilon_{it}|\mu_i, y_{i0}, ..., y_{it-1}) = 0$. They assume y_{i0} is also observed. To derive asymptotic results they assume that ε_{it} for $t = 1, ..., T$ and $i = 1, ..., N$ are independent and identically distributed across time and individuals and independent of μ_i and y_{i0}, with $E(\varepsilon_{it}) = 0$, $Var(\varepsilon_{it}) = \sigma^2$ and finite fourth moments. Additionally they assume that the initial observation are generated as

$$y_{i0} = \frac{\mu_i}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \varepsilon_{i,-j}. \tag{2.1.27}$$

The article than establishes asymptotic properties of the 'Within Group' estimator, the GMM estimator, and the Limited Information Maximum Likelihood (LIML) estimator when both T and N tend to infinity. The WG estimator can be obtained by OLS estimation on the model transformed by the forward orthogonal means transformation (see above Arellano and Bover, 1995). The GMM estimator in their terminology is what I describe above as the first stage GMM estimator on a model transformed by the orthogonal deviations transformation, using the moment conditions of Arellano and Bond (1991). The second stage GMM estimation with an estimated weighting matrix is not considered. Note that my results contain this extension as a special case. See Chapter 4.

The LIML estimator is what has been suggested by Alonso-Borrego and Arellano (1999) as a symmetrically normalized GMM estimator. It can also be regarded as a 'continuously updated GMM estimator' in terminology of Hansen, Heaton and Yaron (1997).[5] The estimator is only an analogue LIML estimator in the sense of the minimax instrumental variable interpretation given by Sargan (1958) to the original LIML estimator. It is defined as

$$\widehat{\boldsymbol{\delta}} = \arg\min_{\boldsymbol{\delta}} \frac{(\mathbf{y}-\mathbf{Z}\boldsymbol{\delta})' (\mathbf{I}_N \otimes \mathbf{C}') \mathbf{H} (\mathbf{H}'\mathbf{H})^{-1} \mathbf{H}' (\mathbf{I}_N \otimes \mathbf{C}) (\mathbf{y}-\mathbf{Z}\boldsymbol{\delta})}{(\mathbf{y}-\mathbf{Z}\boldsymbol{\delta})' (\mathbf{I}_N \otimes \mathbf{C}') (\mathbf{I}_N \otimes \mathbf{C}) (\mathbf{y}-\mathbf{Z}\boldsymbol{\delta})}, \quad (2.1.28)$$

where $\mathbf{H}$ is an instrument matrix.

Alvarez and Arellano show that the asymptotic bias of the WG estimator only disappears when $N/T \rightarrow 0$. When N/T tends to a positive constant, all three estimators are asymptotically biased with negative asymptotic biases of order $1/T$, $1/N$, and $1/(2N-T)$, respectively. When N/T tends to infinity, the fixed T results assumed by the GMM literature remain valid. They also consider a random effects maximum likelihood estimator that leaves the mean and variance of the initial conditions unrestricted and show that this estimator is asymptotically unbiased for all cases.

Keane and Runkle (1992) suggest an alternative estimation procedure that takes into account the variance covariance structure of the disturbances. First the model is estimated by an initial procedure, such as the instrumental variables (IV) with instruments that could, for example, be the instruments suggested by Arellano and Bond (1991). Then an estimate of the inverse of the variance covariance matrix and its Cholesky decomposition is calculated. The model is then transformed and

[5] Instead of keeping σ^2 fixed in the weighting matrix of the GMM criterion, it is continuously updated by making it a function of the argument in the estimating criterion.

estimated with original (untransformed) instruments, i.e.

$$\begin{aligned}\widehat{\boldsymbol{\delta}} &= \left[\mathbf{Z}'\left(\mathbf{I}_N \otimes \widehat{\mathbf{P}}'\right)\mathbf{H}\mathbf{A}^{-1}\mathbf{H}'\left(\mathbf{I}_N \otimes \widehat{\mathbf{P}}\right)\mathbf{Z}\right]^{-1} \\ &\quad \cdot \mathbf{Z}'\left(\mathbf{I}_N \otimes \widehat{\mathbf{P}}'\right)\mathbf{H}\mathbf{A}^{-1}\mathbf{H}'\left(\mathbf{I}_N \otimes \widehat{\mathbf{P}}\right)\mathbf{y},\end{aligned} \tag{2.1.29}$$

where $\widehat{\mathbf{P}}$ is Cholesky decomposition of the estimated inverse of the variance covariance matrix and $\mathbf{A}$ is moment weighting matrix that is chosen analogously to the standard GMM estimators.

2.1.2 Bias Correction

Small sample bias correction procedure of the inconsistent OLS estimation has been proposed by Kiviet (1995). Consider a dynamic panel data model as in (2.1.1). The model in levels can be stacked as in (2.1.6)

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + (\mathbf{I}_N \otimes \mathbf{e}_T)\,\boldsymbol{\mu} + \boldsymbol{\varepsilon}, \tag{2.1.30}$$

where $\mathbf{Z} = [\mathbf{y}_{-1}, \mathbf{X}]$ with

$$\begin{aligned}\mathbf{y} &= (y_{11}, ..., y_{1T}, ..., y_{N1}, ..., y_{NT})', \\ \mathbf{y}_{-1} &= (y_{10}, ..., y_{1,T-1}, ..., y_{N0}, ..., y_{N,T-1})', \\ \mathbf{X} &= (\mathbf{x}_{11}, ..., \mathbf{x}_{1T}, ..., \mathbf{x}_{N1}, ..., \mathbf{x}_{NT})', \\ \boldsymbol{\varepsilon} &= (\varepsilon_{11}, ..., \varepsilon_{1T}, ..., \varepsilon_{N1}, ..., \varepsilon_{NT})', \\ \mu &= (\mu_1, ..., \mu_N)'.\end{aligned} \tag{2.1.31}$$

The within group estimator is defined as

$$\widehat{\boldsymbol{\delta}} = (\mathbf{Z}'\mathbf{A}\mathbf{Z})^{-1}\,\mathbf{Z}'\mathbf{A}\mathbf{y}, \tag{2.1.32}$$

where the $NT \times NT$ within group transformation matrix $\mathbf{A}$ is defined as

$$\mathbf{A} = \mathbf{I}_N \otimes \left(\mathbf{I}_T - \frac{\mathbf{e}_T\mathbf{e}_T'}{T}\right). \tag{2.1.33}$$

Kiviet (1995) calls this estimator Least-Squares Dummy Variables (LSDV) while Anderson and Hsiao (1981) refer to is as Covariance (CV) estimator. This estimator is inconsistent for fixed T due to presence of individual effects in both the disturbances $\boldsymbol{\varepsilon}$ and the regressors $\mathbf{y}_{-1}$. Although consistent estimates can be obtained by IV or GMM procedures, the inconsistent LSDV estimator has a relatively low variance and hence can lead to an estimator with lower root mean square error after the bias is removed. The asymptotic formulae for the bias given in Nickell (1981) for a model with no exogenous regressors has been found to be accurate in small samples, except for large values ϕ. Similar results have been reported by Sevestre and Trognon (1985). Kiviet (1995) provides approximating formulae for the small sample bias that have robust performance over the entire range of parameters.

2.1.3 MD and ML Estimation

Chamberlain's (1982, 1984) proposed to treat each time period as an equation in a multivariate equation framework. Such approach is robust to certain kinds of heteroscedasticity as well as autocorrelation in the errors without imposing a priori restrictions on the variance covariance matrix.

To demonstrate the method assume for simplicity that the model is:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mu_i + \varepsilon_{it} \qquad t = 1, .., T; \quad i = 1, ..., N, \tag{2.1.34}$$

and

$$E\left(\varepsilon_{it} \middle| \mathbf{x}_{i1}, ..., \mathbf{x}_{iT}, \mu_i\right) = 0, \tag{2.1.35}$$

where the $p \times 1$ vector of explanatory variables is assumed to be stochastic and hence the model also covers the lagged dependent variable case. The variables can be stacked by grouping observations for each individual into a vector $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$ and $\mathbf{x}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{iT})'$. Assume that $(\mathbf{y}_i, \mathbf{x}_i)$ is an independent draw from a common unknown multivariate distribution with finite fourth-order moments and with $E\left(\mathbf{x}_i\mathbf{x}_i'\right)$ positive definite. The individual effects are possibly correlated with the explanatory variables. Chamberlain (1984) assumes that the minimum-mean-squared-error linear projection of μ_i onto $\mathbf{x}_i$ is given by[6]

$$E^*\left(\mu_i \middle| \mathbf{x}_i\right) = \overline{\mu} + \sum_{t=1}^{T} \mathbf{a}_t'\mathbf{x}_{it}. \tag{2.1.36}$$

[6]If the conditional expectation of μ_i are linear, we have $E^*\left(\mu_i \middle| \mathbf{x}_i\right) = E\left(\mu_i \middle| \mathbf{x}_i\right)$.

The model can be rewritten as

$$E^*(\mathbf{y}_i|\,\mathbf{x}_i) = E^*\{E^*(\mu_i|\,\mathbf{x}_i,\mu_i)|\,\mathbf{x}_i\} \quad (2.1.37)$$
$$= E^*\{\mu_i\mathbf{e}_T + (\mathbf{I}_T \otimes \boldsymbol{\beta}')\,\mathbf{x}_i|\,\mathbf{x}_i\}$$
$$= \mu_i\mathbf{e}_T + \mathbf{\Pi}\mathbf{x}_i,$$

and

$$\mathbf{y}_i = \mu_i\mathbf{e}_T + (\mathbf{I}_T \otimes \mathbf{x}_i)\,\boldsymbol{\pi} + \boldsymbol{\nu}_i, \quad (2.1.38)$$

where

$$\mathbf{\Pi} = \mathbf{I}_T \otimes \boldsymbol{\beta}' + \mathbf{e}_T\,(\mathbf{a}_1', ..., \mathbf{a}_T')\,, \quad (2.1.39)$$

and $\boldsymbol{\nu}_i = \mathbf{y}_i - E^*(\mathbf{y}_i|\,\mathbf{x}_i)$, and $\boldsymbol{\pi} = vec\,(\mathbf{\Pi})$.

The proposed estimation procedure is then as follows. Treating the coefficients in the above equation as unrestricted, one first obtains initial (usually least-squares) estimate $\widehat{\boldsymbol{\pi}}$ of $\boldsymbol{\pi}$. In the second step, the restrictions on $\mathbf{\Pi}$ in (2.1.39) are incorporated by letting $\boldsymbol{\pi}$ be a function of the parameters of the model $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{a}_1', .., \mathbf{a}_T')$. The restrictions are imposed by using a minimum-distance estimator, specifically

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\,[\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}(\boldsymbol{\theta})]'\,\widehat{\mathbf{\Omega}}\,[\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}(\boldsymbol{\theta})]\,, \quad (2.1.40)$$

where $\widehat{\mathbf{\Omega}}$ is the estimated variance covariance matrix of the asymptotic variance of $\widehat{\boldsymbol{\pi}}$:

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{N}\sum_{i=1}^{N}\left\{\left[(\mathbf{y}_i-\overline{\mathbf{y}})-\widehat{\boldsymbol{\Pi}}\,(\mathbf{x}_i-\overline{\mathbf{x}})\right]\right. \qquad (2.1.41)$$

$$\left.\left[(\mathbf{y}_i-\overline{\mathbf{y}})-\widehat{\boldsymbol{\Pi}}\,(\mathbf{x}_i-\overline{\mathbf{x}})\right]'\otimes\mathbf{S}_{XX}^{-1}\,(\mathbf{x}_i-\overline{\mathbf{x}})\,(\mathbf{x}_i-\overline{\mathbf{x}})'\,\mathbf{S}_{XX}^{-1}\right\},$$

where

$$\mathbf{S}_{XX}^{-1}=\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i-\overline{\mathbf{x}})\,(\mathbf{x}_i-\overline{\mathbf{x}})'. \qquad (2.1.42)$$

Anderson and Hsiao (1981) consider the model (2.1.1) with $|\phi|<1$. They distinguish four different cases based on different assumptions on the initial values of the process (y_{i0}):

- Case I. Fixed initial observations: y_{i0} are fixed observed constants
- Case II. Random initial observations, common mean:

$$y_{i0}=c+\xi_i \qquad (2.1.43)$$

  where ξ has a mean zero and a finite variance and is independent of μ_i and ε_{it}. Here they also suggest that one could assume

$$y_{i0}=c+\mu_i \qquad (2.1.44)$$

  so that the initial endowment affects the level.

- Case III. Random initial observations, different means (in this case there the incidental parameter problem arrises and for fixed T the MLE is inconsistent): the model is

$$y_{it} = w_{it} + \gamma_i \qquad t = 0, 1, .., T, \tag{2.1.45}$$

$$w_{it} = \phi w_{i,t-1} + \varepsilon_{it} \qquad t = 1, .., T, \tag{2.1.46}$$

  where w_{it} and γ_i are unobservable. In this case w_{i0} are unknown constants.

- Case IV. Random initial observations with stationary distribution: same as above but w_{i0} are (a) draws from stationary distribution with mean zero and variance $\frac{var(\varepsilon_{it})}{1-\phi^2}$ or (b) same but the variance is arbitrary. In the subcase (a), the y_{it} come from the stationary distribution of the process.

To derive the likelihood function they assume normality of the error terms ε_{it}, μ_i and when applicable also y_{i0}. Implicit assumption is that $E(\varepsilon_{it}) = 0$ and $Var(\varepsilon_{it}) = \sigma^2$ (uniform over individuals).

Anderson and Hsiao (1982) have

$$y_{it} = \phi y_{i,t-1} + x_{it}\beta + z_i\gamma + \mu_i + \varepsilon_{it} \qquad t = 1, .., T; \quad i = 1, ..., N, \tag{2.1.47}$$

where $|\phi| < 1$ and

$$\begin{aligned} E(\mu_i) &= E(\varepsilon_{it}) = E(\mu_i z_i) = E(\mu_i x_{it}) = E(\mu_i \varepsilon_{it}) = 0 \\ t &= 1,..,T; \quad i = 1,...,N, \end{aligned} \tag{2.1.48}$$

and $E(\mu_i \mu_j) = \sigma_\mu^2$ for $i = j$ and $= 0$ for $i \neq j$,

$$\begin{aligned} E(\varepsilon_{it}\varepsilon_{js}) &= \sigma_\varepsilon^2 \qquad i = j, \quad t = s, \\ &= 0 \qquad otherwise \end{aligned} \tag{2.1.49}$$

They also assume normality of μ_i and ε_{it} and first consider the model with only time-invariant exogenous regressors. Again several cases are distinguished:

- (I) y_{i0} is fixed
- (II) y_{i0} is random with
  - (IIa) y_{i0} independent of μ_i, or
  - (IIb) y_{i0} correlated with μ_i; in their wording "If we wish the initial endowment [y_{i0}] affects the equilibrium level [$\frac{\mu_i}{1-\phi}$] we may let":

$$y_{i0} = z_i\gamma + \mu_i. \tag{2.1.50}$$

- (III) $(y_{i0} - \mu_i)$ is fixed
- (IV) $(y_{i0} - \mu_i)$ is random with

- (IVa) variance $\frac{\sigma_\varepsilon^2}{1-\phi^2}$
- (IVb) unrestricted (but uniform over i) variance

Next Anderson and Hsiao consider the model with only time-varying regressors and they offer two interpretations of the model:

(1) Serial correlation model:

$$y_{it} = \phi y_{i,t-1} + x_{it}\beta - \phi x_{it}\beta + \mu_i + \varepsilon_{it}. \tag{2.1.51}$$

Here they again assume either that $(y_{i0} - x_{i0}\beta - \mu_i)$ is fixed, or random with zero mean and variance $\frac{\sigma_\varepsilon^2}{1-\phi^2}$.

(2) State dependence model:

$$y_{it} = \phi y_{i,t-1} + x_{it}\beta + \mu_i + \varepsilon_{it}. \tag{2.1.52}$$

As before, there is a variety of assumptions concerning y_{i0} considered - the assumption correspond exactly to cases I.-IV above, except that in case of IV they distinguish whether $(y_{i0} - \mu_i)$ is random with

- - (IVa) common mean and variance $\frac{\sigma_\varepsilon^2}{1-\phi^2}$
  - (IVb) common mean and unrestricted variance
  - (IVc) heterogeneous mean and variance $\frac{\sigma_\varepsilon^2}{1-\phi^2}$
  - (IVd) heterogeneous mean and unrestricted variance

Table 1 below summarizes the consistency findings of Anderson and Hsiao:

Table 1. Consistency of ML Estimation

Case	Estimated Parameters	N fixed, $T \to \infty$	T fixed, $N \to \infty$
I.	$\phi, \beta, \sigma_\varepsilon^2$	Consistent	Consistent
	γ, σ_μ^2	Inconsistent	Consistent
II.a	$\phi, \beta, \sigma_\varepsilon^2$	Consistent	Consistent
	$\gamma, \sigma_\mu^2, \sigma_{y_0}^2, E\left(y_{i0}\right)$	Inconsistent	Consistent
II.b	$\phi, \beta, \sigma_\varepsilon^2$	Consistent	Consistent
	$\gamma, \sigma_\mu^2, \sigma_{y_0}^2$	Inconsistent	Consistent
	$E\left(y_{i0}\right), Cov(\varepsilon_{it}, \mu_i)$		
III.	$\phi, \beta, \sigma_\varepsilon^2$	Consistent	Inconsistent
	$\gamma, \sigma_\mu^2, (y_{i0} - \mu_i)$	Inconsistent	Inconsistent
IV.a	$\phi, \beta, \sigma_\varepsilon^2$	Consistent	Consistent
	$\gamma, \sigma_\mu^2, E\left(y_{i0} - \mu_i\right)$	Inconsistent	Consistent
IV.b	$\phi, \beta, \sigma_\varepsilon^2$	Consistent	Consistent
	$\gamma, \sigma_\mu^2, E\left(y_{i0} - \mu_i\right)$	Inconsistent	Consistent
	$Var\left(y_{i0} - \mu_i\right)$		
IV.c	$\phi, \beta, \sigma_\varepsilon^2$	Consistent	Inconsistent
	$\gamma, \sigma_\mu^2, E_i\left(y_{i0} - \mu_i\right)$	Inconsistent	Inconsistent
	$Var\left(y_{i0} - \mu_i\right)$		
IV.d	$\phi, \beta, \sigma_\varepsilon^2$	Consistent	Inconsistent
	$\gamma, \sigma_\mu^2, E_i\left(y_{i0} - \mu_i\right)$	Inconsistent	Inconsistent
	$Var\left(y_{i0} - \mu_i\right)$		

Bhargava and Sargan (1983) consider the dynamic panel data model with exogenous variable of essentially the same form as (2.1.1). They derive the maximum likelihood function under the assumption that the innovations and the individual effects are normally and independently distributed with constant variances, i.e. $\varepsilon_{it} \sim N\left(0, \sigma_{\varepsilon}^2\right)$ and $\mu_i \sim N\left(0, \sigma_{\mu}^2\right)$. The likelihood is derived first treating the initial values y_{i0} as exogenous and then as endogenous by assuming that the initial values are generated from the stationary distribution of the process. In particular, they assume that y_{i0} is generated by a series of equations (2.1.1) and can be written as

$$\begin{aligned} y_{i0} &= \sum_{k=0}^{\infty} \phi^k \left(\mathbf{x}_{i,t-k}\boldsymbol{\beta} + \mu_i + \varepsilon_{i,t-k}\right) \\ &= \overline{y}_{i0} + \frac{\mu_i}{1-\phi} + \sum_{k=0}^{\infty} \phi^k \varepsilon_{i,t-k}, \end{aligned} \tag{2.1.53}$$

where $\overline{y}_{i0}$ is exogenous part of the initial values and is in fact assumed to be stochastic with $\overline{y}_{i0} \sim N\left(\overline{y}_{i0}^*, \sigma_{y_0}^2\right)$, independent of ε_{it} and μ_i.

Hsiao, Pesaran and Tahmiscioglu (2002) consider the model (2.1.1) without exogenous variables,[7] i.e.

$$y_{it} = \phi y_{i,t-1} + \mu_i + \varepsilon_{it} \qquad t = 1, .., T; \quad i = 1, ..., N, \tag{2.1.54}$$

[7] In the second part, the authors extend the model for both strictly and weakly exogenous variables.

with y_{i0} observable. Under the assumption that the process has started at time $-m$ one can express the first difference of the initial observation as

$$\Delta y_{i1} = \phi^m \Delta y_{i,-m+1} + \xi_i, \tag{2.1.55}$$

where $\xi_i = \sum_{j=0}^{m-1} \phi^j \Delta\varepsilon_{i,1-j}$. Hsiao, Pesaran and Tahmiscioglu then distinguish two assumptions for the initial values of the process:

- Case (3.i) $|\phi| < 1$ and the process has been going on for a long time ($m \to \infty$) and $E(\Delta y_{i1}) = 0$, $Var(\Delta y_{i1}) = 2\frac{Var(\varepsilon_{it})}{1+\phi}$, $Cov(\xi_i, \Delta\varepsilon_{i2}) = -Var(\varepsilon_{it})$ and $Cov(\xi_i, \Delta\varepsilon_{it}) = 0$ for $t = 3, 4, ..., T$.
- Case (3.ii) m is finite and $E(\Delta y_{i1}) = b$, $Var(\Delta y_{i1}) = c \cdot var(\varepsilon_{it})$, where $c > 0$, $Cov(\xi_i, \Delta\varepsilon_{i2}) = -Var(\varepsilon_{it})$ and $Cov(\xi_i, \Delta\varepsilon_{it}) = 0$ for $t = 3, 4, ..., T$.

In both cases, the maximum likelihood function is then derived for the model in first differences under the assumption that the error terms are normally distributed with $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$. They also show that the ML function is invariant to the choice of transformation that is used to remove the individual effects from the model.

Hsiao, Pesaran and Tahmiscioglu also define a minimum distance estimator and show that if it ignores the initial conditions, it will be inconsistent when T is fixed. They also study the relationship of the ML estimator the the GMM estimators suggested by Arellano and Bond (1991), Arellano and Bover (1995), and

Ahn and Schmidt (1995). Conditional on σ_ε^2 and the variance of the initial observations, Hsiao, Pesaran and Tahmiscioglu show that the difference between the asymptotic variance covariance matrix of the GMM and the ML (or MD) estimators will be positive definite. They conjecture that the same holds even when σ_ε^2 and the variance of the initial observations is unknown and document this by a Monte Carlo study.

Binder, Hsiao and Pesaran (2000) consider a multivariate extension of the dynamic panel data model. Their specification is

$$\mathbf{w}_{it} = \boldsymbol{\mu}_i + \boldsymbol{\gamma} t + \boldsymbol{\Phi} \left[\mathbf{w}_{i,t-1} - \boldsymbol{\mu}_i - \boldsymbol{\gamma} (t-1) \right] + \boldsymbol{\varepsilon}_{it}, \tag{2.1.56}$$

where $\mathbf{y}_{it}$, $\boldsymbol{\mu}_i$, $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}_{it}$ are $m \times 1$ vectors and $\boldsymbol{\Phi}$ is an $m \times m$ matrix. They define $\mathbf{y}_{it} = \mathbf{w}_{it} - \boldsymbol{\mu}_i - \boldsymbol{\gamma} t$ and hence the model becomes

$$\mathbf{y}_{it} = \boldsymbol{\Phi} \mathbf{y}_{it} + \boldsymbol{\varepsilon}_{it} \tag{2.1.57}$$

They assume that the model started as time $t = -M$, $M \geq 0$ and the initial deviations are given by

$$\mathbf{y}_{i,-M} = \sum_{j=0}^{\infty} \left(\boldsymbol{\Phi}^j - \mathbf{C} \right) \boldsymbol{\varepsilon}_{i,-M-j} + \mathbf{C} \boldsymbol{\xi}_i, \tag{2.1.58}$$

where $\boldsymbol{\varepsilon}_{it}$, $i = 1, 2, ..., N$; $t \leq T$, are i.i.d. across i and over t, and $\boldsymbol{\xi}_i$ are i.i.d. across i with

$$E\begin{pmatrix} \boldsymbol{\varepsilon}_{it} \\ \boldsymbol{\xi}_i \end{pmatrix} = 0 \quad \text{and} \quad Var\begin{pmatrix} \boldsymbol{\varepsilon}_{it} \\ \boldsymbol{\xi}_i \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Omega} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}' & F \end{pmatrix}. \tag{2.1.59}$$

The matrix $\mathbf{C}$ is defined recursively as $\mathbf{C} = \sum_{j=0}^{\infty} \mathbf{C}_j$ where $\mathbf{C}_0 = \mathbf{I}_m$, $\mathbf{C}_1 = \boldsymbol{\Phi} - \mathbf{I}_m$, $\mathbf{C}_j = \mathbf{C}_{j-1}\boldsymbol{\Phi}$, $j \geq 2$. Notice that for $m = 1$, the $\mathbf{C}$ can only be zero or one.

Binder, Hsiao and Pesaran then derive the quasi maximum likelihood function for the model under the assumption the disturbances are $\{\boldsymbol{\varepsilon}_{it}\}$ and $\{\boldsymbol{\xi}_i\}$ are mutually independent and identically distributed. The authors also extend the GMM and MD estimators to the multivariate context and provide simulation evidence that the QML estimator dominates the GMM and MD procedures even when the underlying disturbances are not normal.[8] Binder, Hsiao, Mutl and Pesaran (2002) discuss the same model but with higher order autocorrelation structure and provide further Monte Carlo evidence.

2.2 Modelling Cross-Sectional Dependence

When T is large and N small, one does not have to parametrically specify the cross sectional interdependencies and can allow for arbitrary covariance structure of the disturbances. The model can then be consistently estimated by a general-

[8] The authors consider a case where the underlying disturbances are drawn from a zero mean chi-square distribution.

ized least squares method. This is what Zellner (1962) refers to as the seemingly unrelated regressions (SUR) specification. On the other hand, observe that the dimensions of the variance covariance matrix of the dependent variable (or disturbances) grows with sample size (number of cross-sections). Therefore, when the time dimension of the data is limited or fixed, it becomes impossible to infer the cross-sectional covariance structure of the model without imposing some parametric restrictions.

Typically the interaction among the cross-sectional units is modelled as proportional to some observable distance. The most widely used parameterization are variants of the one considered by Cliff and Ord (1973 and 1981) which I review below. Recent applications include Audretsch and Feldmann (1996), Bernat (1996), Besley and Case (1995), Bollinger and Ihlanfeldt (1997), Buettner (1999), Case (1991), Case, Hines, and Rosen (1993), Dowd and LeSage (1997), Holtz-Eakin (1994), LeSage (1999), Kelejian and Robinson (2000, 1997), Pinkse and Slade (1998), Pinkse, Slade, and Brett (2002), Shroder (1995), and Vigil (1998). See also a host of other papers presented for example at the Spatial Econometrics Workshop in Kiel, 2005 (http://www.uni-kiel.de/ifw/konfer/spatial/spatial-econometrics.htm).

In this thesis, I follow the spatial econometrics literature and study a first order spatial autocorrelation model with a known spatial weighting matrix. The panel spatial autocorrelation model is a generalization of the single cross-section models that include Cliff and Ord (1973, 1981), Whittle (1954), Anselin (1988) or Kelejian and Prucha (1998, 1999 and 2004). See also Lee (2004) who provides asymptotic properties of ML procedure for spatial models. Other recent theoretical developments include Baltagi and Li (2001a,b), Baltagi, Song and Koh (2003), Conley (1999), Das, Kelejian and Prucha (2005), Kelejian and Prucha (2001, 1997), Lee (2003, 2002, 2001a,b), LeSage (2000, 1997), Pace and Barry (1997), Pinkse and Slade (1998), Pinkse, Slade, and Brett (2002), and Rey and Boarnet (2004). An excellent review of the different specifications in spatial econometrics can be found in Anselin (1988). See also Haining (1990) and references therein.

2.2.1 Model Specifications

I will now present the basic specification of spatial dependence suggested in the literature. The Cliff-Ord type model of spatial dependence can be written in the following form. Suppose that we have a panel of observations in space, indexed by $i = 1, ..., N$, and time, indexed by $t = 1, ..., T$. The disturbances[9] $u_{it,N}$ can then be specified to follow a spatial autoregressive process in the form of:

$$u_{it,N} = \rho \sum_{j=1}^{N} w_{ij,N} u_{jt,N} + \varepsilon_{it,N}. \tag{2.2.1}$$

The disturbance $u_{it,N}$ for a cross-section i at a time t consists of a weighted average of contemporaneous disturbances in other cross-sections and a mutually independent innovation term $\varepsilon_{it,N}$. The weights $w_{ij,N}$ are assumed to be observable quantities and, therefore, the extent of correlation in the model is a function of a

[9]Of course spatial lags can also be applied to the endogenous or explanatory variables in the same manner.

single parameter ρ.

This model for spatial correlation was introduced by Cliff and Ord (1973, 1981). Anselin (1988) refers to this model as a first order spatial autoregressive model or SAR(1). The weights $w_{ij,N}$ are referred to as spatial weights and are assumed to be known, ρ is called the spatial autoregressive parameter and $\sum_{j=1}^{N} w_{ij,N} u_{jt,N}$ is referred to as a spatial lag. The spatial weights $w_{ij,N}$ are typically specified to be nonzero if cross sectional unit i relates to unit j in a meaningful way. In such cases, units i and j are said to be neighbors. In practice, the spatial weights are often viewed as normalized in the sense that the summation term in (2.2.1) is an average of neighboring observations. e.g. one postulates that $\sum_{j=1}^{N} w_{ij,N} = 1$.

A more general model can include spatial lags in the disturbances as well as in the endogenous variable, denoted by $y_{it,N}$, e.g.

$$y_{it,N} = \mathbf{x}_{it,N}\boldsymbol{\beta} + \lambda \sum_{j=1}^{N} m_{ij,N} y_{jt,N} + u_{it,N}, \tag{2.2.2}$$

where $\mathbf{x}_{it,N}$ is a vector of exogenous variables, $\boldsymbol{\beta}$ is a vector of parameters, λ is a scalar parameter, $m_{ij,N}$ are spatial weights, and the disturbance $u_{it,N}$ are as in (2.2.1). The term $\sum_{j=1}^{N} m_{ij,N} y_{jt,N}$ is then referred to as a spatial lag of the dependent variable. The weights in the spatial lag of the dependent variable ($m_{ij,N}$) can, but do not necessarily have to, correspond to those in the spatial lag in the disturbances ($w_{ij,N}$).

Observe that all variables are indexed by the sample size N, e.g. they form triangular arrays. This also applies to situations where the spatial weight are specified as fixed constants. Observe that in many cases, it is assumed that each cross-sectional location i has a fixed number of neighbors, say q, for which $w_{ij,N} \neq 0$. Hence each $w_{ij,N}$ is equal either to zero or a fixed number such as $1/q$. Observe that even in such cases, the number of cross-sectional units determines the number of units that enter into the solution of equation (2.2.1). As a result, the disturbances $u_{it,N}$ that are solution to (2.2.1) have to be indexed by the sample size. The fact that the disturbances $u_{it,N}$ are indexed by the sample size leads to certain technical complications and, for example, one has to be careful in applying central limit theorems and make sure that these also hold for triangular arrays.

Contiguity Weights The specifications where each units is, only affected by its neighbors are sometimes referred to as contiguity weights. These could be specified as $w_{ij,N} = 1$, when the two units are neighbors, and $w_{ij,N} = 0$ otherwise. Denoting $\mathbf{W}_N$ the $N \times N$ matrix of the weights $w_{ij,N}$, the row-normalized weights are then given by

$$\mathbf{W}_N^* = \mathbf{W}_N./\left(\mathbf{e}_N' \otimes \mathbf{W}_N \mathbf{e}_N\right), \tag{2.2.3}$$

where $\mathbf{e}_N$ is an $N \times 1$ vector of ones and $./$ denotes element-by-element division.

In practical applications, the definition of a neighbor often follows a natural geographical interpretation. Thus if the space in question is a geographical space and the units of analysis are regions, two regions are classified as neighbors when they share a common border. Other popular specifications of the contiguity weights are rook, queen and related configurations. Suppose that the space is divided in equally sized rectangular units. Below, I depict the rook and queen configuration using one to indicate the units that are neighbors to the unit x and zero to indicate other units that are not direct neighbors (these then correspond to entries on the $x-th$ row of the spatial weighting matrix $\mathbf{W}_N$):

$$\begin{array}{ccccccccccc} & & rook: & & & \qquad & & & queen: & & \\ 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & x & 1 & 0 & & 0 & 1 & x & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & 0 \end{array} \tag{2.2.4}$$

An alternative is to assume that the spatial process has higher order components and use so-called double-rook or double-queen specification, which could be:

$$\begin{array}{ccccccccccc} & & double-rook: & & & \qquad & & & double-queen: & & \\ 0 & 0 & \frac{1}{2} & 0 & 0 & & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} & 0 & & \frac{1}{2} & 1 & 1 & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & x & 1 & \frac{1}{2} & & \frac{1}{2} & 1 & x & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} & 0 & & \frac{1}{2} & 1 & 1 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & 0 & & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array} \tag{2.2.5}$$

Of course the choice of entries 1 and $\frac{1}{2}$ is arbitrary and these can be replaced by some other constants.

Another possibility is to assume that the cross-sectional units can be ordered linearly in space (as an analogy to the linear ordering of observations in time). The specification that is often referred to as $q-$ahead, $r-$behind (in terminology of Kelejian and Prucha, 1999) uses the weights matrix $\mathbf{W}_N^{(q,r)}$ consisting of zeros except for entries of ones on the first q subdiagonals below the main diagonal and entries of ones on the first r subdiagonals above the main diagonal. For example, the $2-$ahead, $2-$behind matrix is:

$$\mathbf{W}_N^{(2,2)} = \begin{pmatrix} 0 & 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \ddots & \ddots & \vdots \\ 1 & 1 & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 1 & 1 \\ \vdots & \ddots & \ddots & 1 & 0 & 1 \\ 0 & \cdots & 0 & 1 & 1 & 0 \end{pmatrix}. \tag{2.2.6}$$

An alternative is to assume a circular ordering of the observation in space. In this case, the $q-$ahead, $r-$behind weights matrices are as above but with added nonzero entries in positions $(i, N-j)$ where $i, j = 0, .., q-1$ and $(N-k, l)$ where $k, l = 0, .., r-1$. For the $2-$ahead, $2-$behind matrix, circularity implies that the first unit is also a neighbor of units N and $N-1$, hence the added entry of one in positions $(N, 1)$, $(N-1, 1)$, $(1, N)$, and $(2, N)$. Additionally the second and last unit (N) as well as the first and $(N-1)-th$ units are neighbors, and hence the entries of one in positions $(N, 2)$ and $(1, N-1)$.

Distance Based Weights When one views the cross-sectional observations as being located in a space, the extent of *direct* correlation between the disturbances at two locations can be interpreted as related to their distance in the space under consideration. Hence the weights can be interpreted as being (inversely) related to some measure of distance among the observations. In practical applications the space does not necessarily have to be a geographical space. The observations can be located in an abstract space in which their proximity is a *known* function of some of their observable characteristics. For example, two industries can be considered to be '*close*' to each other if they use a similar set of inputs, or two countries can be '*close*' if they have received financial flows from the same international lenders.

Under the interpretation of the weights $w_{ij,N}$ being inversely related to a distance measure, one is making an implicit assumption that the weights are symmetric in the sense that $w_{ij,N} = w_{ji,N}$. This is an artefact of the symmetry of distance measures, i.e. the distance from i to j has to be equal to the distance from j to i.[10] Observe, however, that the model considered here is more general. In particular, I do not require the weights to be symmetric and $w_{ij,N}$ does not have to be equal to $w_{ji,N}$. This can be advantageous in situations where the spillover of shocks is not necessarily symmetric. An example is the international transmission of shocks, where a shock originating in a very small country cannot be plausibly assumed to affect a large country in a same way as a shock originating in a large

[10]Observe that the distance based weights can be adjusted (premultiplied) by a factor that accounts for the differences in the direction of the influence. In this case the weights can become asymmetric. Note that the specification in this thesis allows for such asymmetries.

country affects a small country (e.g. US shocks affect say Ecuador much more than Ecuador's shocks can affect the US).

The problem of symmetry of the spatial weights that are based on a distance measure is related to a more general issue of aggregation. Suppose that the data was generated for a larger (disaggregated) sample but is only observed for aggregated spatial units. Mutl (2006) considers such data generating designs in a Monte Carlo study and concludes that only specifications that adjust the spatial weights for the relative size of the units deliver estimates that do not change with the increases in the number of units observed in the sample. The appropriate measure of the size depends on the units of measurement of the endogenous variable. For example, when the dependent variable is expressed as GDP per population, then the spatial weights $w_{ij,N}$ should be a postmultiplied by the population of the region i relative to the entire population of all regions in the sample. Constructing the distance based spatial weights in this fashion takes automatically account of the asymmetrical effects considered above. See also Giacomini and Granger (2004) for related issue of forecasting an aggregate of spatially interrelated observations, and LeSage and Pace (2004) for dealing with missing values in models with spatial dependence.

2.2.2 Estimation

The estimation method for models with spatial autocorrelation suggested by Anselin (1988) or Anselin and Hudak (1992) was maximum likelihood (ML). The asymptotic properties of the ML estimator of a model such as (2.2.1) have been derived only recently by Lee (2004) for one specific Cliff-Ord model. Furthermore, the maximum likelihood function contains a Jacobian term that is a determinant of a matrix that increases with the sample size N. Hence for moderate and large sample sizes, the ML estimation might become infeasible. As an alternative, Kelejian and Prucha (1998) introduced spatial generalized moments (spatial GM) estimator and proved its consistency. The asymptotic distribution of the spatial GM estimator is derived in Kelejian and Prucha (2005). The spatial GM estimator is computationally much simpler and, as a result, is feasible also for large sample sizes.

The OLS estimation of a model with SAR disturbances is inefficient but remains consistent. However, when spatial lags of the dependent variable are included, as in (2.2.2), OLS estimation becomes biased since the stochastic regressor $\sum_{j=1}^{N} w_{ij,N} y_{jt,N}$ on the left hand side is correlated with the error term (endogeneity bias). However, an instrumental variable estimation with spatial lags of the explanatory variable as instruments, will be consistent (Kelejian and Prucha, 1998). Alternative instrument sets are considered in Lee (2003) and Kelejian, Prucha and Yuzefovich (2004).

The stacked version of the model given in (2.2.1) and (2.2.2) is

$$\begin{aligned} \mathbf{y}_N &= \mathbf{X}_N\boldsymbol{\beta} + \lambda\mathbf{M}_N\mathbf{y}_N + \mathbf{u}_N, \\ \mathbf{u}_N &= \rho\mathbf{W}_N\mathbf{u}_N + \boldsymbol{\varepsilon}_N, \end{aligned} \tag{2.2.7}$$

where $\mathbf{y}_N$ is the $N \times 1$ vector of the dependent variable, $\mathbf{X}_N$ is the $N \times p$ matrix of exogenous variables, $\mathbf{M}_N$ and $\mathbf{W}_N$ are $N \times N$ spatial weighting matrices, $\mathbf{u}_N$ and $\boldsymbol{\varepsilon}_N$ are the $N \times 1$ vectors of disturbances and innovations. Under appropriate regularity conditions, the model can be solved as (see, for example, Das, Kelejian and Prucha, 2003, page 4):

$$\mathbf{y}_N = \left(\mathbf{I}_N - \lambda \mathbf{M}_N\right)^{-1} \mathbf{X}_N \boldsymbol{\beta} + \left(\mathbf{I}_N - \lambda \mathbf{M}_N\right)^{-1} \left(\mathbf{I}_N - \rho \mathbf{W}_N\right)^{-1} \boldsymbol{\varepsilon}_N. \quad (2.2.8)$$

Under the assumption that the vector $\boldsymbol{\varepsilon}_N$ is normally distributed with $\boldsymbol{\varepsilon}_N \sim N\left(\mathbf{0}_{N\times 1}, \sigma^2 \mathbf{I}_N\right)$, the likelihood function is:

$$\begin{aligned} \ln(L) &= -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln \left|\boldsymbol{\Omega}_N\right| \qquad (2.2.9) \\ &\quad -\frac{1}{2} \left[\mathbf{y}_N - \left(\mathbf{I}_N - \lambda \mathbf{M}_N\right)^{-1} \mathbf{X}_N \boldsymbol{\beta}\right]' \boldsymbol{\Omega}_N^{-1} \left[\mathbf{y}_N - \left(\mathbf{I}_N - \lambda \mathbf{M}_N\right)^{-1} \mathbf{X}_N \boldsymbol{\beta}\right], \end{aligned}$$

where $\boldsymbol{\Omega}_N$ is the variance covariance matrix of the disturbances $\mathbf{u}_N$ given by

$$\boldsymbol{\Omega}_N = \sigma^2 \left(\mathbf{I}_N - \lambda \mathbf{M}_N\right)^{-1} \left(\mathbf{I}_N - \rho \mathbf{W}_N\right)^{-1} \left(\mathbf{I}_N - \lambda \mathbf{M}'_N\right)^{-1} \left(\mathbf{I}_N - \rho \mathbf{W}'_N\right)^{-1}. \quad (2.2.10)$$

The least squares procedure applied directly to equation (2.2.7) is inconsistent due to correlation of $y_{it,N}$ and $u_{it,N}$. However, there are instrumental variables (IV) procedures that are consistent. Observe that for the current model (see Das, Kelejian and Prucha, 2003, page 7):

$$E\left(\mathbf{y}_N\right) = \left(\mathbf{I}_N - \lambda \mathbf{M}_N\right)^{-1} \mathbf{X}_N \boldsymbol{\beta} = \sum_{k=0}^{\infty} \lambda^k \mathbf{W}_N^k \mathbf{X}_N \boldsymbol{\beta}, \quad (2.2.11)$$

and hence ideal instruments are combinations of matrices $\mathbf{X}_N\boldsymbol{\beta}$, $\mathbf{W}_N\mathbf{X}_N\boldsymbol{\beta}$, $\mathbf{W}_N^2\mathbf{X}_N\boldsymbol{\beta}$, etc. Kelejian and Prucha (1998) show that an IV estimator that uses at least the linearly independent columns of $\mathbf{X}_N$, $\mathbf{W}_N\mathbf{X}_N$, $\mathbf{W}_N^2\mathbf{X}_N$ as instruments is consistent and asymptotically normal.

The spatial autocorrelation parameter ρ can then be estimated with the spatial generalized moments (spatial GM) procedure, suggested by Kelejian and Prucha (1999). Denote $\widehat{\mathbf{u}}_N$ the estimated disturbances based on an initial consistent estimator. Let

$$\begin{aligned} v_{1,N}\left(\underline{\rho},\underline{\sigma}^2\right) &= N^{-1}\left(\mathbf{I}_N-\underline{\rho}\mathbf{W}_N\widehat{\mathbf{u}}_N\right)'\left(\mathbf{I}_N-\underline{\rho}\mathbf{W}_N\widehat{\mathbf{u}}_N\right)-\underline{\sigma}^2, \qquad (2.2.12)\\ v_{2,N}\left(\underline{\rho},\underline{\sigma}^2\right) &= N^{-1}\left(\mathbf{I}_N-\underline{\rho}\mathbf{W}_N^2\widehat{\mathbf{u}}_N\right)'\left(\mathbf{I}_N-\underline{\rho}\mathbf{W}_N^2\widehat{\mathbf{u}}_N\right)-\underline{\sigma}^2N^{-1}tr\left(\mathbf{W}_N'\mathbf{W}_N\right),\\ v_{3,N}\left(\underline{\rho},\underline{\sigma}^2\right) &= N^{-1}\left(\mathbf{I}_N-\underline{\rho}\mathbf{W}_N^2\widehat{\mathbf{u}}_N\right)'\left(\mathbf{I}_N-\underline{\rho}\mathbf{W}_N\widehat{\mathbf{u}}_N\right). \end{aligned}$$

The spatial GM estimator is then defined as

$$\left(\widehat{\rho},\widehat{\sigma}^2\right)=\arg\min\left\{\sum_{k=1}^{3}v_{k,N}'\left(\underline{\rho},\underline{\sigma}^2\right)v_{k,N}\left(\underline{\rho},\underline{\sigma}^2\right):\left(\underline{\rho},\underline{\sigma}^2\right)\in[-a,a]\times\left[0,s^2\right]\right\}, \qquad (2.2.13)$$

where $a\geq 1$ and s^2 is the upper limit considered for σ^2. Kelejian and Prucha (1999) show that the spatial GM estimator is consistent. Kelejian and Prucha (1998) also provide a proof that the spatial autoregressive parameter ρ is a 'nuisance' parameter in the sense that the feasible generalized spatial two stage least squares (FGS2SLS) estimator has the same asymptotic distribution when it is based on a consistent estimator of ρ as when it is based on the true value. Initially, the asymptotic distribution of the spatial GM estimator was not determined. As a result, tests for spatial autocorrelation had to be based on statistics such as the Moran *I*. Kelejian and Prucha (2001) and Pinske and Slade (1998) provide asymptotic distribution of the Moran *I* test statistics. The asymptotic distribution of the spatial GM estimator was then derived for a more general model that includes heteroscedastic disturbances in Kelejian and Prucha (2005).

2.3 Space-Time Models

Time and space is a key feature of almost all human activities. Their interaction has been studied in many disciplines and has received some attention in economics as well. Studies outside economics include many applications in geostatistics (see e.g. Kyriakidis and Journel, 1999 for a review), geography but also in epidemiology, medicine, crime prevention and others. Short overviews can be found in Cressie (1991: 449-452) and Robinson (1998: 319-328).

In economics and econometrics, some interesting cases complementary to the specification in the present thesis are, for example, generalized least squares test to test for unit roots in panel data (although without deriving any asymptotic properties of the estimator) in O'Connell (1998), a two-step sieve least squares procedure to estimate a panel vector autoregression (VAR) model with a nondiagonal cross-sectional covariance matrix that is proportional to an observed economic distance measure in Chen and Conley (2001) who look at asymptotics in the less complicated case when the cross-sectional dimension is fixed, and, finally, Chang (2002) who derives asymptotic properties of a univariate panel model with a general unrestricted form of cross-sectional heterogeneity when the cross-sectional dimension of the panel is also fixed.

In this thesis, I will analyze dynamic model that includes a spatial lag in the disturbance process. This is a special case of the class of stochastic models known as space-time autoregressive (space-time AR) models introduced by Cliff et al. (1975) and generalized by Pfeifer and Deutsch (1980). More recent discussions and applications of the space-time AR model in econometrics are Elhorst (2001), while a generalization of the model to continuous space is proposed by Brown et al. (2000).

Below I review papers that deal with this class of models in more detail. Note that if contemporaneous correlation is present, the observable data become a non-trivial transformation of the underlying random field, resulting in some technical difficulties. Hence I first focus on specifications that do not allow for contemporaneous correlation in the data but instead assume that spatial interactions act with a time lag. In the second subsection I therefore present models that allow for such complications.

2.3.1 Space-Time Autoregressive Moving Average (STARMA) Models

Pfeifer and Deutsch (1980) were the first to propose a STARMA model. Their general STARMA$(p, q; \lambda_1, ..., \lambda_p, m_1, ..., m_q)$ model is:

$$\begin{aligned} y_{it} &= \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \phi_{kl} \sum_{j=1}^{N} w_{ij,l} y_{j,t-k} \\ &\quad - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \theta_{kl} \sum_{j=1}^{N} w_{ij,l} \varepsilon_{j,t-k} + \varepsilon_{it}, \end{aligned} \tag{2.3.1}$$

where p is the autoregressive order, q is the moving average order, λ_k is the spatial order of the $k-th$ autoregressive term, m_k is the spatial order of the $k-th$ moving average term, ϕ_{kl} and θ_{kl} are parameters and the errors are normally distributed with $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}\varepsilon_{j,s}) = \sigma^2$ for $i = j$ and $t = s$, and $E(\varepsilon_{it}\varepsilon_{js}) = 0$ otherwise.

The spatial weights have the usual interpretation (see the previous subsection) and are assumed to be observable and the authors do not impose any restrictions on their structure. Observe that in contrast to Cliff-Ord type model considered in this thesis, their STARMA model does not allow for contemporaneous correlation between spatial units, i.e. for example ε_{it} depends on $\varepsilon_{j,t-1}$ but not on ε_{jt}. As a result, the likelihood function does not involve a Jacobian term in a form of a determinant of an $N \times N$ and, as a result, ML estimation is considerably simpler and it is the estimation method suggested by Pfeifer and Deutsch. The authors derive the likelihood function conditional on initial values of the process and note that it is only appropriate for moderate or large T. However, the restrictions implied by the model on the initial observations are not explicitly derived. The paper also does not provide formal consistency or asymptotic normality results. Abraham (1983) derives the likelihood function for the STARMA model.

Stoffer (1986) outlines different estimation procedure for a spatial STAR model with missing values (spatial ARX in his terminology). The model combines the time series parametrization of an autoregressive moving average process for missing and noisy data with a Cliff and Ord type spatial structure. The data generating process is assumed to be a q-th order autoregressive process where the current observation is influenced by q time lags of its spatial neighbors:

$$y_{it} = \sum_{k=1}^{q} \sum_{j=1}^{N} w_{ij,k} \phi_{kj} y_{j,t-k} + \mathbf{x}_{it}' \boldsymbol{\beta} + \varepsilon_{it}, \tag{2.3.2}$$

where the autoregressive parameters ϕ_{kj} are allowed to vary with spatial location. The spatial weights $w_{ij,k}$ have the usual interpretation (e.g. they are inversely related to a distance) and are allowed to be different at different time lags. The p explanatory variables in $\mathbf{x}_{it}$ are modelled as a stochastic process independent of the innovations ε_{it} and the data sample is observed for $i = 1, ..., N$ and $t = 1, ..., T$.

The estimates are solutions to approximated Yule-Walker equations. For example, with no data problems, $q = 1$ and without explanatory variables, the model can be written as

$$\mathbf{y}_t = \mathbf{W}\mathbf{\Phi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{2.3.3}$$

where $\mathbf{y}_t = (y_{1t}, ..., y_{Nt})'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$, $\mathbf{W}$ is a $N \times N$ matrix of the spatial weights w_{ij} and $\mathbf{\Phi} = diag\,(\phi_1, ..., \phi_N)$. The proposed estimator of $\mathbf{\Phi}$ is then:

$$\widehat{\mathbf{\Phi}} = diag\left(\mathbf{W}^{-1}\widehat{\mathbf{\Gamma}}_{-1}\widehat{\mathbf{\Gamma}}_0^{-1}\right), \tag{2.3.4}$$

where the estimated moments of the data are

$$\widehat{\mathbf{\Gamma}}_0 = \sum_{t=2}^{T} \mathbf{y}_t\mathbf{y}_t', \tag{2.3.5}$$

and

$$\widehat{\mathbf{\Gamma}}_{-1} = \sum_{t=1}^{T} \mathbf{y}_t\mathbf{y}_{t-1}'. \tag{2.3.6}$$

There are no formal asymptotic claims made for the procedure. The method is illustrated with an application to fish catch data at five locations for 240 time periods suggesting that the implicit asymptotic consistency and normality claims are for fixed spatial dimension N and increasing time dimension of the observations.

Pace et al. (1998) model spatial and temporal dependence in housing price data in Fairfax County Virginia between 1961 and 1991. Unlike in standard STAR models, it is not assumed that the autocorrelation in the dependent variable is linearly separable in space and time. Instead an interaction of the space and time lags is considered. In particular, the model is:

$$y_{it} = \sum_{s=1}^{T} \sum_{j=1}^{N} w_{ij,ts} y_{js} + \mathbf{x}'_{it}\boldsymbol{\beta} + \sum_{s=1}^{T} \sum_{j=1}^{N} w_{ij,ts} \mathbf{x}'_{js}\boldsymbol{\beta} + \varepsilon_{it}, \tag{2.3.7}$$

where the observable weights $w_{ij,ts}$ relate observation across time and space simultaneously. It is assumed that $w_{ij,ts} = 0$ for $s \leq t$, meaning that the current and future values of y_{js} and $\mathbf{x}_{js}$ do not influence the process for y_{it}.

Stacking $w_{ij,ts}$ into a $NT \times NT$ matrix $\mathbf{W}$, Pace et al. assume that

$$\mathbf{W} = \rho_s \mathbf{S} - \rho_T \mathbf{T} + \rho_{ST} \mathbf{ST} + \rho_{TS} \mathbf{TS}, \tag{2.3.8}$$

where the $\mathbf{S}$ and $\mathbf{T}$ matrices are interpreted as filters in space and time respectively. Their entries are related to the distance of the of the observation in space and time respectively.

The main limitation of their approach is that it is assumed that there are no concurrent observations and that only past observations have an effect. If the matrix $\mathbf{W}$ is stacked so that the observations are sorted according to time, this assumption implies that both $\mathbf{T}$ and $\mathbf{S}$ are strictly lower (or upper) diagonal. As a result the model can be estimated by OLS. The paper does not provide formal results and does not spell out assumptions on the disturbance process.

Giacomini and Granger (2004) show that the STARMA class of models can be derived as a transformation of vector autoregressive moving average (VARMA) model, where the transformation is a restriction involving spatial weighting matrices. When the number of locations is small, the model can be estimated by an overparametrized VARMA specification. With increasing number of location, the overparameterized VARMA model has a large number of insignificant parameters. Therefore, estimation can be improved in a Bayesian framework by incorporating these as priors. Hence LeSage and Krivelyova (1999) propose a class of prior distributions for a Bayesian VAR model that will approximately constrain the insignificant parameters to zero.

2.3.2 Models with Contemporaneous Spatial Correlation

The papers cited in the above subsection did not allow for contemporaneous dependence of the observations. When such interactions are included, the observation become a nonlinear transformation of the innovations and, as a result, maximum likelihood estimation is more difficult. We next review papers that allow for such complications.

Congdon (1994) considers the spatiotemporal model of the following form:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mu_i + u_{it}, \tag{2.3.9}$$

where $t = 1, ..., T$ and $i = 1, ..., N$ and the error term is both spatially and temporally autocorrelated:

$$u_{it} = \phi u_{i,t-1} + \rho \sum_{j=1}^{N} w_{ij} u_{jt} + \varepsilon_{it}. \tag{2.3.10}$$

It is assumed that y_{i0} and $\mathbf{x}_{i0}$ are known exogenous constants. The first step of the proposed estimation procedure eliminates the individual effects μ_i by subtracting individual means $\overline{y_i}$ and $\overline{\mathbf{x}_i}$ and estimating the slope coefficients $\boldsymbol{\beta}$ by OLS on

$$(y_{it} - \overline{y_i}) = (\mathbf{x}_{it} - \overline{\mathbf{x}_i})' \boldsymbol{\beta} + (v_{it} - \overline{v_i}) . \tag{2.3.11}$$

In the second step, ϕ and ρ are estimated by minimizing

$$g(\phi, \rho) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it}^* - \mathbf{x}_{it}^{*\prime} \widehat{\boldsymbol{\beta}}_{OLS} \right)^2 , \tag{2.3.12}$$

where

$$\begin{aligned} y_{it}^* &= (y_{it} - \overline{y_i}) - \phi (y_{i,t-1} - \overline{y_i}) - \rho \sum_{j=1}^{N} w_{ij} \left(y_{jt} - \overline{y}_j \right) , \qquad (2.3.13) \\ \mathbf{x}_{it}^* &= (\mathbf{x}_{it} - \overline{\mathbf{x}_i}) - \phi (\mathbf{x}_{i,t-1} - \overline{\mathbf{x}_i}) - \rho \sum_{j=1}^{N} w_{ij} (\mathbf{x}_{jt} - \overline{\mathbf{x}}_j) . \end{aligned}$$

Based on Hordijk (1979), the transformation for the first time period is

$$\begin{aligned} \mathbf{y}_1^* &= \left[(\mathbf{I} - \rho \mathbf{W})' (\mathbf{I} - \rho \mathbf{W}) - \phi^2 \mathbf{I}_N \right]^{1/2} (\mathbf{y}_1 - \overline{\mathbf{y}}) , \qquad (2.3.14) \\ \mathbf{X}_1^* &= \left[(\mathbf{I} - \rho \mathbf{W})' (\mathbf{I} - \rho \mathbf{W}) - \phi^2 \mathbf{I}_N \right]^{1/2} \left(\mathbf{X}_1 - \overline{\mathbf{X}} \right) , \end{aligned}$$

where $\mathbf{y}_1 = (y_{11}, ..., y_{1N})'$, $\overline{\mathbf{y}} = (\overline{y}_1, ..., \overline{y}_N)'$, $\mathbf{X}_1 = (\mathbf{x}_{11}', ..., \mathbf{x}_{1N}')'$, $\overline{\mathbf{X}} = (\overline{\mathbf{x}}_1', ..., \overline{\mathbf{x}}_N')'$ and $\mathbf{W}$ is an $N \times N$ matrix with elements w_{ij}. The slope coefficients $\boldsymbol{\beta}$ are estimated by OLS from

$$y_{it}^{*}\left(\widehat{\phi},\widehat{\rho}\right)=\mathbf{x}_{it}^{*}\left(\widehat{\phi},\widehat{\rho}\right)'\boldsymbol{\beta}+\varepsilon_{it}. \tag{2.3.15}$$

In the third step, the variance components $\sigma_{\varepsilon}^{2}=Var\left(\varepsilon_{it}\right)$ and $\sigma_{\mu}^{2}=Var\left(\mu_{i}\right)$ are estimated, e.g.

$$\widehat{\sigma_{\varepsilon}^{2}}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}^{*}-\mathbf{x}_{it}^{*\prime}\widehat{\boldsymbol{\beta}}\right)^{2}, \tag{2.3.16}$$

where $\widehat{\boldsymbol{\beta}}$ is from step 2. [11] The final step is a generalized least squares (GLS) procedure to re-estimate $\boldsymbol{\beta}$.

The paper contains outline and an application of the estimation procedure to mortality rates in London but offers no formal proofs that would support the consistency claims. The estimated GLS procedure is based on suggestion in Anselin (1988), p.111.

Driscoll and Kraay (1995, 1998) Provide a proof of consistency and asymptotic normality of a GMM procedure based on a panel Newey and West (1987) nonparametric heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimator.[12] The limit is taken with respect to the time dimension

[11] The expression for $\widehat{\sigma_{\mu}^{2}}$ in the paper is

$$\widehat{\sigma_{\mu}^{2}}=\frac{1}{N}\sum_{i=1}^{N}\left\{\left(\mathbf{y}_{i}-\widehat{\phi}\mathbf{y}_{i,-1}-\widehat{\rho}\mathbf{W}\overline{\mathbf{y}}\right)-\widehat{\boldsymbol{\beta}}\left(\overline{\mathbf{x}}_{i}-\widehat{\phi}\overline{\mathbf{x}}_{i,-1}-\widehat{\rho}\mathbf{W}\overline{\mathbf{x}}\right)\right\}^{2}-\frac{\widehat{\sigma_{\varepsilon}^{2}}}{T}.$$

This does not seem to have the correct dimensions.

[12] The cross-sectional dimension of the data is collapsed by taking cross-sectional averages. Hence this is not a complete generalization of the HAC estimation to a panel setting.

of the data. Their specification requires that the data is an α-mixing random field of the same size as the number of moment restrictions and hence places only weak restrictions on the form of spatial and temporal correlations.

They consider r orthogonality conditions $E\left[h_{it}\left(z_{it},\boldsymbol{\theta}\right)\right]=0$, where z_{it}, $i=1,...,N$, $t=1,...,T$ is data and $\boldsymbol{\theta}$ is a vector of parameters. The restrictions are assumed to identify the parameters. Their GMM estimator is

$$\widehat{\boldsymbol{\theta}}_T=\arg\min_{\boldsymbol{\theta}}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_t\left(\boldsymbol{\theta},\mathbf{z}_t\right)\right]'\widehat{\mathbf{S}}_T^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_t\left(\boldsymbol{\theta},\mathbf{z}_t\right)\right]\right\}, \qquad (2.3.17)$$

where $\mathbf{z}_t=\left(z_{1t},...,z_{Nt}\right)'$, $\mathbf{h}_t\left(\boldsymbol{\theta},\mathbf{z}_t\right)=N^{-1}\sum_{i=1}^{N}h_{it}\left(z_{it},\boldsymbol{\theta}\right)$, and $\widehat{\mathbf{S}}_T$ is the standard HAC estimator applied to the sequence of cross-sectional averages of $h_{it}\left(z_{it},\boldsymbol{\theta}\right)$.

Bronnenberg and Mahajan (2001) Estimate a model of retailers behavior where the market shares are related to marketing variables. Their model is

$$y_{it}=\alpha_0+\mathbf{x}_{it}'\boldsymbol{\beta}+\mu_i+u_{it}, \qquad (2.3.18)$$

where the disturbances are composed of innovations autocorrelated in time and individual effects autocorrelated in space:

$$\begin{aligned}\mu_i &= \rho\sum_{j=1}^{N}w_{ij}\mu_j+\varepsilon_i, \qquad (2.3.19)\\ u_{it} &= \phi_1 u_{i,t-1}+v_{it}.\end{aligned}$$

The explanatory variables are also modelled as a stochastic process based on the same individual effects μ_i, with the $j-th$ explanatory variable $x_{j,it}$ specified as

$$x_{j,it} = \alpha_{jt} + \gamma_j \mu_i + \delta_{j,it}, \tag{2.3.20}$$

where

$$\delta_{j,it} = \phi_{2j} \delta_{j,i,t-1} + \theta_j \tau_t + \xi_{j,it}. \tag{2.3.21}$$

The model is estimated by Maximum Likelihood under the assumption that the innovations $\mu_i, \varepsilon_i, v_{it}, \xi_{j,it}$ are all jointly normally distributed.

Elhorst (2001) derives a likelihood function for a STAR(1,1) model where he also allows for contemporaneous spatial lags. His general model is

$$\begin{aligned} y_{it} &= \phi y_{i,t-1} + \rho_0 \sum_{j=1}^{N} w_{ij} y_{jt} + \rho_1 \sum_{j=1}^{N} w_{ij} y_{jt,t-1} \\ &\quad + \beta_1 x_{it} + \beta_2 x_{i,t-1} + \beta_3 \sum_{j=1}^{N} w_{ij} x_{it} + \beta_4 \sum_{j=1}^{N} w_{ij} x_{i,t-1} + u_{it}. \end{aligned} \tag{2.3.22}$$

The likelihood is derived under the assumption that the disturbances u_{it} are normally distributed with $E(u_{it}) = 0$, $E(u_{it}^2) = \sigma^2$ and $E(u_{it}u_{sj}) = 0$ if $t \neq s$ or $i \neq j$. The paper assumes that the matrix of the spatial weights $\mathbf{W} = (w_{ij})$ has zeros on the diagonal and that the spatial autoregressive parameter ρ is bounded by the inverse of the largest and smallest eigenvalue of $\mathbf{W}$. It is also implicitly assumed that the matrix $\mathbf{W}$ is symmetric and that the model is dynamically stable (this places a nontrivial condition on the parameters ϕ and ρ_0).[13] The likelihood is not conditionalized on the initial values but instead it is assumed that the initial observations are draws from the stationary distribution of the process.

Kapoor et al. (2005) extend the GM estimator of Kelejian and Prucha to a panel data. The contribution of thesis relative to Kapoor et al. (2005) is to allow for autocorrelation in the time dimension as well. Their specification is

$$y_{it,N} = \mathbf{x}'_{it,N}\boldsymbol{\beta} + u_{it,N}, \tag{2.3.23}$$

where the disturbances are an SAR(1) process with individual effects:

$$u_{it} = \rho \sum_{j=1}^{N} w_{ij} u_{ij,t} + \mu_i + \varepsilon_{it}. \tag{2.3.24}$$

The paper provides formal consistency proof of the spatial GM estimator (with alternative weighting schemes) of ρ, as well as asymptotic normality of a generalized least squares (GLS) estimator of $\boldsymbol{\beta}$.

Baltagi et al. (2003) derive formulae for various Lagrange multiplier tests in a model that includes spatially correlated disturbances. The paper also provides experimental evidence of their performance in small samples. They consider the following model:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mu_i + u_{it}, \tag{2.3.25}$$

[13] Such condition could be, for example $|\phi| + |\rho_0| \cdot \lambda_{\max}(\mathbf{W}) < 1$, where $\lambda_{\max}$ is the largest (in abolute value) eigenvalue of the matrix $\mathbf{W}$ that consists of the spatial weights w_{ij}.

with the disturbances being an SAR(1) process:

$$u_{it} = \rho \sum_{j=1}^{N} w_{ij} u_{jt} + \varepsilon_{it}. \tag{2.3.26}$$

Observe that when the spatial lag does not operate on the individual effects, this specification implies that the covariance between y_{it} and y_{js} is zero for $i \neq j$ and $t \neq s$. This is in contrast to the specification in Kapoor et al. (2005), where the individual effects are spatially correlated and, as a result, the covariance among y_{it} and y_{js} is nonzero for all values of i, j, t and s.

Korniotis (2005) Building on work of Hahn and Kurstiener (2002), Korniotis (2005) considers a bias corrected OLS estimator in a dynamic panel data model that also includes spatial lag of the dependent variable. The specification is

$$y_{it} = \phi y_{i,t-1} + \rho_1 \sum_{j=1}^{N} w_{ij} y_{jt} + \rho_2 \sum_{j=1}^{N} w_{ij} y_{j,t-1} + \mathbf{x}_{it}' \boldsymbol{\beta} + \mu_i + \varepsilon_{it}. \tag{2.3.27}$$

where the disturbances are independent in the time dimension but are allowed to have arbitrary covariance matrix (constant over time) in the cross-sectional dimension. The paper gives the asymptotic formulas for the biases of the OLS estimators when both N and T simultaneously approach infinity.

Yang (2005) extends the proofs of asymptotic normality in Lee (2004) to a static panel data model with random individual and fixed time effects. His model is

$$y_{it} = x_{it}\beta + \eta_t + \mu_i + u_{it}, \tag{2.3.28}$$

where the disturbances u_{it} are an SAR(1) process, i.e.:

$$u_{it} = \rho \sum_{j=1}^{N} w_{ij} u_{ij,t} + \varepsilon_{it}. \tag{2.3.29}$$

The QML function is derived under the assumption that $\{\varepsilon_{it}\}$ and $\{\mu_i\}$ are mutually independent and identically distributed random variables with finite $4 + \delta$ moments for some $\delta > 0$.

3 Model

In this chapter I specify the model and provide a discussion of the maintained assumptions. It proves to be useful to restate the following notational conventions and definitions: I use bold letters for matrices and vectors, and regular font letters to denote scalars. Furthermore, I use lower case letters for vectors and upper case letters for matrices. Let $(\mathbf{A}_N)_{N\epsilon\mathbb{N}}$ be some sequence of $Np \times Np$ matrices where $p \geq 1$ is some fixed positive integer. I denote the (i,j)-th element as $a_{ij,N}$. I say that the row and column sums of the sequence of matrices $\mathbf{A}_N$ are uniformly bounded in absolute value if there exists a positive finite constant c independent of N such that

$$\max_{1\leq i\leq Np}\sum_{j=1}^{Np}|a_{ij,N}| \leq c \qquad and \qquad \max_{1\leq i\leq Np}\sum_{i=1}^{Np}|a_{ij,N}| \leq c. \tag{3.0.1}$$

For future reference, I note that any finite sum and/or product of matrices with row and column sums uniformly bounded in absolute value will also have row and column sums uniformly bounded in absolute value; see Kelejian and Prucha (2004). As a consequence, if $\mathbf{B}$ is a matrix of constants with fixed dimensions and $\mathbf{A}_N$ is a sequence of matrices with row and column sums uniformly bounded in absolute value, then the sequence of matrices $(\mathbf{B} \otimes \mathbf{A}_N)$ will also have row and column sums uniformly bounded in absolute value.

3.1 Model Specification

Consider the following dynamic panel data model ($1 \leq i \leq N$, $1 \leq t \leq T$):

$$y_{it,N} = \phi y_{i,t-1,N} + \mathbf{x}_{it,N}\boldsymbol{\beta} + u_{it,N}, \tag{3.1.1}$$

where $y_{it,N}$ and $\mathbf{x}_{it,N}$ denote the (scalar) dependent variable and the $1 \times p$ vector of exogenous variables corresponding to cross sectional unit i in period t, ϕ and $\boldsymbol{\beta}$ represent corresponding 1×1 and $p \times 1$ parameters, and $u_{it,N}$ denotes the overall disturbance term.

In contrast to the existing dynamic panel data literature I do not assume that the disturbances $u_{it,N}$ are cross-sectionally uncorrelated and I consider potentially heteroscedastic errors. Given the fact that I will derive asymptotic properties of the model when the cross-sectional dimension tends to infinity, the cross-sectional covariance structure will be parametrized with a finite number of parameters. In particular, I assume that the disturbances $u_{it,N}$ follow a spatial autoregressive process in the form of:

$$u_{it,N} = \rho \sum_{j=1}^{N} w_{ij,N} u_{jt,N} + \upsilon_{it,N}, \tag{3.1.2}$$

where the overall disturbance $u_{it,N}$ consists of a spatial lag of contemporaneous disturbances in other cross-sections and an innovation $\upsilon_{it,N}$.

Anselin (1988) refers to this model as a first order spatial autoregressive model or SAR(1). See the previous chapter for more detailed discussion of such specification. The process for the disturbances contains one parameter ρ and N^2 observable spatial weights $w_{ij,N}$. The $\upsilon_{it,N}$ are the innovations that enter the spatial process. They are allowed to be correlated over time and I assume that they have the following error component structure:

$$\upsilon_{it,N} = \mu_{i,N} + \varepsilon_{it,N}, \tag{3.1.3}$$

where $\mu_{i,N}$ are unit specific error components, and $\varepsilon_{it,N}$ are the error components that vary both over cross-sectional units and time periods.

The spatial weights, as well as the endogenous, exogenous and disturbance processes are all allowed to depend on the sample size, i.e., to depend on N. Observe that even if the innovations $\upsilon_{it,N}$ did not depend on the sample size, the disturbances $u_{it,N}$ would still have to be indexed by the sample size due to the presence of the spatial lag $\rho \sum_{j=1}^{N} w_{ij,N} u_{jt,N}$ in (3.1.2).[14]

Stacking across units the model becomes ($1 \leq t \leq T$)

$$\begin{aligned} \underset{N\times 1}{\mathbf{y}_{t,N}} &= \phi \underset{N\times 1}{\mathbf{y}_{t-1,N}} + \underset{N\times p}{\mathbf{X}_{t,N}} \underset{p\times 1}{\boldsymbol{\beta}} + \underset{N\times 1}{\mathbf{u}_{t,N}}, \\ \underset{N\times 1}{\mathbf{u}_{t,N}} &= \rho \underset{N\times N}{\mathbf{W}_N} \underset{N\times 1}{\mathbf{u}_{t,N}} + \underset{N\times 1}{\boldsymbol{\upsilon}_{t,N}}, \end{aligned} \tag{3.1.4}$$

where

$$\underset{N\times 1}{\boldsymbol{\upsilon}_{t,N}} = \underset{N\times 1}{\boldsymbol{\mu}_N} + \underset{N\times 1}{\boldsymbol{\varepsilon}_{t,N}}, \tag{3.1.5}$$

[14]The $N \times 1$ vector of disturbances $\mathbf{u}_{t,N}$ is given by $\mathbf{u}_{t,N} = (\mathbf{I}_N - \rho \mathbf{W}_N)^{-1} \boldsymbol{\upsilon}_{t,N}$ (see equation 3.2.1). Note that the elements of $(\mathbf{I}_N - \rho \mathbf{W}_N)^{-1}$ must depend on the sample size N. This would be true even if the elements $w_{ij,N}$ did not depend on the sample size.

and

$$
\begin{aligned}
\mathbf{y}_{t,N} &= \begin{pmatrix} y_{1t,N} \\ \vdots \\ y_{Nt,N} \end{pmatrix}_{N\times 1}, \quad \mathbf{X}_{t,N} = \begin{pmatrix} \mathbf{x}_{1t,N} \\ \vdots \\ \mathbf{x}_{Nt,N} \end{pmatrix}_{N\times p}, \\
\mathbf{u}_{t,N} &= \begin{pmatrix} u_{1t,N} \\ \vdots \\ u_{Nt,N} \end{pmatrix}_{N\times 1}, \quad \boldsymbol{\mu}_{N} = \begin{pmatrix} \mu_{1,N} \\ \vdots \\ \mu_{N,N} \end{pmatrix}_{N\times 1}, \\
\boldsymbol{\varepsilon}_{t,N} &= \begin{pmatrix} \varepsilon_{1t,N} \\ \vdots \\ \varepsilon_{Nt,N} \end{pmatrix}_{N\times 1}, \quad \mathbf{W}_{N} = \begin{pmatrix} w_{11,N} & \cdots & w_{1N,N} \\ \vdots & \ddots & \vdots \\ w_{N1,N} & \cdots & w_{NN,N} \end{pmatrix}_{N\times N}.
\end{aligned} \tag{3.1.6}
$$

In all of the ensuing discussion T is fixed and $N \to \infty$. I maintain the following assumptions:

Assumption 1 *For each $N > 1$ the innovations $\{\varepsilon_{it,N} : 1 \leq i \leq N, t \leq T\}$ are independently distributed, with zero mean, constant variance $\sigma^2_{\varepsilon,N}$ with $0 < \sigma^2_{\varepsilon,N} < b_\varepsilon < \infty$. Furthermore, the innovations have finite absolute moments of order $4 + \delta_\varepsilon$ for some $\delta_\varepsilon > 0$ and those moments are uniformly bounded by some finite constant.*

Assumption 2 *For each $N > 1$ the individual effects $\left\{\mu_{i,N} : 1 \leq i \leq N\right\}$ are independently distributed, with zero mean, and are independent of the innovations $\{\varepsilon_{it,N} : 1 \leq i \leq N, t \leq T\}$. Furthermore, the individual effects have constant variance $\sigma^2_{\mu,N}$ with $0 < \sigma^2_{\mu,N} < b_\mu < \infty$ and finite absolute moments of* *order $4 + \delta_\mu$ for some $\delta_\mu > 0$ and those moments are uniformly bounded by some finite constant.*

Assumption 3 *The nonstochastic matrix $\mathbf{W}_N$ has the following properties:*

(a) All diagonal elements of $\mathbf{W}_N$ are zero.

(b) The true parameter ρ satisfies $|\rho| < 1$; the matrix $\mathbf{I}_N - r\mathbf{W}_N$ is nonsingular for all $|r| < 1$.

(c) The row and column sums of $\mathbf{W}_N$ and $\mathbf{P}_N(\rho) = (\mathbf{I}_N - \rho\mathbf{W}_N)^{-1}$ are bounded uniformly in absolute value by, respectively, $k_W < \infty$ and $k_P < \infty$ where k_P may depend on ρ.

It will be shown in the next section that the following assumption will guarantee that the variances of the disturbances $u_{it,N}$ are bounded away from zero:

Assumption 4

$$\lambda_{\min}\left(\mathbf{P}_N \mathbf{P}_N'\right) \geq c_P > 0$$

for some c_P where c_P may depend on ρ.

The analysis is conditionalized on the realized values of the exogenous variables and I henceforth view them as constants. I make the following assumptions on the exogenous variables:

Assumption 5 *(a) The matrix of exogenous (nonstochastic) regressors $\mathbf{X}_{t,N}$, $t \leq T$, has a full column rank (for N sufficiently large).*

(b) The elements of $\mathbf{X}_{t,N}$ are uniformly bounded in absolute value.

I complete the model by specifying a process that generates the initial observation of the dependent variable:

Assumption 6 *The model defined in (3.1.4) is dynamically stable, i.e., $|\phi| < 1$, and has been in operation for an infinite period of time.*[15]

The error specification adopted in this thesis corresponds to that of a classical one-way error component model, see e.g. Baltagi (1995, pp. 9). It is also a generalization of the literature on dynamic panel data models with independent innovations. Notice that with $\rho = 0$, my specification becomes, for example, that of Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), Blundell and Bond (1998),[16] or Anderson and Hsiao (1981 and 1981), case IVb.[17] Finally, note that the same error component specification of the disturbance process was adopted in Kapoor et al. (2005), who consider random effect specification in the context of a static panel data model.

3.2 Model Implications

I examine the asymptotic properties of the proposed estimation procedure when the time dimension of the panel is fixed. I assume slope homogeneity of the autoregressive parameters (ϕ does not have an i subscript)[18] and I also assume

[15] Note that Assumptions 1 and 2 have been consistently specified to hold for $-\infty < t \leq T$.

[16] In these papers the exogenous variables are allowed to be stochastic and either strictly exogenous or predetermined while in this thesis I treat the exogenous variables as nonstochastic.

[17] Anderson and Hsiao do not include exogenous variables in their specification.

[18] Note that heterogenous slope coefficients cannot be consistently estimated with a fixed number of observations in the time dimension of the panel.

that the spatial weighting matrices are constant over time.[19] In the rest of this section I explore some implications of the maintained assumptions. Proofs of the claims made in this section are in the Appendix B.

Assumption 1 is a standard restriction for asymptotic results. I do not assume that the innovations are identically distributed and hence a stronger requirements on the existence of moments is necessary. Assumption 2 is a random effects assumption that will be used to prove existence of asymptotic distribution of moment conditions that involve levels of lagged endogenous variables. I conjecture that the estimation procedure suggested in this thesis remains valid also when the individual effects are fixed ($\sigma_\mu^2 = 0$). However, the proofs would have to be modified[20] and hence I choose to concentrate on the random effects case.

Assumption 3(a) is a normalization of the model that also implies that no cross-section is viewed as its own neighbor. Assumption 3(b) implies that the system in (3.1.4) is complete in that it defines endogenous variables in terms of exogenous variables and innovations. In particular, from Assumption 3(b) it follows that

$$\underset{N\times 1}{\mathbf{u}_{t,N}} = \underset{N\times N}{\mathbf{P}_N}\, \underset{N\times 1}{\boldsymbol{v}_{t,N}}. \tag{3.2.1}$$

Furthermore, we can eliminate lagged dependent variables by backward substitution and express the model as a function of lagged disturbance terms and lagged

[19] If the spatial weighting matrices were not constant over time, then first differencing would not remove the individual effects.

[20] I apply central limits theorems to a vector of random variables that includes the individual effects. Hence it is required that $\sigma_\mu^2 > 0$. In the fixed effects case, the central limit theorems would be applied to a vector of random variables that excludes $\boldsymbol{\mu}_N$. Observe that the sequence of vectors μ_N would in this case be required to satisfy some regularity condition such as Assumption A2 in Appendix A.

explanatory variables. From (3.1.4), we have that for $1 \leq t \leq T$

$$\begin{aligned}
\mathbf{y}_{t,N} &= \phi \mathbf{y}_{t-1,N} + \mathbf{X}_{t,N}\boldsymbol{\beta} + \mathbf{u}_{t,N} \qquad (3.2.2)\\
&= \phi\left[\phi \mathbf{y}_{t-2,N} + \mathbf{X}_{t-1,N}\boldsymbol{\beta} + \mathbf{u}_{t-1,N}\right] + \mathbf{X}_{t,N}\boldsymbol{\beta} + \mathbf{u}_{t,N}\\
&\vdots\\
&= \sum_{j=0}^{t-1} \phi^{j}\left[\mathbf{X}_{t-j,N}\boldsymbol{\beta} + \mathbf{u}_{t-j,N}\right] + \phi^{t}\mathbf{y}_{0,N}\\
&= \sum_{j=0}^{t-1} \phi^{j}\left[\mathbf{X}_{t-j,N}\boldsymbol{\beta} + \mathbf{P}_{N}\boldsymbol{\upsilon}_{t-j,N}\right] + \phi^{t}\mathbf{y}_{0,N},
\end{aligned}$$

and hence $\mathbf{y}_{t,N}$ is a well defined transformation of the innovations $\boldsymbol{\upsilon}_{t,N}$, the initial values of the process $\mathbf{y}_{0,N}$, and the exogenous variables $\mathbf{X}_{t,N}$.

Assumption 3(c) restricts the degree of permissible cross-sectional correlation in the sample. Note that some restriction on the correlation is necessary for any large sample results to hold. In practice in the spatial literature, with T fixed and $N \to \infty$, it is often assumed that each cross-sectional unit has a finite number of neighbors, or that the rows of the weight matrices are normalized to sum to unity. It is also often the case that although the matrices may not be sparse, the weights are proportional to an inverse of some distance measure. Therefore, under reasonable conditions, the weight matrices will have row and column sums uniformly bounded in absolute value.

Assumption 4 rules out degenerate weighting matrices that would imply zero variance of the disturbances $\mathbf{u}_{t,N}$. Observe that from Assumption 3, we have $\mathbf{u}_{t,N} = \mathbf{P}_{N}\left(\boldsymbol{\mu}_{N} + \boldsymbol{\varepsilon}_{t,N}\right)$ and hence the variance covariance matrix of the disturbances $\mathbf{u}_{t,N}$ is

$$VC\left(\mathbf{u}_{t,N}\right)=\left(\sigma_{\mu,N}^{2}+\sigma_{\varepsilon,N}^{2}\right)\mathbf{P}_{N}\mathbf{P}_{N}^{\prime}. \tag{3.2.3}$$

In particular, notice that each diagonal element of $VC\left(\mathbf{u}_{t,N}\right)$ is bounded from below by the smallest eigenvalue[21] and hence the assumption implies that each $u_{it,N}$ has variance bounded away from zero. In a model without spatial correlation, $\mathbf{P}_N=\mathbf{I}_N$ and this Assumption is trivially satisfied.

Assumption 5 is an exogeneity assumption of explanatory variables. Finally, under Assumption 6, together with the assumptions on the exogenous variables and the spatial weighting matrix, we have by backward substitution:

$$\begin{aligned}\mathbf{y}_{0,N} &= \sum_{j=0}^{\infty}\phi^{j}\left(\mathbf{X}_{-j,N}\boldsymbol{\beta}+\mathbf{u}_{-j,N}\right) \\ &= \sum_{j=0}^{\infty}\phi^{j}\left[\mathbf{X}_{-j,N}\boldsymbol{\beta}+\mathbf{P_N}\boldsymbol{\varepsilon}_{-j,N}\right]+(1-\phi)^{-1}\mathbf{P}_N\boldsymbol{\mu}_N.\end{aligned} \tag{3.2.4}$$

Hence $\mathbf{y}_{0,N}$ is a random variable that in general depends on N with mean that is not necessarily equal to zero. Notice that $\{u_{it,N}:1\leq i\leq N,-\infty<t\leq 0\}$ is a transformation of $\{\varepsilon_{it,N}:1\leq i\leq N,-\infty<t\leq 0\}$ and $\left\{\mu_{i,N}:1\leq i\leq N\right\}$. Therefore, by Assumptions 1 and 2, the array $\{y_{i0,N}:1\leq i\leq N\}$ is independent of $\{\varepsilon_{it,N}:1\leq i\leq N,1\leq t\leq T\}$. Furthermore, given Assumptions 5 and 6 it also has finite absolute moments of order $4+\delta_{y_o}$ for some $\delta_{y_o}>0$ and those moments are uniformly bounded by some finite constant (see the appendix for a

[21] See e.g. Lemma 2 in Kelejian and Prucha (2003).

proof).[22] For future reference, I note that the variance of $\mathbf{y}_{0,N}$ is

$$VC\left(\mathbf{y}_{0,N}\right) = \left(\frac{\sigma^2_{\varepsilon,N}}{1-\phi^2} + \frac{\sigma^2_{\mu,N}}{\left(1-\phi\right)^2}\right)\mathbf{P}_N\mathbf{P}'_N. \tag{3.2.5}$$

[22]Similarly, it can be shown that the stochastic process $y_{it,N}$ has finite absolute moments of order $4+\delta_y$ for some $\delta_y > 0$ and that those moments are uniformly bounded by some finite constant. The proof of this claim is also in the appendix.

4 Estimation and Inference

This chapter will present the key results of the thesis. I present a procedure to estimate the parameters of the model outlined in Chapter 3 and derive its asymptotic properties. The proposed estimation method consists of three steps. In the first step, I propose to use an instrumental variables (IV) estimator of the slope coefficients ϕ and $\boldsymbol{\beta}$ without efficiently accounting for the spatial correlation of the disturbances.[23] In the second step of the estimation, the estimated disturbances from the first stage are utilized in a spatial generalized moments (GM) estimator to estimate the degree of spatial autocorrelation in the disturbances (ρ). In the last step of the procedure, I propose a GMM estimator of ϕ and $\boldsymbol{\beta}$ with an optimal weighting of the moments that is based on the initial estimators.

For expositional purposes, I choose to present for the first stage an IV estimator that uses a simple set of instruments due to Anderson and Hsiao (1981). Observe, however, that the results on the third stage generalized method of moments (GMM) estimators presented subsequently are sufficiently general to guarantee consistency of IV estimators that use an extended set of instruments, such as the one in Arellano and Bond (1991).

4.1 Initial IV Estimation

In this section I propose a simple estimation procedure to estimate the parameters $\boldsymbol{\theta} = [\phi, \boldsymbol{\beta}']'$ of the model (3.1.1) and demonstrate that the method is consistent and

[23] I do not account for the spatial correlation in formulating the initial IV estimator. However, it is taken into account in the analysis of its properties.

asymptotically normal. Since the model contains individual effects, these cannot be consistently estimated with fixed T. Hence the model is considered after a transformation that removes the individual effects from the dependent variable. I follow the literature on dynamic panels and use first differences. Note that it would also be possible to use other transformations such as central differences. I use moment conditions based on the fact that the first difference of the disturbances is uncorrelated with the level of the endogenous variable lagged twice (or more).[24] In particular, the estimator corresponds to the one suggested by Anderson and Hsiao (1982). Inspection of the proofs reveals that the random effects Assumption 2 is not strictly necessary for the initial estimator to work.[25]

I write the model in first differences as ($t = 2, ..., T$):

$$\underset{N\times 1}{\Delta \mathbf{y}_{t,N}} = \underset{1\times 1}{\phi} \underset{N\times 1}{\Delta \mathbf{y}_{t-1,N}} + \underset{N\times p}{\Delta \mathbf{X}_{t,N}} \underset{p\times 1}{\boldsymbol{\beta}} + \underset{N\times 1}{\Delta \mathbf{u}_{t,N}}, \tag{4.1.1}$$

where Δ is the first difference operator and, in particular, $\Delta \mathbf{y}_{t,N} = \mathbf{y}_{t,N} - \mathbf{y}_{t-1,N}$, $\Delta \mathbf{X}_{t,N} = \mathbf{X}_{t,N} - \mathbf{X}_{t-1,N}$ and $\Delta \mathbf{u}_{t,N} = \mathbf{u}_{t,N} - \mathbf{u}_{t-1,N}$.

Stacking the observations over time yields

$$\underset{(T-1)N\times 1}{\Delta \mathbf{y}_N} = \underset{(T-1)N\times(1+p)}{\Delta \mathbf{Z}_N} \underset{(1+p)\times 1}{\boldsymbol{\theta}} + \underset{(T-1)N\times 1}{\Delta \mathbf{u}_N}, \tag{4.1.2}$$

[24] This claim is formally proved in Lemma 2.

[25] Note that it is not the case that no assumption has to be made on the individual effects, as is often claimed in the literature. Since the lagged endogenous variable is used as an instrument, one still need to maintain that the individual effects are uncorrelated with the idiosyncratic disturbances and satisfy certain moment restrictions as well. This of would of course be satisfied if we view the individual effects as constants.

where

$$\underset{(T-1)N\times p+1}{\Delta\mathbf{Z}_N} = \left[\underset{(T-1)N\times 1}{\Delta\mathbf{y}_{-1,N}}, \underset{(T-1)N\times p}{\Delta\mathbf{X}_N}\right] \tag{4.1.3}$$

and[26]

$$\Delta\mathbf{y}_N = \begin{pmatrix} \Delta\mathbf{y}_{2,N} \\ \vdots \\ \Delta\mathbf{y}_{T,N} \end{pmatrix}_{(T-1)N\times 1}, \quad \Delta\mathbf{y}_{-1,N} = \begin{pmatrix} \Delta\mathbf{y}_{1,N} \\ \vdots \\ \Delta\mathbf{y}_{T-1,N} \end{pmatrix}_{(T-1)N\times 1} \tag{4,1.4}$$

$$\Delta\mathbf{X}_N = \begin{pmatrix} \Delta\mathbf{X}_{2,N} \\ \vdots \\ \Delta\mathbf{X}_{T,N} \end{pmatrix}_{(T-1)N\times p}, \quad \Delta\mathbf{u}_N = \begin{pmatrix} \Delta\mathbf{u}_{2,N} \\ \vdots \\ \Delta\mathbf{u}_{T,N} \end{pmatrix}_{(T-1)N\times 1}.$$

Since $\Delta\mathbf{y}_{t-1,N}$ is correlated with $\Delta\mathbf{u}_{t,N}$ the ordinary least squares estimator for $\boldsymbol{\theta}$ from the above model will generally be inconsistent. However, the level of the dependent variable lagged twice (or more) will not be correlated with the disturbances $\Delta\mathbf{u}_{t,N}$. Motivated by this, I define an instrument matrix

$$\underset{N\times(1+p)}{\mathbf{H}_{t,N}} = [\underset{N\times 1}{\mathbf{y}_{t-2,N}}, \underset{N\times p}{\Delta\mathbf{X}_{t,N}}]. \tag{4.1.5}$$

Given the model assumptions we have, as demonstrated in Lemma 2 below:

$$E\left(\underset{(1+p)\times N}{\mathbf{H}'_{t,N}} \underset{N\times 1}{\Delta\mathbf{u}_{t,N}}\right) = \mathbf{0}_{(1+p)\times 1}, \qquad t = 2, \ldots, T. \tag{4.1.6}$$

[26]Note that most of the dynamic panel data literature stacks the data by first collecting the T observations of each unit in a vector and then stacks those N vectors. The grouping used in this paper is more convenient for modelling spatial correlation via (3.1.2).

The initial IV estimator of $\boldsymbol{\theta}$ utilizes $\mathbf{H}_{t,N}$ as instruments[27] for $\Delta\mathbf{y}_{t-1,N}$ and is defined as

$$\underset{(1+p)\times 1}{\widehat{\boldsymbol{\theta}}_N} = \left[\Delta\widehat{\mathbf{Z}}_N'\Delta\mathbf{Z}_N\right]^{-1}_{(1+p)\times(1+p)} \underset{(1+p)\times(T-1)N}{\Delta\widehat{\mathbf{Z}}_N'} \underset{(T-1)N\times 1}{\Delta\mathbf{y}_N}, \tag{4.1.7}$$

where

$$\underset{(T-1)N\times(1+p)}{\Delta\widehat{\mathbf{Z}}_N} = \underset{(T-1)N\times(T-1)N}{\mathbf{H}_N(\mathbf{H}_N'\mathbf{H}_N)^{-1}\mathbf{H}_N'} \cdot \underset{(T-1)N\times(1+p)}{\Delta\mathbf{Z}_N}, \tag{4.1.8}$$

and

$$\underset{(T-1)N\times(1+p)}{\mathbf{H}_N} = \begin{pmatrix} \mathbf{H}_{2,N} \\ \vdots \\ \mathbf{H}_{T,N} \end{pmatrix}. \tag{4.1.9}$$

is a $(T-1)\,N\times(p+1)$ matrix of instruments.[28]

The initial Anderson and Hsiao IV estimator is a special case of a more general GMM estimator discussed in Section 4.3. However, for expositional purposes I derive its asymptotic properties here. Substituting in the definition of the IV

[27] We note that it is possible to use additional lags and/or levels of the dependent variable as instruments and obtain a consistent initial estimator as well. For example, we could use the instruments suggested in Section 4.3, i.e. $\mathbf{H}_t = [\mathbf{y}_{t-2,N},\ldots,\mathbf{y}_{0,N},\mathbf{X}_{t,N},\ldots,\mathbf{X}_{1,N}]$.

[28] Writing the instruments in this fashion leads to an estimator that is based on moment conditions that are averaged both over N and T. It is also possible to define the $\mathbf{H}_N$ matrix as $\mathbf{H}_N = diag\,(\mathbf{H}_2,..,\mathbf{H}_T)$, and the moment conditions are then only averaged over N. In this case the expressions in Lemmas 1 and 2 have to be modified. Note that these two specifications of the instrument matrix lead to different estimators. The projection matrix $\mathbf{H}_N\left(\mathbf{H}_N'\mathbf{H}_N\right)^{-1}\mathbf{H}_N'$ in the first case has elements in the form $\mathbf{H}_{t,N}\left(\sum_{s=2}^{T}\mathbf{H}_{s,N}'\mathbf{H}_{s,N}\right)^{-1}\mathbf{H}_{t,N}'$ while in the second case they are $\mathbf{H}_{t,N}\left(\mathbf{H}_{t,N}'\mathbf{H}_{t,N}\right)^{-1}\mathbf{H}_{t,N}'$. The case of estimators based on moments averaged only over T will be considered in Section 4.2 below.

estimator in equation (4.1.7) yields

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_N &= \boldsymbol{\theta} + \left[\Delta\widehat{\mathbf{Z}}_N'\Delta\mathbf{Z}_N\right]^{-1}\Delta\widehat{\mathbf{Z}}_N'\Delta\mathbf{u}_N \qquad (4.1.10)\\ &= \boldsymbol{\theta} + \left[\Delta\mathbf{Z}_N'\mathbf{H}_N\left(\mathbf{H}_N'\mathbf{H}_N\right)^{-1}\mathbf{H}_N'\Delta\mathbf{Z}_N\right]^{-1}\Delta\mathbf{Z}_N'\mathbf{H}_N\left(\mathbf{H}_N'\mathbf{H}_N\right)^{-1}\mathbf{H}_N'\Delta\mathbf{u}_N.\end{aligned}$$

For the instruments to be valid, I make the following assumption.

Assumption IV1 *The matrix*

$$\underset{(1+p)\times(1+p)}{\mathbf{M}_{H\Delta Z}} = p\lim\frac{1}{(T-1)\,N}\underset{(1+p)\times(T-1)N}{\mathbf{H}_N'}\quad\underset{(T-1)N\times(1+p)}{\Delta\mathbf{Z}_N},$$

exist and is finite with full column rank. The matrix

$$\underset{(1+p)\times(1+p)}{\mathbf{M}_{HH}} = p\lim\frac{1}{(T-1)\,N}\underset{(1+p)\times(T-1)N}{\mathbf{H}_N'}\quad\underset{(T-1)N\times(1+p)}{\mathbf{H}_N},$$

exists and is nonsingular.

We can also define

$$\underset{(1+p)\times(1+p)}{\mathbf{M}_{\Delta Z}} = p\lim\frac{1}{(T-1)\,N}\underset{(1+p)\times(T-1)N}{\Delta\widehat{\mathbf{Z}}_N}\quad\underset{(T-1)N\times(1+p)}{\Delta\mathbf{Z}_N}, \qquad (4.1.11)$$

Observe that $\Delta\widehat{\mathbf{Z}}_N'\Delta\mathbf{Z}_N{=}\Delta\mathbf{Z}_N'\mathbf{H}_N\left(\mathbf{H}_N'\mathbf{H}_N\right)^{-1}\mathbf{H}_N\Delta\mathbf{Z}_N$ and hence $\mathbf{M}_{\Delta Z} = \mathbf{M}_{H\Delta Z}\mathbf{M}_{HH}^{-1}\mathbf{M}_{H\Delta Z}$. Assumption IV1 thus implies that $\mathbf{M}_{\Delta Z}$ exists and is finite. Also note that the assumption that the $\mathbf{M}$ matrices are finite can be derived from earlier restrictions[29]. However, the existence and invertability of $\mathbf{M}_{\Delta Z}$ and $\mathbf{M}_{HH}$ is not guaranteed by Assumptions 1-6.[30] Observe that one could derive Assumption IV1 from existence and nonsingularity of the limits such as $\lim (TN)^{-1} \sum_{t=j+1}^{T} \mathbf{X}'_{t-j,N} \mathbf{X}_{t,N}$.

To derive the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_N$, I note that given Assumption IV1, it remains be to shown that the term $\mathbf{H}'_N \Delta \mathbf{u}_N$ converges in distribution (when appropriately normalized). It will prove convenient to introduce the following additional notation for lagged exogenous variables

$$\underset{(T-1)N\times p}{\mathbf{X}_{-2,N}} = \left(\mathbf{0}_{p\times N}, \mathbf{X}'_{1,N}, .., \mathbf{X}'_{T-2,N}\right)', \tag{4.1.12}$$

the vector collecting all of the model orthogonal innovations

$$\underset{(T+2)N\times 1}{\boldsymbol{\eta}_N} = \left(\boldsymbol{\mu}'_N, \boldsymbol{\xi}'_N, \boldsymbol{\varepsilon}'_{1,N}, ..., \boldsymbol{\varepsilon}'_{T,N}\right)', \tag{4.1.13}$$

with $\boldsymbol{\xi}_N = \sum_{j=0}^{\infty} \phi^j \boldsymbol{\varepsilon}_{-j,N}$, and a $(T-1)\times T$ difference operator $\mathbf{D}$ and a $(T-1)\times$

[29] For example, the elements of $\mathbf{M}_{HH}$ consist of first and second moments of the stochastic process y_{it} interacted with the exogenous variables. These are bounded by Assumptions 1-6.

[30] For example, Arrelano (1989) examines a univariate AR(1) model with first-order autoregressive exogenous variables, and finds that when the first differences of endogenous variables lagged twice are used as instruments, there exists a significant range of parameters for which there is a singularity point in the estimator. The paper also suggests that the estimator that uses second lags of the levels of the endogenous variables does not have the singularity problem for a reasonable range of parameters. However, this conclusion does not readily generalize for all possible exogenous variables.

$(T-1)$ matrix $\mathbf{\Phi}$

$$\underset{(T-1)\times T}{\mathbf{D}} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}, \underset{(T-1)\times(T-1)}{\mathbf{\Phi}} = \begin{pmatrix} 1 & \phi & \cdots & \phi^{T-2} \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \phi \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{4.1.14}$$

Observe that given Assumptions 1, 2, and 6, the variance covariance matrix of $\boldsymbol{\eta}_N$ is

$$E\left(\boldsymbol{\eta}_N\boldsymbol{\eta}_N'\right)_{(T+2)N\times(T+2)N} = \left(\underset{(T+2)\times(T+2)}{\mathbf{\Sigma}_{\boldsymbol{\eta},N}} \otimes \mathbf{I}_N\right), \tag{4.1.15}$$

where the $(T+2)\times(T+2)$ diagonal matrix $\mathbf{\Sigma}_{\boldsymbol{\eta},N}$ is

$$\underset{(T+2)\times(T+2)}{\mathbf{\Sigma}_{\boldsymbol{\eta},N}} = diag\left(\sigma^2_{\mu,N}, \frac{\sigma^2_{\varepsilon,N}}{1-\phi}, \sigma^2_{\varepsilon,N}, ..., \sigma^2_{\varepsilon,N}\right). \tag{4.1.16}$$

I first express the elements of $\mathbf{H}_N'\Delta\mathbf{u}_N$ (which are $\mathbf{y}_{-2,N}'\Delta\mathbf{u}_N$ and $\Delta\mathbf{X}_N'\Delta\mathbf{u}_N$) in terms of lagged model disturbances and dependent variables:

Lemma 1 *Under the specification (3.1.4) with Assumptions 1-6 and IV1 we have that*

$$\mathbf{y}_{-2,N}'\Delta\mathbf{u}_N = \mathbf{f}_N'\left(\mathbf{I}_{T+2}\otimes\mathbf{P}_N\right)\boldsymbol{\eta}_N + \boldsymbol{\eta}_N'\left(\mathbf{F}\otimes\mathbf{P}_N'\mathbf{P}_N\right)\boldsymbol{\eta}_N,$$

where the $(T+2)\,N\times 1$ *vector* $\mathbf{f}_N$ *is given by*

$$\mathbf{f}_N=\left[\begin{pmatrix}\mathbf{0}_{2\times(T-1)}\\ \underset{T\times(T-1)}{\mathbf{D}'}\end{pmatrix}\otimes\mathbf{I}_N\right]\underset{(T-1)N\times(T-1)N}{(\mathbf{\Phi}'\otimes\mathbf{I}_N)}\left\{\underset{(T-1)N\times p}{\mathbf{X}_{-2,N}}\underset{p\times 1}{\boldsymbol{\beta}}+\underset{(T-1)N\times 1}{\begin{bmatrix}\underset{N\times 1}{E(\mathbf{y}_{0,N})}\\ \mathbf{0}_{(T-2)N\times 1}\end{bmatrix}}\right\}$$

and the $T+2\times T+2$ *matrix* $\mathbf{F}$ *is*

$$\underset{(T+2)\times(T+2)}{\mathbf{F}}=\begin{pmatrix}\frac{1}{1-\phi} & \mathbf{1}_{1\times(T-2)}\\ 1_{1\times 1} & \mathbf{0}_{1\times(T-2)}\\ \mathbf{0}_{(T-2)\times 1} & \mathbf{I}_{T-2}\\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times(T-2)}\end{pmatrix}_{(T+2)\times(T-1)}\underset{(T-1)\times(T-1)}{\mathbf{\Phi}}\underset{(T-1)\times(T+2)}{\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)}.$$

Furthermore $\Delta\mathbf{X}_N'\Delta\mathbf{u}_N$ *can also be expressed as a linear function of* $\boldsymbol{\eta}_N$*:*

$$\Delta\mathbf{X}_N'\Delta\mathbf{u}_N=\Delta\mathbf{X}_N'\left[\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)\otimes\mathbf{P}_N\right]\boldsymbol{\eta}_N.$$

Proof. See the Appendix C.1. ■

Notice that as indicated by the subscript, the size of the $\mathbf{f}_N$ vector depends on the sample size. Since T is fixed, I do not use subscripts for matrices $\mathbf{F}$ and $\mathbf{D}$ whose size and elements only depend on T and not on N.

To determine the asymptotic variance of the estimator, I will make use of the following Lemma that gives an expression for expected value and variance covariance matrix of the moment conditions:

Lemma 2 *Suppose Assumptions 1-6 hold. The expected value of the vector of quadratic forms* $\mathbf{H}_N'\Delta\mathbf{u}_N$ *is zero. Its variance covariance matrix is given by*

$$\begin{aligned}\underset{(1+p)\times(1+p)}{\mathbf{V}_N} &= E\left(\mathbf{H}_N'\Delta\mathbf{u}_N\Delta\mathbf{u}_N'\mathbf{H}_N\right)\\ &= \underset{(1+p)\times(T+2)N}{\mathbf{S}_N'}\underset{(T+2)N\times(T+2)N}{\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}_N\mathbf{P}_N'\right)}\underset{(T+2)N\times(1+p)}{\mathbf{S}_N}+\begin{pmatrix}\nu_N & \mathbf{0}_{1\times p}\\ \mathbf{0}_{p\times 1} & \mathbf{0}_{p\times p}\end{pmatrix},\end{aligned}$$

where

$$\mathbf{S}_N=\underset{(T+2)N\times(1+p)}{\left(\underset{(T+2)N\times 1}{\mathbf{f}_N},\underset{(T-1)N\times(T+2)N}{\left[\underset{(T-1)\times(T+2)}{\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)'}\otimes\mathbf{I}_N\right]}\underset{(T-1)N\times p}{\Delta\mathbf{X}_N}\right)},$$

and

$$\nu_N=2tr\left(\mathbf{F}^S\mathbf{\Sigma}_{\boldsymbol{\eta},N}\mathbf{F}^S\mathbf{\Sigma}_{\boldsymbol{\eta},N}\right)\cdot tr\left(\mathbf{P}_N\mathbf{P}_N'\mathbf{P}_N\mathbf{P}_N'\right),$$

with $\mathbf{F}^S=\frac{1}{2}\left(\mathbf{F}+\mathbf{F}'\right)$.

Proof. See the Appendix C.1. ■

To rule out cases where the moment conditions have zero asymptotic variance, I make the following assumption:

Assumption IV2 *The smallest eigenvalue of* $\left[(T-1)\,N\right]^{-1}\mathbf{S}_N'\mathbf{S}_N$ *is uniformly bounded away from zero for* $T\geq 2$.

Although $\mathbf{S}_N$ depends on the sample size, the dimensions of $\mathbf{S}_N'\mathbf{S}_N$ do not change with N. Furthermore, notice that the assumption also implies that $E\left(\mathbf{H}_N'\mathbf{H}_N\right)$

has eigenvalues uniformly bounded away from zero and, therefore, also implies the invertability of $\mathbf{M}_{HH}$ in Assumption IV1.[31] The above Assumption together with Assumption 4 allows us to prove the following Lemma:

Lemma 3 *Suppose Assumptions 1-4 and IV2 hold. The smallest eigenvalue of* $[(T-1)\,N]^{-1}\,\mathbf{V}_N$ *is uniformly bounded away from zero for* $T \geq 2$*.*

Proof. See the Appendix C.1. ■

The representation of $\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N$ and $\Delta\mathbf{X}'_N\Delta\mathbf{u}_N$ as linear-quadratic forms in $\boldsymbol{\eta}_N$, lets us apply a central limit theorem for quadratic forms of triangular arrays and derive the asymptotic distribution of the IV estimator. The central limit theorem (CLT) I use is given in Appendix A. It is based on a result from Kelejian and Prucha (2005) and is an extension of a CLT in Kelejian and Prucha (2001).

Proposition 1 *Under Assumptions 1-6, IV1 and IV2, we have that*

$$\mathbf{V}_N^{-1/2} \cdot \mathbf{H}'_N\Delta\mathbf{u}_N \xrightarrow{d} N\left(\mathbf{0}, \mathbf{I}_{p+1}\right),$$

where

$$\left(\mathbf{V}_N^{1/2}\right)\left(\mathbf{V}_N^{1/2}\right)' = \mathbf{V}_N.$$

Proof. See the Appendix C.1. ■

[31] However, it does not guarantee the existence of the limit in Assumption IV1.

To be able to write down explicit asymptotic distribution of the estimator, I make the following assumption.

Assumption IV3 $\lim_{N\to\infty}\frac{1}{(T-1)N}\mathbf{V}_N=\mathbf{V}$,*where* $\mathbf{V}$ *is finite.*

We then have the following Theorem:

Theorem 1 *Under Assumptions 1-6, and IV1-IV3, we have that*

$$\sqrt{(T-1)\,N\cdot}\left(\widehat{\boldsymbol{\theta}}_N-\boldsymbol{\theta}\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Psi}\right),$$

with

$$\begin{aligned}\underset{(1+p)\times(1+p)}{\boldsymbol{\Psi}} \quad = \quad & \underset{(1+p)\times(1+p)}{\mathbf{M}_{\Delta Z}^{-1}}\ \underset{(1+p)\times(1+p)}{\mathbf{M}_{H\Delta Z}'}\ \underset{(1+p)\times(1+p)}{\mathbf{M}_{HH}^{-1}}\ * \\ & \underset{(1+p)\times(1+p)}{\mathbf{V}}\ \underset{(1+p)\times(1+p)}{\mathbf{M}_{HH}^{-1}}\ \underset{(1+p)\times(1+p)}{\mathbf{M}_{H\Delta Z}}\ \underset{(1+p)\times(1+p)}{\mathbf{M}_{\Delta Z}^{-1}}\end{aligned}$$

Proof. See the Appendix C.1. ■

I do not provide an estimate of $\boldsymbol{\Psi}$ since it would depend on an estimate of the $\mathbf{P}_N=(\mathbf{I}_N-\rho\mathbf{W}_N)^{-1}$ matrix which includes an unknown parameter ρ. I will provide small sample guidance for the second stage estimator in Section 4.3. Note that by Theorem 17 in Pötcher and Prucha (2001), the result in the above Theorem implies that $\sqrt{(T-1)\,N}\,\widehat{\boldsymbol{\theta}}_N$ is $O_p(1)$ and the initial estimator IV satisfies the conditions required in the following section and hence can be used in the subsequent estimation steps.

4.2 Estimation of the Degree of Spatial Autocorrelation

The specification in this thesis reduces to that of Kapoor et al. (2005) in the static case ($\phi = 0$) which is in turn a generalization of the single cross-section case in Kelejian and Prucha (1999). In this section, I will show that the procedure adopted in Kapoor et al. (2005) provides a consistent estimate of the spatial autoregressive parameter in a dynamic panel data model as well. To do that, I define the generalized moments (GM) estimator following Kapoor et al. (2005) and then extend their proofs for the dynamic case. For simplicity, I only consider one of the weighting schemes for the moment condition in Kapoor et al. (2005).

Observe that the spatial GM estimator in this section is essentially the same as the estimator in Kapoor et al. (2005). However, the presence of stochastic regressors (lagged dependent variable) renders the proofs in that paper inapplicable to the specification considered in this thesis. Nevertheless, the proofs in this section, with small exceptions (most notably Lemmas C4 and C6 in the Appendix C.2), are a direct analogy of those in Kapoor et al. (2005).

I take an initial consistent estimate of the spatially correlated errors and use it to estimate the spatial autocorrelation parameter based on a set of moment conditions. The initial consistent estimate of the errors can be, for example based on the IV estimator in the previous section. The moment conditions are chosen so that the estimator will have an Analysis of Variance interpretation.

Consider an estimator $\underset{p+1\times 1}{\widehat{\boldsymbol{\theta}}_N}$ of the parameter vector $\underset{p+1\times 1}{\boldsymbol{\theta}}$ such that

$\sqrt{(T-1)\,N}\;\widehat{\boldsymbol{\theta}}_N = O_p(1)$ and denote the predictors of $\mathbf{u}_t$ by $\widehat{\mathbf{u}}_t$:

$$\underset{N\times 1}{\widehat{\mathbf{u}}_{t,N}} = \underset{N\times 1}{\mathbf{y}_{t,N}} - \left(\mathbf{y}_{t-1,N}, \mathbf{X}_{t,N}\right)_{N\times p+1} \cdot \underset{p+1\times 1}{\widehat{\boldsymbol{\theta}}_N}, \qquad 1 \leq t \leq T. \tag{4.2.1}$$

The model implies that (see equation 3.1.2 in Chapter 3)

$$\underset{N\times 1}{\mathbf{u}_{t,N}} = \rho \underset{N\times N}{\mathbf{W}_N} \underset{N\times 1}{\mathbf{u}_{t,N}} + \underset{N\times 1}{\boldsymbol{\upsilon}_{t,N}}, \qquad 1 \leq t \leq T, \tag{4.2.2}$$

where $\boldsymbol{\upsilon}_{t,N} = \boldsymbol{\varepsilon}_{t,N} + \boldsymbol{\mu}_N$. In a stacked notation this becomes

$$\underset{NT\times 1}{\mathbf{u}_N} = \rho \left(\mathbf{I}_T \otimes \mathbf{W}_N\right)_{NT\times NT} \underset{NT\times 1}{\mathbf{u}_N} + \underset{NT\times 1}{\boldsymbol{\upsilon}_N}, \tag{4.2.3}$$

where $\mathbf{u}_N = \left[\mathbf{u}_{1,N}', ..., \mathbf{u}_{T,N}'\right]'$ and $\boldsymbol{\upsilon}_N = \boldsymbol{\varepsilon}_N + \left(\mathbf{e}_T \otimes \boldsymbol{\mu}_N\right)$, with $\boldsymbol{\varepsilon}_N = \left[\boldsymbol{\varepsilon}_{1,N}', ..., \boldsymbol{\varepsilon}_{T,N}'\right]'$, $\mathbf{e}_T$ being a $T \times 1$ vector of unit elements, and $\boldsymbol{\mu}_N$ the $N \times 1$ vector of individual effects. It will prove convenient to introduce the following notation:

$$\begin{aligned}
\overline{\mathbf{u}}_N &= \left(\mathbf{I}_T \otimes \mathbf{W}_N\right) \mathbf{u}_N, \qquad (4.2.4)\\
\overline{\overline{\mathbf{u}}}_N &= \left(\mathbf{I}_T \otimes \mathbf{W}_N\right) \overline{\mathbf{u}}_N,\\
\overline{\boldsymbol{\upsilon}}_N &= \left(\mathbf{I}_T \otimes \mathbf{W}_N\right) \boldsymbol{\upsilon}_N.
\end{aligned}$$

I will also use the following transformation matrices that are utilized in the error component literature:

$$\underset{NT\times NT}{\mathbf{Q}_{0,N}} = \left(\mathbf{I}_T - \frac{\mathbf{J}_T}{T}\right) \otimes \mathbf{I}_N, \tag{4.2.5}$$
$$\underset{NT\times NT}{\mathbf{Q}_{1,N}} = \frac{\mathbf{J}_T}{T} \otimes \mathbf{I}_N,$$

where $\mathbf{J}_T = \mathbf{e}_T\mathbf{e}_T'$ is a $T \times T$ matrix of unit elements.[32] Note that using the transformation matrices, we can express the variance-covariance matrix of the innovations as

$$\begin{aligned} E\left(\boldsymbol{\upsilon}_N\boldsymbol{\upsilon}_N'\right)_{NT\times NT} &= \sigma^2_{\varepsilon,N}\mathbf{I}_{NT} + \sigma^2_{\mu,N}\left(\mathbf{J}_T \otimes \mathbf{I}_N\right) \\ &= \sigma^2_{\varepsilon,N}\mathbf{Q}_{0,N} + \sigma^2_{1,N}\mathbf{Q}_{1,N}, \end{aligned} \tag{4.2.6}$$

where $\sigma^2_{1,N} = \sigma^2_{\varepsilon,N} + T \cdot \sigma^2_{\mu,N}$.

The spatial GM estimator is based on the following moment conditions:

$$\begin{aligned} E\left(\boldsymbol{\upsilon}_N'\mathbf{Q}_{0,N}\boldsymbol{\upsilon}_N\right) &= N\left(T-1\right)\sigma^2_{\varepsilon,N}, \\ E\left(\overline{\boldsymbol{\upsilon}}_N'\mathbf{Q}_{0,N}\overline{\boldsymbol{\upsilon}}_N\right) &= \left(T-1\right)\sigma^2_{\varepsilon,N} \cdot tr\left(\mathbf{W}_N'\mathbf{W}_N\right), \\ E\left(\overline{\boldsymbol{\upsilon}}_N'\mathbf{Q}_{0,N}\boldsymbol{\upsilon}_N\right) &= 0, \\ E\left(\boldsymbol{\upsilon}_N'\mathbf{Q}_{1,N}\boldsymbol{\upsilon}_N\right) &= N\sigma^2_{1,N}, \\ E\left(\overline{\boldsymbol{\upsilon}}_N'\mathbf{Q}_{1,N}\overline{\boldsymbol{\upsilon}}_N\right) &= \sigma^2_{1,N} \cdot tr\left(\mathbf{W}_N'\mathbf{W}_N\right), \\ E\left(\overline{\boldsymbol{\upsilon}}_N'\mathbf{Q}_{1,N}\boldsymbol{\upsilon}_N\right) &= 0. \end{aligned} \tag{4.2.7}$$

[32] The $\mathbf{Q}_1$ transformation calculates unit specific sample means while the $\mathbf{Q}_0$ transformation substracts them from the original variable.

For derivation of the moment conditions see Kapoor et al. (2005). Notice that based on (4.2.3), the moment conditions can be rewritten in terms of the transformed (by $\mathbf{Q}_{j,N}$) disturbance vectors $\mathbf{u}_N$, $\overline{\mathbf{u}}_N$ and $\overline{\overline{\mathbf{u}}}_N$:

$$\boldsymbol{\gamma}_N = \boldsymbol{\Gamma}_N \boldsymbol{\alpha}, \tag{4.2.8}$$

where $\boldsymbol{\alpha} = \left(\rho, \rho^2, \sigma^2_{\varepsilon,N}, \sigma^2_{1,N}\right)'$, and

$$\underset{6\times4}{\boldsymbol{\Gamma}_N} = E\begin{pmatrix} \gamma^0_{11,N} & \gamma^0_{12,N} & \gamma^0_{13,N} & 0 \\ \gamma^0_{21,N} & \gamma^0_{22,N} & \gamma^0_{23,N} & 0 \\ \gamma^0_{31,N} & \gamma^0_{32,N} & \gamma^0_{33,N} & 0 \\ \gamma^1_{11,N} & \gamma^1_{12,N} & 0 & \gamma^1_{13,N} \\ \gamma^1_{21,N} & \gamma^1_{22,N} & 0 & \gamma^1_{23,N} \\ \gamma^1_{31,N} & \gamma^1_{32,N} & 0 & \gamma^1_{33,N} \end{pmatrix}, \qquad \underset{6\times1}{\boldsymbol{\gamma}_N} = E\begin{pmatrix} \gamma^0_{1,N} \\ \gamma^0_{2,N} \\ \gamma^0_{3,N} \\ \gamma^1_{1,N} \\ \gamma^1_{2,N} \\ \gamma^1_{3,N} \end{pmatrix}, \tag{4.2.9}$$

with ($j=0,1$)

$$
\begin{aligned}
\gamma_{11,N}^{j} &= \frac{2}{N\left(T-1\right)^{1-j}}\mathbf{u}_N'\mathbf{Q}_{j,N}\overline{\mathbf{u}}_N, \qquad \gamma_{12}^{j}=\frac{-1}{N\left(T-1\right)^{1-j}}\overline{\mathbf{u}}_N'\mathbf{Q}_{j,N}\overline{\mathbf{u}}_N,\\
\gamma_{21,N}^{j} &= \frac{2}{N\left(T-1\right)^{1-j}}\overline{\overline{\mathbf{u}}}_N'\mathbf{Q}_{j,N}\overline{\mathbf{u}}_N, \qquad \gamma_{22}^{j}=\frac{-1}{N\left(T-1\right)^{1-j}}\overline{\overline{\mathbf{u}}}_N'\mathbf{Q}_{j,N}\overline{\overline{\mathbf{u}}}_N,\\
\gamma_{31,N}^{j} &= \frac{1}{N\left(T-1\right)^{1-j}}\left(\mathbf{u}_N'\mathbf{Q}_{j,N}\overline{\overline{\mathbf{u}}}_N+\overline{\mathbf{u}}_N'\mathbf{Q}_{j,N}\overline{\mathbf{u}}_N\right),\\
\gamma_{32,N}^{j} &= \frac{-1}{N\left(T-1\right)^{1-j}}\overline{\mathbf{u}}_N'\mathbf{Q}_{j,N}\overline{\overline{\mathbf{u}}}_N, \qquad\qquad\qquad\qquad (4.2.10)\\
\gamma_{13,N}^{j} &= 1, \qquad \gamma_{1}^{j}=\frac{1}{N\left(T-1\right)^{1-j}}\mathbf{u}_N'\mathbf{Q}_{j,N}\mathbf{u}_N,\\
\gamma_{23,N}^{j} &= \frac{1}{N}tr\left(\mathbf{W}_N'\mathbf{W}_N\right), \qquad \gamma_{2}^{j}=\frac{1}{N\left(T-1\right)^{1-j}}\overline{\mathbf{u}}_N'\mathbf{Q}_{j,N}\overline{\mathbf{u}}_N,\\
\gamma_{33,N}^{j} &= 0, \qquad \gamma_{3}^{j}=\frac{1}{N\left(T-1\right)^{1-j}}\mathbf{u}_N'\mathbf{Q}_{j,N}\overline{\mathbf{u}}_N.
\end{aligned}
$$

The sample counterparts of the six equations in (4.2.9) replace $\mathbf{u}_N$ with $\widehat{\mathbf{u}}_N=\left(\widehat{\mathbf{u}}_{1,N}',...,\widehat{\mathbf{u}}_{T,N}'\right)'$ based on (4.2.1) with the implied notation $\widehat{\overline{\mathbf{u}}}_N=\left(\mathbf{I}_T\otimes\mathbf{W}_N\right)\widehat{\mathbf{u}}_N$ and $\widehat{\overline{\overline{\mathbf{u}}}}_N=\left(\mathbf{I}_T\otimes\mathbf{W}_N\right)\widehat{\overline{\mathbf{u}}}_N$:

$$\underset{6\times1}{\mathbf{g}_N}=\underset{6\times4}{\mathbf{G}_N}\underset{4\times1}{\boldsymbol{\alpha}}+\underset{6\times1}{\boldsymbol{\vartheta}_N}, \qquad (4.2.11)$$

where $\boldsymbol{\vartheta}_N$ can be viewed as a vector of regression residuals and

$$\underset{6\times 4}{\mathbf{G}_N} = \begin{pmatrix} g^0_{11,N} & g^0_{12,N} & g^0_{13,N} & 0 \\ g^0_{21,N} & g^0_{22,N} & g^0_{23,N} & 0 \\ g^0_{31,N} & g^0_{32,N} & g^0_{33,N} & 0 \\ g^1_{11,N} & g^1_{12,N} & 0 & g^1_{13,N} \\ g^1_{21,N} & g^1_{22,N} & 0 & g^1_{23,N} \\ g^1_{31,N} & g^1_{32,N} & 0 & g^1_{33,N} \end{pmatrix}, \qquad \underset{6\times 1}{\mathbf{g}_N} = \begin{pmatrix} g^0_{1,N} \\ g^0_{2,N} \\ g^0_{3,N} \\ g^1_{1,N} \\ g^1_{2,N} \\ g^1_{3,N} \end{pmatrix}, \tag{4.2.12}$$

with $(j = 0, 1)$

$$\begin{aligned} g^j_{11,N} &= \frac{2}{N(T-1)^{1-j}} \widehat{\mathbf{u}}'_N \mathbf{Q}_{j,N} \widehat{\overline{\mathbf{u}}}_N, \qquad g^j_{12} = \frac{-1}{N(T-1)^{1-j}} \widehat{\overline{\mathbf{u}}}'_N \mathbf{Q}_{j,N} \widehat{\overline{\mathbf{u}}}_N, \\ g^j_{21,N} &= \frac{2}{N(T-1)^{1-j}} \widehat{\overline{\overline{\mathbf{u}}}}'_N \mathbf{Q}_{j,N} \widehat{\overline{\mathbf{u}}}_N, \qquad g^j_{22} = \frac{-1}{N(T-1)^{1-j}} \widehat{\overline{\overline{\mathbf{u}}}}'_N \mathbf{Q}_{j,N} \widehat{\overline{\overline{\mathbf{u}}}}_N, \\ g^j_{31,N} &= \frac{1}{N(T-1)^{1-j}} \left(\widehat{\mathbf{u}}'_N \mathbf{Q}_{j,N} \widehat{\overline{\overline{\mathbf{u}}}}_N + \widehat{\overline{\mathbf{u}}}'_N \mathbf{Q}_{j,N} \widehat{\overline{\mathbf{u}}}_N \right), \\ g^j_{32,N} &= \frac{-1}{N(T-1)^{1-j}} \widehat{\overline{\mathbf{u}}}'_N \mathbf{Q}_{j,N} \widehat{\overline{\overline{\mathbf{u}}}}_N, \\ g^j_{13,N} &= 1, \qquad g^j_1 = \frac{1}{N(T-1)^{1-j}} \widehat{\mathbf{u}}'_N \mathbf{Q}_{j,N} \widehat{\mathbf{u}}_N, \\ g^j_{23,N} &= \frac{1}{N} tr\left(\mathbf{W}'_N \mathbf{W}_N\right), \qquad g^j_2 = \frac{1}{N(T-1)^{1-j}} \widehat{\overline{\mathbf{u}}}'_N \mathbf{Q}_{j,N} \widehat{\overline{\mathbf{u}}}_N, \\ g^j_{33,N} &= 0, \qquad g^j_3 = \frac{1}{N(T-1)^{1-j}} \widehat{\mathbf{u}}'_N \mathbf{Q}_{j,N} \widehat{\overline{\mathbf{u}}}_N. \end{aligned} \tag{4.2.13}$$

The generalized moments (GM) estimator of $\boldsymbol{\delta} = \left(\rho, \sigma^2_{\varepsilon,N}, \sigma^2_{1,N}\right)'$ say $\widehat{\boldsymbol{\delta}}_N$ can be written as

$$\underset{3\times 1}{\widehat{\boldsymbol{\delta}}_N} = \arg\min_{\boldsymbol{\delta}\in\boldsymbol{\Theta}} \left\{ \left(\underset{6\times 1}{\mathbf{g}_N} - \underset{6\times 4}{\mathbf{G}_N} \underset{4\times 1}{\boldsymbol{\alpha}} \right)' \underset{6\times 6}{\mathbf{A}_N} \left(\underset{6\times 1}{\mathbf{g}_N} - \underset{6\times 4}{\mathbf{G}_N} \underset{4\times 1}{\boldsymbol{\alpha}} \right) \right\}, \tag{4.2.14}$$

where Θ is the admissible optimization space; in particular it is assumed that $\Theta = \left\{ \rho, \sigma^2_{\varepsilon,N}, \sigma^2_{1,N} \middle| \rho \in [0, b_1], \sigma^2_{\varepsilon,N} \in [0, b_2], \sigma^2_{1,N} \in [0, b_3] \right\}$ with b_1, b_2 and b_3 being predetermined constants. The moments are weighted by a sequence of weighting matrices $\mathbf{A}_N$. Following Kapoor et al. (2005), two choices for $\mathbf{A}_N$ are considered. An initial unweighted spatial GM estimators uses $\mathbf{A}_N = \mathbf{I}_6$. The second choice is to use an approximation to variance covariance matrix of the moments. In particular, Kapoor et al. (2005) show that under normality the variance covariance matrix of the six moment conditions in (4.2.7) is given by

$$\underset{6\times 6}{\boldsymbol{\Xi}_N} = \begin{bmatrix} \frac{1}{T-1}\sigma^2_{\varepsilon N} & 0 \\ 0 & \sigma^2_{1N} \end{bmatrix} \otimes \underset{2\times 2}{\mathbf{T}_{W,N}}, \tag{4.2.15}$$

where

$$\underset{2\times 2}{\mathbf{T}_{W,N}} = \begin{bmatrix} 2 & 2tr\left(\frac{\mathbf{W}'_N\mathbf{W}_N}{N}\right) & 0 \\ 2tr\left(\frac{\mathbf{W}'_N\mathbf{W}_N}{N}\right) & 2tr\left(\frac{\mathbf{W}'_N\mathbf{W}_N\mathbf{W}'_N\mathbf{W}_N}{N}\right) & tr\left(\frac{\mathbf{W}'_N\mathbf{W}_N\left(\mathbf{W}_N+\mathbf{W}'_N\right)}{N}\right) \\ 0 & tr\left(\frac{\mathbf{W}'_N\mathbf{W}_N\left(\mathbf{W}_N+\mathbf{W}'_N\right)}{N}\right) & tr\left(\frac{\mathbf{W}_N\mathbf{W}_N+\mathbf{W}'_N\mathbf{W}_N}{N}\right) \end{bmatrix}. \tag{4.2.16}$$

The weighted spatial GM estimator then replaces $\sigma^2_{\varepsilon N}$ and σ^2_{1N} by their initial estimators and utilizes the weighting matrices $\mathbf{A}_N = \widehat{\mathbf{\Xi}}_N^{-1}\left(\widehat{\sigma^2_{\varepsilon}}_N, \widehat{\sigma^2_1}_N\right)$ where

$$\underset{6\times 6}{\widehat{\mathbf{\Xi}}_N\left(\widehat{\sigma^2_{\varepsilon}}_N, \widehat{\sigma^2_1}_N\right)} = \begin{bmatrix} \frac{1}{T-1}\widehat{\sigma^2_{\varepsilon}}_N & 0 \\ 0 & \widehat{\sigma^2_1}_N \end{bmatrix} \otimes \underset{2\times 2}{\mathbf{T}_{W,N}}, \tag{4.2.17}$$

and the estimators $\widehat{\sigma^2_{\varepsilon}}_N, \widehat{\sigma^2_1}_N$ are based on the initial unweighted spatial GM estimator.

The following additional assumption is required in order to establish consistency of $\widehat{\boldsymbol{\delta}}_{GM,N}$ (the assumption is used to demonstrate that the estimator is identifiably unique):

Assumption GM1 *The smallest eigenvalue of $\mathbf{\Gamma}_N'\mathbf{\Gamma}_N$ is uniformly bounded away from zero. Furthermore,* $0 < \underline{\lambda} \leq \lambda_{\min}\left(\mathbf{\Xi}_N^{-1}\right) \leq \lambda_{\max}\left(\mathbf{\Xi}_N^{-1}\right) \leq \overline{\lambda} < \infty$.

The following theorem establishes the consistency of the GM estimator.

Theorem 2 *Suppose Assumptions 1-6 and GM1 hold.*
If $\widehat{\boldsymbol{\theta}}_N$ is a consistent estimator of $\boldsymbol{\theta}$ with $\sqrt{(T-1)\,N}\,\widehat{\boldsymbol{\theta}}_N = O_p(1)$, then

$$\widehat{\boldsymbol{\delta}}_N \xrightarrow{P} \boldsymbol{\delta} \quad as \quad N \to \infty.$$

Proof. See the Appendix C.2. ■

4.3 Second Stage GMM Estimation

In this section I define a second stage generalized method of moments (GMM) estimator of the slope coefficients $\boldsymbol{\theta} = (\phi, \boldsymbol{\beta}')'$ and derive its asymptotic distribution. I base the estimator on a set of weighted moment conditions. In the first part of this section, I consider a general case of stochastic instruments of a certain form and show that the normalized GMM estimator based on these moment conditions converges (under the assumptions maintained in this thesis and under additional assumptions spelled out in this section) in distribution. Next, I consider the choice of an optimal weighting matrix for a given set of instruments. I close the section with an application of these results to a feasible GMM estimator based on moment conditions utilized in the literature (see Chapter 2 for a review).

Consider again the model (4.1.2)

$$\underset{(T-1)N\times 1}{\Delta \mathbf{y}_N} = \underset{(T-1)N\times(1+p)}{\Delta \mathbf{Z}_N} \underset{(1+p)\times 1}{\boldsymbol{\theta}} + \underset{(T-1)N\times 1}{\Delta \mathbf{u}_N}, \tag{4.3.1}$$

where the explanatory variable $\Delta \mathbf{Z}_N = (\Delta \mathbf{y}_{-1,N}, \Delta \mathbf{X}_N)$ contains lagged endogenous variables. Let $\mathbf{H}_N$ be a $(T-1)\,N \times k$ set of instruments (to be determined later) such that

$$E\left(\underset{k\times(T-1)N}{\mathbf{H}'_N} \underset{(T-1)N\times 1}{\Delta \mathbf{u}_N}\right) = \mathbf{0}_{k\times 1}. \tag{4.3.2}$$

Also, let $\mathbf{A}_N$ be a sequence of nonsingular symmetric $k \times k$ matrices with nonsingular limit

$$p \lim_{N\to\infty} \underset{k\times k}{\mathbf{A}_N} = \underset{k\times k}{\mathbf{A}}. \tag{4.3.3}$$

Consider the GMM estimator $\widetilde{\boldsymbol{\theta}}_N$ based on instruments $\mathbf{H}_N$ and weights $\mathbf{A}_N$ defined as a minimizer of

$$(\Delta \mathbf{y}_N - \Delta \mathbf{Z}_N \boldsymbol{\theta})'_{1\times(T-1)N} \underset{(T-1)N\times k}{\mathbf{H}_N} \underset{k\times k}{\mathbf{A}_N^{-1}} \underset{k\times(T-1)N}{\mathbf{H}'_N} (\Delta \mathbf{y}_N - \Delta \mathbf{Z}_N \boldsymbol{\theta})_{(T-1)N}\,, \tag{4.3.4}$$

i.e.,

$$\begin{aligned} \underset{(1+p)\times 1}{\widetilde{\boldsymbol{\theta}}_N} &= \left[\underset{(1+p)\times(T-1)N}{\Delta \mathbf{Z}'_N} \underset{(T-1)N\times k}{\mathbf{H}_N} \underset{k\times k}{\mathbf{A}_N^{-1}} \underset{k\times(T-1)N}{\mathbf{H}'_N} \underset{(T-1)N\times(1+p)}{\Delta \mathbf{Z}_N} \right]^{-1} * \\ & \underset{(1+p)\times(T-1)N}{\Delta \mathbf{Z}'_N} \underset{(T-1)N\times k}{\mathbf{H}_N} \underset{k\times k}{\mathbf{A}_N^{-1}} \underset{k\times(T-1)N}{\mathbf{H}'_N} \underset{(T-1)N\times 1}{\Delta \mathbf{y}_N}\,. \end{aligned} \tag{4.3.5}$$

Note that it is possible to define an initial IV estimator is of this form, with $\mathbf{A}_N = [(T-1)\,N]^{-1}\,\mathbf{H}'_N\mathbf{H}_N$. The initial IV estimator in Section 4.1 utilized lagged levels of the endogenous variable as instruments and the instrument matrix $\mathbf{H}_N$ was given in (4.1.5) and (4.1.9).

In the literature (e.g. Arellano and Bond, 1991; see Chapter 2 for a review) the instrument set at time t is expanded to include all available lags of the endogenous (and possibly also the exogenous) variable. As a result the number of the moment condition is different at different time periods and the instrument matrix $\mathbf{H}_N$ can be, for example, as in (4.3.20) below. Observe that under the specification considered in this thesis, the endogenous variable can be expressed as linear forms of the (mutually independent) innovations of the model:

Lemma 4 *Under Assumptions 1-6 we can express the dependent variable as*

$$\underset{N\times 1}{\mathbf{y}_{t,N}} = \underset{N\times 1}{\mathbf{a}_{t,N}} + \left(\underset{1\times(T+2)}{\mathbf{b}_{t}} \otimes \underset{N\times N}{\mathbf{P}_{N}} \right) \underset{(T+2)N\times 1}{\boldsymbol{\eta}_{N}},$$

where the sequence of nonstochastic $N \times 1$ *vectors* $\mathbf{a}_{t,N}$ *and the sequence of nonstochastic* $1 \times (T+2)$ *vectors* $\mathbf{b}_t$ *have elements uniformly bounded in absolute value.*

Proof. See the Appendix C.3. ■

Motivated by the expression in the above Lemma, I consider a set of k_t stochastic instruments in each time period $\underset{N\times k_t}{\mathbf{H}_{t,N}} = (\mathbf{h}_{1,t,N}, ..., \mathbf{h}_{k_t,t,N})$ and assume that each instrument can be expressed as a linear combination of the model disturbances,

$$\underset{N\times 1}{\mathbf{h}_{r,t,N}} = \underset{N\times 1}{\mathbf{a}_{rt,N}} + \left(\underset{1\times(T+2)}{\mathbf{b}_{rt}} \otimes \underset{N\times N}{\mathbf{P}_{N}} \right) \underset{(T+2)N\times 1}{\boldsymbol{\eta}_{N}}, \qquad r = 1, ..., k_t, \tag{4.3.6}$$

where the sequence of nonstochastic $N \times 1$ vectors $\mathbf{a}_{t,N}$ and the sequence of nonstochastic $1 \times (T+2)$ vectors $\mathbf{b}_t$ have elements uniformly bounded in absolute value. The total number of instruments is $k = k_2 + ... + k_T$ and the instruments can be collected in a $(T-1)\,N \times k$ block-diagonal matrix[33]

$$\underset{(T-1)N\times k}{\mathbf{H}_N} = \begin{pmatrix} \underset{N\times k_2}{\mathbf{H}_{2,N}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \underset{N\times k_T}{\mathbf{H}_{T,N}} \end{pmatrix}. \tag{4.3.7}$$

Observe that the disturbances $\Delta\mathbf{u}_N$ can also be expressed as a linear form of the innovations $\boldsymbol{\eta}_N$,

$$\underset{(T-1)N\times 1}{\Delta\mathbf{u}_N} = \left[\underset{(T-1)\times(T+2)}{\left(\mathbf{0}_{(T-1)\times 2}, \underset{(T-1)\times T}{\mathbf{D}} \right)} \otimes \underset{N\times N}{\mathbf{P}_N} \right] \underset{(T+2)N\times 1}{\boldsymbol{\eta}_N}, \tag{4.3.8}$$

where $\mathbf{D}$ is the first difference operator matrix defined in e.g. (4.1.14). Furthermore, the t-th period disturbances can be expressed as

$$\underset{N\times 1}{\Delta\mathbf{u}_{t,N}} = \left(\underset{1\times(T+2)}{\mathbf{d}_t} \otimes \underset{N\times N}{\mathbf{P}_N} \right) \underset{(T+2)N\times 1}{\boldsymbol{\eta}_N}, \tag{4.3.9}$$

with $\mathbf{d}_t$ consisting of $(t-1)$-th row of $\left(\mathbf{0}_{(T-1)\times 2}, \mathbf{D}\right)$.

As a result, the moment conditions collected in $\mathbf{H}_N'\Delta\mathbf{u}_N$ are quadratic forms

[33] This definition of the instrument matrix is based on moment condititons that are only averaged over N and not over T. An alternative is to average over both N and T as in the initial IV estimator in Section 4.1.

in $\boldsymbol{\eta}_N$:

$$\underset{1\times N}{\mathbf{h}'_{rt,N}}\underset{N\times 1}{\Delta\mathbf{u}_{t,N}} = \underset{1\times N}{\mathbf{a}'_{rt,N}}\left(\underset{1\times(T+2)}{\mathbf{d}_t}\otimes\underset{N\times N}{\mathbf{P}_N}\right)\underset{(T+2)N\times 1}{\boldsymbol{\eta}_N} + \underset{1\times(T+2)N}{\boldsymbol{\eta}'_N}\left(\underset{(T+2)\times 1}{\mathbf{b}'_{rt}}\underset{1\times(T+2)}{\mathbf{d}_t}\otimes\underset{N\times N}{\mathbf{P}'_N\mathbf{P}_N}\right)\underset{(T+2)N\times 1}{\boldsymbol{\eta}_N}. \tag{4.3.10}$$

Below, I will apply the central limit theorem for triangular arrays of quadratic forms stated in Theorem A1 in Appendix A.[34] From Assumptions 1 and 2, it follows that the random variables $\boldsymbol{\eta}_N$ satisfy Assumptions A1 and A3. Observe that when the instruments are chosen to be lagged levels of the endogenous variable (i.e. $\mathbf{h}_{r,t,N} = \mathbf{y}_{t-s,N},\ s > 1$), Lemma 4 and Assumption 3 guarantee that $(\mathbf{d}'_t \otimes \mathbf{P}'_N)\,\mathbf{a}_{rt,N}$ and $(\mathbf{b}'_{rt}\mathbf{d}_t \otimes \mathbf{P}'_N\mathbf{P}_N)$ satisfy Assumption A2.

The condition

$$E\left(\underset{k\times(T-1)N}{\mathbf{H}'_N}\underset{(T-1)N\times 1}{\Delta\mathbf{u}_N}\right) = \mathbf{0}_{k\times 1}, \tag{4.3.11}$$

then implies that the matrix $(\mathbf{b}'_{rt}\mathbf{d}_t \otimes \mathbf{P}'_N\mathbf{P}_N)$ has zeros on the main diagonal and, therefore, the quadratic forms satisfy conditions of Lemma A1. In particular, their variances and covariances can be derived using the expressions in that Lemma. The following Lemma shows that under regularity conditions the quadratic forms $\mathbf{h}'_{rt,N}\Delta\mathbf{u}_{t,N}$ converge in distribution when normalized by their standard errors.

Lemma 5 *Consider a set of k instruments $\mathbf{H}_N$ given in (4.3.7), with the diagonal blocks $\mathbf{H}_{t,N} = (\mathbf{h}_{1t,N}, ..., \mathbf{h}_{k_t t,N})$ being $N \times k_t$ matrices ($k = k_2 + .. + k_T$) with*

[34] Observe that $(\mathbf{d}'_t \otimes \mathbf{P}'_N)\,\mathbf{a}_{rt,N}$ then corresponds to the sequence of vectors $\mathbf{b}_t$, while $(\mathbf{b}'_{rt}\mathbf{d}_t \otimes \mathbf{P}'_N\mathbf{P}_N)$ corresponds to the sequence of matrices $\mathbf{A}_n$, and $\boldsymbol{\eta}_N$ corresponds to the sequence of vectors of random variables $\boldsymbol{\varsigma}_n$ in Theorem A1 in Appendix A.

columns $\mathbf{h}_{rt,N} = \mathbf{a}_{rt,N} + (\mathbf{b}_{rt} \otimes \mathbf{P}_N)\,\boldsymbol{\eta}_N$,*where the sequence of nonstochastic* $N \times 1$ *vectors* $\mathbf{a}_{rt,N}$ *and the sequence of nonstochastic* $1 \times (T+2)$ *vectors* $\mathbf{b}_{rt}$ *have elements uniformly bounded in absolute value. Under Assumptions 1-6, and given that the instruments are such that* $E\left(\mathbf{H}'_N \Delta \mathbf{u}_N\right) = \mathbf{0}_{k\times 1}$, $E\left(\mathbf{H}'_N \Delta \mathbf{u}_N \Delta \mathbf{u}'_N \mathbf{H}_N\right) = \mathbf{V}_N$ *and*
$\left[(T-1)\,N\right]^{-1} \lambda_{\min}\left(\mathbf{V}_N\right) \geq c > 0$, *we have that*

$$\mathbf{V}_N^{-1/2} \mathbf{H}'_N \Delta \mathbf{u}_N \xrightarrow{d} N\left(\mathbf{0}, \mathbf{I}_k\right),$$

where $\mathbf{V}_N = E\left(\mathbf{H}'_N \Delta \mathbf{u}_N \Delta \mathbf{u}'_N \mathbf{H}_N\right) = \mathbf{V}_N^{1/2} \mathbf{V}_N^{1/2}$.

Proof. See the Appendix C.3. ■

Given that the moment conditions converge in distribution, the GMM estimator defined in (4.3.5) will under appropriate regularity conditions also converge in distribution:

Lemma 6 *Consider a set of stochastic instruments* $\mathbf{H}_N$ *such that*

$$\mathbf{V}_N^{-1/2} \mathbf{H}'_N \Delta \mathbf{u}_N \xrightarrow{d} N\left(\mathbf{0}, \mathbf{I}_k\right),$$

where $\mathbf{V}_N = E\left(\mathbf{H}'_N \Delta \mathbf{u}_N \Delta \mathbf{u}'_N \mathbf{H}_N\right) = \mathbf{V}_N^{1/2} \mathbf{V}_N^{1/2}$, *with*

$$p \lim_{N\to\infty} \left[(T-1)\,N\right]^{-1} \mathbf{V}_N = \mathbf{V},$$

where $\mathbf{V}$ *is finite. Furthermore, consider a sequence of weighting (possibly sto-*

chastic) matrices $\mathbf{A}_N$ with nonsingular (probability) limit

$$p\lim_{N\rightarrow\infty}\mathbf{A}_N=\mathbf{A}.$$

Under Assumptions 1-6 and given that

$$\mathbf{M}_{H\Delta Z}=p\lim_{N\rightarrow\infty}\left[\left(T-1\right)N\right]^{-1}\mathbf{H}_N'\Delta\mathbf{Z},$$

exists and has full column rank, we have that the GMM estimator defined in (4.3.5) converges in distribution and

$$\sqrt{\left(T-1\right)N}\left(\widetilde{\boldsymbol{\theta}}_N-\boldsymbol{\theta}\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Psi}\right),$$

where

$$\boldsymbol{\Psi}=\left(\mathbf{M}_{\Delta ZH}\mathbf{A}^{-1}\mathbf{M}_{\Delta ZH}'\right)^{-1}\mathbf{M}_{\Delta ZH}\mathbf{A}^{-1}\mathbf{V}\mathbf{A}^{-\mathbf{1}}\mathbf{M}_{\Delta ZH}\left(\mathbf{M}_{\Delta ZH}\mathbf{A}^{-1}\mathbf{M}_{\Delta ZH}'\right)^{-1}.$$

Proof. See the Appendix C.3. ■

I give a small sample approximation for $\boldsymbol{\Psi}$ for the specific GMM estimator considered below. Note that given Lemmas 4 and 5, the asymptotic result in the above Lemma 6 applies to a general class of GMM estimators which includes the initial IV estimator discussed in Section 4.1,[35] as well as the different

[35] The lemma is directly applicable when the moment conditions in the initial IV estimator are averaged only over the cross-sectional units. Note that in Section 4.1, the moment conditions are averaged over both cross-sectional units and time. I have provided the asymptotic results for this

variants of the GMM estimators in Arellano and Bond (1991) and, in particular, the feasible GMM estimator discussed below. Note that in applying the above Lemma to these estimators it remains to be checked whether in the particular application the instruments satisfy the stipulated regularity conditions, e.g. that $p\lim_{N\rightarrow\infty}\left[(T-1)\,N\right]^{-1}\mathbf{H}_N'\Delta\mathbf{Z}_N$ exists and has full column rank and that the variance covariance matrix of the moment conditions has the smallest eigenvalue uniformly bounded away from zero.

I now consider the issue of an optimal choice of the sequence of the weighting matrices, given a set of instruments. I close this section with proving consistency, asymptotic normality and providing a small sample guidance for a feasible second stage GMM estimator based on moment conditions considered in the literature.

4.3.1 Optimal Weighting Matrix

Consider now the optimal choice of the sequence of the weighting matrices $\mathbf{A}_N$. It can be shown[36] that given a set of instruments, the asymptotic variance covariance matrix of an estimator defined as a minimizer of (4.3.4) is minimized[37] when

$$p\lim_{N\rightarrow\infty}\left[(T-1)\,N\right]^{-1}\mathbf{A}_N=\mathbf{V}. \tag{4.3.12}$$

initial IV estimator in Theorem 1 above.

[36]See e.g. Hansen (1982), Bates and White (1993), Newey and McFadden (1994), or Wooldridge (2002), Ch. 8 and 14.

[37]In the sense that the difference with respect to any other VC matrix of an estimator that is a minimizer of (4.3.4) is positive semi-definite.

Given that $p\lim_{N\to\infty}\left[(T-1)\,N\right]^{-1}\mathbf{V}_N=\mathbf{V}$, the small sample weighting matrices $\mathbf{A}_N$ can be chosen to be estimators of the small sample variance covariance matrix $\mathbf{V}_N = E\left(\mathbf{H}_N'\Delta\mathbf{u}_N\Delta\mathbf{u}_N'\mathbf{H}_N\right)$. Observe that the matrix $\mathbf{V}_N$ can be partitioned as

$$\mathbf{V}_N=\begin{pmatrix}\mathbf{V}_{22,N} & & \mathbf{V}_{2T,N}\\ & \ddots & \\ \mathbf{V}_{T2,N} & & \mathbf{V}_{TT,N}\end{pmatrix}, \tag{4.3.13}$$

where $\mathbf{V}_{ts,N} = E\left(\mathbf{H}_{t,N}'\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}_{s,N}'\mathbf{H}_{s,N}\right)$. I denote the ij-th element of $\mathbf{V}_{ts,N}$ as $v_{ij,ts,N} = E\left(\mathbf{h}_{it,N}'\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}_{s,N}'\mathbf{h}_{js,N}\right)$. Given the structure of the instruments assumed in this section, the moment conditions are quadratic forms in $\boldsymbol{\eta}_N$ and satisfy conditions of Lemma A1 in Appendix A - see the discussion preceding Lemma 5. In particular, we have as in (4.3.10) above:

$$\mathbf{h}_{it,N}'\Delta\mathbf{u}_{t,N}=\mathbf{a}_{it,N}'\left(\mathbf{d}_t\otimes\mathbf{P}_N\right)\boldsymbol{\eta}_N+\boldsymbol{\eta}_N'\left(\mathbf{b}_{it}'\mathbf{d}_t\otimes\mathbf{P}_N'\mathbf{P}_N\right)\boldsymbol{\eta}_N, \tag{4.3.14}$$

and hence from Lemma A1 in Appendix A, the covariance of $\mathbf{h}_{it,N}'\Delta\mathbf{u}_{t,N}$ and $\mathbf{h}_{js,N}'\Delta\mathbf{u}_{s,N}$ denoted as $v_{ij,ts,N}$ is given by:

$$\begin{aligned} v_{ij,ts,N} &= \mathbf{a}_{it,N}'\left(\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}_s'\otimes\mathbf{P}_N\mathbf{P}_N'\right)\mathbf{a}_{js,N} \\ &\quad +2tr\left(\mathbf{b}_{it}'\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}_s'\mathbf{b}_{js}\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}_N'\mathbf{P}_N\mathbf{P}_N'\mathbf{P}_N\right), \end{aligned} \tag{4.3.15}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}$ is defined in (4.1.16).

Observe that for $|s-t|>1$, we have $\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}_s'=0$ and hence the above covariance is zero. An expectations based estimator, say $\widehat{\mathbf{V}}_N^E$, of $\mathbf{V}_N$ would then replace the true values of the parameters in the above expression by their initial consistent estimates. Note that in addition to $\mathbf{\Sigma}_{\boldsymbol{\eta},N}$ and $\mathbf{P}_N$, the terms $\mathbf{a}_{it,N}$ and $\mathbf{b}_{it}$ also potentially depend on the parameters of the model (compare e.g. the expressions for $\mathbf{a}_{t,N}$ and $\mathbf{b}_t$ in the proof of Lemma 4 in Appendix C.3). The exact form depends on the choice of the instruments. In Section 4.3.3 below, I consider a set of instruments utilized in the literature (e.g. Arellano and Bond, 1991) and I also provide an expression for such expectation based variance covariance matrix estimator given such choice of instruments. Note that the instruments considered in Section 4.1 are also of the form assumed here; see the proof of Lemma 1. The expression for $\mathbf{V}_N$ is then given by Lemma 2.

As an alternative to $\widehat{\mathbf{V}}_N^E$, the small sample weighting matrices can be constructed based on approximations to $\mathbf{H}_N' E\left(\Delta\mathbf{u}_N \Delta\mathbf{u}_N'\right)\mathbf{H}_N$. For stochastic instruments, such estimator will not in general be consistent estimator of $E\left(\mathbf{H}_N'\Delta\mathbf{u}_N\Delta\mathbf{u}_N'\mathbf{H}_N\right)$. Nevertheless, based on Lemma 6, the resultant second stage GMM estimator is consistent. It is also computationally simpler and has reasonable small sample properties (see Chapter 5).

This estimator denoted by $\widehat{\mathbf{V}}_N^{mix}$ ignores the fact that the instruments collected in $\mathbf{H}_N$ are stochastic and replaces the disturbances $\Delta\mathbf{u}_N\Delta\mathbf{u}_N'$ by an estimate of their expected value:

$$\widehat{\mathbf{V}}_N^{mix} = \left[\left(T-1\right)N\right]^{-1}\mathbf{H}_N'\widehat{\mathbf{\Omega}}_{\Delta u,N}\mathbf{H}_N, \qquad (4.3.16)$$

where $\widehat{\mathbf{\Omega}}_{\Delta u,N}$ is an estimator of the variance covariance matrix of the disturbances. In our case this could be:

$$\widehat{\mathbf{\Omega}}_{\Delta u,N} = \widehat{\sigma^2_{\varepsilon N}} \left(\mathbf{D} \otimes \widehat{\mathbf{P}}_N \right) \mathbf{Q}_{0,N} \left(\mathbf{D}' \otimes \widehat{\mathbf{P}}'_N \right), \tag{4.3.17}$$

where $\widehat{\rho}_N$ and $\widehat{\sigma^2_{\varepsilon N}}$ are initial estimates and

$$\widehat{\mathbf{P}}_N = \left(\mathbf{I}_N - \widehat{\rho}_N \mathbf{W}_N \right)^{-1}. \tag{4.3.18}$$

4.3.2 Feasible GMM Estimator

Consider now a GMM estimator based on the moment conditions of the form

$$E \left[\widetilde{\mathbf{H}}'_N \Delta \mathbf{u}_N \right] = 0, \tag{4.3.19}$$

where

$$\widetilde{\mathbf{H}}_N = \begin{pmatrix} \widetilde{\mathbf{H}}_{2,N} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \widetilde{\mathbf{H}}_{T,N} \end{pmatrix}_{N(T-1)\times k}, \tag{4.3.20}$$

with $\widetilde{\mathbf{H}}_{t,N} = (\mathbf{y}_{t-2,N}, ..., \mathbf{y}_{0,N}, \mathbf{X}_{t,N}, ..., \mathbf{X}_{1,N})$ being a $N \times k_t$ matrix of instruments at time t. Note that $k_t = (t-1) + t \cdot p$ and $k = k_2 + .. + k_T$. Let

$$\widetilde{\mathbf{V}}_N = E \left(\widetilde{\mathbf{H}}'_N \Delta \mathbf{u}_N \Delta \mathbf{u}'_N \widetilde{\mathbf{H}}_N \right), \tag{4.3.21}$$

then the estimator is given by

$$\widetilde{\boldsymbol{\theta}}_N = \left[\Delta\mathbf{Z}'_N\widetilde{\mathbf{H}}_N\widetilde{\mathbf{V}}_N^{-1}\widetilde{\mathbf{H}}'_N\Delta\mathbf{Z}_N\right]^{-1}\Delta\mathbf{Z}'_N\widetilde{\mathbf{H}}_N\widetilde{\mathbf{V}}_N^{-1}\widetilde{\mathbf{H}}'_N\Delta\mathbf{y}_N. \tag{4.3.22}$$

The instrument matrix in (4.3.20) utilizes moment conditions of the form

$$\begin{aligned} E\left[\left(u_{t,i}-u_{t-1,i}\right)y_{t-1-s,i}\right] &= 0 \qquad s=1,..,t-1, \\ E\left[\left(u_{t,i}-u_{t-1,i}\right)\mathbf{X}_{t-s,i}\right] &= \mathbf{0}_{1\times p}, \\ E\left[\left(u_{t,i}-u_{t-1,i}\right)\mathbf{X}_{t,i}\right] &= \mathbf{0}_{1\times p}. \end{aligned} \tag{4.3.23}$$

Observe that the instruments consist of $\mathbf{y}_{t-1-s,N}$, $\mathbf{X}_{t,N}$ and $\mathbf{X}_{t-s,N}$, $s=1,...,t-1$ and hence by Lemma 4 they are linear forms in the innovations of the form considered above, e.g. they satisfy the conditions in Lemma 5. To complete the verification of the conditions stipulated in Lemma 5, I consider the smallest eigenvalues of the sequence of matrices $\widetilde{\mathbf{V}}_N = E\left(\widetilde{\mathbf{H}}'_N\Delta\mathbf{u}_N\Delta\mathbf{u}'_N\widetilde{\mathbf{H}}_N\right)$.

Note that from Lemma 4 it follows that $\mathbf{y}_{t,N} = \mathbf{a}_{t,N} + \left(\mathbf{b}_t\otimes\mathbf{P}_N\right)\boldsymbol{\eta}_N$. Let us denote

$$\widetilde{\mathbf{S}}_{t,N} = \left[\mathbf{a}_{t-2,N},...,\mathbf{a}_{0,N},\mathbf{X}_{t,N},...,\mathbf{X}_{1,N}\right]_{N\times k_t}, \tag{4.3.24}$$

and

$$\boldsymbol{\Upsilon}_{t,N} = \left[\left(\underset{1\times(t-1)(T+2)}{(\mathbf{b}_{t-2},...,\mathbf{b}_0)}\otimes\underset{N\times N}{\mathbf{P}_N}\right)\left(\underset{(t-1)\times(t-1)}{\mathbf{I}_{t-1}}\otimes\underset{N(T+2)\times 1}{\boldsymbol{\eta}_N}\right),\mathbf{0}_{N\times tp}\right]. \tag{4.3.25}$$

The instruments can then be expressed as

$$\underset{N\times k_t}{\widetilde{\mathbf{H}}_{t,N}} = \underset{N\times k_t}{\widetilde{\mathbf{S}}_{t,N}} + \underset{N\times k_t}{\boldsymbol{\Upsilon}_{t,N}}, \tag{4.3.26}$$

As a result the full matrix of instruments is

$$\widetilde{\mathbf{H}}_N = \widetilde{\mathbf{S}}_N + \boldsymbol{\Upsilon}_N, \tag{4.3.27}$$

where the matrix $\widetilde{\mathbf{S}}_N$ contains the nonstochastic elements of the instruments and is defined as

$$\widetilde{\mathbf{S}}_N = \begin{pmatrix} \widetilde{\mathbf{S}}_{2,N} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \widetilde{\mathbf{S}}_{T,N} \end{pmatrix}_{N(T-1)\times k}, \tag{4.3.28}$$

while the stochastic components of the instrument matrix are

$$\boldsymbol{\Upsilon}_N = \begin{pmatrix} \boldsymbol{\Upsilon}_{2,N} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\Upsilon}_{T,N} \end{pmatrix}_{N(T-1)\times k}. \tag{4.3.29}$$

To guarantee that the smallest eigenvalue of $[(T-1)\,N]^{-1}\,\widetilde{\mathbf{V}}_N$ is uniformly bounded away from zero, I make the following assumption:

Assumption GMM1 *The smallest eigenvalue of* $[(T-1)\,N]^{-1}\,\widetilde{\mathbf{S}}'_N\widetilde{\mathbf{S}}_N$ *is uniformly bounded away from zero.*

Given the above Assumption, we have by Lemma 5 that the normalized moment conditions converge in distribution. I next show that the estimator $\widetilde{\boldsymbol{\theta}}_N$, where the weighting matrix for the moment conditions is based on the true value of the parameters is consistent and asymptotically normal. Corresponding to Assumptions IV1 and IV3 for the first stage estimator, I introduce the following assumptions. Let

$$\underset{(1+p)\times(1+p)}{\widetilde{\mathbf{M}}_{H\Delta Z}} = p\lim \frac{1}{(T-1)\,N} \widetilde{\mathbf{H}}_N' \Delta \mathbf{Z}_N. \tag{4.3.30}$$

Assumption GMM2 *The matrix $\widetilde{\mathbf{M}}_{H\Delta Z}$ exist and is finite with full column rank.*

Assumption GMM3 *The matrix $\widetilde{\mathbf{V}} = p\lim_{N\to\infty} \left[(T-1)\,N\right]^{-1} \widetilde{\mathbf{V}}_N$ exists and is finite and invertible.*

As a consequence of Lemma 6, we now have the following Theorem.

Theorem 3 *Under Assumptions 1-6, and GMM1-GMM3, we have that*

$$\sqrt{(T-1)\,N}\left(\widetilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right) \xrightarrow{d} N\left[\mathbf{0}, \boldsymbol{\Psi}\right],$$

where

$$\boldsymbol{\Psi} = \left[\widetilde{\mathbf{M}}_{H\Delta Z}' \widetilde{\mathbf{V}}^{-1} \widetilde{\mathbf{M}}_{H\Delta Z}\right]^{-1}.$$

Proof. See the appendix. ■

The above estimator is based on the true value of the parameters which are unknown and have to be estimated. I now provide an expression for the expectation based estimator of the variance covariance matrix of the moment conditions $\widetilde{\mathbf{V}}_N$, denoted by $\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)$ where $\widehat{\boldsymbol{\delta}}_N$ is an initial consistent estimator of $\boldsymbol{\delta} = \left(\rho_N, \sigma^2_{\varepsilon N}, \sigma^2_{\mu}, \phi\right)$. I then show that when the feasible GMM estimator uses $\left[\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)\right]^{-1}$ as the moment weighting matrix, the parameters collected in the vector $\boldsymbol{\delta}$ are nuisance parameters.

The variance covariance matrix of the moment conditions collected in $\widetilde{\mathbf{H}}'_N \Delta \mathbf{u}_N$ with $\widetilde{\mathbf{H}}$ defined in (4.3.20), can be written analogically to (4.3.13) as

$$\widetilde{\mathbf{V}}_N = \begin{pmatrix} \widetilde{\mathbf{V}}_{22,N} & & \widetilde{\mathbf{V}}_{2T,N} \\ & \ddots & \\ \widetilde{\mathbf{V}}_{T2,N} & & \widetilde{\mathbf{V}}_{TT,N} \end{pmatrix}, \tag{4.3.31}$$

where $\widetilde{\mathbf{V}}_{ts,N} = E\left(\widetilde{\mathbf{H}}'_{t,N} \Delta \mathbf{u}_{t,N} \Delta \mathbf{u}'_{s,N} \widetilde{\mathbf{H}}_{s,N}\right)$. Since $\widetilde{\mathbf{H}}_{t,N}$ consists of stochastic part $\left(\mathbf{y}_{t-2,N}, .., \mathbf{y}_{0,N}\right)$ and nonstochastic part $\left(\mathbf{X}_{t,N}, ..., \mathbf{X}_{1,N}\right)$, the matrix $\widetilde{\mathbf{V}}_{ts,N}$ is partitioned accordingly:[38]

$$\widetilde{\mathbf{V}}_{ts,N} = \begin{pmatrix} \widetilde{\mathbf{V}}^{\mathbf{y}}_{ts,N} & \mathbf{0}_{(t-1)\times tp} \\ \mathbf{0}_{sp\times(s-1)} & \widetilde{\mathbf{V}}^{\mathbf{X}}_{ts,N} \end{pmatrix}_{k_t \times k_s}, \tag{4.3.32}$$

[38] I show that the off-diagonal blocks of $\widetilde{\mathbf{V}}_{ts,N}$ are matrices of zeros as a part of the proof of the Lemma 7 below.

where the upper block is

$$\widetilde{\mathbf{V}}^{\mathbf{y}}_{ts,N}=\left(\widetilde{v}^{\mathbf{y}}_{qr,ts,N}\right)^{q=1,..,t-1}_{r=1,..,s-1}, \tag{4.3.33}$$

with $\widetilde{v}^{\mathbf{y}}_{qr,ts,N}=E\left(\mathbf{y}'_{t-1-q}\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}'_{s,N}\mathbf{y}_{s-1-r}\right)$. Given expressions for $\mathbf{y}_{t-1-q}$ and $\mathbf{y}_{s-1-r}$ in Lemma 4 and the expressions for $\Delta\mathbf{u}_{t,N}$ and $\Delta\mathbf{u}_{s,N}$ in (4.3.9), the moment conditions $\mathbf{y}'_{t-1-q}\Delta\mathbf{u}_{t,N}$ and $\mathbf{y}'_{s-1-r}\Delta\mathbf{u}_{s,N}$ are quadratic forms in $\boldsymbol{\eta}_N$ and their covariance is (using Lemma A1 in Appendix A) given by

$$\begin{aligned}\widetilde{v}^{\mathbf{y}}_{qr,ts,N} &= E\left(\mathbf{y}_{t-1-q}\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}'_{s,N}\mathbf{y}_{s-1-r}\right) \\ &= \mathbf{a}'_{t-1-q,N}\left(\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}'_s\otimes\mathbf{P}_N\mathbf{P}'_N\right)\mathbf{a}_{s-1-r,N} \\ &\quad +2tr\left(\mathbf{b}'_{t-1-q,N}\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}'_s\mathbf{b}_{s-1-r,N}\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}'_N\mathbf{P}_N\mathbf{P}'_N\mathbf{P}_N\right).\end{aligned} \tag{4.3.34}$$

Note that by (4.3.9), the disturbances $\Delta\mathbf{u}_{t,N}$ are linear forms in the innovations $\boldsymbol{\eta}_N$. From Lemma A1 in Appendix it then follows that the variance covariance matrix of $\Delta\mathbf{u}_{t,N}$ and $\Delta\mathbf{u}_{s,N}$ is $\left(\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}'_s\otimes\mathbf{P}_N\mathbf{P}'_N\right)$. Hence the second block of $\widehat{\mathbf{V}}_{ts,N}$ is:

$$\begin{aligned}\widetilde{\mathbf{V}}^{\mathbf{X}}_{ts,N} &= \left(\mathbf{X}_{t,N},...,\mathbf{X}_{1,N}\right)'E\left(\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}'_{s,N}\right)\left(\mathbf{X}_{s,N},...,\mathbf{X}_{1,N}\right) \\ &= \left(\mathbf{X}_{t,N},...,\mathbf{X}_{1,N}\right)'\left(\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}'_s\otimes\mathbf{P}_N\mathbf{P}'_N\right)\left(\mathbf{X}_{s,N},...,\mathbf{X}_{1,N}\right).\end{aligned} \tag{4.3.35}$$

The estimator $\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)$ replaces the true values in the expressions (4.3.31)-(4.3.35) by their initial estimates collected in the vector $\widehat{\boldsymbol{\delta}}_N=\left(\widehat{\rho}_N,\widehat{\sigma^2_{\varepsilon N}},\widehat{\sigma^2_{\mu}},\widehat{\phi}\right)'$.

In particular, it replaces $\mathbf{\Sigma}_{\boldsymbol{\eta},N}$, $\mathbf{P}_N$, $\mathbf{a}_{t,N}$, and $\mathbf{b}_{t,N}$ with

$$\begin{aligned}
\widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N} &= diag\left(\widehat{\sigma}^2_{\mu,N}, \frac{\widehat{\sigma}^2_{\varepsilon,N}}{1-\widehat{\phi}}, \widehat{\sigma}^2_{\varepsilon,N}, ..., \widehat{\sigma}^2_{\varepsilon,N}\right), \\
\widehat{\mathbf{P}}_N &= \left(\mathbf{I}_N - \widehat{\rho}_N \mathbf{W}_N\right)^{-1}, \\
\widehat{\mathbf{a}}_{t,N} &= \sum_{j=0}^{t-1} \widehat{\phi}^j_N \mathbf{X}_{t-j,N} \widehat{\boldsymbol{\beta}}_N, \\
\widehat{\mathbf{b}}_{t,N} &= \left(\frac{1}{1-\widehat{\phi}_N}, 1, \widehat{\phi}^{t-1}_N, .., \widehat{\phi}^0_N, \mathbf{0}_{1\times(T-t)}\right).
\end{aligned} \tag{4.3.36}$$

Note that in order to for the estimator of the variance covariance matrix of the moment conditions to be feasible, this implicitly assumes that the past values of the exogenous variables are so that $\sum_{j=0}^{\infty} \phi^j \mathbf{X}_{-j,N} \boldsymbol{\beta} = 0$, i.e. there are no individual effects other than those contained in μ_i.[39] The following Lemma shows that the estimator $\widehat{\mathbf{V}}_N$ is consistent.

Lemma 7 *Under Assumptions 1-6, and GMM1-GMM3, and given that* $\widehat{\boldsymbol{\delta}}_N \xrightarrow{p} \boldsymbol{\delta}$ *as* $N \to \infty$*, the row and column sums of the matrix* $r\mathbf{W}_N$ *are uniformly bounded in absolute value for some* r *with* $|\rho| < r < 1$*, and that* $\sum_{j=0}^{\infty} \phi^j \mathbf{X}_{-j,N} \boldsymbol{\beta} = 0$*, we have that* $\frac{1}{(T-1)N} \widehat{\mathbf{V}}_N \left(\widehat{\boldsymbol{\delta}}_N\right) \xrightarrow{p} \widetilde{\mathbf{V}}$.

Proof. See the Appendix C.3. ■

Consider now the feasible GMM estimator that uses $\left[\widehat{\mathbf{V}}_N \left(\widehat{\boldsymbol{\delta}}_N\right)\right]^{-1}$ as the

[39]This will not be satisfied when the model contains a deterministic constant terms. In this case, it is necessary to assume that the past values of the exogenous variables are observable and replace the expression for $\widehat{\mathbf{a}}_{t,N}$ with $\widehat{\mathbf{a}}_{t,N} = \sum_{j=0}^{\infty} \widehat{\phi}^j_N \mathbf{X}_{t-j,N} \widehat{\boldsymbol{\beta}}_N$.

moment weighting matrix and is defined as

$$\overset{\vee}{\boldsymbol{\theta}}_N\left(\widehat{\boldsymbol{\delta}}_N\right) = \left[\Delta\mathbf{Z}'_N\widetilde{\mathbf{H}}_N\left(\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)\right)^{-1}\widetilde{\mathbf{H}}'_N\Delta\mathbf{Z}_N\right]^{-1} * \\ \Delta\mathbf{Z}'_N\widetilde{\mathbf{H}}_N\left(\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)\right)^{-1}\widetilde{\mathbf{H}}'_N\Delta\mathbf{y}_N. \tag{4.3.37}$$

The following Theorem establishes that the parameters collected in the vector $\boldsymbol{\delta}$ are nuisance parameters.

Theorem 4 *Under Assumptions 1-6, and GMM1-GMM3, and given that* $\widehat{\boldsymbol{\delta}}_N \xrightarrow{p} \boldsymbol{\delta}$ *as* $N \to \infty$*, the row and column sums of the matrix* $r\mathbf{W}_N$ *are uniformly bounded in absolute value for some* r *with* $|\rho| < r < 1$*, and that* $\sum_{j=0}^{\infty} \phi^j \mathbf{X}_{-j,N}\boldsymbol{\beta} = 0$*, we have that*

$$\sqrt{N(T-1)}\left[\overset{\vee}{\boldsymbol{\theta}}_N\left(\widehat{\boldsymbol{\delta}}_N\right) - \widetilde{\boldsymbol{\theta}}_N\right] \xrightarrow{p} \mathbf{0},$$

and hence

$$\sqrt{N(T-1)}\left[\overset{\vee}{\boldsymbol{\theta}}_N\left(\widehat{\boldsymbol{\delta}}_N\right) - \boldsymbol{\theta}\right] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}).$$

Proof. See the Appendix C.3. ■

The small sample approximation to the variance covariance matrix $\boldsymbol{\Psi}$ can be based on the following Lemma.

Lemma 8 *Under Assumptions 1-6, and GMM1-GMM3, and given that* $\widehat{\boldsymbol{\delta}}_N \xrightarrow{p} \boldsymbol{\delta}$ *as* $N \to \infty$*, the row and column sums of the matrix* $r\mathbf{W}_N$ *are uniformly bounded for some* r *with* $|\rho| < r < 1$*, and that* $\sum_{j=0}^{\infty} \phi^j \mathbf{X}_{-j,N}\boldsymbol{\beta} = 0$*, we have that*

$$\widehat{\boldsymbol{\Psi}}_N\left(\widehat{\boldsymbol{\delta}}_N\right) \xrightarrow{p} \boldsymbol{\Psi},$$

as $N \to \infty$*, where*

$$\widehat{\mathbf{\Psi}}_N\left(\widehat{\boldsymbol{\delta}}_N\right) = \frac{1}{(T-1)\,N}\left[\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N\widehat{\mathbf{V}}_N^{-1}\widetilde{\mathbf{H}}_N'\Delta\mathbf{Z}_N\right]^{-1}.$$

Proof. See the Appendix C.3. ■

5 Monte Carlo Study

I consider the same dynamic panel data model as specified in Chapter 3. Here I will first define the estimators that I examine in the Monte Carlo study. I then describe how I generated the artificial data samples, briefly describe the range of the parameters I considered and finally present the results of the experiments.

5.1 Estimators Considered

I consider the following estimators in my simulations. The first group of estimators, labeled 'Initial Estimators', ignores the spatial autocorrelation of the disturbances and estimates only the slope coefficients of the model (i.e. $\boldsymbol{\beta}$ and ϕ). The second group of estimators uses some initial estimator of the slope coefficients (and the projected disturbances it implies) and provides an estimate of the spatial autocorrelation parameter (ρ) and the variances of the innovations and the individual effects (σ_{ε}^2 and σ_{μ}^2). Finally, the third group, labeled as 'Second Stage GMM Estimators', are estimators that use different weighting schemes to weight the same moment conditions as the initial estimators. The weights are based on initial estimators of ρ, σ_{ε}^2 and σ_{μ}^2. For comparison, I also include results a for two stage GMM estimator that ignores spatial correlation. The rest of this section will introduce the different estimators. For clarity of the exposition, I will drop the sample size subscript in this section.

5.1.1 Initial Estimators

I consider the instrumental variable (IV) estimators suggested by Anderson and Hsiao (1981) as well as IV estimators that use a larger instrument set, corresponding to the initial estimators suggested by Arellano and Bond (1991) and others. All these estimators can be written as IV estimators but with a different instrument matrix. In particular, they are of the form

$$\widehat{\boldsymbol{\theta}}_{IV} = \left(\widehat{\phi}_{IV}, \widehat{\boldsymbol{\beta}}'_{IV}\right)' = \left[\Delta\mathbf{Z}'\mathbf{H}\left(\mathbf{H}\mathbf{H}'\right)^{-1}\mathbf{H}'\Delta\mathbf{Z}\right]^{-1}\Delta\mathbf{Z}'\mathbf{H}\left(\mathbf{H}\mathbf{H}'\right)^{\mathbf{-1}}\mathbf{H}'\Delta\mathbf{y}, \tag{5.1.1}$$

where the right-hand side variables of the first differenced model (3.1.1 or 4.1.1) are stacked in a matrix $\Delta\mathbf{Z}$ as in, e.g. (4.1.3), the dependent variable is $\Delta\mathbf{y}$ as in (4.1.2) and the matrix $\mathbf{H}$ collects the instruments used. The instrument matrix is block diagonal with each block containing the set of instruments for the relevant time period:

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{H}_T \end{pmatrix}.$$

Different choices of $\mathbf{H}$ will lead to different initial estimators. In particular, the following estimators are considered in the experiments: the IV estimators suggested by Anderson and Hsiao (1981), the initial IV estimators suggested by Arellano and Bond (1991), as well as the IV estimator with the instrument set discussed in this thesis in Chapter 4, Section 4.3.2.

The two Anderson and Hsiao (AH) estimators use respectively lagged first difference of the endogenous variable ($\mathbf{y}_{t-2} - \mathbf{y}_{t-3}$) and level of the endogenous variable lagged twice ($\mathbf{y}_{t-2}$) as instruments for the lagged difference of the endogenous variable ($\mathbf{y}_{t-1} - \mathbf{y}_{t-2}$). The instrument matrices hence have the following form:

$$\mathbf{H}_{AH1} = \begin{bmatrix} (\mathbf{X}_2 - \mathbf{X}_1) & & & \\ & (\mathbf{y}_1 - \mathbf{y}_0, \mathbf{X}_3 - \mathbf{X}_2) & & \\ & & \ddots & \\ & & & (\mathbf{y}_{T-2} - \mathbf{y}_{T-3}, \mathbf{X}_T - \mathbf{X}_{T-1}) \end{bmatrix}, \tag{5.1.2}$$

and

$$\mathbf{H}_{AH2} = \begin{bmatrix} (\mathbf{y}_0, \mathbf{X}_2 - \mathbf{X}_1) & & & \\ & (\mathbf{y}_1, \mathbf{X}_3 - \mathbf{X}_2) & & \\ & & \ddots & \\ & & & (\mathbf{y}_{T-2}, \mathbf{X}_T - \mathbf{X}_{T-1}) \end{bmatrix}. \tag{5.1.3}$$

In addition to the moment condition ($i = 1, ..., N$)

$$E\left[(u_{it} - u_{i,t-1})(\mathbf{x}_{it} - \mathbf{x}_{i,t-1})\right] = \mathbf{0}_{p\times 1} \qquad t = 1, ..., T, \tag{5.1.4}$$

the AH estimators each utilize at each time period one additional moment condition:

$$E\left[(u_{it} - u_{i,t-1})(y_{i,t-2} - y_{i,t-3})\right] = 0 \qquad t = 2, ..., T \tag{5.1.5}$$

and

$$E\left[\left(u_{it}-u_{i,t-1}\right)y_{i,t-1}\right]=0 \qquad t=1,...,T \tag{5.1.6}$$

respectively. However, as pointed out by Arellano and Bond (1991), there are additional moment conditions, not utilized by the AH estimators. In particular, for the observation at a time t, we have the following additional moments:

$$E\left[\left(u_{it}-u_{i,t-1}\right)y_{i,t-1-k}\right]=0 \qquad k=1,..t-1.$$

Similarly, there are additional moment conditions involving lags of the exogenous variables in addition to the condition utilized by the AH estimators. Therefore, based on Arellano and Bond (AB), I consider an instrument matrix discussed in Section 4.3:

$$\mathbf{H}_{AB}=\begin{bmatrix} (\mathbf{y}_0,\mathbf{X}_1,\mathbf{X}_2) & & & \\ & (\mathbf{y}_0,\mathbf{y}_1,\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3) & & \\ & & \ddots & \\ & & & (\mathbf{y}_0,..,\mathbf{y}_{T-2},\mathbf{X}_1,..,\mathbf{X}_T) \end{bmatrix}. \tag{5.1.7}$$

The table below summarizes the initial estimators, their instrument matrices and the moment conditions that the instruments are based on.

Table 2. Estimators Considered

Estimator (Instrument Matrix)	Moment Conditions $i = 1, .., N$ and $t = 1, .., T$
AH difference ($\mathbf{H}_{AH1}$)	$E\left[(u_{it} - u_{i,t-1})(y_{i,t-2} - y_{i,t-3})\right] = 0,$ $t = 1$ not considered $E\left[(u_{it} - u_{i,t-1})(\mathbf{x}_{it} - \mathbf{x}_{i,t-1})\right] = \mathbf{0}_{p\times 1}$
AH level ($\mathbf{H}_{AH2}$)	$E\left[(u_{it} - u_{i,t-1})\, y_{i,t-2}\right] = 0$ $E\left[(u_{it} - u_{i,t-1})(\mathbf{x}_{it} - \mathbf{x}_{i,t-1})\right] = \mathbf{0}_{p\times 1}$
AB ($\mathbf{H}_{AB}$)	$E\left[(u_{it} - u_{i,t-1})\, y_{i,t-2-k}\right] = 0,$ $k = 0, ..., t-1$ $E\left[(u_{it} - u_{i,t-1})\, \mathbf{x}_{is}\right] = \mathbf{0}_{p\times 1},$ $s = 1, .., t$

5.1.2 Spatial Parameter Estimators

I consider the spatial generalized moments (GM) estimators of the spatial autoregressive parameter ρ suggested by Kapoor et al. (2005) and discussed in Chapter 4. The spatial GM estimator was defined in (4.2.14).

The estimators differ along two dimensions. First, they differ with respect to how the estimated disturbances were calculated. I consider the three initial estimators from the previous section as well as the true value of the disturbances. Secondly, the estimators differ with respect to how the moments are weighted. The first estimator is referred to as 'Unweighted Spatial GM Estimator' and weights the moment conditions equally, e.g. by setting $\mathbf{A}_N = \mathbf{I}_6$ in (4.2.14). The second estimator I consider is based on the full set of weighted moments and utilizes the weighting matrix $\mathbf{A}_N = \widehat{\mathbf{\Xi}}_N$ defined in (4.2.17). I refer to this estimator as the 'Weighted Spatial GM Estimator'.

Altogether, there are four different possibilities to calculate the estimated disturbances (three initial estimators and the true values) and two types of GM estimator (unweighted and weighted moment conditions), i.e. altogether eight possible combinations.

5.1.3 Second Stage GMM Estimators

The second stage GMM estimators utilize the same moment conditions as the initial AB estimator but with a weighting matrix. The estimators are of the form:

$$\widehat{\boldsymbol{\theta}}_{GMM} = \left(\widehat{\phi}_{GMM}, \widehat{\boldsymbol{\beta}}'_{GMM}\right)' = \left[\Delta\mathbf{Z}'\mathbf{H}\mathbf{A}_k^{-1}\mathbf{H}'\Delta\mathbf{Z}\right]^{-1}\Delta\mathbf{Z}'\mathbf{H}\mathbf{A}_k^{-1}\mathbf{H}'\Delta\mathbf{y}, \tag{5.1.8}$$

where the weighting matrix $\mathbf{A}_k$, $k = 1, 2, 3$ is calculated in three different ways. The first case is a weighting matrix that ignores the spatial autocorrelation of the disturbances but uses an estimators for σ_ε^2 and σ_1^2 that are consistent even for nonzero values of ρ. In particular, the first weighting scheme uses:

$$\mathbf{A}_1 = \mathbf{H}'\widetilde{\mathbf{\Omega}}\mathbf{H}, \tag{5.1.9}$$

with

$$\widetilde{\mathbf{\Omega}} = (\mathbf{D} \otimes \mathbf{I}_N) \left(\widetilde{\sigma}_\varepsilon^2 \mathbf{Q}_0 + \widetilde{\sigma}_1^2 \mathbf{Q}_1 \right) (\mathbf{D}' \otimes \mathbf{I}_N) \, , \tag{5.1.10}$$

where the estimators $\widetilde{\sigma}_\varepsilon^2$, and $\widetilde{\sigma}_1^2$ are the spatial GM estimators (with weighted moment conditions) described above and, based on an initial IV estimator with $\mathbf{H}_{AB}$ as the instrument matrix.

The second weighting scheme uses $\widehat{\mathbf{V}}^{mix}$ as an estimate of the variance covariance matrix of the moment conditions (see Section 4.3), i.e. it employs $\mathbf{A}_2 = \widehat{\mathbf{V}}^{mix}$, where

$$\widehat{\mathbf{V}}^{mix} = \mathbf{H}'\widehat{\mathbf{\Omega}}\mathbf{H}, \tag{5.1.11}$$

with

$$\widehat{\mathbf{\Omega}} = \left(\mathbf{D} \otimes \widetilde{\mathbf{P}}\right) \left(\widetilde{\sigma}_\varepsilon^2 \mathbf{Q}_0 + \widetilde{\sigma}_1^2 \mathbf{Q}_1 \right) \left(\mathbf{D}' \otimes \widetilde{\mathbf{P}}'\right) , \tag{5.1.12}$$

where

$$\widetilde{\mathbf{P}} = (\mathbf{I}_N - \widetilde{\rho}\mathbf{W})^{-1} \, , \tag{5.1.13}$$

$\mathbf{D}$ is the $(T-1) \times (T-1)$ first difference transformation matrix: $\mathbf{D}$:

$$\mathbf{D} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix} , \tag{5.1.14}$$

and estimators $\widetilde{\sigma}_\varepsilon^2$, $\widetilde{\sigma}_1^2$ and $\widetilde{\rho}$ are the spatial GM estimators (with weighted moment conditions) based on the initial IV estimator with $\mathbf{H}_{AB}$ as the instrument matrix.

Finally, the last weighting scheme uses a consistent estimate of the variance covariance matrix of the moment conditions $\widehat{\mathbf{V}}^E$, i.e. it employs $\mathbf{A}_3 = \widehat{\mathbf{V}}^E$, where $\widehat{\mathbf{V}}^E$ is defined in Section 4.3.2 (equations 4.3.31 - 4.3.35) and is based on the same initial estimators $\widetilde{\sigma}_\varepsilon^2$, $\widetilde{\sigma}_1^2$ and $\widetilde{\rho}$, as well as on the initial IV estimators $\widehat{\phi}$ and $\widehat{\boldsymbol{\beta}}$.

5.2 Data Generation

I first generate the exogenous variables so that these are common across the different replications. The exogenous variables consist of a nonstochastic intercept (equal to unity) and a second stochastic exogenous variable. I generate the second exogenous variable as independent (of all the other random variables in the model) jointly normally distributed random variables, i.e.

$$\left[(vec\mathbf{X}_1)', ..., (vec\mathbf{X}_T)'\right]' \sim N\left(\mathbf{0}, \mathbf{I}_{NT}\right). \tag{5.2.1}$$

The exogenous variables are generated once and are used in all replications of the model.

In each replication, I then draw $(T+2)N$ independent jointly normally distributed random numbers that are used to construct draws of the vector $\boldsymbol{\eta}$. The first N draws are scaled by $\sigma_\mu = \sqrt{\eta_\mu \sigma_\varepsilon^2}$ and are used for creating the $N \times 1$ vector of individual effects $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)'$.[40] The next N draws are scaled by

[40]We find that altering the ratio of the variance of the individual effects and model disturbances does not qualitativelly affect our results and hence we only consider $\eta_\mu = 1$.

$\sqrt{\frac{\sigma_\varepsilon^2}{(1-\phi)^2}}$ and are used for creating the $N \times 1$ vector of the initial disturbances

$$\boldsymbol{\xi} = \sum_{j=0}^{\infty} \phi^j \boldsymbol{\varepsilon}_{-j}. \tag{5.2.2}$$

Finally, the last TN draws are scaled by σ_ε and are used for creating the $NT \times 1$ vector of disturbances $(\boldsymbol{\varepsilon}_1', ..., \boldsymbol{\varepsilon}_T')'$

I construct the $N \times 1$ vector of initial observations as[41]

$$\mathbf{y}_0 = (\mathbf{I}_N - \rho \mathbf{W})^{-1} \left[\boldsymbol{\xi} + (1-\phi)^{-1} \boldsymbol{\mu}\right]. \tag{5.2.3}$$

The subsequent observations for $t = 1, .., T$ are then generated according to the our model as

$$\mathbf{y}_t = \phi \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + (\mathbf{I}_N - \rho \mathbf{W})^{-1} (\boldsymbol{\varepsilon}_t + \boldsymbol{\mu}). \tag{5.2.4}$$

5.3 Designs Considered

In all experiments I set $N = 100$ and $T = 5$. I consider three specifications for the spatial weighting matrix, as in Kelejian and Prucha (1999) and Kapoor et al. (2005). The matrices differ in the degree of their sparseness. The first matrix has in its i-th row, $1 < i < N$, nonzero elements in positions $i-1$ and $i+1$, so that the i-th unit is directly related to its immediate neighbors. I define this matrix in a circular world so that the nonzero elements in the first and last rows are at

[41] This specification implicitly assumes that the contribution of initial values of the exogenous variables is zero. i.e. that $\sum_{j=0}^{\infty} \mathbf{X}_{-j} \beta = 0$.

positions $(1, 2)$, $(1, N)$, $(N, 1)$ and $(N, N-1)$. This matrix is row normalized and hence all the nonzero elements are equal to 1/2. As in Kelejian and Prucha (1999), I refer to this matrix as ”1 ahead and 1 behind”. The next weighting matrices are defined in a corresponding way as ”3 ahead and 3 behind” and ”5 ahead and 5 behind” with nonzero entries 1/6 and 1/10 respectively. In the tables of results below I reference the matrices by $W = 1, 2, 3$.

The exogenous variables were generated once prior to the Monte Carlo experiments and the process is described above in the Data Generation section. For simplicity I always set $\boldsymbol{\beta} = (1, 1)'$. The rest of the coefficients of the model take on the following values:

$$\phi \in \{-0.9, -0.75, -0.25, 0, 0.25, 0.75, 0.9\}, \tag{5.3.1}$$

and

$$\rho \in \{-0.9, -0.5, -0.25, 0, 0.25, 0.5, 0.9\}. \tag{5.3.2}$$

I find that the results do not qualitatively change with the ratio of the variances of the independent innovations and the individual effects and hence I always set $\eta_\mu = \frac{\sigma_\mu^2}{\sigma_\varepsilon^2} = 1$. The variance of the independent innovations is always set to one. As a result the different specifications will have different overall average R^2 of the data. The variance of the dependent variable conditional on the explanatory variables (equal to variance of the disturbances) is given by

$$VC\left(\mathbf{y}_t | \mathbf{X}_t, \mathbf{y}_{t-1}\right) = \left(\sigma_\varepsilon^2 + \sigma_\mu^2\right) \cdot \mathbf{PP}', \tag{5.3.3}$$

where as before I define $\mathbf{P} = (\mathbf{I}_N - \rho\mathbf{W})^{-1}$. Furthermore, the unconditional variance of the dependent variable is[42]

$$VC(\mathbf{y}_t) = \left(\frac{\sigma_\varepsilon^2}{1-\phi^2} + \frac{\sigma_\mu^2}{(1-\phi)^2}\right) \cdot \mathbf{PP}'. \tag{5.3.4}$$

The expected R^2 of the data is then equal to the ratio of the conditional and unconditional variance of the dependent variable and hence is a function of the true values of the parameters ϕ, $\mathbf{W}$ and ρ (as well as σ_ε^2 and σ_μ^2).

In particular, consider the vector of observation of the dependent variable as $\mathbf{y} = (\mathbf{y}_0', ..., \mathbf{y}_T')'$ and its mean vector denoted as

$$\overline{\mathbf{y}} = E\left[(\mathbf{y}_1', ..., \mathbf{y}_T')'\right] = \left[\boldsymbol{\beta}'\mathbf{X}_1', ..., \sum_{t=1}^{T} \phi^{T-t}\boldsymbol{\beta}'\mathbf{X}_t'\right]', \tag{5.3.5}$$

The sample correlation coefficient between $\mathbf{y}$ and $\overline{\mathbf{y}}$ is then defined as

$$r = \frac{(\mathbf{y} - \mathbf{e}_{TN} \cdot \mathbf{y}'\mathbf{y})'(\overline{\mathbf{y}} - \mathbf{e}_{TN} \cdot \overline{\mathbf{y}}'\overline{\mathbf{y}})}{\sqrt{(\mathbf{y} - \mathbf{e}_{TN} \cdot \mathbf{y}'\mathbf{y})'(\mathbf{y} - \mathbf{e}_{TN} \cdot \mathbf{y}'\mathbf{y})(\overline{\mathbf{y}} - \mathbf{e}_{TN} \cdot \overline{\mathbf{y}}'\overline{\mathbf{y}})'(\overline{\mathbf{y}} - \mathbf{e}_{TN} \cdot \overline{\mathbf{y}}'\overline{\mathbf{y}})}}, \tag{5.3.6}$$

where $\mathbf{e}_{TN}$ is a $TN \times 1$ vector of unit elements. The designs considered in the Monte Carlo experiments are such that average (over the replications of a particular design) r is between 0.54 and 0.78.

To summarize, we have 7 values for ϕ, 7 values for ρ and 3 different weighting matrices $\mathbf{W}$, that is 147 different parameter designs.

[42]Note that the expression is derived analogously to the variance-covariance matrix of the initial observations, given in equation (3.2.5)

5.4 Tables of Results

The tables of results D1-D4 contain bias and a measure of the root mean square errors of the different estimators for the 147 designs considered. For each constellation of parameters, the random numbers were generated 1000 times and the estimators calculated and their values saved. For each estimator, I report the median and root mean square error calculated as in Kapoor (2005); that is using the interquantile based measure:

$$RMSE = \left[bias^2 + \left(\frac{IQ}{1.35} \right)^2 \right]^{1/2}, \tag{5.4.1}$$

where bias is the true value of the parameter minus the median of the estimators, and IQ is the difference between .75 and .25 quantiles.

Observe that the comparison of the different estimation procedures in Tables D1-D4 is only based on comparing the .25, .50 and .75 quantiles of their distributions. Note that hypothesis tests are often based on the .05 and the .95 quantiles and hence it might be of interest to consider quantiles other than those used in constructing the bias and RMSE measure.

To make such comparison feasible, I present in Figures 1 through 6 the quantile-to-quantile plots that compare the small sample distribution of the estimated slope coefficient ϕ with the Gaussian normal distribution. The plots depict the sample cumulative distribution of the estimator (over the 1,000 replications of each design). The left hand side axis of the plots has a nonlinear scale so that if the data was exactly normally distributed, the plot would be linear. Therefore, any nonlinearity in the plot represents deviations from normality at the appropriate quantiles.

I superimpose the 147 design on top of each other[43] in each Figure, so that the deviations from the straight line represent the worst-case scenarios over the entire parameter space. For illustration purposes, Figure 7 shows the quantile-to-quantile plot of 1,000 replications of $N(0,1)$ distribution, and Figure 8 show the same plot where the sample was drawn from a student-t distribution with 5 degrees of freedom. Observe that the quantile-to-quantile plot allows an easy detection of even such small deviations from normality.

5.5 Conclusions and Comparison with Other Studies

The results of the experiments confirm the finding in the literature that for some parameter values the performance of the Anderson-Hsiao estimator AH1 is not very satisfactory (see Table D1). However, the second initial estimator AH2 (using the twice lagged level of the endogenous variable as an instrument) performs quite well and in fact for most parameter values it is better (in terms of lower bias and/or lower RMSE) than the estimator AB that uses a larger instrument set. Note that if the model did not contain individual effects, the instruments used by the estimator AH2 would be the conditional expectations of the right-hand side variables. This might explain its relativelly good performance. Note that the AH1 and AH2 estimators are exactly identified and hence their performance cannot be

[43] To maintain compatibility over different designs, the small sample distributions were normalized by their medians and the difference between the .25 and .75 quantiles.

improved by weighting the moment conditions.

In contrast, once the moment conditions are weighted, the performance of the AB estimator improves (see Table D2) and becomes better than that of the AH2 estimator. Observe that ignoring spatial autocorrelation in constructing the weights involves a moderate loss of efficiency relative to the other weighting schemes especially when the spatial autocorrelation is high and positive. On the other hand for low or negative values of ρ, this weighting scheme performs as well as the more computationally involved alternatives and hence is a viable option in case where the calculation of the inverse in $(\mathbf{I}_N - \rho\mathbf{W}_N)^{-1}$ is computationally prohibitive.

The second weighting scheme (labeled mix) uses an inconsistent estimate of the variance covariance matrix of the moment conditions. However, this does not negatively affects the small sample performance of the GMM estimator and the performance is for most parameter values in fact better than that of the other two alternatives.

The last weighting scheme has for many parameter values clearly the smallest bias but its RMSE is about the same as that of the alternatives. Overall there seems to be no clear best choice of the weighting scheme and all of the weights lead to a second stage GMM estimator that performs satisfactory over the entire range of parameter (which is not true for any of the initial estimators).

Examining other quantiles of the small sample distributions of the estimators in Figures 1-6 shows that the distributions of the initial IV estimators are not are not well approximated by the normal distribution. The Anderson Hsiao estimators

(AH1 and AH2) exhibit large deviations after the .20 quantiles and although the extended instrument set employed by the AB estimator alleviates this, there are still deviations from normality at the .10 quantile.

On the other hand, the second stage GMM estimators show no dramatic deviations from normality up to their .10 quantile. The weighting scheme that ignores the spatial correlation shows some deviations from normality at the .05 quantile and hence the resultant estimator might not perform well in the usual hypothesis tests. The weights based on $\widehat{\mathbf{V}}^{mix}$ and $\widehat{\mathbf{V}}^{E}$ perform better at the .05 quantile, with the estimator based on $\widehat{\mathbf{V}}^{E}$ being marginally better than the one based on $\widehat{\mathbf{V}}^{mix}$. Nevertheless, for both weighting schemes there is still some size distortion of tests based on the .05 and .95 quantiles. This is in line with finding of other studies that looked at the performance of GMM estimators and found that often the use of asymptotic distributions of the GMM estimators as a small sample guidance was not satisfactory, suggesting the use of ML estimation (e.g. Binder et al. 2000).

Turning to the estimator of the spatial autocorrelation parameter ρ in Tables D3 and D4, it is remarkable that the spatial GM procedure works well even when based on inefficient initial estimators. The loss of efficiency in terms of RMSE is for many parameter values negligible. Observe that with $\phi = 0$, the simulations in this study are comparable to those in Kapoor et al. (2005). To check whether this is indeed the case, Figures 9-11 present the comparison of the values of RMSE for the unweighted spatial GM estimator based on the true values of the disturbances obtained in this simulation study with the comparable RMSE values reported in an earlier draft of the Kapoor et al. paper. The value for W in the labels corresponds to the type of the weighting matrix used and is the same as in the Tabled D1-D4.

6 Directions for Future Research

In this thesis I have concentrated on studying a specific model and deriving formal results on the properties of the suggested estimation procedure under a particular set of maintained assumptions. In the future this approach can obviously be extended along several dimensions.

Firstly, the model under consideration can be extended to include other elements. In particular it would be of interest to consider a spatial lag in the dependent variable in addition to the spatial lag in the disturbance process.

Secondly, the estimation procedure under consideration can be altered. In this respect it could be interesting to consider potentially more efficient estimation procedures such as GMM estimators based on an extended set of moment conditions as suggested by, for example Ahn and Schmidt (1995), or some form of continuously updating GMM estimator.

Finally, the set of maintained assumptions can be made more general. Here the first extension that can be tackled is to allow for the exogenous variables to be stochastic.

A Appendix: Central Limit Theorem for Vectors of Linear Quadratic Forms

For the convenience of the reader I first give explicit formulae for the mean and covariances of linear quadratic forms. I focus on the case where the diagonal elements of the quadratic forms are zero and the innovations have zero mean.[44] The following lemma is a special case of a Lemma A.1 in Kelejian and Prucha (2005).

Lemma A1 *Let* $\boldsymbol{\varsigma}_N = (\varsigma_1, \ldots, \varsigma_n)' \sim (0, \boldsymbol{\Sigma}_n)$ *where* $\boldsymbol{\Sigma}_n$ *is diagonal and positive definite, and let* $\mathbf{A}_n = (a_{ij,n})$ *and* $\mathbf{B}_n = (b_{ij,n})$ *be* $n \times n$ *nonstochastic symmetric matrices where* $a_{ii,n} = b_{ii,n} = 0$*. Let* $\mathbf{a}_n$ *and* $\mathbf{b}_n$ *be* $n \times 1$ *nonstochastic vectors. Consider the decomposition* $\boldsymbol{\Sigma}_n = \mathbf{P}_n\mathbf{P}_n'$*, and let* $\boldsymbol{\zeta}_n = (\zeta_{1,n}, \ldots, \zeta_{n,n})' = \mathbf{P}_n^{-1}\boldsymbol{\varsigma}_n$*. Then assuming that the elements of* $\boldsymbol{\zeta}_n$ *are independently distributed with zero mean, variance one fourth moments* $E(\zeta_{i,n}^4) = \mu_{\zeta_i}^{(4)}$ *we have*

$$\begin{aligned}
E\left(\mathbf{a}_n'\boldsymbol{\varsigma}_n + \boldsymbol{\varsigma}_n'\mathbf{A}_n\boldsymbol{\varsigma}_n\right) &= 0, \\
VC(\mathbf{a}_n'\boldsymbol{\varsigma}_n + \boldsymbol{\varsigma}_n'\mathbf{A}_n\boldsymbol{\varsigma}_n) &= 2tr(\mathbf{A}_n\boldsymbol{\Sigma}_n\mathbf{A}_n\boldsymbol{\Sigma}_n) + \mathbf{a}_n'\boldsymbol{\Sigma}_n\mathbf{a}_n, \\
Cov(\mathbf{a}_n'\boldsymbol{\varsigma}_n + \boldsymbol{\varsigma}_n'\mathbf{A}_n\boldsymbol{\varsigma}_n, \mathbf{b}_n'\boldsymbol{\varsigma}_n + \boldsymbol{\varsigma}_n'\mathbf{B}_n\boldsymbol{\varsigma}_n) &= 2tr(\mathbf{A}_n\boldsymbol{\Sigma}_n\mathbf{B}_n\boldsymbol{\Sigma}_n) + \mathbf{a}_n'\boldsymbol{\Sigma}_n\mathbf{b}_n.
\end{aligned}$$

[44] In general the variance and covariance of quadratic forms will depend on the second, third and fourth moments of the innovations. However, since we specialize to the case where the diagonal elements of the quadratic forms are zero, the variance and covariance of the quadratic forms will only depend on the second moments of the innovations.

For proof see Kelejian and Prucha (2005). The expressions also correspond to those given in Kelejian and Prucha (2001). Obviously, in case $\mathbf{A}_n$ and $\mathbf{B}_n$ are not symmetric the above formulae apply with $\mathbf{A}_n$ and $\mathbf{B}_n$ replaced by $(\mathbf{A}_n + \mathbf{A}_n')/2$ and $(\mathbf{B}_n + \mathbf{B}_n')/2$.

For convenience of the reader, I next state a Central Limit Theorem (CLT) for vectors of quadratic forms of triangular arrays based on Theorem A.1 in Kelejian and Prucha (2005).

Let $\boldsymbol{\varsigma}_n = (\varsigma_{1,n}, \ldots, \varsigma_{n,n})'$ be an $n \times 1$ random vector, let

$$\mathbf{A}_{r,n} = (a_{ij,r,n})_{i,j=1,\ldots,n}, \tag{A.1}$$

be nonstochastic matrices, and let $\mathbf{b}_{r,n} = (b_{1,r,n}, \ldots b_{n,r,n})'$ be nonstochastic vectors ($r = 1, \ldots, m$). Consider the following assumptions:

Assumption A1 *The real valued random variables of the array* $\{\varsigma_{i,n} : 1 \leq i \leq n,$ $n \geq 1\}$ *satisfy* $E\varsigma_{i,n} = 0$*. Furthermore, for each* $n \geq 1$ *the random variables* $\varsigma_{1,n}, \ldots, \varsigma_{n,n}$ *are totally independent.*

Assumption A2 *For* $r = 1, \ldots, m$ *the elements of the array of real numbers* $\{a_{ij,r,n} : 1 \leq i, j \leq n, n \geq 1\}$ *satisfy* $a_{ij,r,n} = a_{ji,r,n}$ *and*[45]

$$\sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^{n} |a_{ij,r,n}| < \infty.$$

[45]The assumption of symmetry of the elements of A_n is maintained w.l.o.g. since $\varepsilon_n' A_n \varepsilon_n = \varepsilon_n' \left[(A_n + A_n')/2\right] \varepsilon_n$.

The elements of the array of real numbers $\{b_{i,r,n} : 1 \leq i \leq n, n \geq 1\}$ *satisfy*

$$\sup_{n \geq 1} n^{-1} \sum_{i=1}^{n} |b_{i,r,n}|^{2+\delta_1} < \infty$$

for some $\delta_1 > 0$.

Note that a sufficient condition for Assumption A2 is that the row and column sums of $\mathbf{A}_n$ and the elements of $\mathbf{b}_n$ are uniformly bounded in absolute value.

Assumption A3 *For* $r = 1, \ldots, m$ *we assume that one of the following two conditions holds.*

(a) $\sup_{1 \leq i \leq n, n \geq 1} E\,|\varsigma_{i,n}|^{2+\delta_2} < \infty$ *for some* $\delta_2 > 0$ *and* $a_{ii,r,n} = 0$.

(b) $\sup_{1 \leq i \leq n, n \geq 1} E\,|\varsigma_{i,n}|^{4+\delta_2} < \infty$ *for some* $\delta_2 > 0$ *(but possibly* $a_{ii,r,n} \neq 0$*).*

Consider the quadratic forms

$$q_{r,n} = \boldsymbol{\varsigma}_n' \mathbf{A}_{r,n} \boldsymbol{\varsigma}_n + \mathbf{b}_{r,n}' \boldsymbol{\varsigma}_n \tag{A.2}$$

and define the vector of linear quadratic forms

$$\mathbf{q}_n = \begin{bmatrix} q_{1,n} \\ \vdots \\ q_{m,n} \end{bmatrix}, \tag{A.3}$$

and let

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{q}_n} &= E\mathbf{q}_n, \\ \boldsymbol{\Sigma}_{\mathbf{q}_n} &= E(\mathbf{q}_n - E\mathbf{q}_n)(\mathbf{q}_n - E\mathbf{q}_n)' \end{aligned} \tag{A.4}$$

denote the mean vector and the variance covariance matrix of $\mathbf{q}_n$. Then

$$\boldsymbol{\mu}_{\mathbf{q}_n} = \begin{bmatrix} \mu_{q_{1,n}} \\ \vdots \\ \mu_{q_{m,n}} \end{bmatrix}, \qquad \boldsymbol{\Sigma}_{\mathbf{q}_n} = \begin{bmatrix} \sigma_{q_{11,n}} & \cdots & \sigma_{q_{1m,n}} \\ \vdots & \ddots & \vdots \\ \sigma_{q_{m1,n}} & \cdots & \sigma_{q_{mm,n}} \end{bmatrix}. \tag{A.5}$$

where $\mu_{q_{r,n}}$ and $\sigma_{q_{rs,n}}$ denote the mean of $q_{r,n}$ and the covariance between $q_{r,n}$ and $q_{s,n}$, respectively, for $r, s = 1, \ldots, m$. We now have the following CLT.

Theorem A1 *Suppose Assumptions A1-A3 hold and* $n^{-1}\lambda_{\min}(\boldsymbol{\Sigma}_{\mathbf{q}_n}) \geq c$ *for some* $c > 0$. *Let* $\boldsymbol{\Sigma}_{\mathbf{q}_n} = \left(\boldsymbol{\Sigma}_{\mathbf{q}_n}^{1/2}\right)\left(\boldsymbol{\Sigma}_{\mathbf{q}_n}^{1/2}\right)'$*, then*

$$\boldsymbol{\Sigma}_{\mathbf{q}_n}^{-1/2}\left(\mathbf{q}_n - \boldsymbol{\mu}_{\mathbf{q}_n}\right) \xrightarrow{d} N\left(0, \mathbf{I}_m\right).$$

Of course, the theorem remains valid, if all assumptions are assumed to hold for $n > n_0$ where n_0 is finite. The above theorem can also be applied to situations where $n = TN$ with T finite and $N \to \infty$; see footnote 13 in Kelejian and Prucha (2001).

I now illustrate this in more detail. Suppose, we have sample sizes $T, 2T, 3T, ..., NT, ..., \infty$ as $N \to \infty$ and the random variables are triangular arrays is

$$
\begin{aligned}
\boldsymbol{\varepsilon}_1 &= (\varsigma_{11,1}, \dots, \varsigma_{T1,1})' \\
\boldsymbol{\varepsilon}_2 &= (\varsigma_{11,2}, \varsigma_{12,2}, \dots, \varsigma_{T1,2}, \varsigma_{T2,2})' \\
&\vdots \\
\boldsymbol{\varepsilon}_N &= (\varsigma_{11,N}, \dots, \varsigma_{1N,N}, \varsigma_{21,N}, \dots, \varsigma_{2N,N}, \dots, \varsigma_{T1,N}, \dots, \varsigma_{TN,N})',
\end{aligned}
\tag{A.6}
$$

Consider the sequence of vectors of linear quadratic forms and the vectors of linear quadratic forms

$$\mathbf{v}_N = (v_{1,N}, ..., v_{m,N})', \tag{A.7}$$

with

$$v_{r,N} = \boldsymbol{\varepsilon}_N' \mathbf{A}_{r,TN} \boldsymbol{\varepsilon}_N + \mathbf{b}_{r,TN}' \boldsymbol{\varepsilon}_N. \tag{A.8}$$

As above, we denote by $\boldsymbol{\mu}_{\mathbf{v}_N}$ and $\boldsymbol{\Sigma}_{\mathbf{v}_N}$ the mean vector and variance covariance matrix of the vector $\mathbf{v}_N$.

Suppose that the random variables collected in $\boldsymbol{\varepsilon}_N$ satisfy Assumptions A1 and A3, and the sequences of matrices $\mathbf{A}_{r,TN}$ and vectors $\mathbf{b}_{r,TN}$ satisfy Assumption A2.

We can define additional triangular arrays of sizes between tN and $(t+1)\,N$ to obtain a sequence

$$\begin{aligned}
\boldsymbol{\varsigma}_1 &= (\varsigma_{11,1}) && \text{(A.9)}\\
\boldsymbol{\varsigma}_2 &= (\varsigma_{11,1}, \varsigma_{21,1})' \\
&\vdots \\
\boldsymbol{\varsigma}_T &= (\varsigma_{11,1}, \ldots, \varsigma_{T1,1})' \\
\boldsymbol{\varsigma}_{T+1} &= (\varsigma_{11,2}, \ldots, \varsigma_{T1,2}, \varsigma_{12,2})' && \text{(A.10)}\\
\boldsymbol{\varsigma}_{T+2} &= (\varsigma_{11,2}, \ldots, \varsigma_{T1,2}, \varsigma_{12,2}, \varsigma_{12,2}, \varsigma_{22,2})' \\
&\vdots \\
\boldsymbol{\varsigma}_{2T} &= (\varsigma_{11,2}, \ldots, \varsigma_{T1,2}, \varsigma_{12,2}, \ldots, \varsigma_{T2,2})' \\
&\vdots \\
\boldsymbol{\varsigma}_{NT} &= (\varsigma_{11,N}, \ldots, \varsigma_{1N,N}, \varsigma_{21,N}, \ldots, \varsigma_{2N,N}, \ldots, \varsigma_{T1,N}, \ldots, \varsigma_{TN,N})'\,.
\end{aligned}$$

Observe that the new sequence $\boldsymbol{\varsigma}_n$ satisfies Assumptions A1 and A3 and that for $n = NT$ we have $\boldsymbol{\varsigma}_n = \boldsymbol{\varepsilon}_N$.

Similarly, we can extend the sequence of vectors of linear quadratic forms to

$$\mathbf{q}_n = (q_{1,n}, ..., q_{m,n})'\,, \tag{A.11}$$

where

$$q_{r,n} = \boldsymbol{\varsigma}_n' \mathbf{A}_{r,n} \boldsymbol{\varsigma}_n + \mathbf{b}_{r,n}' \boldsymbol{\varsigma}_n, \tag{A.12}$$

with

$$
\begin{aligned}
\mathbf{A}_{r,n} &= \begin{pmatrix} \mathbf{A}_{r,\left[\frac{n}{T}\right]T} & \mathbf{0}_{\left[\frac{n}{T}\right]T\times 1} & \cdots & \mathbf{0}_{\left[\frac{n}{T}\right]T\times 1} \\ \mathbf{0}_{1\times\left[\frac{n}{T}\right]T} & a_{11,r,\left[\frac{n}{T}\right]T+1} & \cdots & a_{k1,r,\left[\frac{n}{T}\right]T+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times\left[\frac{n}{T}\right]T} & a_{1k,r,\left[\frac{n}{T}\right]T+1} & \cdots & a_{kk,r,\left[\frac{n}{T}\right]T+1} \end{pmatrix}, \qquad \text{(A.13)} \\
\mathbf{b}_{r,n} &= \begin{pmatrix} \mathbf{b}_{r,\left[\frac{n}{T}\right]T} \\ b_{1,r,\left[\frac{n}{T}\right]T+1} \\ \vdots \\ b_{k,r,\left[\frac{n}{T}\right]T+1} \end{pmatrix},
\end{aligned}
$$

and $k = n - \left[\frac{n}{T}\right]T$, where I use $\left[\frac{r}{s}\right]$ to denote the whole part of a rational number $\frac{r}{s}$.

Observe that by definition for $n = NT$, we have $\mathbf{q}_n = \mathbf{v}_N$. Furthermore, since $\mathbf{A}_{r,n}$ and $\mathbf{b}_{r,n}$ satisfy Assumption A2 for $n = NT$, it follows from the construction of $\mathbf{A}_{r,NT}$ and $\mathbf{b}_{r,NT}$ that they satisfy Assumption A2 for all n. As a result, quadratic forms $\mathbf{q}_n$ fulfill conditions of Theorem A1 and $\boldsymbol{\Sigma}_{\mathbf{q}_n}^{-1/2}\left(\mathbf{q}_n - \boldsymbol{\mu}_{\mathbf{q}_n}\right) \xrightarrow{d} N\left(0, \mathbf{I}_m\right)$ as $n \to \infty$, where as before $\boldsymbol{\mu}_{\mathbf{q}_n}$ and $\boldsymbol{\Sigma}_{\mathbf{q}_n}$ denote the mean vector and variance covariance matrix of the vector $\mathbf{q}_n$. Hence the sequence of distribution functions of $\boldsymbol{\Sigma}_{\mathbf{q}_n}^{-1/2}\left(\mathbf{q}_n - \boldsymbol{\mu}_{\mathbf{q}_n}\right)$ converges weakly to the distribution function of $N\left(0, \mathbf{I}_m\right)$. We now select a subsequence from the distribution functions of $\boldsymbol{\Sigma}_{\mathbf{q}_n}^{-1/2}\left(\mathbf{q}_n - \boldsymbol{\mu}_{\mathbf{q}_n}\right)$ for $n = NT$ (we treat T as a fixed constant) and observe that these are equivalent to the sequence of distribution functions of $\boldsymbol{\Sigma}_{\mathbf{v}_N}^{-1/2}\left(\mathbf{v}_N - \boldsymbol{\mu}_{\mathbf{v}_N}\right)$.

This subsequence must have the same limit and, as a consequence, we have that

$$\boldsymbol{\Sigma}_{\mathbf{v}_N}^{-1/2}\left(\mathbf{v}_N - \boldsymbol{\mu}_{\mathbf{v}_N}\right) \xrightarrow{d} N\left(0, \mathbf{I}_m\right), \tag{A.14}$$

as $N \to \infty$.

B Appendix: Proof of Claims in Chapter 3

Lemma B1 *: Let ς_j, $j \in \mathbb{N}$, be a sequence of totally independent real valued random variables with $E\left|\varsigma_j\right|^p \leq k_\varsigma < \infty$ for some $2 \leq p < \infty$. Let a_j be a sequence of real numbers such that $\sum_{j=0}^{\infty}\left|a_j\right| \leq k_a < \infty$. (a) Consider the random variables $\chi_m = \sum_{j=0}^{m} a_j\varsigma_j$, then there exists a random variable χ, which we denote as $\sum_{j=0}^{\infty} a_j\varsigma_j$, such that $\chi_m \xrightarrow{r} \chi$ for $0 < r \leq p$. (b) Furthermore, $E\left|\chi\right|^r \leq k_a^r k_\varsigma^{r/p} < \infty$, for $0 < r \leq p$.*

Proof: To prove part (a) I first show that each χ_m has finite p-th absolute moments and hence belongs to the L^p space of random variables with finite absolute p-th moments. I then demonstrate that the sequence χ_m is a Cauchy sequence. By invoking the completeness property of the L^p space we will then have that the limiting random variable χ also belongs to L^p. I now turn to each of these steps in detail.

Let $p \geq 2$ and $1/q + 1/p = 1$. Then, using the triangle and Hölder's inequalities

$$\begin{aligned}
\left|\sum_{i=1}^{m} a_i\varsigma_i\right| &\leq \sum_{i=1}^{m}\left|a_i\right|\left|\varsigma_i\right| = \sum_{i=1}^{m}\left|a_i\right|^{1/q}\left|a_i\right|^{1/p}\left|\varsigma_i\right| \\
&\leq \left[\sum_{i=1}^{m}\left|a_i\right|\right]^{1/q}\left[\sum_{i=1}^{m}\left|a_i\right|\left|\varsigma_i\right|^p\right]^{1/p} \\
&\leq k_a^{1/q}\left[\sum_{i=1}^{m}\left|a_i\right|\left|\varsigma_i\right|^p\right]^{1/p},
\end{aligned} \tag{B.1}$$

and further

$$
\begin{aligned}
E\left|\chi_m\right|^p &= E\left|\sum_{i=1}^{m} a_i \varsigma_i\right|^p \leq k_a^{p/q} \sum_{i=1}^{m} |a_i|\, E\left|\varsigma_i\right|^p \\
&\leq k_a^{p/q} k_\varsigma \sum_{i=1}^{m} |a_i|\, k_a^{p/q+1} k_\varsigma = k_a^p k_\varsigma < \infty,
\end{aligned}
\tag{B.2}
$$

and hence each χ_m belongs to L^p.

I now demonstrate that the sequence χ_m is Cauchy in L^p, or in the terminology of Shiryayev (1984, p.251) that it is fundamental in L^p. Since $\sum_{i=1}^{\infty} |a_i| < \infty$ it follows from the Cauchy Test (Neylor and Sell, 1982, p.225) that for every $\varepsilon > 0$ there exist and index N_ε such that

$$
\sum_{i=m+1}^{m+k} |a_i| < \varepsilon, \tag{B.3}
$$

for all $m \geq N_\varepsilon$ and $k \geq 0$. Now choose some $\varepsilon_* > 0$ and $\varepsilon = \varepsilon_*/(k_a^p k_\varsigma)$, then by argumentation analogous to above

$$
\begin{aligned}
E\left|\chi_{m+k} - \chi_m\right|^p &= E\left|\sum_{i=1}^{m+k} a_i \varsigma_i - \sum_{i=1}^{m} a_i \varsigma_i\right|^p \\
&\leq k_a^{p/q} k_\varsigma \sum_{i=m+1}^{m+k} |a_i| \leq k_a^p k_\varsigma \varepsilon = \varepsilon_*,
\end{aligned}
\tag{B.4}
$$

for all $m \geq N_\varepsilon$ and $k \geq 0$. Thus under the maintained assumptions the sequence χ_m is Cauchy in L^p. By Theorem 7 in Shiryayev (1984, p.258) we then have that the sequence χ_m converges in p-th mean to a random variable in L^p, which implies that χ exists as a limit in p-th mean. Of course, since for $r \leq p$

$$\|\chi_m - \chi\|_r \leq \|\chi_m - \chi\|_p \,, \tag{B.5}$$

by Lyapunov's inequality it follows that χ_m converges to χ also in r-th mean for $0 < r \leq p$.

To prove part (b) observe that from the above $E\,|\chi|^r \leq c$ for some $c < \infty$. Hence $E\,|\chi|^r \leq (E\,|\chi|^p)^{\frac{r}{p}} \leq c^{\frac{r}{p}} < \infty$. ■

Lemma B2 *: Let $\mathbf{A}_n$ be a sequence of nonstochastic matrices of dimensions $n \times n$ where $n \in \mathbb{N}$ such that $\max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}| \leq k_A < \infty$. Consider a sequence of $n \times 1$ random vectors $\boldsymbol{\varsigma}_n$, with elements $\varsigma_{i,n}$ that are real valued random variables with $E\,|\varsigma_{i,n}|^p \leq k_\varsigma < \infty$ for some $2 \leq p < \infty$. Then the elements of the random vector $\overline{\boldsymbol{\varsigma}}_n = \mathbf{A}_n \boldsymbol{\varsigma}_n$ have finite r-th moments with $E\,|\overline{\varsigma}_{i,n}|^r \leq k_A^r k_\varsigma^{r/p} < \infty$, for $0 < r \leq p$.*

Proof: Let $p \geq 2$ and $1/q + 1/p = 1$. Using the triangle and Hölder's inequalities, we have

$$\begin{aligned} |\overline{\varsigma}_{i,n}| &= \left| \sum_{j=1}^{n} a_{ij,n} \varsigma_{j,n} \right| \leq \left[\sum_{j=1}^{n} |a_{ij,n}|\, |\varsigma_{j,n}| \right] = \left[\sum_{j=1}^{n} |a_{ij,n}|^{\frac{1}{q}}\, |a_{ij,n}|^{\frac{1}{p}}\, |\varsigma_{j,n}| \right] \\ &\leq \left[\sum_{j=1}^{n} |a_{ij,n}| \right]^{\frac{1}{q}} \left[\sum_{j=1}^{n} |a_{ij,n}|\, |\varsigma_{j,n}|^p \right]^{\frac{1}{p}} . \end{aligned} \tag{B.6}$$

and further

$$
\begin{aligned}
E\left|\overline{\varsigma}_{i,n}\right|^{p} &\leq \left[\sum_{j=1}^{n}\left|a_{ij,n}\right|\right]^{\frac{p}{q}}\left[\sum_{j=1}^{n}\left|a_{ij,n}\right|E\left|\varsigma_{j,n}\right|^{p}\right] \\
&\leq k_A^{p/q}\sum_{j=1}^{n}\left|a_{ij,n}\right|k_{\varsigma} \leq k_A^{p}k_{\varsigma}.
\end{aligned} \tag{B.7}
$$

Observe that by Lyapunov's inequality for $0 < r \leq p$,

$$
\left\|\overline{\varsigma}_{i,n}\right\|_{r} \leq \left\|\overline{\varsigma}_{i,n}\right\|_{p} = \left[E\left|\overline{\varsigma}_{i,n}\right|^{p}\right]^{1/p} \leq k_A k_{\varsigma}^{1/p}, \tag{B.8}
$$

and hence $E\left|\overline{\varsigma}_{i,n}\right|^{r} = \left\|\overline{\varsigma}_{i,n}\right\|_{r}^{r} \leq k_A^{r}k_{\varsigma}^{r/p}$. ■

Lemma B3 *: Suppose Assumptions 1, 2, 3 and 5 hold.*

(a) Let $\boldsymbol{\omega}_{t,N} = \mathbf{X}_{t,N}\boldsymbol{\beta} + \mathbf{u}_{t,N}$*, and let* $\omega_{it,N}$ *denote the* i*-th element of* $\boldsymbol{\omega}_{t,N}$*, then*

$$
E\left|\omega_{it,N}\right|^{4+\delta} \leq k_{\omega} < \infty,
$$

where k_{ω} *does not depend* i, t, N*.*

(b) The random vector

$$
\mathbf{y}_{t,N} = \sum_{j=0}^{\infty}\phi^{j}\boldsymbol{\omega}_{t-j,N},
$$

is well defined as the limit of the finite sums in quadratic means and there is $\delta > 0$ *such that* $E\left|y_{it,N}\right|^{r} \leq k_y < \infty$ *for all* $r \leq 4+\delta$*, where* k_y *does not depend* i, t, N*.*

Proof: In the following let $p = 4 + \delta$ with $\delta = \min\{\delta_\varepsilon, \delta_\mu\}$. I first prove part (a). Denoting $x_{ikt,N}$ the k-th element of $\mathbf{x}_{it,N}$, we have from Assumption 5(b) that $|x_{ikt,N}| \leq k_X < \infty$ and thus

$$\left|\mathbf{x}'_{it,N}\boldsymbol{\beta}\right|^p \leq k_X^p (\boldsymbol{\beta}'\boldsymbol{\beta})^{p/2} < \infty. \tag{B.9}$$

Next, from Assumptions 1 and 2, we have

$$E\left|\upsilon_{it,N}\right|^p \leq 2^{p-1}\left(E\left|\varepsilon_{it,N}\right|^p + E\left|\mu_{i,N}\right|^p\right) \leq 2^{p-1}\left(k_\varepsilon + k_\mu\right) < \infty, \tag{B.10}$$

by inequality (1.4.3) in Bierens (1994). Now observe that $\mathbf{u}_{t,N} = \mathbf{P}_N \boldsymbol{\upsilon}_{t,N}$. By Assumption 3(c) we have

$$\max_i \sum_{j=1}^{N} p_{ij,N} \leq k_P < \infty, \tag{B.11}$$

and hence by Lemma B2 we have $E\left|u_{it,N}\right|^p \leq k_P^p 2^{p-1}\left(k_\varepsilon + k_\mu\right)$. Hence

$$\begin{aligned} E\left|\omega_{it,N}\right|^p &\leq 2^{p-1}\left\{\left[E\left|\mathbf{x}'_{it,N}\boldsymbol{\beta}\right|^p + E\left|u_{it,N}\right|^p\right]\right\} \\ &\leq 2^{p-1}\left\{k_X^p (\boldsymbol{\beta}'\boldsymbol{\beta})^{p/2} + k_P^p 2^{p-1}\left(k_\varepsilon + k_\mu\right)\right\} < \infty, \end{aligned} \tag{B.12}$$

i.e., the p-th absolute moment of $\omega_{it,N}$ is uniformly bounded by a finite constant that does not depend i, t, N.

To prove part (b) observe that $\sum_{i=0}^{\infty} |\phi|^i = 1/(1 - |\phi|) < \infty$. Given part (a) of the Lemma, part (b) now follows immediately from Lemma B1. ∎

Equation (3.2.4): The vector of endogenous variables is defined by a stochastic difference equation:

$$\mathbf{y}_{t,N} = \phi \mathbf{y}_{t-1,N} + \boldsymbol{\omega}_{t,N}. \tag{B.13}$$

From Lemma B3 above it immediately follows that the random variables

$$\begin{aligned} y_{it}^p &= \sum_{j=0}^{\infty} \phi^j \omega_{i,t-j}, \\ y_{i,t-1}^p &= \sum_{j=0}^{\infty} \phi^j \omega_{i,t-1-j}, \end{aligned} \tag{B.14}$$

are well defined as limits of the finite sums in quadratic means.

I now show that they are a particular solution. Substituting into the RHS of the difference equation defining $\mathbf{y}_{t,N}$, we have (using Theorems 2.6 and 2.7 in Prucha, 2004):

$$\begin{aligned} \phi y_{i,t-1,N}^p + \omega_{it,N} &= \phi \sum_{j=0}^{\infty} \phi^j \omega_{i,t-1-j} + \omega_{it,N} \\ &= \sum_{j=1}^{\infty} \phi^j \omega_{i,t-j} + \omega_{it,N} \\ &= \sum_{j=0}^{\infty} \phi^j \omega_{i,t-j} = y_{it,N}^p, \end{aligned} \tag{B.15}$$

and hence $\mathbf{y}_{t,N}^p$ is a particular solution. The homogeneous part of the difference equation is

$$\mathbf{y}_{t,N}^h - \phi \mathbf{y}_{t-1,N}^h = 0. \tag{B.16}$$

and its solution is of the form $\mathbf{y}_{t,N}^{h} = \boldsymbol{\gamma}\phi^{t+m}$, where $\boldsymbol{\gamma}$ is a $N \times 1$ vector of (finite) constants and $-m$ is the starting point of the process. Since I assume that the process has started in an infinite past ($m = +\infty$), we have that $\mathbf{y}_{t,N}^{h} = \lim_{m\to\infty} \boldsymbol{\gamma}\phi^{t+m} = 0$ and, as a result, the unique solution is

$$\mathbf{y}_{t,N} = \mathbf{y}_{t,N}^{p} + \mathbf{y}_{t,N}^{h} = \sum_{j=0}^{\infty} \phi^{j} \boldsymbol{\omega}_{t-j,N}.$$

Substituting for the definition of $\boldsymbol{\omega}_{t-j,N}$ and utilizing Theorem 2.6 in Prucha (2004) yields

$$\begin{aligned}
\mathbf{y}_{t,N} &= \sum_{j=0}^{\infty} \phi^{j} \boldsymbol{\omega}_{t-j,N} && \text{(B.17)} \\
&= \sum_{j=0}^{\infty} \phi^{j} \left(\mathbf{X}_{t-j,N}\boldsymbol{\beta} + \mathbf{P}_{N}\boldsymbol{\varepsilon}_{t-j,N} + \mathbf{P}_{N}\boldsymbol{\mu}_{N} \right) \\
&= \sum_{j=0}^{\infty} \phi^{j} \left(\mathbf{X}_{t-j,N}\boldsymbol{\beta} + \mathbf{P}_{N}\boldsymbol{\varepsilon}_{t-j,N} \right) + \sum_{j=0}^{\infty} \phi^{j} \mathbf{P}_{N}\boldsymbol{\mu}_{N} \\
&= \sum_{j=0}^{\infty} \phi^{j} \left(\mathbf{X}_{t-j,N}\boldsymbol{\beta} + \mathbf{P}_{N}\boldsymbol{\varepsilon}_{t-j,N} \right) + (1-\phi)^{-1} \mathbf{P}_{N}\boldsymbol{\mu}_{N}.
\end{aligned}$$

The claim in Chapter 3 then follows from specializing the above expression for $t = 0$.

Equation (3.2.5): By Lemma B3 we have that

$$\mathbf{y}_{0,N} = \sum_{j=0}^{\infty} \phi^{j} \boldsymbol{\omega}_{t-j}, \tag{B.18}$$

with $E\left(\omega_{it,N}^2\right) \leq k_\omega < \infty$ and $|\phi| < 1$. Using Theorems 2.6 and 2.7 in Prucha (2004) we can write

$$\begin{aligned}
\mathbf{y}_{0,N} &= \sum_{j=0}^{\infty} \phi^j \boldsymbol{\omega}_{t-j} \qquad \text{(B.19)} \\
&= \sum_{j=0}^{\infty} \phi^j \left(\mathbf{X}_{-j,N}\boldsymbol{\beta} + \mathbf{u}_{t,N}\right) \\
&= \left(\sum_{j=0}^{\infty} \phi^j \mathbf{X}_{-j,N}\boldsymbol{\beta}\right) + \left(\sum_{j=0}^{\infty} \phi^j \mathbf{u}_{t,N}\right) \\
&= \left(\sum_{j=0}^{\infty} \phi^j \mathbf{X}_{-j,N}\boldsymbol{\beta}\right) + \left(\sum_{j=0}^{\infty} \phi^j \mathbf{P}_N \boldsymbol{\varepsilon}_{-j,N}\right) + \left(\sum_{j=0}^{\infty} \phi^j \mathbf{P}_N \boldsymbol{\mu}_N\right) \\
&= \mathbf{c}_N + \mathbf{P}_N \boldsymbol{\xi}_N + (1-\phi)^{-1} \mathbf{P}_N \boldsymbol{\mu}_N,
\end{aligned}$$

where $\mathbf{c}_N$ is nonstochastic and the vectors of random variables are $\boldsymbol{\xi}_N = \sum_{j=0}^{\infty} \phi^j \boldsymbol{\varepsilon}_{-j,N}$ and $\boldsymbol{\mu}_N$. Notice that by Lemma B1 the random variable $\boldsymbol{\xi}_N$ is well defined. From Assumption 1 and Theorem 2.2 in Prucha (2004) we have that

$$E\left(\boldsymbol{\xi}_N\right) = \sum_{j=0}^{\infty} \phi^j E\left(\boldsymbol{\varepsilon}_{-j,N}\right) = \mathbf{0}_{N\times 1} \qquad \text{(B.20)}$$

and

$$\begin{aligned}
VC\left(\boldsymbol{\xi}_N\right) &= E\left(\boldsymbol{\xi}_N \boldsymbol{\xi}_N'\right) = \sum_{j=0}^{\infty} \left(\phi^2\right)^j \sigma_\varepsilon^2 \mathbf{I}_N \qquad \text{(B.21)} \\
&= \sigma_\varepsilon^2 \left(1-\phi^2\right)^{-1} \mathbf{I}_N.
\end{aligned}$$

Furthermore,

$$\begin{aligned} VC\left(\mathbf{P}_N \boldsymbol{\xi}_N\right) &= E\left(\mathbf{P}_N \boldsymbol{\xi}_N \boldsymbol{\xi}_N' \mathbf{P}_N'\right) = \mathbf{P}_N E\left(\boldsymbol{\xi}_N \boldsymbol{\xi}_N'\right) \mathbf{P}_N' \\ &= \sigma_\varepsilon^2 \left(1-\phi^2\right)^{-1} \mathbf{P}_N \mathbf{P}_N'. \end{aligned} \tag{B.22}$$

Observe that by Assumption 2, the random variables $\mathbf{P}_N \boldsymbol{\xi}_N$ and $\boldsymbol{\mu}_N$ are independent. Thus we have

$$\begin{aligned} VC\left(\mathbf{y}_{0,N}\right) &= VC\left(\mathbf{P}_N \boldsymbol{\xi}_N\right) + VC\left[(1-\phi)^{-1} \mathbf{P}_N \boldsymbol{\mu}_N\right] \qquad \text{(B.23)} \\ &= \sigma_\varepsilon^2 \left(1-\phi^2\right)^{-1} \mathbf{P}_N \mathbf{P}_N' + (1-\phi)^{-2} \mathbf{P}_N VC\left(\boldsymbol{\mu}_N\right) \mathbf{P}_N' \\ &= \left(\frac{\sigma_\varepsilon^2}{1-\phi^2} + \frac{\sigma_\mu^2}{(1-\phi)^2}\right) \mathbf{P}_N \mathbf{P}_N'. \end{aligned}$$

C Appendix: Proofs for Chapter 4

I will make repeated use of the following facts:

Lemma C1 *Let* $\mathbf{C} = \mathbf{A} + \mathbf{B}$ *be square real valued symmetric matrices of same dimensions. Then*

$$\lambda_{\min}(\mathbf{C}) \geq \lambda_{\min}(\mathbf{A}) + \lambda_{\min}(\mathbf{B}).$$

For proof see, e.g., Rao and Rao (1998), Proposition 10.1.1.

Lemma C2 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $n \times m$ *and* $n \times n$ *matrices. If* $\mathbf{B}$ *is symmetric then*

$$\lambda_{\min}(\mathbf{A}'\mathbf{B}\mathbf{A}) \geq \lambda_{\min}(\mathbf{A}'\mathbf{A}) \cdot \lambda_{\min}(\mathbf{B}).$$

Proof: By Rayleigh-Ritz Theorem (see, e.g. Proposition 4.2.2 in Horn and Johnson 1985) we have that the smallest eigenvalue of a symmetric matrix can be obtained as:

$$\lambda_{\min}(\mathbf{C}) = \inf_{\boldsymbol{\alpha} \neq \mathbf{0}} \left[(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1} (\boldsymbol{\alpha}'\mathbf{C}\boldsymbol{\alpha}) \right] = \inf_{\boldsymbol{\alpha};\boldsymbol{\alpha}'\boldsymbol{\alpha}=1} (\boldsymbol{\alpha}'\mathbf{C}\boldsymbol{\alpha}). \tag{C.0.1}$$

Since $\mathbf{B}$ is symmetric, we can decompose it as $\mathbf{B} = \mathbf{U}'\boldsymbol{\Lambda}\mathbf{U}$ where $\mathbf{U}$ is orthogonal and $\boldsymbol{\Lambda} = diag(\lambda_1, .., \lambda_n)$ is diagonal with eigenvalues of $\mathbf{B}$ on the diagonal (cp. Proposition 52 in Dhrymes 1984). Hence we have

$$\begin{aligned}
\lambda_{\min}\left(\mathbf{A}'\mathbf{B}\mathbf{A}\right) &= \lambda_{\min}\left(\mathbf{A}'\mathbf{U}'\mathbf{\Lambda}\mathbf{U}\mathbf{A}\right) \qquad \text{(C.0.2)}\\
&= \inf_{\boldsymbol{\alpha};\boldsymbol{\alpha}'\boldsymbol{\alpha}=1}\left[\boldsymbol{\alpha}'\mathbf{A}'\mathbf{U}'\mathbf{\Lambda}\mathbf{U}\mathbf{A}\boldsymbol{\alpha}\right]\\
&\geq \inf_j \lambda_j \inf_{\boldsymbol{\alpha};\boldsymbol{\alpha}'\boldsymbol{\alpha}=1}\left[\boldsymbol{\alpha}'\mathbf{A}'\mathbf{U}'\mathbf{U}\mathbf{A}\boldsymbol{\alpha}\right]\\
&= \lambda_{\min}\left(\mathbf{B}\right)\cdot \inf_{\boldsymbol{\alpha};\boldsymbol{\alpha}'\boldsymbol{\alpha}=1}\left[\boldsymbol{\alpha}'\mathbf{A}'\mathbf{A}\boldsymbol{\alpha}\right]\\
&= \lambda_{\min}\left(\mathbf{B}\right)\cdot\lambda_{\min}\left(\mathbf{A}'\mathbf{A}\right).
\end{aligned}$$

■

Lemma C3 *Let $\mathbf{a}_n$, and $\mathbf{b}_n$be sequences of $n \times 1$ vectors and $\mathbf{C}_n$ be a sequence of $n \times n$ matrices. Suppose that the elements of $\mathbf{a}_n$ and $\mathbf{b}_n$ are uniformly bounded in absolute value, and that the matrix $\mathbf{C}_n$ has uniformly bounded absolute row (or column) sums. Then $n^{-1}\mathbf{a}_n'\mathbf{C}_n\mathbf{b}_n$ is uniformly bounded in absolute value.*

Proof: Denote the uniform bounds of the elements of the vectors $\mathbf{a}_n$ and $\mathbf{b}_n$ as k_a and k_b and the uniform bound of the absolute row sums of the matrices $\mathbf{C}_n$ as k_c. We have by the triangle inequality

$$\begin{aligned}
n^{-1}\left|\mathbf{a}_n'\mathbf{C}_n\mathbf{b}_n\right| &= n^{-1}\left|\sum_{i=1}^n\sum_{i=1}^n a_{i,n}c_{ij,n}b_{j,n}\right| \leq n^{-1}\sum_{i=1}^n\sum_{i=1}^n \left|a_{i,n}\right|\left|c_{ij,n}\right|\left|b_{j,n}\right|\\
&\leq n^{-1}\sum_{i=1}^n\sum_{i=1}^n k_a\left|c_{ij,n}\right|k_b = k_ak_bn^{-1}\sum_{i=1}^n\sum_{i=1}^n\left|c_{ij,n}\right| \qquad \text{(C.0.3)}\\
&\leq k_ak_bn^{-1}\sum_{i=1}^n k_c = k_ak_bk_c < \infty.
\end{aligned}$$

$\blacksquare$

C.1 Proofs for Section 4.1

Proof of Lemma 1: By backward substitution we can eliminate lagged dependent variables and express $\mathbf{y}_{-2}$ as a function of lagged disturbance terms and lagged explanatory variables. From (3.2.2), we have that $\mathbf{y}_{-2,N}$ is

$$\begin{aligned}
\mathbf{y}_{-2,N} &= \begin{pmatrix} \mathbf{y}_{0,N} \\ \mathbf{y}_{1,N} \\ \vdots \\ \mathbf{y}_{T-2,N} \end{pmatrix} = \begin{pmatrix} \mathbf{y}_{0,N} \\ \mathbf{X}_{1,N}\boldsymbol{\beta} + \mathbf{u}_{1,N} + \phi \mathbf{y}_{0,N} \\ \vdots \\ \sum_{j=0}^{T-3} \phi^j \left[\mathbf{X}_{T-2-j,N}\boldsymbol{\beta} + \mathbf{u}_{T-2-j,N}\right] + \phi^{T-2}\mathbf{y}_{0,N} \end{pmatrix} \\
&= (\boldsymbol{\Phi}' \otimes \mathbf{I}_N) \left[\begin{pmatrix} \mathbf{0}_{N\times p} \\ \mathbf{X}_{1,N} \\ \vdots \\ \mathbf{X}_{T-2,N} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \mathbf{y}_{0,N} \\ \mathbf{u}_{1,N} \\ \vdots \\ \mathbf{u}_{T-2,N} \end{pmatrix} \right], \qquad \text{(C.1.4)}
\end{aligned}$$

where

$$\boldsymbol{\Phi} = \begin{pmatrix} 1 & \phi & \cdots & \phi^{T-2} \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \phi \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \qquad \text{(C.1.5)}$$

Next I express $\left(\mathbf{y}'_{0,N}, \mathbf{u}'_{1,N}, ..., \mathbf{u}'_{T-2,N}\right)'$ as a linear form of

$$\boldsymbol{\eta}_N = \left(\boldsymbol{\mu}_N, \boldsymbol{\xi}_N, \boldsymbol{\varepsilon}'_{1,N}, ..., \boldsymbol{\varepsilon}'_{T,N}\right)'. \tag{C.1.6}$$

Observe that $\mathbf{u}_{t,N} = \mathbf{P}_N\left(\boldsymbol{\varepsilon}_{t,N} + \boldsymbol{\mu}_N\right)$ and from equation (3.2.4), $\mathbf{y}_{0,N} = E\left(\mathbf{y}_{0,N}\right) + \mathbf{P}_N\left[\boldsymbol{\xi}_N + \frac{\boldsymbol{\mu}_N}{1-\phi}\right]$, with $\boldsymbol{\xi}_N = \sum_{j=0}^{\infty} \phi^j \boldsymbol{\varepsilon}_{-j,N}$ well defined by Lemma B1. Therefore,

$$\begin{aligned} & \left(\mathbf{y}'_{0,N}, \mathbf{u}'_{1,N}, ..., \mathbf{u}'_{T-2,N}\right)' \\ = & \left[\begin{pmatrix} [1-\phi]^{-1} & 1 & \mathbf{0}_{1\times T-2} & \mathbf{0}_{1\times 2} \\ \mathbf{1}_{T-2\times 1} & \mathbf{0}_{T-2\times 1} & \mathbf{I}_{T-2} & \mathbf{0}_{T-2\times 2} \end{pmatrix}_{T-1\times T+2} \otimes \mathbf{P}_N\right]\boldsymbol{\eta}_N \\ & + \left[E\left(\mathbf{y}'_{0,N}\right), \mathbf{0}_{1\times(T-2)N}\right]'. \end{aligned} \tag{C.1.7}$$

Hence with the notation $\mathbf{X}_{-2,N} = \left(\mathbf{0}'_{N\times p}, \mathbf{X}'_{1,N}, ..., \mathbf{X}'_{T-2,N}\right)'$ we have

$$\begin{aligned} \mathbf{y}_{-2,N} = & \left(\boldsymbol{\Phi}' \otimes \mathbf{I}_N\right)\left\{\mathbf{X}_{-2,N}\boldsymbol{\beta} + \left[E\left(\mathbf{y}'_{0,N}\right), \mathbf{0}_{T-2\times 1}\right]'\right\} \\ & + \left[\boldsymbol{\Phi}'\begin{pmatrix} [1-\phi]^{-1} & 1 & \mathbf{0}_{1\times T-2} & \mathbf{0}_{1\times 2} \\ \mathbf{1}_{T-2\times 1} & \mathbf{0}_{T-2\times 1} & \mathbf{I}_{T-2} & \mathbf{0}_{T-2} \end{pmatrix} \otimes \mathbf{P}_N\right]\boldsymbol{\eta}_N. \end{aligned} \tag{C.1.8}$$

Therefore, given that

$$\Delta\mathbf{u}_N = \left[\left(\mathbf{0}_{(T-1)\times 2}, \mathbf{D}\right) \otimes \mathbf{P}_N\right]\boldsymbol{\eta}_N, \tag{C.1.9}$$

we can express $\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N$ as a function of the model disturbances and explanatory variables:

$$
\begin{aligned}
& \mathbf{y}'_{-2,N}\Delta\mathbf{u}_N && \text{(C.1.10)}\\
= \; & \left(\boldsymbol{\beta}'\mathbf{X}'_{-2,N} + \left[E\left(\mathbf{y}'_{0,N}\right), \mathbf{0}_{T-2\times 1}\right]\right)\left(\boldsymbol{\Phi}\otimes\mathbf{I}_N\right)\left[\left(\mathbf{0}_{(T-1)\times 2}, \mathbf{D}\right)\otimes\mathbf{P}_N\right]\boldsymbol{\eta}_N \\
& +\boldsymbol{\eta}'_N \left[\begin{pmatrix} \frac{1}{1-\phi} & \mathbf{1}_{1\times T-2} \\ 1_{1\times 1} & \mathbf{0}_{1\times T-2} \\ \mathbf{0}_{T-2\times 1} & \mathbf{I}_{T-2} \\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times T-2} \end{pmatrix} \boldsymbol{\Phi}\left(\mathbf{0}_{(T-1)\times 2}, \mathbf{D}\right)\otimes\mathbf{P}'_N\mathbf{P}_N\right]\boldsymbol{\eta}_N \\
= \; & \mathbf{f}'_N\left(\mathbf{I}_{T+2}\otimes\mathbf{P}_N\right)\boldsymbol{\eta}_N + \boldsymbol{\eta}'_N\left(\mathbf{F}\otimes\mathbf{P}'_N\mathbf{P}_N\right)\boldsymbol{\eta}_N,
\end{aligned}
$$

where

$$
\mathbf{f}'_N = \left(\boldsymbol{\beta}'\mathbf{X}'_{-2,N} + \left[E\left(\mathbf{y}'_{0,N}\right), \mathbf{0}_{T-2\times 1}\right]\right)\left(\boldsymbol{\Phi}\otimes\mathbf{I}_N\right)\left[\left(\mathbf{0}_{(T-1)\times 2}, \mathbf{D}\right)\otimes\mathbf{I}_N\right] \tag{C.1.11}
$$

and

$$
\mathbf{F}_{T+2\times T+2} = \begin{pmatrix} \frac{1}{1-\phi} & \mathbf{1}_{1\times T-2} \\ 1_{1\times 1} & \mathbf{0}_{1\times T-2} \\ \mathbf{0}_{T-2\times 1} & \mathbf{I}_{T-2} \\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times T-2} \end{pmatrix} \boldsymbol{\Phi}\left(\mathbf{0}_{(T-1)\times 2}, \mathbf{D}\right) \tag{C.1.12}
$$

The expression for $\Delta\mathbf{X}'_N\Delta\mathbf{u}_N$ follows from a trivial substitution of $\Delta\mathbf{u}_N = \left[\left(\mathbf{0}_{(T-1)\times 2}, \mathbf{D}\right)\otimes\mathbf{P}_N\right]\boldsymbol{\eta}_N$. ■

Proof of Lemma 2: To obtain the expected value and variance of the two quadratic forms, I use the expression from Lemma 1:

$$\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N = \mathbf{f}'_N\left[\mathbf{I}_{T+2}\otimes\mathbf{P}_N\right]\boldsymbol{\eta}_N + \boldsymbol{\eta}'_N\left(\mathbf{F}\otimes\mathbf{P}'_N\mathbf{P}_N\right)\boldsymbol{\eta}_N, \tag{C.1.13}$$

and

$$\Delta\mathbf{X}'_N\Delta\mathbf{u}_N = \Delta\mathbf{X}'_N\left[\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)\otimes\mathbf{P}_N\right]\boldsymbol{\eta}_N, \tag{C.1.14}$$

where $\boldsymbol{\eta}_N = \left[\boldsymbol{\mu}'_N, \boldsymbol{\xi}'_N, \boldsymbol{\varepsilon}'_{1,N}, ..., \boldsymbol{\varepsilon}'_{T,N}\right]'$ is a vector of independent zero mean random variables with uniformly bounded fourth moments. Next I verify that assumptions of Lemma A1 in Appendix A are satisfied. Given Assumption 1 and 2, it remains to be verified that diagonal elements of $(\mathbf{F}\otimes\mathbf{P}'_N\mathbf{P}_N)$ are zero. Observe that from Lemma 1 we have $\mathbf{F} = \mathbf{A}\boldsymbol{\Phi}\mathbf{B}$, where

$$\underset{(T+2)\times(T-1)}{\mathbf{A}} = \begin{pmatrix} \frac{1}{1-\phi} & \mathbf{1}_{1\times(T-2)} \\ 1_{1\times 1} & \mathbf{0}_{1\times(T-2)} \\ \mathbf{0}_{(T-2)\times 1} & \mathbf{I}_{T-2} \\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times(T-2)} \end{pmatrix}, \quad \underset{(T-1)\times(T+2)}{\mathbf{B}} = \left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right). \tag{C.1.15}$$

The diagonal elements of $\mathbf{F}$ are then

$$\begin{aligned} \mathbf{F}_{ii} &= \left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{B}\right\}_{ii} = \sum_{j=1}^{T-1}\mathbf{A}_{ij}\left\{\boldsymbol{\Phi}\mathbf{B}\right\}_{ji} \qquad (\text{C.1.16}) \\ &= \sum_{j=1}^{T-1}\mathbf{A}_{ij}\sum_{k=1}^{T-1}\boldsymbol{\Phi}_{jk}\mathbf{B}_{ki}, \end{aligned}$$

where $\mathbf{A}_{ij}$ and $\mathbf{B}_{ij}$ denote the $ij-th$ elements of matrices $\mathbf{A}$ and $\mathbf{B}$ respectively. Note that $\mathbf{B}_{ki}=0$ for $k<i+2$ and $\mathbf{\Phi}_{jk}=0$ for $k<j$, and, therefore, $\{\mathbf{\Phi B}\}_{ji}=\sum_{k=1}^{T-1}\mathbf{\Phi}_{jk}\mathbf{B}_{ki}=0$ for $i<j+2$.[46] Furthermore, the elements $\mathbf{A}_{ij}$ are zero for $i>j+1$ and hence $\mathbf{F}_{ii}=\sum_{j=1}^{T-1}\mathbf{A}_{ij}\{\mathbf{\Phi B}\}_{ji}=0$.

Hence I can use Lemma A1 to derive the mean and variances and covariances of $\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N$ and $\Delta\mathbf{X}'_N\Delta\mathbf{u}_N$. In particular, we have that

$$E\left(\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N\right)=E\left(\Delta\mathbf{X}'_N\Delta\mathbf{u}_N\right)=0, \tag{C.1.17}$$

and

$$\begin{aligned} VC\left(\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N\right) &= \mathbf{f}'_N\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}_N\mathbf{P}'_N\right)\mathbf{f}_N \\ &\quad +2tr\left(\mathbf{F}^S\mathbf{\Sigma}_{\boldsymbol{\eta},N}\mathbf{F}^S\mathbf{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}'_N\mathbf{P}_N\mathbf{P}'_N\mathbf{P}_N\right) \\ &= \mathbf{f}'_N\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}_N\mathbf{P}'_N\right)\mathbf{f}_N+\nu_N, \end{aligned} \tag{C.1.18}$$

$$\begin{aligned} VC\left(\Delta\mathbf{X}'_N\Delta\mathbf{u}_N\right) &= \Delta\mathbf{X}'_N\left[\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)\otimes\mathbf{I}_N\right]\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}_N\mathbf{P}'_N\right)* \\ &\quad \left[\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)'\otimes\mathbf{I}_N\right]\Delta\mathbf{X}_N. \end{aligned} \tag{C.1.19}$$

[46] Note that the both matrices $\mathbf{\Phi}$ and $\mathbf{D}$ are upper diagonal (in the sense that their $ij-th$ elements are zero for $i<j$) and hence their $\mathbf{\Phi D}$ product also has the same property. As a result, the matrix $\mathbf{\Phi B}=\left(\mathbf{0}_{(T-1)\times 2},\mathbf{\Phi D}\right)$ will have its $ij-th$ elements equal to zero for $i<j+2$.

and finally

$$Cov\left(\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N,\Delta\mathbf{X}'_N\Delta\mathbf{u}_N\right)=\mathbf{f}'_N\left[\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)'\otimes\mathbf{P}'_N\right]\Delta\mathbf{X}_N, \tag{C.1.20}$$

where I defined

$$\begin{aligned}\nu_N &= 2tr\left(\mathbf{F}^S\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{F}^S\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}'_N\mathbf{P}_N\mathbf{P}'_N\mathbf{P}_N\right) \qquad \text{(C.1.21)}\\ &= 2tr\left(\mathbf{F}^S\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{F}^S\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\right)\cdot tr\left(\mathbf{P}'_N\mathbf{P}_N\mathbf{P}'_N\mathbf{P}_N\right).\end{aligned}$$

Together we have that

$$\begin{aligned}\mathbf{V}_N &= \begin{pmatrix} VC\left(\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N\right) & Cov\left(\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N,\Delta\mathbf{X}'_N\Delta\mathbf{u}_N\right)\\ Cov\left(\mathbf{y}'_{-2,N}\Delta\mathbf{u}_N,\Delta\mathbf{X}'_N\Delta\mathbf{u}_N\right)' & VC\left(\Delta\mathbf{X}'_N\Delta\mathbf{u}_N\right)\end{pmatrix}\\ &= \mathbf{S}'_N\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}_N\mathbf{P}'_N\right)\mathbf{S}_N+\begin{pmatrix}\nu_N & \mathbf{0}_{1\times p}\\ \mathbf{0}_{p\times 1} & \mathbf{0}_{p\times p}\end{pmatrix}, \qquad \text{(C.1.22)}\end{aligned}$$

where $\mathbf{S}_N=\left(\mathbf{f}_N,\left[\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)'\otimes\mathbf{I}_N\right]\Delta\mathbf{X}_N\right)$.

■

Proof of Lemma 3: From Lemma C1, we have that

$$\lambda_{\min}\left(\mathbf{V}_N\right)\geq\lambda_{\min}\left[\mathbf{S}'_N\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{P}_N\mathbf{P}'_N\right)\mathbf{S}_N\right]+\min\left(\nu_N,0\right). \tag{C.1.23}$$

Note that since $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ is symmetric, by Proposition 52 in Dhrymes (1984) we can express it as $\mathbf{\Sigma}_{\boldsymbol{\eta}} = \mathbf{\Lambda}'\mathbf{\Lambda}$. Hence

$$\begin{aligned} tr\left(\mathbf{F}^{S}\mathbf{\Sigma}_{\boldsymbol{\eta}}\mathbf{F}^{S}\mathbf{\Sigma}_{\boldsymbol{\eta}}\right) &= tr\left(\mathbf{F}^{S}\mathbf{\Lambda}'\mathbf{\Lambda}\mathbf{F}^{S}\mathbf{\Lambda}'\mathbf{\Lambda}\right) \\ &= tr\left(\mathbf{\Lambda}\mathbf{F}^{S}\mathbf{\Lambda}'\mathbf{\Lambda}\mathbf{F}^{S}\mathbf{\Lambda}'\right) \\ &= tr\left(\mathbf{A}'\mathbf{A}\right) \geq 0 \end{aligned}$$

with $\mathbf{A} = \mathbf{\Lambda}\mathbf{F}^{S}\mathbf{\Lambda}'$, since $\mathbf{F}^{S}$ is also symmetric. Therefore,

$$tr\left(\mathbf{F}^{S}\mathbf{\Sigma}_{\boldsymbol{\eta},N}\mathbf{F}^{S}\mathbf{\Sigma}_{\boldsymbol{\eta},N}\right) \geq 0. \tag{C.1.24}$$

Furthermore,

$$tr\left(\mathbf{P}'_N\mathbf{P}_N\mathbf{P}'_N\mathbf{P}_N\right) = tr\left[\left(\mathbf{P}'_N\mathbf{P}_N\right)\left(\mathbf{P}'_N\mathbf{P}_N\right)\right] \geq 0, \tag{C.1.25}$$

and, therefore, $\nu_N \geq 0$.

By Lemma C2 the smallest eigenvalue of $\mathbf{V}_N$ is then

$$\begin{aligned} \lambda_{\min}\left(\mathbf{V}_N\right) &\geq \lambda_{\min}\left[\mathbf{S}_N\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N} \otimes \mathbf{P}_N\mathbf{P}'_N\right)\mathbf{S}'_N\right] \qquad \text{(C.1.26)} \\ &\geq \lambda_{\min}\left(\mathbf{S}'_N\mathbf{S}_N\right) \cdot \lambda_{\min}\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N} \otimes \mathbf{P}_N\mathbf{P}'_N\right). \end{aligned}$$

From Theorem 4.2.12 in Horn and Johnson (1991) we have

$$\lambda_{\min}\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N} \otimes \mathbf{P}_N\mathbf{P}'_N\right) = \lambda_{\min}\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N}\right) \cdot \lambda_{\min}\left(\mathbf{P}_N\mathbf{P}'_N\right), \tag{C.1.27}$$

and hence

$$
\begin{aligned}
\left[(T-1)N\right]^{-1}\lambda_{\min}\left(\mathbf{V}_N\right) &\geq \left[(T-1)N\right]^{-1}\lambda_{\min}\left(\mathbf{S}'_N\mathbf{S}_N\right)\cdot \\
&\quad \cdot\lambda_{\min}\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\right)\cdot\lambda_{\min}\left(\mathbf{P}_N\mathbf{P}'_N\right) \\
&= \lambda_{\min}\left(\left[(T-1)N\right]^{-1}\mathbf{S}'_N\mathbf{S}_N\right)\cdot \\
&\quad \cdot\lambda_{\min}\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\right)\cdot\lambda_{\min}\left(\mathbf{P}_N\mathbf{P}'_N\right).
\end{aligned} \tag{C.1.28}
$$

By Assumptions 4 we have that $\lambda_{\min}\left(\mathbf{P}_N\mathbf{P}'_N\right) \geq c_P > 0$, by Assumption IV2 we have that $\lambda_{\min}\left(\left[(T-1)N\right]^{-1}\mathbf{S}'_N\mathbf{S}_N\right) \geq c_S > 0$. Since $\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}$ is diagonal, we have $\lambda_{\min}\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\right) = \min\left[\sigma_\mu^2, var\left(\boldsymbol{\xi}_{i,N}\right), \sigma_\varepsilon^2\right] = \min\left[\sigma_\mu^2, \frac{\sigma_\varepsilon^2}{1-\phi^2}, \sigma_\varepsilon^2\right] \geq c_\Sigma > 0$ and hence $\left[(T-1)N\right]^{-1}\lambda_{\min}\left(\mathbf{V}_N\right) \geq c_S c_\Sigma c_P > 0$. ■

Proof of Proposition 1: The result in the Proposition is a special case of the general result in Lemma 5 in Section 4.3,[47] which is in turn based on the CLT in Theorem A1 in Appendix A. Here I verify directly that the conditions of Theorem A1 hold.

[47]The conditions of that Lemma are satisfied since by Lemma 1 (and also Lemma 4 in Section 4.3), the instruments $\mathbf{y}_{-2,N}$ and $\Delta\mathbf{X}_N$ are linear forms in the innovations of the form assumed in Lemma 5. Furthermore, by Lemma 3, the smallest eigenvalue $\mathbf{V}_N$ is uniformly bounded away from zero. Finally, the moment conditions are valid since by Lemma 2, we have $E\left(\mathbf{H}'_N\Delta\mathbf{u}_N\right) = \mathbf{0}$. Therefore, conditions of Lemma 5 are satisfied and we have that $\mathbf{V}_N^{-1/2}\mathbf{H}'_N\Delta\mathbf{u}_N \xrightarrow{d} N\left(\mathbf{0}, \mathbf{I}\right)$.

The moment conditions are

$$\begin{aligned}
\mathbf{H}_N'\Delta\mathbf{u}_N &= \begin{pmatrix} \mathbf{H}_{2,N} \\ \vdots \\ \mathbf{H}_{T,N} \end{pmatrix}' \Delta\mathbf{u}_N = \begin{bmatrix} (\mathbf{y}_{0,N}, \Delta\mathbf{X}_{2,N}) \\ \vdots \\ (\mathbf{y}_{T-2,N}, \Delta\mathbf{X}_{T,N}) \end{bmatrix}' \Delta\mathbf{u}_N \qquad \text{(C.1.29)} \\
&= \left[\begin{pmatrix} \mathbf{y}_{0,N} \\ \vdots \\ \mathbf{y}_{T-2,N} \end{pmatrix}, \begin{pmatrix} \Delta\mathbf{X}_{2,N} \\ \vdots \\ \Delta\mathbf{X}_{T,N} \end{pmatrix}\right]' \Delta\mathbf{u}_N = \begin{pmatrix} \mathbf{y}_{-2,N}'\Delta\mathbf{u}_N \\ \Delta\mathbf{X}_N'\Delta\mathbf{u}_N \end{pmatrix}
\end{aligned}$$

Observe that by Lemma 1, the instruments $\mathbf{y}_{-2,N}$ and $\Delta\mathbf{X}_N$ are linear forms in the innovations and, as a result, the moment conditions collected in $\mathbf{H}_N'\Delta\mathbf{u}_N$ are linear quadratic form in the innovations

$$\boldsymbol{\eta}_N = \left(\boldsymbol{\mu}_N', \boldsymbol{\xi}_N', \boldsymbol{\varepsilon}_{1,N}', ..., \boldsymbol{\varepsilon}_{T,N}'\right), \qquad \text{(C.1.30)}$$

where $\boldsymbol{\xi}_N = \sum_{j=0}^{\infty} \phi^j \boldsymbol{\varepsilon}_{-j,N}$. By Assumptions 1 and 6 it follows from Lemma B1 in Appendix B that the random variable $\boldsymbol{\xi}_N$ satisfies condition A3 in Appendix A. Therefore, by Assumptions 1 and 2, the elements of the innovations $\boldsymbol{\eta}_N$ satisfy conditions A1 and A3 in Appendix A.

By Lemma 2, the variance covariance matrix of the moment conditions collected in $\mathbf{H}_N'\Delta\mathbf{u}_N$ is $\mathbf{V}_N$ and by Lemma 3, the smallest eigenvalue of $[(T-1)\,N]^{-1}\,\mathbf{V}_N$ is uniformly bounded away from zero. Hence it remains to be shown that the linear quadratic forms collected in $\mathbf{H}_N'\Delta\mathbf{u}_N$ satisfy condition A2 in Appendix A.

Note that from Lemma 1 we have that the elements of $\mathbf{H}_N'\Delta\mathbf{u}_N$ are

$$\mathbf{y}_{-2,N}'\Delta\mathbf{u}_N = \mathbf{f}_N'\left(\mathbf{I}_{T+2}\otimes\mathbf{P}_N\right)\boldsymbol{\eta}_N + \boldsymbol{\eta}_N'\left(\mathbf{F}\otimes\mathbf{P}_N'\mathbf{P}_N\right)\boldsymbol{\eta}_N, \tag{C.1.31}$$

and

$$\Delta\mathbf{X}_N'\Delta\mathbf{u}_N = \Delta\mathbf{X}_N'\left[\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)\otimes\mathbf{P}_N\right]\boldsymbol{\eta}_N. \tag{C.1.32}$$

Observe that any finite sum, product or Kronecker product of matrices with row and column sums uniformly bounded in absolute value will also have row and column sums uniformly bounded in absolute value; see Kelejian and Prucha (2001d) for details.

From Lemma 1, we have that

$$\mathbf{f}_N' = \left\{\boldsymbol{\beta}'\mathbf{X}_{-2,N}' + \left[E\left(\mathbf{y}_{0,N}'\right),\mathbf{0}_{1\times(T-2)N}\right]\right\}\left[\boldsymbol{\Phi}\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)\otimes\mathbf{I}_N\right]. \tag{C.1.33}$$

Elements and dimensions of $\boldsymbol{\Phi}\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)$ do not depend on N and hence trivially $\left[\boldsymbol{\Phi}\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)\otimes\mathbf{I}_N\right]$ has row and column sums uniformly bounded in absolute value. Elements of the vector $\boldsymbol{\beta}'\mathbf{X}_{-2,N}'$ are uniformly bounded in absolute value by Assumption 5 and elements of $\left[E\left(\mathbf{y}_{0,N}'\right),\mathbf{0}_{1\times(T-2)N}\right]$ are uniformly bounded in absolute value since, as demonstrated by Lemma B3 in Appendix B, y_{it} has uniformly bounded $4+\delta$ moments for some $\delta>0$. Together we then have that $\mathbf{f}_N$ has elements uniformly bounded in absolute value. The sequence of matrices $\mathbf{P}_N$ has row and column sums uniformly bounded in absolute value (Assumption 3) and hence elements of $\mathbf{f}_N'\left(\mathbf{I}_{T+2}\otimes\mathbf{P}_N\right)$ are uniformly bounded in absolute value. Similarly, by Assumptions 5 and 3, $\Delta \mathbf{X}_N' \left[\left(\mathbf{0}_{(T-1)\times 2}, \mathbf{D} \right) \otimes \mathbf{P}_N \right]$ has row and column sums uniformly bounded in absolute value. Finally, since dimensions of $\mathbf{F}$ do not change with N and its elements are also independent of N, the matrix $(\mathbf{F} \otimes \mathbf{P}_N' \mathbf{P}_N)$ has row and column sums uniformly bounded in absolute value.

This completes the verification of conditions of Theorem A1 and, therefore, we have that $\mathbf{V}_N^{-1/2} \mathbf{H}_N' \Delta \mathbf{u}_N \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$. ■

Proof of Theorem 1: From equation (4.1.10) we have

$$
\begin{aligned}
& \sqrt{(T-1)N} \left(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta} \right) \qquad \text{(C.1.34)} \\
= \; & \sqrt{(T-1)N} \left[\Delta \mathbf{Z}_N' \mathbf{H}_N \left(\mathbf{H}_N' \mathbf{H}_N \right)^{-1} \mathbf{H}_N' \Delta \mathbf{Z}_N \right]^{-1} * \\
& \Delta \mathbf{Z}_N' \mathbf{H}_N \left(\mathbf{H}_N' \mathbf{H}_N \right)^{-1} \mathbf{H}_N' \Delta \mathbf{u}_N \\
= \; & \sqrt{(T-1)N} \left[\frac{\Delta \mathbf{Z}_N' \mathbf{H}_N}{(T-1)N} \left(\frac{\mathbf{H}_N' \mathbf{H}_N}{(T-1)N} \right)^{-1} \frac{\mathbf{H}_N' \Delta \mathbf{Z}_N}{(T-1)N} \right]^{-1} * \\
& \frac{\Delta \mathbf{Z}_N' \mathbf{H}_N}{(T-1)N} \left(\frac{\mathbf{H}_N' \mathbf{H}_N}{(T-1)N} \right)^{-1} \frac{\mathbf{H}_N' \Delta \mathbf{u}_N}{(T-1)N} \\
= \; & \left[\frac{\Delta \mathbf{Z}_N' \mathbf{H}_N}{(T-1)N} \left(\frac{\mathbf{H}_N' \mathbf{H}_N}{(T-1)N} \right)^{-1} \frac{\mathbf{H}_N' \Delta \mathbf{Z}_N}{(T-1)N} \right]^{-1} * \\
& \frac{\Delta \mathbf{Z}_N' \mathbf{H}_N}{(T-1)N} \left(\frac{\mathbf{H}_N' \mathbf{H}_N}{(T-1)N} \right)^{-1} \frac{\mathbf{H}_N' \Delta \mathbf{u}_N}{\sqrt{(T-1)N}}.
\end{aligned}
$$

Given Assumptions IV1 and IV3, our result follows from Proposition 1 in this thesis and Corollary 5 in Pötcher and Prucha (2001). ■

C.2 Proofs for Section 4.2

I now give a sequence of Lemmas that will be used to prove Theorem 2. I use the notation $\|.\|$ to denote the matrix norm $\|\mathbf{M}\| := [tr(\mathbf{M}'\mathbf{M})]^{1/2}$.

Lemma C4 *Let $\widehat{\mathbf{u}}_N$ be based on a $N^{1/2}$ consistent estimate of $\boldsymbol{\theta}$. Then under Assumptions 1-6 we can write*

$$\mathbf{u}_N - \widehat{\mathbf{u}}_N = \mathbf{D}_N \boldsymbol{\Delta}_N.$$

where the random matrix $\mathbf{D}_N$ has elements $d_{ij,N}$ that have uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$, i.e. $E|d_{ij,N}|^{4+\delta} \leq c_d < \infty$ where c_d does not depend on N, and the random vector $\boldsymbol{\Delta}$ is such that $N^{1/2}\|\boldsymbol{\Delta}_N\| = O_p(1)$.

Proof: Note that from (4.2.1) we can write $\mathbf{u}_{t,N} - \widehat{\mathbf{u}}_{t,N}$ as

$$\mathbf{u}_{t,N} - \widehat{\mathbf{u}}_{t,N} = (\mathbf{y}_{t-1,N}, \mathbf{X}_{t,N})\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_N\right), \tag{C.2.1}$$

I define $\mathbf{D}_{t,N} = (\mathbf{y}_{t-1,N}, \mathbf{X}_{t,N})$ and $\boldsymbol{\Delta}_N = \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_N\right)$. Hence we have

$$\mathbf{u}_N - \widehat{\mathbf{u}}_N = \mathbf{D}_N \boldsymbol{\Delta}_N, \tag{C.2.2}$$

where $\mathbf{D}_N = \left(\mathbf{D}'_{1,N}, ..., \mathbf{D}'_{T,N}\right)'$.

Since $\widehat{\boldsymbol{\theta}}_N$ is $\sqrt{N}$ consistent, it follows that $N^{1/2}\|\boldsymbol{\Delta}_N\| = O_p(1)$. By Lemma B3, elements of $\mathbf{y}_{t-1,N}$ have finite $4+\delta$ absolute moments for some $\delta > 0$.

The nonstochastic elements of $\mathbf{D}_N$ are uniformly bounded in absolute value by Assumption 5 and hence also their $4+\delta$ power is uniformly bounded in absolute value. Thus $\mathbf{D}_N$ has uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. ■

Note that the claim in the above lemma also holds for $2+\delta$ moments since by Lyapunov's inequality,

$$E\left|y_{i,t-1,N}\right|^{2+\delta} \leq \left[E\left|y_{i,t-1,N}\right|^{(4+\delta)}\right]^{(2+\delta)/(4+\delta)} \leq k_y^{(2+\delta)/(4+\delta)} < \infty. \quad \text{(C.2.3)}$$

Lemma C5 *Given Assumptions 1-6, the moment conditions converge to their expectations in probability, i.e.*

$$\gamma_{kl,N}^j - E\left(\gamma_{kl,N}^j\right) \xrightarrow{p} 0 \textit{ and } \gamma_{k,N}^j - E\left(\gamma_{k,N}^j\right) \xrightarrow{p} 0$$

as $N \to \infty$ *for* $j = 0, 1$, $k, l = 1, 2, 3$.

Proof: The moment conditions correspond to those considered in Kapoor et al. (2005) and, in particular, Assumptions 1,2 and 4 of their paper are satisfied,[48] and hence the lemma is their Lemma A2. ■

Lemma C6 *The sample counterparts of the moment conditions converge in probability to the true moments, i.e.*

$$g_{kl,N}^j - E\left(\gamma_{kl,N}^j\right) \xrightarrow{p} 0 \textit{ and } g_{k,N}^j - E\left(\gamma_{k,N}^j\right) \xrightarrow{p} 0$$

[48] Assumption 1 is directly implied by our Assumptions 1 and 2. Assumptions 2 and 4 are contained in our Assumption 3.

as $N \to \infty$ *for* $j = 0, 1$, $k, l = 1, 2, 3$.

Proof: In light of Lemma C5, it suffices to show that $g^j_{kl,N} - \gamma^j_{kl,N} \xrightarrow{p} 0$ and $g^j_{k,N} - \gamma^j_{k,N} \xrightarrow{p} 0$. These can be expressed as quadratic forms:

$$
\begin{aligned}
g^j_{kl,N} - \gamma^j_{kl,N} &= \frac{1}{N}\left[\widehat{\mathbf{u}}'_N \mathbf{C}^j_{kl,N} \widehat{\mathbf{u}}_N - \mathbf{u}'_N \mathbf{C}^j_{kl,N} \mathbf{u}_N\right], \qquad \text{(C.2.4)}\\
g^j_{k,N} - \gamma^j_{k,N} &= \frac{1}{N}\left[\widehat{\mathbf{u}}'_N \mathbf{C}^j_{k,N} \widehat{\mathbf{u}}_N - \mathbf{u}'_N \mathbf{C}^j_{k,N} \mathbf{u}_N\right],
\end{aligned}
$$

where the the $NT \times NT$ matrices $\mathbf{C}^j_{kl,N}$ and $\mathbf{C}^j_{k,N}$ are defined for $j = 0, 1$, $k = 1, 2, 3$ and $l = 1, 2$. Explicit expressions are given below. Note that for $l = 3$ we have (see 4.2.10 and 4.2.13):

$$
\begin{aligned}
g^j_{13,N} &= \gamma^j_{13,N} = 1, \qquad \text{(C.2.5)}\\
g^j_{23,N} &= \gamma^j_{23,N} = N^{-1} tr\left(\mathbf{W}'_N \mathbf{W}_N\right),\\
g^j_{33,N} &= \gamma^j_{33,N} = 0,
\end{aligned}
$$

and hence trivially $g^j_{k3,N} - \gamma^j_{k3,N} \xrightarrow{p} 0$ for $j = 0, 1$ and $k = 1, 2, 3$.

For $j = 0, 1$, $k = 1, 2, 3$ and $l = 1, 2$, the $\mathbf{C}^j_{kl,N}$ and $\mathbf{C}^j_{k,N}$ matrices are products of (some of) the matrices $(\mathbf{I}_T \otimes \mathbf{W}'_N)$, $\mathbf{Q}_{j,N}$, and $(\mathbf{I}_T \otimes \mathbf{W}_N)$. In particular, from

(4.2.10) and (4.2.13), $j=0,1$:

$$\begin{aligned}
\mathbf{C}_{11,N}^{j} &= 2\left(T-1\right)^{j-1}\mathbf{Q}_{j,N}\left(\mathbf{I}_T\otimes\mathbf{W}_N\right), \qquad \text{(C.2.6)}\\
\mathbf{C}_{12,N}^{j} &= -\left(T-1\right)^{j-1}\left(\mathbf{I}_T\otimes\mathbf{W}_N'\right)\mathbf{Q}_{j,N}\left(\mathbf{I}_T\otimes\mathbf{W}_N\right),\\
\mathbf{C}_{21,N}^{j} &= 2\left(T-1\right)^{j-1}\left(\mathbf{I}_T\otimes\mathbf{W}_N'\right)\left(\mathbf{I}_T\otimes\mathbf{W}_N'\right)\mathbf{Q}_{j,N}\left(\mathbf{I}_T\otimes\mathbf{W}_N\right),\\
\mathbf{C}_{22,N}^{j} &= -\left(T-1\right)^{j-1}\left(\mathbf{I}_T\otimes\mathbf{W}_N'\right)\left(\mathbf{I}_T\otimes\mathbf{W}_N'\right)\mathbf{Q}_{j,N}\left(\mathbf{I}_T\otimes\mathbf{W}_N\right)\left(\mathbf{I}_T\otimes\mathbf{W}_N\right),\\
\mathbf{C}_{31,N}^{j} &= \left(T-1\right)^{j-1}\mathbf{Q}_{j,N}\left(\mathbf{I}_T\otimes\mathbf{W}_N\right)\left(\mathbf{I}_T\otimes\mathbf{W}_N\right)\\
&\quad +\left(T-1\right)^{j-1}\left(\mathbf{I}_T\otimes\mathbf{W}_N'\right)\mathbf{Q}_{j,N}\left(\mathbf{I}_T\otimes\mathbf{W}_N\right),\\
\mathbf{C}_{32,N}^{j} &= -\left(T-1\right)^{j-1}\left(\mathbf{I}_T\otimes\mathbf{W}_N'\right)\mathbf{Q}_{j,N}\left(\mathbf{I}_T\otimes\mathbf{W}_N\right)\left(\mathbf{I}_T\otimes\mathbf{W}_N\right),\\
\mathbf{C}_{1,N}^{j} &= \left(T-1\right)^{j-1}\mathbf{Q}_{j,N},\\
\mathbf{C}_{2,N}^{j} &= \left(T-1\right)^{j-1}\left(\mathbf{I}_T\otimes\mathbf{W}_N'\right)\mathbf{Q}_{j,N}\left(\mathbf{I}_T\otimes\mathbf{W}_N\right),\\
\mathbf{C}_{3,N}^{j} &= \left(T-1\right)^{j-1}\mathbf{Q}_{j,N}\left(\mathbf{I}_T\otimes\mathbf{W}_N\right).
\end{aligned}$$

By their definition (see equation 4.2.5), the row and column sums of the $\mathbf{Q}_{j,N}$ matrices ($j=0,1$) are less than two in absolute value.[49] The row and column sums of $(\mathbf{I}_T\otimes\mathbf{W}_N)$ and $(\mathbf{I}_T\otimes\mathbf{W}_N')$ are uniformly bounded in absolute value by Assumption 3. Therefore, for $j=0,1$, $k=1,2,3$ and $l=1,2$, each $\mathbf{C}_{kl,N}^{j}$ and $\mathbf{C}_{k,N}^{j}$ matrix has row and column sums uniformly bounded in absolute value.

[49]The row and column sums of $|\mathbf{Q}_{0,N}|$ are equal to $2\frac{T-1}{T}$, while the row and column sums of $|\mathbf{Q}_{1,N}|$ are equal to one.

By Lemma C4 we have $\mathbf{u}_N - \widehat{\mathbf{u}}_N = \mathbf{D}_N \boldsymbol{\Delta}_N$. Utilizing this expression I can write for $j = 0, 1$, $k = 1, 2, 3$ and $l = 1, 2$:

$$
\begin{aligned}
g^j_{kl,N} - \gamma^j_{kl,N} &= \psi^j_{kl,N} + \varphi^j_{kl,N}, \\
g^j_{k,N} - \gamma^j_{k,N} &= \psi^j_{k,N} + \varphi^j_{k,N},
\end{aligned} \tag{C.2.7}
$$

with

$$
\begin{aligned}
\varphi^j_{kl,N} &= \frac{1}{N} \underset{1\times(p+1)}{\boldsymbol{\Delta}'_N} \underset{(p+1)\times NT}{\mathbf{D}'_N} \underset{NT\times NT}{\left(\mathbf{C}^j_{kl,N} + \mathbf{C}^{j\prime}_{kl,N}\right)} \underset{NT\times 1}{\mathbf{u}_N}, \\
\varphi^j_{k,N} &= \frac{1}{N} \underset{1\times(p+1)}{\boldsymbol{\Delta}'_N} \underset{(p+1)\times NT}{\mathbf{D}'_N} \underset{NT\times NT}{\left(\mathbf{C}^j_{k,N} + \mathbf{C}^{j\prime}_{k,N}\right)} \underset{NT\times 1}{\mathbf{u}_N}, \\
\psi^j_{kl,N} &= \frac{1}{N} \underset{1\times(p+1)}{\boldsymbol{\Delta}'_N} \underset{(p+1)\times NT}{\mathbf{D}'_N} \underset{NT\times NT}{\mathbf{C}^j_{kl,N}} \underset{NT\times(p+1)}{\mathbf{D}_N} \underset{(p+1)\times 1}{\boldsymbol{\Delta}_N}, \\
\psi^j_{k,N} &= \frac{1}{N} \underset{1\times(p+1)}{\boldsymbol{\Delta}'_N} \underset{(p+1)\times NT}{\mathbf{D}'_N} \underset{NT\times NT}{\mathbf{C}^j_{k,N}} \underset{NT\times(p+1)}{\mathbf{D}_N} \underset{(p+1)\times 1}{\boldsymbol{\Delta}_N}.
\end{aligned} \tag{C.2.8}
$$

To prove the claim, I show that all the terms $\varphi^j_{kl,N}$, $\varphi^j_{k,N}$, $\psi^j_{kl,N}$ and $\psi^j_{k,N}$ are all $o_p(1)$. To simplify notation, I consider a sequence of $NT \times NT$ matrices $\mathbf{C}_N$ that have row and column sums uniformly bounded in absolute value. I define

$$
\begin{aligned}
\varphi_N &= \frac{1}{N} \underset{1\times(p+1)}{\boldsymbol{\Delta}'_N} \underset{(p+1)\times NT}{\mathbf{D}'_N} \underset{NT\times NT}{\left(\mathbf{C}^j_N + \mathbf{C}'_N\right)} \underset{NT\times 1}{\mathbf{u}_N} \\
&= \frac{1}{N} \underset{1\times(p+1)}{\boldsymbol{\Delta}'_N} \underset{(p+1)\times NT}{\mathbf{D}'_N} \underset{NT\times NT}{\left(\mathbf{C}^j_N + \mathbf{C}'_N\right)} \underset{NT\times NT}{\left(\mathbf{I}_T \otimes \mathbf{P}_N\right)} \underset{NT\times 1}{\mathbf{v}_N}, \\
\psi_N &= \frac{1}{N} \underset{1\times(p+1)}{\boldsymbol{\Delta}'_N} \underset{(p+1)\times NT}{\mathbf{D}'_N} \underset{NT\times NT}{\mathbf{C}_N} \underset{NT\times(p+1)}{\mathbf{D}_N} \underset{(p+1)\times 1}{\boldsymbol{\Delta}_N},
\end{aligned} \tag{C.2.9}
$$

and show that both φ_N, and ψ_N are $o_p(1)$. By substituting $\mathbf{C}_N = \mathbf{C}^j_{kl,N}$ for $j = 0, 1$, $k = 1, 2, 3$ and $l = 1, 2$, and $\mathbf{C}_N = \mathbf{C}^j_{k,N}$ for $k = 1, 2, 3$ and $j = 0, 1$, we then obtain that $\varphi^j_{kl,N}$, $\varphi^j_{k,N}$, $\psi^j_{kl,N}$ and $\psi^j_{k,N}$ are all $o_p(1)$.

Observe that φ_N and ψ_N correspond to ϕ_N and ψ_N in the proof of Lemma C.1 in Kelejian and Prucha (2005), with $\mathbf{C}_n = \mathbf{C}_N$, $\mathbf{A}_n = \mathbf{C}_N(\mathbf{I}_T \otimes \mathbf{P}_N)$ and $\boldsymbol{\varepsilon}_n = \mathbf{v}_N$. Inspection of their proof of $\phi_N = o_p(1)$ and $\psi_N = o_p(1)$ reveals that it only utilizes Assumption 4 of that paper, the fact that the matrices $\mathbf{C}_n$ and $\mathbf{A}_n$ have uniformly bounded absolute row and column sums and that $n^{-1}\sum_{i=1}^n \varepsilon_{i,n} = O_p(1)$.

I assume that the row and column sums of $\mathbf{C}_N$ are uniformly bounded in absolute value. Given Lemma C4, Assumption 4 in that paper holds and hence ψ_N is by their proof $o_p(1)$. Note that by Assumption 3, $\mathbf{C}_N(\mathbf{I}_T \otimes \mathbf{P}_N)$ has uniformly bounded absolute row and column sums. Instead of $\varepsilon_{i,n}$, I consider the random variables $v_{it,N} = \varepsilon_{it,N} + \mu_{i,N}$. By the triangle inequality

$$\begin{aligned}(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T |v_{ij,N}| &\leq (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T |\varepsilon_{it,N}| + (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \left|\mu_{i,N}\right| \\ &= (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T |\varepsilon_{it,N}| + N^{-1}\sum_{i=1}^N \left|\mu_{i,N}\right|. \quad \text{(C.2.10)}\end{aligned}$$

Since by Assumption 1, the random variables $\varepsilon_{it,N}$ are independent with uniformly bounded second moments, it follows that $(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T |\varepsilon_{it,N}| = O_p(1)$. Similarly, by Assumption 2, the random variables $\mu_{i,N}$ are independent with uniformly bounded second moments, and hence it follows that $N^{-1}\sum_{i=1}^N \left|\mu_{i,N}\right| =$

$O_p(1)$. As a result $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}|v_{ij,N}| = O_p(1)$. Hence the proof that $\phi_N = o_p(1)$ in Kelejian and Prucha (2005) also applies for the structure considered in this thesis and $\varphi_N = o_p(1)$. ■

Proof of Theorem 2: Given Lemma C6, the proof is identical to the proof of Theorem 2 in Kapoor et al. (2001). ■

C.3 Proofs for Section 4.3

Proof of Lemma 4: The dependent variable can be expressed as in equation (3.2.4):

$$\begin{aligned}
\mathbf{y}_{t,N} &= \sum_{j=0}^{\infty}\phi^j\left[\mathbf{X}_{t-j,N}\boldsymbol{\beta}+\mathbf{u}_{t-j,N}\right] \qquad \text{(C.3.1)}\\
&= \sum_{j=0}^{\infty}\phi^j\mathbf{X}_{t-j,N}\boldsymbol{\beta}+\sum_{j=0}^{t-1}\phi^j\mathbf{u}_{t-j,N}+\sum_{j=0}^{\infty}\phi^{t+j}\mathbf{u}_{-j,N}\\
&= \left(\sum_{j=0}^{\infty}\phi^j\mathbf{X}_{t-j,N}\boldsymbol{\beta}\right)+\mathbf{P}_N\left(\sum_{j=0}^{t-1}\phi^j\boldsymbol{\varepsilon}_{t-j,N}+\sum_{j=0}^{\infty}\phi^{t+j}\boldsymbol{\varepsilon}_{-j,N}\right)+\frac{\mathbf{P}_N\mu}{1-\phi}\\
&= \left(\sum_{j=0}^{\infty}\phi^j\mathbf{X}_{t-j,N}\boldsymbol{\beta}\right)+\left[\left(\frac{1}{1-\phi},1,\phi^{t-1},..,\phi^0,\mathbf{0}_{1\times(T-t)}\right)\otimes\mathbf{P}_N\right]\boldsymbol{\eta}_N.
\end{aligned}$$

Hence we can define

$$\mathbf{a}_{t,N}=\sum_{j=0}^{\infty}\phi^j\mathbf{X}_{t-j,N}\boldsymbol{\beta}, \qquad \text{(C.3.2)}$$

and

$$\mathbf{b}_t=\left(\frac{1}{1-\phi},1,\phi^{t-1},..,\phi^0,\mathbf{0}_{1\times(T-t)}\right). \qquad \text{(C.3.3)}$$

Given Assumptions 5 and 6, we have that $\sum_{j=0}^{\infty} \phi^j \mathbf{X}_{t-j,N} \boldsymbol{\beta}$ is uniformly bounded in absolute value and hence the sequence of vectors $\mathbf{a}_{t,N}$ has elements uniformly bounded in absolute value. Note that the elements (as well as dimensions) of $\mathbf{b}_t$ do not depend on N, and hence they are trivially uniformly bounded in absolute value. ■

Proof of Lemma 5: The claim is a consequence of Theorem A1 in Appendix A. I now verify that its conditions are met. As in equation (4.3.10), we have that the elements of $\mathbf{H}'_N \Delta \mathbf{u}_N$ are quadratic forms in the innovations:

$$\mathbf{h}'_{rt,N} \Delta \mathbf{u}_{t,N} = \mathbf{a}'_{rt,N} \left(\mathbf{d}_t \otimes \mathbf{P}_N\right) \boldsymbol{\eta}_N + \boldsymbol{\eta}'_N \left(\mathbf{b}'_{rt} \mathbf{d}_t \otimes \mathbf{P}'_N \mathbf{P}_N\right) \boldsymbol{\eta}_N, \qquad \text{(C.3.4)}$$

where $\boldsymbol{\eta}_N = \left(\boldsymbol{\mu}'_N, \boldsymbol{\xi}'_N, \boldsymbol{\varepsilon}'_{1,N}, ..., \boldsymbol{\varepsilon}'_{T,N}\right)$, with $\boldsymbol{\xi}_N = \sum_{j=0}^{\infty} \phi^j \boldsymbol{\varepsilon}_{-j,N}$. By Assumptions 1 and 6 it follows from Lemma B1 in Appendix B that the random variable $\boldsymbol{\xi}_N$ satisfies condition A3 in Appendix A. Therefore, by Assumptions 1 and 2, the innovations $\boldsymbol{\eta}_N$ satisfy conditions A1 and A3 in Appendix A. The Lemma stipulates that the vectors $\mathbf{a}_{rt,N}$ have elements uniformly bounded in absolute value. Observe that by Assumption 3, the matrix $(\mathbf{d}_t \otimes \mathbf{P}_N)$ has row sums uniformly bounded in absolute value and hence the vector $\mathbf{a}'_{rt,N} \left(\mathbf{d}_t \otimes \mathbf{P}_N\right)$ has elements uniformly bounded in absolute value and thus satisfies condition A2 in Appendix A. Furthermore, given that the dimensions and elements of $\mathbf{b}'_{rt} \mathbf{d}_t$ do not change with N, we have that Assumption 3 implies that the matrix $\left(\mathbf{b}'_{rt} \mathbf{d}_t \otimes \mathbf{P}'_N \mathbf{P}_N\right)$ fulfills condition A2 as well. Finally, $\left[(T-1) N\right]^{-1} \lambda_{\min} \left(\mathbf{V}_N\right) \geq c > 0$ is a condition stipulated in the Lemma. ■

Proof of Lemma 6: Substituting the model (equation 4.3.1) into the definition of the GMM estimator in (4.3.5) leads to:

$$\begin{aligned}
& \sqrt{(T-1)N}\left(\widetilde{\boldsymbol{\theta}}_N-\boldsymbol{\theta}\right) \qquad \text{(C.3.5)}\\
= & \sqrt{(T-1)N}\left[\Delta\mathbf{Z}_N'\mathbf{H}_N\mathbf{A}_N^{-1}\mathbf{H}_N'\Delta\mathbf{Z}_N\right]^{-1} * \\
& \Delta\mathbf{Z}_N'\mathbf{H}_N\mathbf{A}_N^{-1}\mathbf{H}_N'\Delta\mathbf{u}_N \\
= & \left[\frac{\Delta\mathbf{Z}_N'\mathbf{H}_N}{(T-1)N}\left(\frac{\mathbf{A}_N}{(T-1)N}\right)^{-1}\frac{\mathbf{H}_N'\Delta\mathbf{Z}_N}{(T-1)N}\right]^{-1} * \\
& \frac{\Delta\mathbf{Z}_N'\mathbf{H}_N}{(T-1)N}\left(\frac{\mathbf{A}_N}{(T-1)N}\right)^{-1}\frac{\mathbf{H}_N'\Delta\mathbf{u}_N}{\sqrt{(T-1)N}}.
\end{aligned}$$

By assumption in the lemma we have that $\mathbf{V}_N^{-1/2}\mathbf{H}_N'\Delta\mathbf{u}_N \xrightarrow{d} N(\mathbf{0},\mathbf{I}_k)$ with $[(T-1)N]^{-1}\mathbf{V}_N \xrightarrow{p} \mathbf{V}$ finite. Hence by Corollary 5 in Pötcher and Prucha (2001), we have

$$\left(\frac{\mathbf{V}_N}{(T-1)N}\right)^{1/2}\mathbf{V}_N^{-1/2}\mathbf{H}_N'\Delta\mathbf{u}_N = \frac{\mathbf{H}_N'\Delta\mathbf{u}_N}{\sqrt{(T-1)N}} \xrightarrow{d} N(\mathbf{0},\mathbf{V}). \qquad \text{(C.3.6)}$$

Furthermore, the lemma assumes that

$$\begin{aligned}
\frac{\Delta\mathbf{Z}_N'\mathbf{H}_N}{(T-1)N} &\xrightarrow{p} \mathbf{M}_{\Delta ZH}, \qquad \text{(C.3.7)}\\
\frac{\mathbf{A}_N}{(T-1)N} &\xrightarrow{p} \mathbf{A},
\end{aligned}$$

where $\mathbf{M}_{\Delta ZH}$ is finite with full column rank and $\mathbf{A}$ is finite and invertible. Hence, by Corollary 5 in Pötcher and Prucha (2001), we have the desired result. ■

Proof of Theorem 3: Observe that the instruments collected in $\widetilde{\mathbf{H}}_N$ consist of $\mathbf{y}_{t,N}$ and columns of $\mathbf{X}_{t,N}$ and hence by Lemma 4 are linear forms of the innovations of the form assumed in Lemma 5 and satisfy its conditions. Below I verify that $[(T-1)N]^{-1}\widetilde{\mathbf{V}}_N$ has the smallest eigenvalue uniformly bounded away from zero. This will complete verification of conditions of Lemma 5 and hence we will have that $\widetilde{\mathbf{V}}_N^{-1/2}\widetilde{\mathbf{H}}_N'\Delta\mathbf{u}_N \xrightarrow{d} N\left(\mathbf{0},\widetilde{\mathbf{V}}\right)$.

Observe that using the expression $\widetilde{\mathbf{H}}_N = \widetilde{\mathbf{S}}_N + \boldsymbol{\Upsilon}_N$, where $\widetilde{\mathbf{S}}_N$ is the nonstochastic part of the instruments (see Section 4.3.3), we have

$$\begin{aligned}
[(T-1)N]^{-1}\widetilde{\mathbf{V}}_N &= [(T-1)N]^{-1}E\left(\widetilde{\mathbf{H}}_N'\Delta\mathbf{u}_N\Delta\mathbf{u}_N'\widetilde{\mathbf{H}}_N\right) \qquad \text{(C.3.8)}\\
&= [(T-1)N]^{-1}E\left[\left(\widetilde{\mathbf{S}}_N' + \boldsymbol{\Upsilon}_N'\right)\Delta\mathbf{u}_N\Delta\mathbf{u}_N'\left(\widetilde{\mathbf{S}}_N + \boldsymbol{\Upsilon}_N\right)\right]\\
&= [(T-1)N]^{-1}\left(\widetilde{\mathbf{V}}_{1,N} + \widetilde{\mathbf{V}}_{2,N} + \widetilde{\mathbf{V}}_{3,N} + \widetilde{\mathbf{V}}_{4,N}\right),
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\mathbf{V}}_{1,N} &= \widetilde{\mathbf{S}}_N'E\left(\Delta\mathbf{u}_N\Delta\mathbf{u}_N'\right)\widetilde{\mathbf{S}}_N \qquad \text{(C.3.9)}\\
\widetilde{\mathbf{V}}_{2,N} &= \widetilde{\mathbf{S}}_N'E\left(\Delta\mathbf{u}_N\Delta\mathbf{u}_N'\boldsymbol{\Upsilon}_N\right)\\
\widetilde{\mathbf{V}}_{3,N} &= E\left(\boldsymbol{\Upsilon}_N'\Delta\mathbf{u}_N\Delta\mathbf{u}_N'\right)\widetilde{\mathbf{S}}_N\\
\widetilde{\mathbf{V}}_{4,N} &= E\left(\boldsymbol{\Upsilon}_N'\Delta\mathbf{u}_N\Delta\mathbf{u}_N'\boldsymbol{\Upsilon}_N\right).
\end{aligned}$$

In the following I show that the smallest eigenvalue of $[(T-1)N]^{-1}\widetilde{\mathbf{V}}_{1,N}$ is uniformly bounded away from zero. I also show that $\widetilde{\mathbf{V}}_{2,N} = \mathbf{0}$, and $\widetilde{\mathbf{V}}_{3,N} = \mathbf{0}$. Since the eigenvalues of $\widetilde{\mathbf{V}}_{4,N}$ are nonnegative it then follows from Lemma C1 that the smallest eigenvalue of $[(T-1)N]^{-1}\widetilde{\mathbf{V}}_N$ is uniformly bounded away from zero.

Using

$$\Delta\mathbf{u}_N = \left[\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)\otimes\mathbf{P}_N\right]\boldsymbol{\eta}_N, \tag{C.3.10}$$

where as in (4.1.15) $E\left(\boldsymbol{\eta}_N\boldsymbol{\eta}_N'\right) = \left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\otimes\mathbf{I}_N\right)$, it follows that

$$\widetilde{\mathbf{V}}_{1,N} = \widetilde{\mathbf{S}}_N'\left[\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)\boldsymbol{\Sigma}_{\eta,N}\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)'\otimes\mathbf{P}_N\mathbf{P}_N'\right]\widetilde{\mathbf{S}}_N. \tag{C.3.11}$$

By Lemma C2 the smallest eigenvalue of $\widetilde{\mathbf{V}}_{1,N}$ is then

$$\begin{aligned}
\lambda_{\min}\left(\widetilde{\mathbf{V}}_{1,N}\right) &\geq \lambda_{\min}\left(\widetilde{\mathbf{S}}_N'\widetilde{\mathbf{S}}_N\right)* \qquad\qquad (C.3.12)\\
&\quad\lambda_{\min}\left[\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)\boldsymbol{\Sigma}_{\eta,N}\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)'\otimes\left(\mathbf{P}_N\mathbf{P}_N'\right)\right]\\
&= \lambda_{\min}\left(\widetilde{\mathbf{S}}_N'\widetilde{\mathbf{S}}_N\right)\cdot\lambda_{\min}\left[\left(\mathbf{D}\boldsymbol{\Sigma}_{\boldsymbol{\eta},\mathbf{N}}\mathbf{D}'\right)\otimes\left(\mathbf{P}_N\mathbf{P}_N'\right)\right]\\
&= \lambda_{\min}\left(\widetilde{\mathbf{S}}_N'\widetilde{\mathbf{S}}_N\right)\cdot\lambda_{\min}\left(\mathbf{D}\boldsymbol{\Sigma}_{\boldsymbol{\eta},\mathbf{N}}\mathbf{D}'\right)\cdot\lambda_{\min}\left(\mathbf{P}_N\mathbf{P}_N'\right)\\
&= \lambda_{\min}\left(\widetilde{\mathbf{S}}_N'\widetilde{\mathbf{S}}_N\right)\cdot\lambda_{\min}\left(\mathbf{D}\mathbf{D}'\right)\cdot\lambda_{\min}\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},\mathbf{N}}\right)\cdot\lambda_{\min}\left(\mathbf{P}_N\mathbf{P}_N'\right),,
\end{aligned}$$

where I also used Theorem 4.2.12 in Horn and Johnson (1991). Observe that from the definition of the first difference operator matrix $\mathbf{D}$ (see 4.1.14), it follows that $\mathbf{D}\mathbf{D}' = 2\mathbf{I}_{T-1}$ and hence $\lambda_{\min}\left(\mathbf{D}\mathbf{D}'\right) = 2$. Since $\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}$ is diagonal, we have $\lambda_{\min}\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\right) = \min\left[\sigma_\mu^2, var\left(\boldsymbol{\xi}_{i,N}\right),\sigma_\varepsilon^2\right] = \min\left[\sigma_\mu^2,\frac{\sigma_\varepsilon^2}{1-\phi^2},\sigma_\varepsilon^2\right] \geq c_\Sigma > 0$. By Assumption 4 we have that $\lambda_{\min}\left(\mathbf{P}_N\mathbf{P}_N'\right)\geq c_P > 0$ and, therefore

$$\lambda_{\min}\left(\widetilde{\mathbf{V}}_{1,N}\right)\geq 2c_\Sigma c_P\lambda_{\min}\left(\widetilde{\mathbf{S}}_N'\widetilde{\mathbf{S}}_N\right). \tag{C.3.13}$$

From Assumption GMM1 we have that $\lambda_{\min}\left(\left[(T-1)N\right]^{-1}\widetilde{\mathbf{S}}'_{t,N}\widetilde{\mathbf{S}}_{t,N}\right)\geq c_S>0$ and hence

$$\left[(T-1)N\right]^{-1}\lambda_{\min}\left(\widetilde{\mathbf{V}}_{1,N}\right)\geq 2c_{\Sigma}c_P c_S>0. \tag{C.3.14}$$

Next, I show that $\widetilde{\mathbf{V}}_{2,N}$ and $\widetilde{\mathbf{V}}_{3,N}$ are matrices of zeros. Recall that $\boldsymbol{\Upsilon}_N$ consists of blocks $\boldsymbol{\Upsilon}_{t,N}$ on the main diagonal and zeros elsewhere. Thus $\boldsymbol{\Upsilon}'_N\Delta\mathbf{u}_N\Delta\mathbf{u}'_N$ consists of blocks $\boldsymbol{\Upsilon}'_{t,N}\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}'_{t,N}$ on the main diagonal and zeros elsewhere. Observe that

$$\boldsymbol{\Upsilon}_{t,N}=\left[\left((\mathbf{b}_{t-2},...,\mathbf{b}_0)\otimes\mathbf{P}_N\right)\left(\mathbf{I}_{t-1}\otimes\boldsymbol{\eta}_N\right),\mathbf{0}_{N\times tp}\right], \tag{C.3.15}$$

and thus

$$\boldsymbol{\Upsilon}'_{t,N}\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}'_{t,N}=\begin{pmatrix}\boldsymbol{\eta}'_N\left(\mathbf{b}'_0\otimes\mathbf{P}'_N\right)\\ \vdots\\ \boldsymbol{\eta}'_N\left(\mathbf{b}'_{t-2}\otimes\mathbf{P}'_N\right)\\ \mathbf{0}_{tp\times N}\end{pmatrix}\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}'_{t,N}. \tag{C.3.16}$$

Observe that $\Delta\mathbf{u}_{t,N}=\left(\mathbf{d}_t\otimes\mathbf{P}_N\right)\boldsymbol{\eta}_N$ (as in 4.3.9) and thus

$$\boldsymbol{\eta}'_N\left(\mathbf{b}'_{t-s}\otimes\mathbf{P}'_N\right)\Delta\mathbf{u}_{t,N}=\boldsymbol{\eta}'_N\left(\mathbf{b}'_{t-s}\mathbf{d}_t\otimes\mathbf{P}'_N\mathbf{P}_N\right)\boldsymbol{\eta}_N, \tag{C.3.17}$$

where $\mathbf{d}_t$ is a $(t+1)-th$ row of $\left(\mathbf{0}_{(T-1)\times 2},\mathbf{D}\right)$, with the $(T-1)\times T$ matrix $\mathbf{D}$ is defined in (4.1.14). Hence the $1\times(T+2)$ vector $\mathbf{d}_t$ is a row vector with zeros in the first t positions. Furthermore, the $1\times(T+2)$ vector $\mathbf{b}_{t-s}$ (defined in the proof of Lemma 4 above) has zero entries starting from position $(t-2+s)$. As a result, for $s>1$, the product $\mathbf{b}'_{t-s}\mathbf{d}_t$ is a $(T+2)\times(T+2)$ matrix with zeros on the main diagonal. Hence $\boldsymbol{\eta}'_N\left(\mathbf{b}'_{t-s}\otimes\mathbf{P}'_N\right)\Delta\mathbf{u}_{t,N}$ is a quadratic form in the innovations $\boldsymbol{\eta}_N$ with zeros on the main diagonal (and no linear component). Each element of $\Delta\mathbf{u}_{t,N}$ is a linear form in innovations $\boldsymbol{\eta}_N$ and hence can also be treated as a linear-quadratic form in $\boldsymbol{\eta}_N$ where the matrix defining the quadratic component consists of zeros. As a result, we can apply Lemma A1 in Appendix A to obtain that the covariance of $\boldsymbol{\eta}'_N\left(\mathbf{b}'_{t-s}\otimes\mathbf{P}'_N\right)\Delta\mathbf{u}_{t,N}$ and $\Delta u_{it,N}$ is zero. Thus it follows that

$$E\left[\boldsymbol{\eta}'_N\left(\mathbf{b}'_{t-s}\otimes\mathbf{P}'_N\right)\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}'_{t,N}\right]=0, \tag{C.3.18}$$

where $s>1$, implying that $E\left(\boldsymbol{\Upsilon}'_N\Delta\mathbf{u}_N\Delta\mathbf{u}'_N\right)$ is a matrix of zeros. As a consequence

$$\widetilde{\mathbf{V}}_{2,N}=E\left(\boldsymbol{\Upsilon}'_N\Delta\mathbf{u}_N\Delta\mathbf{u}'_N\right)\widetilde{\mathbf{S}}_N=\mathbf{0}_{k\times k}. \tag{C.3.19}$$

The same argument implies that $\widetilde{\mathbf{V}}_{3,N}$ is a matrix of zeros. Finally, observe that the matrix $\widetilde{\mathbf{V}}_{4,N}$ is itself a variance covariance matrix (i.e. symmetric positive semidefinite) and thus it has non-negative eigenvalues.

This completes the verification of the conditions of Lemma 5 and hence we have that $\widetilde{\mathbf{V}}_N^{-1/2}\widetilde{\mathbf{H}}'_N\Delta\mathbf{u}_N\xrightarrow{d}N\left(\mathbf{0},\widetilde{\mathbf{V}}\right)$. We can now write the estimator as

$$\widetilde{\boldsymbol{\theta}}_N=\boldsymbol{\theta}+\left[\Delta\mathbf{Z}'_N\widetilde{\mathbf{H}}_N\widetilde{\mathbf{V}}_N^{-1}\widetilde{\mathbf{H}}'_N\Delta\mathbf{Z}_N\right]^{-1}\Delta\mathbf{Z}'_N\widetilde{\mathbf{H}}_N\widetilde{\mathbf{V}}_N^{-1}\widetilde{\mathbf{H}}'_N\Delta\mathbf{u}_N, \tag{C.3.20}$$

where by Assumptions GMM2 and GMM3,

$$p\lim_{N\to\infty}\frac{1}{(T-1)\,N}\Delta\mathbf{Z}'_N\widetilde{\mathbf{H}}_N=\widetilde{\mathbf{M}}_{H\Delta Z}, \tag{C.3.21}$$

and

$$p\lim_{N\to\infty}\frac{1}{(T-1)\,N}\widetilde{\mathbf{V}}_N=\widetilde{\mathbf{V}}. \tag{C.3.22}$$

Therefore by Lemma 6, the estimator converges in distribution with

$$\sqrt{(T-1)\,N}\left(\widetilde{\boldsymbol{\theta}}_N-\boldsymbol{\theta}\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Psi}\right), \tag{C.3.23}$$

where

$$\begin{aligned}\boldsymbol{\Psi} &= \left(\widetilde{\mathbf{M}}_{\Delta ZH}\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{M}}'_{\Delta ZH}\right)^{-1}\widetilde{\mathbf{M}}_{\Delta ZH}\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{V}}\widetilde{\mathbf{V}}^{-\mathbf{1}}\widetilde{\mathbf{M}}_{\Delta ZH}\,* \\ &\quad\left(\widetilde{\mathbf{M}}_{\Delta ZH}\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{M}}'_{\Delta ZH}\right)^{-1} \\ &= \left(\widetilde{\mathbf{M}}_{\Delta ZH}\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{M}}'_{\Delta ZH}\right)^{-1}=\boldsymbol{\Psi},\end{aligned} \tag{C.3.24}$$

which is the claim in the Theorem. ■

To prove Lemma 7, I will use Lemma C.6 in Kelejian and Prucha (2005). For convenience of the reader, I restate a simplified version of that lemma:

Lemma C7 *Let $\mathbf{a}_n$ and $\mathbf{b}_n$ be sequences of $n\times 1$ vectors and let $\mathbf{W}_n$ be a sequence of $n\times n$ matrices. Assume that the vectors $\mathbf{a}_n$ and $\mathbf{b}_n$ have elements uniformly bounded in absolute value and that the matrices $(r\mathbf{W}_n)$ have row and* *column sums uniformly bounded in absolute value for $r < 1$ by one and some finite constant respectively. Consider a sequence of random variables $\widetilde{\rho}_n$ converging in probability to ρ as $n \to \infty$, where $|\rho| < r$. Denote $\mathbf{P}_n(r) = (\mathbf{I}_n - r\mathbf{W}_n)^{-1}$. Then*

$$n^{-1}\mathbf{a}_n'\mathbf{P}_n(\rho)'\mathbf{P}_n(\rho)\mathbf{b}_n - n^{-1}\mathbf{a}_n'\mathbf{P}_n(\widetilde{\rho}_n)'\mathbf{P}_n(\widetilde{\rho}_n)\mathbf{b}_n = o_p(1), \tag{C.3.25}$$

and

$$\begin{aligned} &n^{-1}tr\left[\mathbf{P}_n(\rho)'\mathbf{P}_n(\rho)\mathbf{P}_n(\rho)'\mathbf{P}_n(\rho)\right] \\ &-n^{-1}tr\left[\mathbf{P}_n(\widetilde{\rho}_n)'\mathbf{P}_n(\widetilde{\rho}_n)\mathbf{P}_n(\widetilde{\rho}_n)'\mathbf{P}_n(\widetilde{\rho}_n)\right] = o_p(1). \end{aligned} \tag{C.3.26}$$

Proof: The proof of the first claim follows from Lemma C.6 in Kelejian and Prucha (2005) by choosing (in their notation) $\mathbf{\Sigma}_n = \widetilde{\widetilde{\mathbf{\Sigma}}}_n = \mathbf{I}_n$ and $\mathbf{H}_n = (\mathbf{a}_n, \mathbf{b}_n)$.

The second claim is not a direct consequence of the Lemma C.6, however, its proof follows the same structure. Denote

$$\begin{aligned} v_n \;=\; & n^{-1}tr\left[\mathbf{P}_n(\rho)'\mathbf{P}_n(\rho)\mathbf{P}_n(\rho)'\mathbf{P}_n(\rho)\right] \\ & -n^{-1}tr\left[\mathbf{P}_n(\widetilde{\rho}_n)'\mathbf{P}_n(\widetilde{\rho}_n)\mathbf{P}_n(\widetilde{\rho}_n)'\mathbf{P}_n(\widetilde{\rho}_n)\right]. \end{aligned} \tag{C.3.27}$$

Using the same argument as on p.39 in Kelejian and Prucha (2005), it follows that for every subsequence (n_m) there exists a subsequence (n_m') such that for $\omega \in A$, $P(A) = 1$, there is critical index N_ω such that for all $n_m' \geq N_\omega : \; \left|\widetilde{\rho}_{n_m'}(\omega)\right| \leq r_*$, where $r_* = (r + |\rho|)/2$. Furthermore, it also follows from the argument on the same page that for $n_m' \geq N_\omega$ the row sums of $\widetilde{\rho}_{n_m'}(\omega)\,\mathbf{W}_n$ are less than unity in absolute value and that $\left[\mathbf{I}_{n_m'} - \widetilde{\rho}_{n_m'}(\omega)\,\mathbf{W}_n\right]$ and $\left[\mathbf{I}_{n_m'} - \rho_{n_m'}\mathbf{W}_n\right]$ are invertible with

$$\begin{aligned}\left[\mathbf{I}_{n_m'} - \widetilde{\rho}_{n_m'}(\omega)\,\mathbf{W}_n\right]^{-1} &= \sum_{l=1}^{\infty}\left[\widetilde{\rho}_{n_m'}(\omega)\right]^l \mathbf{W}_n^l, \\ \left[\mathbf{I}_{n_m'} - \rho_{n_m'}\mathbf{W}_n\right]^{-1} &= \sum_{l=1}^{\infty}\left[\rho_{n_m'}\right]^l \mathbf{W}_n^l.\end{aligned} \tag{C.3.28}$$

Hence we have that

$$\begin{aligned} v_{n_m'}(\omega) &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=1}^{\infty}\left(\rho_{n_m'}\right)^{k+l+p+q}\left(\mathbf{W}_n'\right)^k\left(\mathbf{W}_n\right)^l\left(\mathbf{W}_n'\right)^p\left(\mathbf{W}_n\right)^q \\ &\quad -\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=1}^{\infty}\left[\widetilde{\rho}_{n_m'}(\omega)\right]^{k+l+p+q}\left(\mathbf{W}_n'\right)^k\left(\mathbf{W}_n\right)^l\left(\mathbf{W}_n'\right)^p\left(\mathbf{W}_n\right)^q \\ &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=1}^{\infty}\chi_{n_m'}^{(k,l,p,q)}(\omega), \end{aligned} \tag{C.3.29}$$

where

$$\chi_{n_m'}^{(k,l,p,q)}(\omega) = \left[\frac{\rho^{k+l+p+q} - \widetilde{\rho}_{n_m'}^{k+l+p+q}(\omega)}{r^{k+l+p+q}}\right]\varkappa_{n_m'}^{(k,l,p,q)},$$

with

$$\varkappa_{n_m'}^{(k,l,p,q)} = \left(n_m'\right)^{-1} tr\left[r^{k+l+p+q}\left(\mathbf{W}_{n_m'}'\right)^k\left(\mathbf{W}_{n_m'}\right)^l\left(\mathbf{W}_{n_m'}'\right)^p\left(\mathbf{W}_{n_m'}\right)^q\right]. \tag{C.3.30}$$

Given that the row and column sums of the matrix $r\mathbf{W}_n$ are uniformly bounded in absolute value by one and some finite constant respectively, it follows that

$\varkappa_{n'_m}^{(k,l,p,q)} = O(1)$. Furthermore, observe that

$$\frac{\left|\rho^{k+l+p+q} - \widetilde{\rho}_{n'_m}^{k+l+p+q}\right|}{r^{k+l+p+q}} \leq 2\left(\frac{r_*}{r}\right)^{k+l+p+q}, \tag{C.3.31}$$

(see Kelejian and Prucha, 2005, p. 40) and hence $\left|\chi_{n'_m}^{(k,l,p,q)}(\omega)\right| \leq B^{(k,l,p,q)} = 2c(r_*/r)^{k+l+p+q}$ where c is the uniform bound for $\left|\varkappa_{n'_m}^{(k,l,p,q)}\right|$. Since $r_*/r < 1$ by construction, clearly $\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=1}^{\infty} B^{(k,l,p,q)} < \infty$. By dominated convergence it follows that $\upsilon_{n'_m}(\omega) \to 0$ as $n'_m \to \infty$, and as a result $\upsilon_n \to \infty$ by the subsequence argument (Kelejian and Prucha, 2005, p. 39; Gänsler and Slute, 1977, pp. 61-62). ■

Proof of Lemma 7: Recall that based on the expression for the covariance of the quadratic forms in $\widetilde{\mathbf{V}}_N$ and $\widehat{\mathbf{V}}_N$, the elements of the first diagonal block of $\widetilde{\mathbf{V}}_{ts,N} - \widehat{\mathbf{V}}_{ts,N}$ are (see 4.3.34):

$$\begin{aligned}
\widetilde{\upsilon}^{\mathbf{y}}_{qr,ts,N} - \widehat{\upsilon}^{\mathbf{y}}_{qr,ts,N} &= \mathbf{a}'_{t-1-q,N}\left(\mathbf{d}_t\mathbf{\Sigma}_{\boldsymbol{\eta},\mathbf{N}}\mathbf{d}'_s \otimes \mathbf{P}_N\mathbf{P}'_N\right)\mathbf{a}_{s-1-r,N} \\
&\quad -\widehat{\mathbf{a}}'_{t-1-q,N}\left(\mathbf{d}_t\widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N}\mathbf{d}'_s \otimes \widehat{\mathbf{P}}_N\widehat{\mathbf{P}}'_N\right)\widehat{\mathbf{a}}_{s-1-r,N} \\
&\quad +2tr\left(\mathbf{b}'_{t-1-q}\mathbf{d}_t\mathbf{\Sigma}_{\boldsymbol{\eta},\mathbf{N}}\mathbf{d}'_s\mathbf{b}_{s-1-r}\mathbf{\Sigma}_{\boldsymbol{\eta},N} \otimes \mathbf{P}'_N\mathbf{P}_N\mathbf{P}'_N\mathbf{P}_N\right) \\
&\quad -2tr\left(\widehat{\mathbf{b}}'_{t-1-q,N}\mathbf{d}_t\widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N}\mathbf{d}'_s\widehat{\mathbf{b}}_{s-1-r,N}\widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N} \otimes \widehat{\mathbf{P}}'_N\widehat{\mathbf{P}}_N\widehat{\mathbf{P}}'_N\widehat{\mathbf{P}}_N\right).
\end{aligned} \tag{C.3.32}$$

Note that from (C.3.2) and since the lemma assumes $\sum_{k=0}^{-\infty}\phi^k\mathbf{X}_{-k,N}\boldsymbol{\beta} = \mathbf{0}$, it follows that

$$\begin{aligned}\left(\mathbf{a}'_{t-1-q,N}-\widehat{\mathbf{a}}'_{t-1-q,N}\right) &= \sum_{k=0}^{t-2-q}\left(\phi^k\boldsymbol{\beta}'-\widehat{\phi}^k\widehat{\boldsymbol{\beta}}'_N\right)\mathbf{X}'_{t-1-q-k,N}, \quad \text{(C.3.33)}\\ \left(\mathbf{a}_{s-1-r,N}-\widehat{\mathbf{a}}_{s-1-r,N}\right) &= \sum_{k=0}^{s-2-r}\mathbf{X}_{s-1-r-k,N}\left(\phi^k\boldsymbol{\beta}-\widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N\right).\end{aligned}$$

Since $\mathbf{d}_t\boldsymbol{\Sigma}_{\eta,\mathbf{N}}\mathbf{d}'_s$ is a scalar and we can then rearrange the above expression as[50]

$$\widetilde{v}^{\mathbf{y}}_{qr,ts,N}-\widehat{v}^{\mathbf{y}}_{qr,ts,N}=\sum_{m=1}^{18}v^{y}_{m,N}, \tag{C.3.34}$$

[50]I use the following, rather tedious algebraic rule: let a,b,c,d and $\widehat{a},\widehat{b},\widehat{c},\widehat{d}$ be matrices (and/or scalars or vectors) of conformable dimensions. It is then easy to verify that:

$$\begin{aligned} abcd-\widehat{a}\widehat{b}\widehat{c}\widehat{d} &= (a-\widehat{a})\,bcd+a\left(b-\widehat{b}\right)cd+ab\,(c-\widehat{c})\,d+abc\left(d-\widehat{d}\right)\\ &\quad-(a-\widehat{a})\left(b-\widehat{b}\right)cd-(a-\widehat{a})\,b\,(c-\widehat{c})\,d-(a-\widehat{a})\,bc\left(d-\widehat{d}\right)\\ &\quad-a\left(b-\widehat{b}\right)(c-\widehat{c})\,d-a\left(b-\widehat{b}\right)c\left(d-\widehat{d}\right)-ab\,(c-\widehat{c})\left(d-\widehat{d}\right)\\ &\quad+(a-\widehat{a})\left(b-\widehat{b}\right)(c-\widehat{c})\,d+(a-\widehat{a})\left(b-\widehat{b}\right)c\left(d-\widehat{d}\right)\\ &\quad+(a-\widehat{a})\,b\,(c-\widehat{c})\left(d-\widehat{d}\right)+a\left(b-\widehat{b}\right)(c-\widehat{c})\left(d-\widehat{d}\right)\\ &\quad-(a-\widehat{a})\left(b-\widehat{b}\right)(c-\widehat{c})\left(d-\widehat{d}\right),\end{aligned}$$

and

$$ab-\widehat{a}\widehat{b}=(a-\widehat{a})\,b+a\left(b-\widehat{b}\right)-(a-\widehat{a})\left(b-\widehat{b}\right).$$

where

$$
\begin{aligned}
v_{1,N}^{y} &= \mathbf{d}_t\left(\mathbf{\Sigma}_{\boldsymbol{\eta},\mathbf{N}} - \widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\left(\mathbf{a}_{t-1-q,N}'\mathbf{P}_N\mathbf{P}_N'\mathbf{a}_{s-1-r,N}\right), \qquad \text{(C.3.35)}\\
v_{2,N}^{y} &= \mathbf{d}_t\mathbf{\Sigma}_{\boldsymbol{\eta},\mathbf{N}}\mathbf{d}_s'\left(\mathbf{a}_{t-1-q,N}' - \widehat{\mathbf{a}}_{t-1-q,N}'\right)\mathbf{P}_N\mathbf{P}_N'\mathbf{a}_{s-1-r,N}\\
&= \mathbf{d}_t\mathbf{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}_s'\left[\sum_{k=0}^{t-2-q}\left(\phi^k\boldsymbol{\beta}' - \widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N'\right)\mathbf{X}_{t-1-q-k,N}'\right]\mathbf{P}_N\mathbf{P}_N'\mathbf{a}_{s-1-r,N},\\
v_{3,N}^{y} &= \mathbf{d}_t\mathbf{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}_s'\mathbf{a}_{t-1-q,N}'\left(\mathbf{P}_N\mathbf{P}_N' - \widehat{\mathbf{P}}_N\widehat{\mathbf{P}}_N'\right)\mathbf{a}_{s-1-r,N},\\
v_{4,N}^{y} &= \mathbf{d}_t\mathbf{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}_s'\mathbf{a}_{t-1-q,N}'\mathbf{P}_N\mathbf{P}_N'\left(\mathbf{a}_{s-1-r,N} - \widehat{\mathbf{a}}_{s-1-r,N}\right)\\
&= \mathbf{d}_t\mathbf{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}_s'\mathbf{a}_{t-1-q,N}'\mathbf{P}_N\mathbf{P}_N'\left[\sum_{k=0}^{s-2-r}\mathbf{X}_{s-1-r-k,N}\left(\phi^k\boldsymbol{\beta} - \widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
v_{5,N}^{y} &= -\mathbf{d}_t\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\left(\mathbf{a}_{t-1-q,N}' - \widehat{\mathbf{a}}_{t-1-q,N}'\right)\mathbf{P}_N\mathbf{P}_N'\mathbf{a}_{s-1-r,N}\\
&= -\mathbf{d}_t\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\left[\sum_{k=0}^{s-2-r}\mathbf{X}_{s-1-r-k,N}\left(\phi^k\boldsymbol{\beta} - \widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N\right)\right] *\\
&\quad \mathbf{P}_N\mathbf{P}_N'\mathbf{a}_{s-1-r,N},
\end{aligned}
$$

$$
v_{6,N}^{y} = -\mathbf{d}_t\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\mathbf{a}_{t-1-q,N}'\left(\mathbf{P}_N\mathbf{P}_N' - \widehat{\mathbf{P}}_N\widehat{\mathbf{P}}_N'\right)\mathbf{a}_{s-1-r,N},
$$

$$
\begin{aligned}
v_{7,N}^{y} &= -\mathbf{d}_t\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\mathbf{a}_{t-1-q,N}'\mathbf{P}_N\mathbf{P}_N'\left(\mathbf{a}_{s-1-r,N} - \widehat{\mathbf{a}}_{s-1-r,N}\right)\\
&= -\mathbf{d}_t\left(\mathbf{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\mathbf{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\mathbf{a}_{t-1-q,N}' *\\
&\quad \mathbf{P}_N\mathbf{P}_N'\left[\sum_{k=0}^{s-2-r}\mathbf{X}_{s-1-r-k,N}\left(\phi^k\boldsymbol{\beta} - \widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N\right)\right],
\end{aligned}
$$

$$\begin{aligned} v^y_{8,N} &= -\mathbf{d}_t \boldsymbol{\Sigma}_{\boldsymbol{\eta},N} \mathbf{d}'_s \left(\mathbf{a}'_{t-1-q,N} - \widehat{\mathbf{a}}'_{t-1-q,N}\right) \left(\mathbf{P}_N \mathbf{P}'_N - \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N\right) \mathbf{a}_{s-1-r,N} \\ &= -\mathbf{d}_t \boldsymbol{\Sigma}_{\boldsymbol{\eta},N} \mathbf{d}'_s \left[\sum_{k=0}^{t-2-q} \left(\phi^k \boldsymbol{\beta}' - \widehat{\phi}^k \widehat{\boldsymbol{\beta}}'_N\right) \mathbf{X}'_{t-1-q-k,N}\right] * \\ &\quad \left(\mathbf{P}_N \mathbf{P}'_N - \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N\right) \mathbf{a}_{s-1-r,N}, \end{aligned}$$

$$\begin{aligned} v^y_{9,N} &= -\mathbf{d}_t \boldsymbol{\Sigma}_{\boldsymbol{\eta},N} \mathbf{d}'_s \left(\mathbf{a}'_{t-1-q,N} - \widehat{\mathbf{a}}'_{t-1-q,N}\right) \mathbf{P}_N \mathbf{P}'_N \left(\mathbf{a}_{s-1-r,N} - \widehat{\mathbf{a}}_{s-1-r,N}\right) \\ &= -\mathbf{d}_t \boldsymbol{\Sigma}_{\boldsymbol{\eta},N} \mathbf{d}'_s \left[\sum_{k=0}^{t-2-q} \left(\phi^k \boldsymbol{\beta}' - \widehat{\phi}^k \widehat{\boldsymbol{\beta}}'_N\right) \mathbf{X}'_{t-1-q-k,N}\right] * \\ &\quad \mathbf{P}_N \mathbf{P}'_N \left[\sum_{k=0}^{s-2-r} \mathbf{X}_{s-1-r-k,N} \left(\phi^k \boldsymbol{\beta} - \widehat{\phi}^k \widehat{\boldsymbol{\beta}}_N\right)\right], \end{aligned}$$

$$\begin{aligned} v^y_{10,N} &= -\mathbf{d}_t \boldsymbol{\Sigma}_{\boldsymbol{\eta},N} \mathbf{d}'_s \mathbf{a}'_{t-1-q,N} \left(\mathbf{P}_N \mathbf{P}'_N - \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N\right) \left(\mathbf{a}_{s-1-r,N} - \widehat{\mathbf{a}}_{s-1-r,N}\right) \\ &= -\mathbf{d}_t \boldsymbol{\Sigma}_{\boldsymbol{\eta},N} \mathbf{d}'_s \mathbf{a}'_{t-1-q,N} * \\ &\quad \left(\mathbf{P}_N \mathbf{P}'_N - \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N\right) \left[\sum_{k=0}^{s-2-r} \mathbf{X}_{s-1-r-k,N} \left(\phi^k \boldsymbol{\beta} - \widehat{\phi}^k \widehat{\boldsymbol{\beta}}_N\right)\right], \end{aligned}$$

$$\begin{aligned} v^y_{11,N} &= -\mathbf{d}_t \left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right) \mathbf{d}'_s \left(\mathbf{a}'_{t-1-q,N} - \widehat{\mathbf{a}}'_{t-1-q,N}\right) * \\ &\quad \left(\mathbf{P}_N \mathbf{P}'_N - \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N\right) \mathbf{a}_{s-1-r,N} \\ &= -\mathbf{d}_t \left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right) \mathbf{d}'_s \left[\sum_{k=0}^{t-2-q} \left(\phi^k \boldsymbol{\beta}' - \widehat{\phi}^k \widehat{\boldsymbol{\beta}}'_N\right) \mathbf{X}'_{t-1-q-k,N}\right] * \\ &\quad \left(\mathbf{P}_N \mathbf{P}'_N - \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N\right) \mathbf{a}_{s-1-r,N}, \end{aligned}$$

$$
\begin{aligned}
v_{12,N}^{y} &= -\mathbf{d}_t\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}-\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\left(\mathbf{a}_{t-1-q,N}'-\widehat{\mathbf{a}}_{t-1-q,N}'\right)\\
&\quad\cdot\mathbf{P}_N\mathbf{P}_N'\left(\mathbf{a}_{s-1-r,N}-\widehat{\mathbf{a}}_{s-1-r,N}\right)\\
&= -\mathbf{d}_t\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}-\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\left[\sum_{k=0}^{t-2-q}\left(\phi^k\boldsymbol{\beta}'-\widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N'\right)\mathbf{X}_{t-1-q-k,N}'\right]\\
&\quad\cdot\mathbf{P}_N\mathbf{P}_N'\left(\mathbf{a}_{s-1-r,N}-\widehat{\mathbf{a}}_{s-1-r,N}\right),
\end{aligned}
$$

$$
\begin{aligned}
v_{13,N}^{y} &= -\mathbf{d}_t\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}-\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\mathbf{a}_{t-1-q,N}' *\\
&\quad\left(\mathbf{P}_N\mathbf{P}_N'-\widehat{\mathbf{P}}_N\widehat{\mathbf{P}}_N'\right)\left(\mathbf{a}_{s-1-r,N}-\widehat{\mathbf{a}}_{s-1-r,N}\right)\\
&= -\mathbf{d}_t\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}-\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}_s'\mathbf{a}_{t-1-q,N}' *\\
&\quad\left(\mathbf{P}_N\mathbf{P}_N'-\widehat{\mathbf{P}}_N\widehat{\mathbf{P}}_N'\right)\left[\sum_{k=0}^{s-2-r}\mathbf{X}_{s-1-r-k,N}\left(\phi^k\boldsymbol{\beta}-\widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
v_{14,N}^{y} &= -\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}_s'\left(\mathbf{a}_{t-1-q,N}'-\widehat{\mathbf{a}}_{t-1-q,N}'\right)\\
&\quad\cdot\left(\mathbf{P}_N\mathbf{P}_N'-\widehat{\mathbf{P}}_N\widehat{\mathbf{P}}_N'\right)\left(\mathbf{a}_{s-1-r,N}-\widehat{\mathbf{a}}_{s-1-r,N}\right)\\
&= -\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}_s'\left[\sum_{k=0}^{t-2-q}\left(\phi^k\boldsymbol{\beta}'-\widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N'\right)\mathbf{X}_{t-1-q-k,N}'\right]\\
&\quad\cdot\left(\mathbf{P}_N\mathbf{P}_N'-\widehat{\mathbf{P}}_N\widehat{\mathbf{P}}_N'\right)\left[\sum_{k=0}^{s-2-r}\mathbf{X}_{s-1-r-k,N}\left(\phi^k\boldsymbol{\beta}-\widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
v^y_{15,N} &= -\mathbf{d}_t \left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right) \mathbf{d}'_s \left(\mathbf{a}'_{t-1-q,N} - \widehat{\mathbf{a}}'_{t-1-q,N}\right) \\
&\quad \cdot \left(\mathbf{P}_N \mathbf{P}'_N - \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N\right) \left(\mathbf{a}_{s-1-r,N} - \widehat{\mathbf{a}}_{s-1-r,N}\right) \\
&= -\mathbf{d}_t \left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right) \mathbf{d}'_s \left[\sum_{k=0}^{t-2-q} \left(\phi^k \boldsymbol{\beta}' - \widehat{\phi}^k \widehat{\boldsymbol{\beta}}'_N\right) \mathbf{X}'_{t-1-q-k,N}\right] \\
&\quad \cdot \left(\mathbf{P}_N \mathbf{P}'_N - \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N\right) \left[\sum_{k=0}^{s-2-r} \mathbf{X}_{s-1-r-k,N} \left(\phi^k \boldsymbol{\beta} - \widehat{\phi}^k \widehat{\boldsymbol{\beta}}_N\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
v^y_{16,N} &= 2\left[tr\left(\mathbf{b}'_{t-1-q} \mathbf{d}_t \boldsymbol{\Sigma}_{\boldsymbol{\eta},N} \mathbf{d}'_s \mathbf{b}_{s-1-r} \boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\right)\right. \\
&\quad \left. -tr\left(\widehat{\mathbf{b}}'_{t-1-q,N} \mathbf{d}_t \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N} \mathbf{d}'_s \widehat{\mathbf{b}}_{s-1-r,N} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right)\right] \cdot tr\left(\mathbf{P}'_N \mathbf{P}_N \mathbf{P}'_N \mathbf{P}_N\right), \\
v^y_{17,N} &= 2tr\left(\mathbf{b}'_{t-1-q} \mathbf{d}_t \boldsymbol{\Sigma}_{\boldsymbol{\eta},N} \mathbf{d}'_s \mathbf{b}_{s-1-r} \boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\right) * \\
&\quad tr\left(\mathbf{P}'_N \mathbf{P}_N \mathbf{P}'_N \mathbf{P}_N - \widehat{\mathbf{P}}'_N \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N \widehat{\mathbf{P}}_N\right), \\
v^y_{18,N} &= -2\left[tr\left(\mathbf{b}'_{t-1-q} \mathbf{d}_t \boldsymbol{\Sigma}_{\boldsymbol{\eta},N} \mathbf{d}'_s \mathbf{b}_{s-1-r} \boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\right)\right. \\
&\quad \left. -tr\left(\widehat{\mathbf{b}}'_{t-1-q,N} \mathbf{d}_t \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N} \mathbf{d}'_s \widehat{\mathbf{b}}_{s-1-r,N} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right)\right] * \\
&\quad tr\left(\mathbf{P}'_N \mathbf{P}_N \mathbf{P}'_N \mathbf{P}_N - \widehat{\mathbf{P}}'_N \widehat{\mathbf{P}}_N \widehat{\mathbf{P}}'_N \widehat{\mathbf{P}}_N\right).
\end{aligned}
$$

Observe that for notational convenience I drop the dependence of the scalars $v^y_{m,N}$ on the values of the indexes q, r, s, t.

I now examine the nonstochastic elements of the scalars $v^y_{m,N}$. Note that the elements and dimensions of $\mathbf{d}_t$ and $\mathbf{d}'_s$ do not depend on N and hence they are trivially uniformly bounded in absolute value. The dimensions of $\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}$ (defined in 4.1.16) do not depend on N and its elements are uniformly bounded in absolute value by Assumptions 1, 2 and 6. I now show that the other nonstochastic components are uniformly bounded in absolute value when scaled by N^{-1}. Note that since $|\phi| < 1$, it follows from Assumption 5 that $\mathbf{a}_{t-1-q,N}$ as well as $\mathbf{a}_{s-1-r,N}$ have elements uniformly bounded in absolute value. By Assumption 3, the matrix $\mathbf{P}_N$ has row and column sums uniformly bounded in absolute value. As a result, it follows from Lemma C3 that $N^{-1}\mathbf{a}'_{t-1-q,N}\mathbf{P}_N\mathbf{P}'_N\mathbf{a}_{s-1-r,N}$ is uniformly bounded in absolute value. Similarly, given Assumptions 3 and 5, it follows from Lemma C3 that $N^{-1}\mathbf{X}'_{t-1-q-k,N}\mathbf{P}_N\mathbf{P}'_N\mathbf{a}_{s-1-r,N}$ (where $k = 0, ..t-1-q$) and $N^{-1}\mathbf{a}'_{t-1-q,N}\mathbf{P}_N\mathbf{P}'_N\mathbf{X}_{s-1-r-k,N}$ (where $k = 0, .., s-1-r$) have elements that are uniformly bounded in absolute value.

Next I show that the stochastic components of $v^y_{m,N}$ with dimensions that do not depend on N are $o_p(1)$. Recall that

$$\mathbf{\Sigma}_{\boldsymbol{\eta},N} = diag\left(\sigma^2_{\mu,N}, \frac{\sigma^2_{\varepsilon,N}}{1-\phi}, \sigma^2_{\varepsilon,N}, ..., \sigma^2_{\varepsilon,N}\right), \tag{C.3.36}$$

is a $(T+2) \times (T+2)$ diagonal matrix and that

$$\mathbf{b}_{t-1-q} = \left(\frac{1}{1-\phi}, 1, \phi^{t-2-q}, .., \phi^0, \mathbf{0}_{1\times(T-t-1-q)}\right), \tag{C.3.37}$$

is a $1 \times (T+2)$ vector. Since $\widehat{\delta}_N \xrightarrow{p} \delta$ and $|\phi| < 1$, we then have by Theorem 14 in Pötscher and Prucha (2001) that

$$\begin{aligned}\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta}}-\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right) &= o_p\left(1\right), \\ \left(\phi^k\boldsymbol{\beta}-\widehat{\phi}^k\widehat{\boldsymbol{\beta}}_N\right) &= o_p\left(1\right),\ k\geq 0 \\ \left(\mathbf{b}_{t-1-q}-\widehat{\mathbf{b}}_{t-1-q,N}\right) &= o_p\left(1\right), \\ \left(\mathbf{b}_{s-1-r}-\widehat{\mathbf{b}}_{s-1-r,N}\right) &= o_p\left(1\right),\end{aligned} \tag{C.3.38}$$

and

$$\begin{aligned}&tr\left(\mathbf{b}'_{t-1-q}\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta}}\mathbf{d}'_s\mathbf{b}_{s-1-r}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}\right) \\ &-tr\left(\widehat{\mathbf{b}}'_{t-1-q,N}\mathbf{d}_t\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\mathbf{d}'_s\widehat{\mathbf{b}}_{s-1-r,N}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right) = o_p\left(1\right).\end{aligned} \tag{C.3.39}$$

Thus it follows that for $m=1,2,4,5,7,9,11,12,13,$ and 14, all elements of $N^{-1}v^y_{m,N}$ are either $o_p\left(1\right)$ or uniformly bounded in absolute value. Hence, $N^{-1}v^y_{m,N}=o_p\left(1\right)$ for $m=1,2,4,5,7,9,11,12,13,14$ and 16.

Finally, I examine the remaining scalars $v^y_{m,N}$ that contain stochastic elements with dimensions that depend on N. Observe that by assumption in the lemma, the parameter ρ and the matrix $\mathbf{P}_N\left(\rho\right)=\left(\mathbf{I}_N-\rho\mathbf{W}_N\right)^{-1}$ satisfy the condition in Lemma C7. Thus

$$\begin{aligned}\mathbf{a}'_{t-1-q,N}\left(\mathbf{P}_N\mathbf{P}'_N-\widehat{\mathbf{P}}_N\widehat{\mathbf{P}}'_N\right)\mathbf{a}_{s-1-r,N} &= o_p\left(1\right), \\ \mathbf{X}'_{t-1-q-k,N}\left(\mathbf{P}_N\mathbf{P}'_N-\widehat{\mathbf{P}}_N\widehat{\mathbf{P}}'_N\right)\mathbf{a}_{s-1-r,N} &= o_p\left(1\right), \\ \mathbf{X}'_{t-1-q-k,N}\left(\mathbf{P}_N\mathbf{P}'_N-\widehat{\mathbf{P}}_N\widehat{\mathbf{P}}'_N\right)\mathbf{X}_{s-1-r-k,N} &= o_p\left(1\right), \\ tr\left(\mathbf{P}'_N\mathbf{P}_N\mathbf{P}'_N\mathbf{P}_N-\widehat{\mathbf{P}}'_N\widehat{\mathbf{P}}_N\widehat{\mathbf{P}}'_N\widehat{\mathbf{P}}_N\right) &= o_p\left(1\right).\end{aligned} \tag{C.3.40}$$

Hence $N^{-1}v^{y}_{m,N} = o_p(1)$ for $m = 3, 6, 8, 10, 15, 17$, and 18. As a result, we have that

$$[N(T-1)]^{-1}\left(\widetilde{v}^{\mathbf{y}}_{qr,ts,N} - \widehat{v}^{\mathbf{y}}_{qr,ts,N}\right) = o_p(1). \tag{C.3.41}$$

Next I consider the lower-diagonal block of $\widehat{\mathbf{V}}_{ts,N}$. As above, I express the difference between the typical element of $\widetilde{\mathbf{V}}^{\mathbf{X}}_{ts,N}$ and $\widehat{\mathbf{V}}^{\mathbf{X}}_{ts,N}$ as

$$\begin{aligned}\widetilde{\mathbf{V}}^{\mathbf{X}}_{qr,ts,N} - \widehat{\mathbf{V}}^{\mathbf{X}}_{qr,ts,N} &= \mathbf{X}'_{t-q,N}\left(\mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}'_s \otimes \mathbf{P}_N\mathbf{P}'_N\right)\mathbf{X}_{s-r,N} \\ &\quad -\mathbf{X}'_{t-q,N}\left(\mathbf{d}_t\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\mathbf{d}'_s \otimes \widehat{\mathbf{P}}_N\widehat{\mathbf{P}}'_N\right)\mathbf{X}_{s-r,N} \\ &= \mathbf{A}_{1,N} + \mathbf{A}_{2,N} + \mathbf{A}_{3,N},\end{aligned} \tag{C.3.42}$$

where[51]

$$\begin{aligned}\mathbf{A}_{1,N} &= \mathbf{d}_t\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}'_s\mathbf{X}'_{t-q,N}\mathbf{P}_N\mathbf{P}'_N\mathbf{X}_{s-r,N}, \\ \mathbf{A}_{2,N} &= \mathbf{d}_t\boldsymbol{\Sigma}_{\boldsymbol{\eta},N}\mathbf{d}'_s\mathbf{X}'_{t-q,N}\left(\mathbf{P}_N\mathbf{P}'_N - \widehat{\mathbf{P}}_N\widehat{\mathbf{P}}'_N\right)\mathbf{X}_{s-r,N}, \\ \mathbf{A}_{3,N} &= -\mathbf{d}_t\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right)\mathbf{d}'_s\mathbf{X}'_{t-q,N}\left(\mathbf{P}_N\mathbf{P}'_N - \widehat{\mathbf{P}}_N\widehat{\mathbf{P}}'_N\right)\mathbf{X}_{s-r,N}.\end{aligned} \tag{C.3.43}$$

I again do not explicitly denote the dependence of the $p \times p$ matrices $\mathbf{A}_{1,N}$, $\mathbf{A}_{2,N}$ and $\mathbf{A}_{3,N}$ on the value of the indexes q, r, s, t.

As above we have that $\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta},N} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta},N}\right) = o_p(1)$. Since by Assumption

[51] Similarly to the decomposition above this uses the the following algebraic rule: let a,b and $\widehat{a}$,$\widehat{b}$ be matrices of conformable dimensions. Then

$$ab - \widehat{a}\widehat{b} = (a - \widehat{a})\,b + a\left(b - \widehat{b}\right) - (a - \widehat{a})\left(b - \widehat{b}\right).$$

5, the elements of and $\mathbf{X}_{t-q,N}$ and $\mathbf{X}_{s-r,N}$ are uniformly bounded in absolute value and from Assumption 3, the matrix $\mathbf{P}_N\mathbf{P}_N'$ has row and column sums uniformly bounded in absolute value, it follows from Lemma C3 that the elements of $N^{-1}\mathbf{X}_{t-q,N}'\mathbf{P}_N\mathbf{P}_N'\mathbf{X}_{s-r,N}$ are uniformly bounded in absolute value and, therefore, $N^{-1}\mathbf{A}_{1,N} = o_p(1)$. Similarly, from Lemma C7 it follows that $N^{-1}\mathbf{X}_{t-q,N}'\left(\mathbf{P}_N\mathbf{P}_N' - \widehat{\mathbf{P}}_N\widehat{\mathbf{P}}_N'\right)\mathbf{X}_{s-r,N} = o_p(1)$, and hence $N^{-1}\mathbf{A}_{2,N}$ and $N^{-1}\mathbf{A}_{3,N}$ are $o_p(1)$. As a result,

$$[N(T-1)]^{-1}\left(\widetilde{\mathbf{V}}_{ts,N}^{\mathbf{X}} - \widehat{\mathbf{V}}_{ts,N}^{\mathbf{X}}\right) = o_p(1)\,. \tag{C.3.44}$$

Finally, I show that the off-diagonal blocks in $\widetilde{\mathbf{V}}_{ts,N}$ are matrices of zeros. Observe that from Lemma 4 it follows that the moments $\Delta\mathbf{u}_{s,N}'\mathbf{y}_{s-1-r,N}$ are linear-quadratic forms in the innovations $\boldsymbol{\eta}_N$. Since $E\left(\Delta\mathbf{u}_{s,N}'\mathbf{y}_{s-1-r,N}\right) = 0$ (as $r > 0$), it follows that the diagonal elements of the quadratic forms are zeros. Because elements of $\mathbf{X}_{t-q,N}'\Delta\mathbf{u}_{t,N}$ are linear forms in $\boldsymbol{\eta}_N$, it follows from Lemma A1 that

$$E\left(\mathbf{X}_{t-q,N}'\Delta\mathbf{u}_{t,N}\Delta\mathbf{u}_{s,N}'\mathbf{y}_{s-1-r,N}\right) = \mathbf{0}_{p\times 1},$$

and hence the off-diagonal blocks in both $\widetilde{\mathbf{V}}_{ts,N}$ and $\widehat{\mathbf{V}}_{ts,N}$ are matrices of zeros. Thus we have together that

$$[N(T-1)]^{-1}\left(\widetilde{\mathbf{V}}_{ts,N} - \widehat{\mathbf{V}}_{ts,N}\right) \xrightarrow{p} \mathbf{0}_{k_t\times k_s}, \tag{C.3.45}$$

or by repeating the above arguments for other values of t and s that

$$[N(T-1)]^{-1}\left(\widetilde{\mathbf{V}}_N-\widehat{\mathbf{V}}_N\right)\xrightarrow{p}\mathbf{0}_{k\times k}. \tag{C.3.46}$$

From $[(T-1)N]^{-1}\widetilde{\mathbf{V}}_N\xrightarrow{p}\widetilde{\mathbf{V}}_N$ (Assumption GMM3) it now follows that $[(T-1)N]^{-1}\widehat{\mathbf{V}}_N\xrightarrow{p}\widetilde{\mathbf{V}}$. ■

Proof of Theorem 4: The feasible second stage GMM estimator is

$$\check{\boldsymbol{\theta}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)=\left[\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N\widehat{\mathbf{V}}_N^{-1}\left(\widehat{\boldsymbol{\delta}}_N\right)\widetilde{\mathbf{H}}_N'\Delta\mathbf{Z}_N\right]^{-1}\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N\widehat{\mathbf{V}}_N^{-1}\left(\widehat{\boldsymbol{\delta}}_N\right)\widetilde{\mathbf{H}}_N'\Delta\mathbf{y}_N. \tag{C.3.47}$$

To prove the claim it suffices to show that, see e.g. Schmidt (1976), p. 71:

$$\begin{aligned}\boldsymbol{\Delta}_{1,N}= &\quad [N(T-1)]^{-1}\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N\widehat{\mathbf{V}}_N^{-1}\left(\widehat{\boldsymbol{\delta}}_N\right)\widetilde{\mathbf{H}}_N'\Delta\mathbf{Z}_N\\ &-[N(T-1)]^{-1}\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N\widetilde{\mathbf{V}}_N^{-1}\widetilde{\mathbf{H}}_N'\Delta\mathbf{Z}_N\xrightarrow{p}0,\end{aligned} \tag{C.3.48}$$

and

$$\begin{aligned}\boldsymbol{\Delta}_{2,N}= &\quad [N(T-1)]^{-1/2}\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N\widehat{\mathbf{V}}_N^{-1}\left(\widehat{\boldsymbol{\delta}}_N\right)\Delta\widetilde{\mathbf{H}}_N'\Delta\mathbf{u}_N\\ &-[N(T-1)]^{-1/2}\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N\widetilde{\mathbf{V}}_N^{-1}\widetilde{\mathbf{H}}_N'\Delta\mathbf{u}_N\xrightarrow{p}0.\end{aligned} \tag{C.3.49}$$

Note that

$$\begin{aligned}\boldsymbol{\Delta}_{1,N} &= [N(T-1)]^{-1}\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N * \\ &* \left[\left([N(T-1)]^{-1}\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)\right)^{-1}-\left([N(T-1)]^{-1}\widetilde{\mathbf{V}}_N\right)^{-1}\right]\\ &* [N(T-1)]^{-1}\widetilde{\mathbf{H}}_N'\Delta\mathbf{Z}_N.\end{aligned} \tag{C.3.50}$$

From Lemma 7 and Assumption GMM3, it folows that the matrices $[(T-1)\,N]^{-1}\,\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)$ and $[N\,(T-1)]^{-1}\,\widetilde{\mathbf{V}}_N$ both converge to $\widetilde{\mathbf{V}}$ in probability. Since by Assumption GMM3 the matrix $\widetilde{\mathbf{V}}$ is finite and nonsingular, it follows from Theorem 14 in Pötscher and Prucha (2001) that

$$\left[\left([N\,(T-1)]^{-1}\,\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)\right)^{-1}-\left([N\,(T-1)]^{-1}\,\widetilde{\mathbf{V}}_N\right)^{-1}\right]=o_p\,(1)\,. \quad \text{(C.3.51)}$$

Given Assumption GMM2, it then follows that $\boldsymbol{\Delta}_{1,N}\xrightarrow{p}0$.

Similarly we have for $\boldsymbol{\Delta}_{2,N}$:

$$\begin{aligned}\boldsymbol{\Delta}_{2,N} &= [N\,(T-1)]^{-1}\,\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N\,* \qquad \text{(C.3.52)}\\ &\quad *\left[\left([N\,(T-1)]^{-1}\,\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)\right)^{-1}-\left([N\,(T-1)]^{-1}\,\widetilde{\mathbf{V}}_N\right)^{-1}\right]\\ &\quad *[N\,(T-1)]^{-1/2}\,\widetilde{\mathbf{H}}_N'\Delta\mathbf{u}_N,\end{aligned}$$

where as above

$$[N\,(T-1)]^{-1}\,\Delta\mathbf{Z}_N'\widetilde{\mathbf{H}}_N\xrightarrow{p}\widetilde{\mathbf{M}}_{H\Delta Z}', \quad \text{(C.3.53)}$$

and

$$\left[\left([N\,(T-1)]^{-1}\,\widehat{\mathbf{V}}_N\left(\widehat{\boldsymbol{\delta}}_N\right)\right)^{-1}-\left([N\,(T-1)]^{-1}\,\widetilde{\mathbf{V}}_N\right)^{-1}\right]\xrightarrow{p}\mathbf{0}_{k\times k}. \quad \text{(C.3.54)}$$

Note that from Lemma 5, it follows that $\widetilde{\mathbf{V}}_N^{-1/2}\widetilde{\mathbf{H}}_N'\Delta\mathbf{u}\xrightarrow{d}N\,(\mathbf{0},\mathbf{I}_k)$. Given Assumption GMM3, it follows from Theorem 15 in Pötscher and Prucha (2001) that

$$\frac{\widetilde{\mathbf{H}}'_N \Delta \mathbf{u}_N}{[N(T-1)]^{1/2}} = \left(\frac{\widetilde{\mathbf{V}}_N}{N(T-1)} \right)^{1/2} \cdot \widetilde{\mathbf{V}}_N^{-1/2} \widetilde{\mathbf{H}}'_N \Delta \mathbf{u}_N \xrightarrow{d} N\left(\mathbf{0}, \widetilde{\mathbf{V}}\right) \quad \text{(C.3.55)}$$

Hence by Corollary 5, part (a), in Pötscher and Prucha (2001), we have that $\mathbf{\Delta}_{2,N} \xrightarrow{p} 0$. ■

Proof of Lemma 8: Given Assumption GMM3, the claim follows directly from C.3.48. ■

D Appendix: Tables of Monte Carlo Results

Table D1

Initial IV Estimators of Φ

Estimator			AH1		AH2		AB	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	-0.90	1	0.0977	0.0057	0.0948	0.0088	0.0883	-0.0169
-0.75	-0.90	1	0.1213	0.0079	0.1094	0.0087	0.1030	-0.0252
-0.25	-0.90	1	0.2367	0.0139	0.1601	0.0110	0.1634	-0.0858
0.00	-0.90	1	0.3246	0.0151	0.1885	0.0110	0.2089	-0.1369
0.25	-0.90	1	0.5168	0.0140	0.2223	0.0100	0.2758	-0.2090
0.75	-0.90	1	0.9098	-0.1899	0.2126	0.0099	0.3447	-0.2852
0.90	-0.90	1	0.3725	-0.0178	0.1594	0.0050	0.2694	-0.2165
-0.90	-0.50	1	0.0452	0.0014	0.0428	0.0016	0.0411	-0.0066
-0.75	-0.50	1	0.0554	0.0021	0.0502	0.0020	0.0477	-0.0097
-0.25	-0.50	1	0.1004	0.0044	0.0784	0.0013	0.0731	-0.0284
0.00	-0.50	1	0.1386	0.0047	0.0927	0.0041	0.0862	-0.0427
0.25	-0.50	1	0.2112	0.0070	0.1098	0.0059	0.1030	-0.0638
0.75	-0.50	1	1.0751	-0.2148	0.1156	0.0055	0.1167	-0.0864
0.90	-0.50	1	0.2612	-0.0031	0.0877	0.0043	0.0988	-0.0734
-0.90	-0.25	1	0.0372	0.0011	0.0362	0.0025	0.0343	-0.0046
-0.75	-0.25	1	0.0458	0.0008	0.0421	0.0021	0.0410	-0.0068
-0.25	-0.25	1	0.0866	0.0026	0.0655	0.0014	0.0606	-0.0199
0.00	-0.25	1	0.1187	0.0022	0.0781	0.0024	0.0701	-0.0296
0.25	-0.25	1	0.1711	0.0030	0.0902	0.0037	0.0816	-0.0441
0.75	-0.25	1	1.2234	-0.3342	0.0948	0.0052	0.0914	-0.0606
0.90	-0.25	1	0.2557	-0.0095	0.0733	0.0019	0.0783	-0.0533
-0.90	0.00	1	0.0364	0.0018	0.0362	0.0011	0.0346	-0.0048
-0.75	0.00	1	0.0446	0.0015	0.0413	0.0010	0.0396	-0.0064
-0.25	0.00	1	0.0845	0.0016	0.0644	0.0020	0.0591	-0.0181
0.00	0.00	1	0.1135	0.0022	0.0736	0.0014	0.0651	-0.0257
0.25	0.00	1	0.1660	0.0028	0.0856	0.0008	0.0766	-0.0391
0.75	0.00	1	1.3519	-0.3439	0.0884	0.0058	0.0849	-0.0538
0.90	0.00	1	0.2572	-0.0050	0.0713	0.0050	0.0725	-0.0487
-0.90	0.25	1	0.0385	0.0000	0.0365	0.0018	0.0357	-0.0055
-0.75	0.25	1	0.0477	0.0012	0.0430	0.0022	0.0427	-0.0084
-0.25	0.25	1	0.0884	0.0012	0.0669	0.0030	0.0629	-0.0196
0.00	0.25	1	0.1229	0.0022	0.0804	0.0032	0.0714	-0.0295
0.25	0.25	1	0.1824	0.0022	0.0927	0.0019	0.0825	-0.0423
0.75	0.25	1	1.2421	-0.3463	0.0979	0.0032	0.0912	-0.0622
0.90	0.25	1	0.2583	-0.0029	0.0768	0.0046	0.0792	-0.0544

Table D1 cont.

Initial IV Estimators of Φ

Estimator			AH1		AH2		AB	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.50	1	0.0473	0.0002	0.0443	0.0013	0.0426	-0.0071
-0.75	0.50	1	0.0574	0.0010	0.0528	0.0016	0.0497	-0.0107
-0.25	0.50	1	0.1058	0.0028	0.0807	0.0047	0.0732	-0.0264
0.00	0.50	1	0.1449	0.0019	0.0992	0.0040	0.0885	-0.0428
0.25	0.50	1	0.2197	0.0029	0.1153	0.0064	0.1037	-0.0616
0.75	0.50	1	1.0355	-0.2428	0.1204	0.0041	0.1190	-0.0896
0.90	0.50	1	0.2672	-0.0110	0.0948	0.0042	0.1001	-0.0751
-0.90	0.90	1	0.0950	-0.0029	0.0960	0.0013	0.0916	-0.0229
-0.75	0.90	1	0.1178	-0.0037	0.1145	0.0013	0.1100	-0.0321
-0.25	0.90	1	0.2298	-0.0047	0.1761	0.0073	0.1691	-0.0896
0.00	0.90	1	0.3335	-0.0058	0.2008	0.0084	0.2131	-0.1363
0.25	0.90	1	0.5477	-0.0176	0.2251	0.0143	0.2764	-0.2062
0.75	0.90	1	0.9974	-0.1566	0.2144	0.0061	0.3543	-0.2889
0.90	0.90	1	0.3929	-0.0086	0.1662	-0.0005	0.2672	-0.2119
-0.90	-0.90	2	0.0408	0.0006	0.0379	0.0002	0.0372	-0.0067
-0.75	-0.90	2	0.0498	-0.0002	0.0448	0.0004	0.0434	-0.0091
-0.25	-0.90	2	0.0937	0.0001	0.0676	0.0001	0.0655	-0.0235
0.00	-0.90	2	0.1300	-0.0027	0.0821	-0.0015	0.0788	-0.0353
0.25	-0.90	2	0.1905	0.0008	0.0923	0.0003	0.0881	-0.0509
0.75	-0.90	2	1.0960	-0.2450	0.0989	0.0011	0.0985	-0.0708
0.90	-0.90	2	0.2442	-0.0024	0.0770	0.0018	0.0853	-0.0604
-0.90	-0.50	2	0.0367	0.0015	0.0356	0.0009	0.0349	-0.0052
-0.75	-0.50	2	0.0451	0.0003	0.0413	0.0010	0.0412	-0.0074
-0.25	-0.50	2	0.0864	-0.0011	0.0638	0.0004	0.0603	-0.0202
0.00	-0.50	2	0.1170	0.0020	0.0770	-0.0002	0.0696	-0.0283
0.25	-0.50	2	0.1750	0.0060	0.0859	-0.0002	0.0806	-0.0423
0.75	-0.50	2	1.1873	-0.3030	0.0921	0.0035	0.0897	-0.0599
0.90	-0.50	2	0.2555	-0.0040	0.0723	0.0025	0.0762	-0.0521
-0.90	-0.25	2	0.0364	0.0015	0.0349	0.0011	0.0347	-0.0043
-0.75	-0.25	2	0.0441	0.0010	0.0403	0.0009	0.0400	-0.0064
-0.25	-0.25	2	0.0849	-0.0002	0.0640	0.0017	0.0583	-0.0180
0.00	-0.25	2	0.1141	0.0037	0.0747	0.0000	0.0677	-0.0270
0.25	-0.25	2	0.1697	0.0049	0.0844	0.0015	0.0782	-0.0406
0.75	-0.25	2	1.3170	-0.3437	0.0901	0.0032	0.0863	-0.0561
0.90	-0.25	2	0.2563	-0.0088	0.0703	0.0040	0.0732	-0.0489

Table D1 cont.

Initial IV Estimators of Φ

Estimator			AH1		AH2		AB	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.00	2	0.0364	0.0018	0.0362	0.0011	0.0346	-0.0048
-0.75	0.00	2	0.0446	0.0015	0.0413	0.0010	0.0396	-0.0064
-0.25	0.00	2	0.0845	0.0016	0.0644	0.0020	0.0591	-0.0181
0.00	0.00	2	0.1135	0.0022	0.0736	0.0014	0.0651	-0.0257
0.25	0.00	2	0.1660	0.0028	0.0856	0.0008	0.0766	-0.0391
0.75	0.00	2	1.3519	-0.3439	0.0884	0.0058	0.0849	-0.0538
0.90	0.00	2	0.2572	-0.0050	0.0713	0.0050	0.0725	-0.0487
-0.90	0.25	2	0.0370	0.0011	0.0360	0.0015	0.0345	-0.0047
-0.75	0.25	2	0.0465	0.0011	0.0427	0.0017	0.0399	-0.0069
-0.25	0.25	2	0.0855	0.0013	0.0656	0.0034	0.0604	-0.0202
0.00	0.25	2	0.1192	0.0034	0.0783	0.0033	0.0693	-0.0289
0.25	0.25	2	0.1760	0.0026	0.0886	0.0029	0.0794	-0.0413
0.75	0.25	2	1.2863	-0.3508	0.0909	0.0043	0.0880	-0.0584
0.90	0.25	2	0.2585	0.0024	0.0755	0.0047	0.0759	-0.0516
-0.90	0.50	2	0.0416	0.0011	0.0409	0.0024	0.0391	-0.0063
-0.75	0.50	2	0.0507	0.0012	0.0479	0.0030	0.0460	-0.0090
-0.25	0.50	2	0.0958	0.0016	0.0721	0.0041	0.0679	-0.0250
0.00	0.50	2	0.1344	0.0075	0.0894	0.0060	0.0797	-0.0354
0.25	0.50	2	0.1997	0.0094	0.1063	0.0060	0.0935	-0.0526
0.75	0.50	2	1.2256	-0.2925	0.1091	0.0093	0.1046	-0.0755
0.90	0.50	2	0.2678	-0.0126	0.0885	0.0055	0.0907	-0.0659
-0.90	0.90	2	0.1252	-0.0041	0.1163	0.0077	0.1105	-0.0267
-0.75	0.90	2	0.1538	-0.0030	0.1380	0.0121	0.1285	-0.0365
-0.25	0.90	2	0.2908	-0.0019	0.2099	0.0137	0.2005	-0.1060
0.00	0.90	2	0.4118	-0.0013	0.2549	0.0156	0.2457	-0.1611
0.25	0.90	2	0.6497	-0.0186	0.2939	0.0166	0.3227	-0.2403
0.75	0.90	2	1.2519	-0.3148	0.2742	0.0071	0.4062	-0.3263
0.90	0.90	2	0.5361	-0.0507	0.2255	0.0068	0.3408	-0.2655
-0.90	-0.90	3	0.0392	0.0016	0.0370	0.0020	0.0364	-0.0052
-0.75	-0.90	3	0.0474	0.0021	0.0431	0.0023	0.0419	-0.0075
-0.25	-0.90	3	0.0900	0.0035	0.0664	0.0026	0.0635	-0.0184
0.00	-0.90	3	0.1228	0.0058	0.0790	0.0015	0.0728	-0.0291
0.25	-0.90	3	0.1857	0.0068	0.0916	0.0021	0.0834	-0.0442
0.75	-0.90	3	1.1327	-0.3200	0.0931	0.0023	0.0901	-0.0631
0.90	-0.90	3	0.2562	-0.0151	0.0741	0.0041	0.0777	-0.0534

Table D1 cont.

Initial IV Estimators of Φ

Estimator			AH1		AH2		AB	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	-0.50	3	0.0372	0.0014	0.0359	0.0012	0.0344	-0.0042
-0.75	-0.50	3	0.0455	0.0021	0.0417	0.0007	0.0411	-0.0067
-0.25	-0.50	3	0.0886	0.0012	0.0642	0.0015	0.0605	-0.0181
0.00	-0.50	3	0.1181	0.0045	0.0764	0.0016	0.0685	-0.0278
0.25	-0.50	3	0.1757	0.0043	0.0863	0.0009	0.0791	-0.0404
0.75	-0.50	3	1.2685	-0.3750	0.0893	0.0046	0.0855	-0.0572
0.90	-0.50	3	0.2589	-0.0104	0.0714	0.0056	0.0736	-0.0499
-0.90	-0.25	3	0.0363	0.0019	0.0358	0.0005	0.0348	-0.0038
-0.75	-0.25	3	0.0454	0.0014	0.0413	0.0008	0.0397	-0.0062
-0.25	-0.25	3	0.0858	0.0008	0.0641	0.0015	0.0596	-0.0183
0.00	-0.25	3	0.1163	0.0041	0.0744	0.0018	0.0668	-0.0270
0.25	-0.25	3	0.1672	0.0051	0.0857	0.0020	0.0764	-0.0384
0.75	-0.25	3	1.3017	-0.3856	0.0886	0.0052	0.0843	-0.0555
0.90	-0.25	3	0.2542	-0.0107	0.0702	0.0051	0.0720	-0.0484
-0.90	0.00	3	0.0364	0.0018	0.0362	0.0011	0.0346	-0.0048
-0.75	0.00	3	0.0446	0.0015	0.0413	0.0010	0.0396	-0.0064
-0.25	0.00	3	0.0845	0.0016	0.0644	0.0020	0.0591	-0.0181
0.00	0.00	3	0.1135	0.0022	0.0736	0.0014	0.0651	-0.0257
0.25	0.00	3	0.1660	0.0028	0.0856	0.0008	0.0766	-0.0391
0.75	0.00	3	1.3519	-0.3439	0.0884	0.0058	0.0849	-0.0538
0.90	0.00	3	0.2572	-0.0050	0.0713	0.0050	0.0725	-0.0487
-0.90	0.25	3	0.0374	0.0009	0.0364	0.0015	0.0344	-0.0045
-0.75	0.25	3	0.0456	0.0013	0.0417	0.0016	0.0395	-0.0054
-0.25	0.25	3	0.0869	0.0003	0.0651	0.0021	0.0597	-0.0193
0.00	0.25	3	0.1154	0.0028	0.0772	0.0018	0.0693	-0.0276
0.25	0.25	3	0.1709	0.0037	0.0889	0.0025	0.0795	-0.0396
0.75	0.25	3	1.3535	-0.3464	0.0878	0.0023	0.0864	-0.0568
0.90	0.25	3	0.2564	0.0049	0.0739	0.0034	0.0743	-0.0494
-0.90	0.50	3	0.0405	0.0006	0.0397	0.0020	0.0379	-0.0052
-0.75	0.50	3	0.0493	0.0014	0.0472	0.0022	0.0447	-0.0081
-0.25	0.50	3	0.0940	0.0036	0.0720	0.0021	0.0674	-0.0227
0.00	0.50	3	0.1316	0.0046	0.0855	0.0039	0.0774	-0.0328
0.25	0.50	3	0.1919	0.0072	0.1013	0.0039	0.0889	-0.0471
0.75	0.50	3	1.2627	-0.3083	0.1014	0.0098	0.0954	-0.0670
0.90	0.50	3	0.2714	-0.0080	0.0802	0.0042	0.0847	-0.0588

Table D1 cont.

Initial IV Estimators of Φ

Estimator			AH1		AH2		AB	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.90	3	0.1371	-0.0009	0.1252	0.0076	0.1180	-0.0288
-0.75	0.90	3	0.1699	-0.0027	0.1503	0.0126	0.1369	-0.0393
-0.25	0.90	3	0.3219	0.0020	0.2256	0.0187	0.2058	-0.1003
0.00	0.90	3	0.4354	0.0032	0.2673	0.0204	0.2551	-0.1572
0.25	0.90	3	0.6794	0.0100	0.3249	0.0148	0.3273	-0.2381
0.75	0.90	3	1.3234	-0.3345	0.3122	0.0082	0.4046	-0.3253
0.90	0.90	3	0.5756	-0.0721	0.2463	0.0068	0.3432	-0.2676

Table D2

Second Stage GMM Estimators of Φ

Estimator			ignoring		mix		exp	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	-0.90	1	0.0853	-0.0065	0.0713	-0.0082	0.0850	-0.0016
-0.75	-0.90	1	0.0987	-0.0093	0.0845	-0.0147	0.0987	-0.0070
-0.25	-0.90	1	0.1419	-0.0425	0.1333	-0.0536	0.1468	-0.0411
0.00	-0.90	1	0.1676	-0.0678	0.1616	-0.0822	0.1736	-0.0735
0.25	-0.90	1	0.1989	-0.0998	0.1934	-0.1158	0.2113	-0.1165
0.75	-0.90	1	0.1773	-0.0866	0.1757	-0.1082	0.2431	-0.1575
0.90	-0.90	1	0.1279	-0.0562	0.1291	-0.0787	0.1783	-0.0911
-0.90	-0.50	1	0.0417	-0.0030	0.0404	-0.0027	0.0426	0.0003
-0.75	-0.50	1	0.0490	-0.0039	0.0462	-0.0040	0.0499	-0.0022
-0.25	-0.50	1	0.0703	-0.0140	0.0681	-0.0138	0.0693	-0.0113
0.00	-0.50	1	0.0845	-0.0224	0.0783	-0.0186	0.0773	-0.0162
0.25	-0.50	1	0.0983	-0.0317	0.0927	-0.0306	0.0883	-0.0236
0.75	-0.50	1	0.0901	-0.0319	0.0868	-0.0378	0.0928	-0.0410
0.90	-0.50	1	0.0668	-0.0221	0.0688	-0.0274	0.0777	-0.0279
-0.90	-0.25	1	0.0347	-0.0017	0.0349	-0.0017	0.0367	0.0005
-0.75	-0.25	1	0.0408	-0.0027	0.0406	-0.0025	0.0424	-0.0008
-0.25	-0.25	1	0.0585	-0.0094	0.0589	-0.0085	0.0589	-0.0078
0.00	-0.25	1	0.0688	-0.0147	0.0667	-0.0146	0.0658	-0.0118
0.25	-0.25	1	0.0796	-0.0232	0.0765	-0.0210	0.0737	-0.0150
0.75	-0.25	1	0.0744	-0.0260	0.0771	-0.0281	0.0784	-0.0309
0.90	-0.25	1	0.0584	-0.0180	0.0594	-0.0188	0.0683	-0.0202
-0.90	0.00	1	0.0338	-0.0012	0.0338	-0.0013	0.0349	0.0012
-0.75	0.00	1	0.0386	-0.0021	0.0388	-0.0024	0.0403	-0.0003
-0.25	0.00	1	0.0572	-0.0091	0.0568	-0.0090	0.0556	-0.0058
0.00	0.00	1	0.0649	-0.0127	0.0646	-0.0126	0.0634	-0.0086
0.25	0.00	1	0.0738	-0.0189	0.0734	-0.0191	0.0707	-0.0133
0.75	0.00	1	0.0744	-0.0252	0.0742	-0.0252	0.0764	-0.0284
0.90	0.00	1	0.0572	-0.0167	0.0573	-0.0168	0.0657	-0.0205
-0.90	0.25	1	0.0345	-0.0024	0.0344	-0.0024	0.0378	-0.0005
-0.75	0.25	1	0.0410	-0.0029	0.0403	-0.0035	0.0422	-0.0008
-0.25	0.25	1	0.0594	-0.0099	0.0585	-0.0099	0.0572	-0.0072
0.00	0.25	1	0.0696	-0.0143	0.0688	-0.0149	0.0641	-0.0086
0.25	0.25	1	0.0770	-0.0228	0.0776	-0.0213	0.0739	-0.0118
0.75	0.25	1	0.0765	-0.0271	0.0787	-0.0291	0.0790	-0.0319
0.90	0.25	1	0.0619	-0.0197	0.0611	-0.0206	0.0696	-0.0236

Table D2 cont.

Second Stage GMM Estimators of Φ

Estimator			ignoring		mix		exp	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.50	1	0.0400	-0.0016	0.0390	-0.0034	0.0432	-0.0021
-0.75	0.50	1	0.0465	-0.0032	0.0471	-0.0047	0.0502	-0.0021
-0.25	0.50	1	0.0699	-0.0115	0.0678	-0.0131	0.0642	-0.0105
0.00	0.50	1	0.0792	-0.0191	0.0779	-0.0191	0.0736	-0.0131
0.25	0.50	1	0.0928	-0.0306	0.0890	-0.0314	0.0821	-0.0203
0.75	0.50	1	0.0938	-0.0340	0.0962	-0.0396	0.0923	-0.0426
0.90	0.50	1	0.0710	-0.0237	0.0723	-0.0284	0.0788	-0.0329
-0.90	0.90	1	0.0899	-0.0101	0.0765	-0.0138	0.0863	-0.0055
-0.75	0.90	1	0.1042	-0.0139	0.0882	-0.0171	0.1009	-0.0072
-0.25	0.90	1	0.1466	-0.0369	0.1290	-0.0453	0.1445	-0.0418
0.00	0.90	1	0.1733	-0.0592	0.1529	-0.0719	0.1709	-0.0700
0.25	0.90	1	0.1917	-0.0890	0.1819	-0.1069	0.2045	-0.1085
0.75	0.90	1	0.1767	-0.0865	0.1862	-0.1129	0.2410	-0.1530
0.90	0.90	1	0.1372	-0.0623	0.1417	-0.0823	0.1722	-0.0912
-0.90	-0.90	2	0.0367	-0.0028	0.0372	-0.0028	0.0399	-0.0005
-0.75	-0.90	2	0.0421	-0.0038	0.0439	-0.0040	0.0456	-0.0018
-0.25	-0.90	2	0.0611	-0.0108	0.0604	-0.0112	0.0595	-0.0069
0.00	-0.90	2	0.0713	-0.0175	0.0696	-0.0163	0.0681	-0.0126
0.25	-0.90	2	0.0834	-0.0265	0.0828	-0.0262	0.0780	-0.0185
0.75	-0.90	2	0.0812	-0.0278	0.0828	-0.0304	0.0871	-0.0365
0.90	-0.90	2	0.0615	-0.0175	0.0631	-0.0196	0.0703	-0.0230
-0.90	-0.50	2	0.0342	-0.0024	0.0345	-0.0021	0.0373	0.0007
-0.75	-0.50	2	0.0400	-0.0031	0.0407	-0.0033	0.0432	-0.0014
-0.25	-0.50	2	0.0579	-0.0093	0.0579	-0.0085	0.0561	-0.0060
0.00	-0.50	2	0.0655	-0.0131	0.0658	-0.0137	0.0628	-0.0088
0.25	-0.50	2	0.0763	-0.0211	0.0767	-0.0211	0.0731	-0.0148
0.75	-0.50	2	0.0752	-0.0246	0.0766	-0.0261	0.0802	-0.0308
0.90	-0.50	2	0.0586	-0.0171	0.0596	-0.0181	0.0666	-0.0208
-0.90	-0.25	2	0.0339	-0.0016	0.0340	-0.0013	0.0347	0.0011
-0.75	-0.25	2	0.0399	-0.0027	0.0400	-0.0025	0.0408	-0.0008
-0.25	-0.25	2	0.0563	-0.0086	0.0574	-0.0093	0.0560	-0.0057
0.00	-0.25	2	0.0645	-0.0123	0.0649	-0.0130	0.0615	-0.0093
0.25	-0.25	2	0.0762	-0.0188	0.0763	-0.0185	0.0693	-0.0140
0.75	-0.25	2	0.0756	-0.0255	0.0753	-0.0256	0.0779	-0.0303
0.90	-0.25	2	0.0577	-0.0168	0.0585	-0.0171	0.0651	-0.0206

Table D2 cont.

Second Stage GMM Estimators of Φ

Estimator			ignoring		mix		exp	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.00	2	0.0338	-0.0012	0.0337	-0.0016	0.0349	0.0013
-0.75	0.00	2	0.0386	-0.0021	0.0387	-0.0021	0.0401	-0.0005
-0.25	0.00	2	0.0572	-0.0091	0.0565	-0.0090	0.0549	-0.0060
0.00	0.00	2	0.0649	-0.0127	0.0650	-0.0127	0.0621	-0.0084
0.25	0.00	2	0.0738	-0.0189	0.0734	-0.0192	0.0704	-0.0133
0.75	0.00	2	0.0744	-0.0252	0.0745	-0.0253	0.0766	-0.0284
0.90	0.00	2	0.0572	-0.0167	0.0570	-0.0166	0.0655	-0.0203
-0.90	0.25	2	0.0341	-0.0007	0.0346	-0.0013	0.0362	0.0001
-0.75	0.25	2	0.0401	-0.0028	0.0397	-0.0026	0.0411	-0.0009
-0.25	0.25	2	0.0581	-0.0084	0.0573	-0.0085	0.0571	-0.0059
0.00	0.25	2	0.0674	-0.0137	0.0680	-0.0137	0.0635	-0.0089
0.25	0.25	2	0.0754	-0.0203	0.0766	-0.0208	0.0723	-0.0138
0.75	0.25	2	0.0748	-0.0275	0.0741	-0.0277	0.0773	-0.0299
0.90	0.25	2	0.0589	-0.0191	0.0584	-0.0191	0.0673	-0.0213
-0.90	0.50	2	0.0384	-0.0023	0.0379	-0.0029	0.0383	-0.0016
-0.75	0.50	2	0.0453	-0.0030	0.0447	-0.0030	0.0449	-0.0023
-0.25	0.50	2	0.0661	-0.0106	0.0614	-0.0109	0.0611	-0.0080
0.00	0.50	2	0.0736	-0.0155	0.0708	-0.0173	0.0673	-0.0121
0.25	0.50	2	0.0850	-0.0258	0.0814	-0.0235	0.0752	-0.0156
0.75	0.50	2	0.0859	-0.0305	0.0858	-0.0350	0.0830	-0.0357
0.90	0.50	2	0.0676	-0.0223	0.0684	-0.0258	0.0724	-0.0271
-0.90	0.90	2	0.1070	-0.0068	0.0657	-0.0109	0.0730	-0.0073
-0.75	0.90	2	0.1243	-0.0118	0.0759	-0.0150	0.0837	-0.0103
-0.25	0.90	2	0.1726	-0.0500	0.1142	-0.0420	0.1173	-0.0326
0.00	0.90	2	0.2035	-0.0835	0.1381	-0.0656	0.1349	-0.0479
0.25	0.90	2	0.2406	-0.1138	0.1691	-0.0929	0.1492	-0.0674
0.75	0.90	2	0.2403	-0.1234	0.2314	-0.1464	0.1812	-0.1127
0.90	0.90	2	0.1807	-0.0827	0.1811	-0.1088	0.1585	-0.0905
-0.90	-0.90	3	0.0356	-0.0014	0.0359	-0.0015	0.0390	0.0004
-0.75	-0.90	3	0.0413	-0.0038	0.0417	-0.0028	0.0443	-0.0011
-0.25	-0.90	3	0.0589	-0.0095	0.0603	-0.0088	0.0595	-0.0063
0.00	-0.90	3	0.0688	-0.0148	0.0691	-0.0138	0.0679	-0.0116
0.25	-0.90	3	0.0821	-0.0215	0.0819	-0.0206	0.0762	-0.0152
0.75	-0.90	3	0.0779	-0.0252	0.0791	-0.0270	0.0840	-0.0339
0.90	-0.90	3	0.0610	-0.0172	0.0613	-0.0177	0.0712	-0.0232

Table D2 cont.

Second Stage GMM Estimators of Φ

Estimator			ignoring		mix		exp	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	-0.50	3	0.0344	-0.0019	0.0342	-0.0018	0.0370	0.0011
-0.75	-0.50	3	0.0395	-0.0034	0.0397	-0.0027	0.0418	0.0000
-0.25	-0.50	3	0.0574	-0.0088	0.0569	-0.0078	0.0574	-0.0066
0.00	-0.50	3	0.0658	-0.0128	0.0664	-0.0121	0.0643	-0.0098
0.25	-0.50	3	0.0783	-0.0189	0.0783	-0.0187	0.0711	-0.0138
0.75	-0.50	3	0.0742	-0.0249	0.0747	-0.0249	0.0781	-0.0307
0.90	-0.50	3	0.0587	-0.0170	0.0603	-0.0174	0.0670	-0.0217
-0.90	-0.25	3	0.0338	-0.0016	0.0339	-0.0015	0.0357	0.0015
-0.75	-0.25	3	0.0391	-0.0023	0.0396	-0.0021	0.0407	-0.0004
-0.25	-0.25	3	0.0564	-0.0085	0.0569	-0.0083	0.0562	-0.0064
0.00	-0.25	3	0.0659	-0.0130	0.0665	-0.0135	0.0626	-0.0094
0.25	-0.25	3	0.0756	-0.0179	0.0745	-0.0182	0.0700	-0.0140
0.75	-0.25	3	0.0744	-0.0247	0.0746	-0.0244	0.0752	-0.0292
0.90	-0.25	3	0.0582	-0.0172	0.0588	-0.0173	0.0662	-0.0208
-0.90	0.00	3	0.0338	-0.0012	0.0336	-0.0012	0.0349	0.0013
-0.75	0.00	3	0.0386	-0.0021	0.0387	-0.0022	0.0401	-0.0006
-0.25	0.00	3	0.0572	-0.0091	0.0569	-0.0089	0.0554	-0.0059
0.00	0.00	3	0.0649	-0.0127	0.0651	-0.0129	0.0621	-0.0091
0.25	0.00	3	0.0738	-0.0189	0.0738	-0.0190	0.0706	-0.0136
0.75	0.00	3	0.0744	-0.0252	0.0741	-0.0253	0.0763	-0.0286
0.90	0.00	3	0.0572	-0.0167	0.0573	-0.0167	0.0658	-0.0204
-0.90	0.25	3	0.0347	-0.0009	0.0349	-0.0012	0.0354	0.0000
-0.75	0.25	3	0.0395	-0.0023	0.0398	-0.0023	0.0405	-0.0008
-0.25	0.25	3	0.0574	-0.0094	0.0582	-0.0090	0.0551	-0.0058
0.00	0.25	3	0.0658	-0.0137	0.0665	-0.0137	0.0614	-0.0091
0.25	0.25	3	0.0758	-0.0201	0.0777	-0.0208	0.0713	-0.0136
0.75	0.25	3	0.0741	-0.0270	0.0741	-0.0274	0.0776	-0.0284
0.90	0.25	3	0.0588	-0.0186	0.0591	-0.0186	0.0666	-0.0203
-0.90	0.50	3	0.0381	-0.0015	0.0364	-0.0029	0.0378	-0.0015
-0.75	0.50	3	0.0449	-0.0027	0.0439	-0.0039	0.0429	-0.0021
-0.25	0.50	3	0.0633	-0.0097	0.0604	-0.0100	0.0580	-0.0079
0.00	0.50	3	0.0720	-0.0162	0.0688	-0.0156	0.0646	-0.0117
0.25	0.50	3	0.0809	-0.0219	0.0805	-0.0232	0.0729	-0.0157
0.75	0.50	3	0.0816	-0.0309	0.0850	-0.0340	0.0808	-0.0325
0.90	0.50	3	0.0650	-0.0225	0.0677	-0.0253	0.0698	-0.0244

Table D2 cont.

Second Stage GMM Estimators of Φ

Estimator			ignoring		mix		exp	
True Values								
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.90	3	0.1131	-0.0056	0.0570	-0.0086	0.0666	-0.0091
-0.75	0.90	3	0.1307	-0.0109	0.0674	-0.0131	0.0740	-0.0130
-0.25	0.90	3	0.1833	-0.0499	0.0986	-0.0335	0.1008	-0.0252
0.00	0.90	3	0.2080	-0.0836	0.1128	-0.0459	0.1135	-0.0371
0.25	0.90	3	0.2467	-0.1224	0.1455	-0.0736	0.1227	-0.0489
0.75	0.90	3	0.2560	-0.1356	0.2234	-0.1442	0.1462	-0.0799
0.90	0.90	3	0.1988	-0.0946	0.1803	-0.1092	0.1385	-0.0702

Table D3

Unweighted Spatial GM Estimators of ρ

Initial Estimator / True Values			AH1		AH2		AB		True	
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	-0.90	1	0.028	0.013	0.028	0.013	0.033	0.016	0.018	-0.001
-0.75	-0.90	1	0.028	0.014	0.028	0.013	0.033	0.016	0.018	-0.001
-0.25	-0.90	1	0.033	0.019	0.030	0.014	0.036	0.019	0.018	-0.001
0.00	-0.90	1	0.039	0.024	0.030	0.016	0.040	0.024	0.018	-0.001
0.25	-0.90	1	0.052	0.032	0.030	0.017	0.048	0.030	0.018	-0.001
0.75	-0.90	1	0.069	0.046	0.029	0.017	0.055	0.037	0.018	-0.001
0.90	-0.90	1	0.039	0.024	0.029	0.015	0.045	0.030	0.018	-0.001
-0.90	-0.50	1	0.047	0.003	0.047	0.003	0.047	0.005	0.048	-0.001
-0.75	-0.50	1	0.047	0.004	0.047	0.004	0.048	0.005	0.048	-0.001
-0.25	-0.50	1	0.048	0.006	0.047	0.004	0.047	0.007	0.048	-0.001
0.00	-0.50	1	0.047	0.008	0.048	0.005	0.048	0.007	0.048	-0.001
0.25	-0.50	1	0.052	0.014	0.047	0.005	0.050	0.011	0.048	-0.001
0.75	-0.50	1	0.115	0.067	0.047	0.006	0.055	0.018	0.048	-0.001
0.90	-0.50	1	0.057	0.020	0.047	0.005	0.052	0.014	0.048	-0.001
-0.90	-0.25	1	0.057	0.001	0.056	0.000	0.057	0.001	0.057	-0.001
-0.75	-0.25	1	0.056	0.001	0.056	0.000	0.057	0.001	0.057	-0.001
-0.25	-0.25	1	0.058	0.001	0.057	0.001	0.056	0.002	0.057	-0.001
0.00	-0.25	1	0.057	0.002	0.057	0.001	0.058	0.003	0.057	-0.001
0.25	-0.25	1	0.057	0.005	0.057	0.001	0.057	0.004	0.057	-0.001
0.75	-0.25	1	0.088	0.041	0.056	0.002	0.057	0.007	0.057	-0.001
0.90	-0.25	1	0.061	0.011	0.057	0.001	0.058	0.006	0.057	-0.001
-0.90	0.00	1	0.061	-0.001	0.061	-0.001	0.060	-0.001	0.061	-0.001
-0.75	0.00	1	0.061	-0.001	0.061	-0.001	0.060	-0.001	0.061	-0.001
-0.25	0.00	1	0.061	-0.001	0.060	-0.002	0.061	-0.001	0.061	-0.001
0.00	0.00	1	0.062	-0.001	0.060	-0.002	0.061	-0.001	0.061	-0.001
0.25	0.00	1	0.061	-0.001	0.060	-0.002	0.061	-0.001	0.061	-0.001
0.75	0.00	1	0.075	-0.001	0.062	-0.001	0.060	-0.001	0.061	-0.001
0.90	0.00	1	0.063	0.001	0.061	-0.001	0.060	0.000	0.061	-0.001
-0.90	0.25	1	0.059	-0.003	0.060	-0.003	0.058	-0.003	0.058	-0.001
-0.75	0.25	1	0.059	-0.003	0.059	-0.003	0.057	-0.003	0.058	-0.001
-0.25	0.25	1	0.059	-0.004	0.059	-0.004	0.057	-0.004	0.058	-0.001
0.00	0.25	1	0.058	-0.005	0.058	-0.005	0.058	-0.005	0.058	-0.001
0.25	0.25	1	0.061	-0.007	0.059	-0.005	0.060	-0.006	0.058	-0.001
0.75	0.25	1	0.100	-0.049	0.060	-0.005	0.061	-0.008	0.058	-0.001
0.90	0.25	1	0.064	-0.011	0.060	-0.005	0.061	-0.006	0.058	-0.001

Table D3 cont.

Unweighted Spatial GM Estimators of ρ

Initial Estimator / True Values			AH1		AH2		AB		True	
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.50	1	0.051	-0.006	0.051	-0.006	0.051	-0.006	0.049	-0.001
-0.75	0.50	1	0.051	-0.006	0.050	-0.005	0.051	-0.007	0.049	-0.001
-0.25	0.50	1	0.050	-0.007	0.051	-0.007	0.051	-0.008	0.049	-0.001
0.00	0.50	1	0.052	-0.009	0.050	-0.007	0.052	-0.010	0.049	-0.001
0.25	0.50	1	0.055	-0.013	0.051	-0.008	0.054	-0.013	0.049	-0.001
0.75	0.50	1	0.120	-0.075	0.051	-0.009	0.058	-0.020	0.049	-0.001
0.90	0.50	1	0.059	-0.018	0.051	-0.008	0.054	-0.014	0.049	-0.001
-0.90	0.90	1	0.028	-0.013	0.027	-0.013	0.031	-0.015	0.019	0.000
-0.75	0.90	1	0.028	-0.014	0.028	-0.013	0.031	-0.016	0.019	0.000
-0.25	0.90	1	0.034	-0.019	0.029	-0.016	0.035	-0.019	0.019	0.000
0.00	0.90	1	0.039	-0.023	0.031	-0.017	0.038	-0.023	0.019	0.000
0.25	0.90	1	0.051	-0.032	0.031	-0.018	0.049	-0.030	0.019	0.000
0.75	0.90	1	0.070	-0.048	0.031	-0.017	0.059	-0.040	0.019	0.000
0.90	0.90	1	0.040	-0.025	0.028	-0.015	0.045	-0.029	0.019	0.000
-0.90	-0.90	2	0.132	0.005	0.132	0.006	0.131	0.007	0.132	-0.002
-0.75	-0.90	2	0.132	0.005	0.132	0.006	0.130	0.007	0.132	-0.002
-0.25	-0.90	2	0.131	0.011	0.133	0.007	0.134	0.012	0.132	-0.002
0.00	-0.90	2	0.136	0.016	0.133	0.008	0.134	0.015	0.132	-0.002
0.25	-0.90	2	0.143	0.032	0.136	0.009	0.140	0.025	0.132	-0.002
0.75	-0.90	2	0.325	0.223	0.132	0.012	0.149	0.044	0.132	-0.002
0.90	-0.90	2	0.154	0.046	0.133	0.009	0.135	0.028	0.132	-0.002
-0.90	-0.50	2	0.128	-0.001	0.127	0.000	0.127	0.001	0.126	-0.003
-0.75	-0.50	2	0.127	-0.001	0.126	0.001	0.126	0.001	0.126	-0.003
-0.25	-0.50	2	0.124	0.002	0.126	0.000	0.123	0.002	0.126	-0.003
0.00	-0.50	2	0.125	0.005	0.125	0.001	0.124	0.003	0.126	-0.003
0.25	-0.50	2	0.126	0.012	0.125	0.002	0.127	0.007	0.126	-0.003
0.75	-0.50	2	0.209	0.118	0.128	0.001	0.131	0.016	0.126	-0.003
0.90	-0.50	2	0.136	0.020	0.127	0.000	0.127	0.008	0.126	-0.003
-0.90	-0.25	2	0.120	-0.004	0.119	-0.003	0.119	-0.002	0.116	-0.003
-0.75	-0.25	2	0.120	-0.003	0.118	-0.002	0.118	-0.002	0.116	-0.003
-0.25	-0.25	2	0.117	-0.001	0.117	-0.002	0.115	-0.002	0.116	-0.003
0.00	-0.25	2	0.118	-0.001	0.116	-0.003	0.118	-0.001	0.116	-0.003
0.25	-0.25	2	0.115	0.004	0.117	-0.002	0.118	0.000	0.116	-0.003
0.75	-0.25	2	0.161	0.055	0.119	-0.003	0.123	0.003	0.116	-0.003
0.90	-0.25	2	0.129	0.005	0.118	-0.004	0.120	0.001	0.116	-0.003

Table D3 cont.

Unweighted Spatial GM Estimators of ρ

Initial Estimator True Values			AH1		AH2		AB		True	
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.00	2	0.106	-0.005	0.107	-0.004	0.106	-0.004	0.103	-0.003
-0.75	0.00	2	0.107	-0.004	0.106	-0.005	0.104	-0.003	0.103	-0.003
-0.25	0.00	2	0.107	-0.003	0.107	-0.006	0.104	-0.004	0.103	-0.003
0.00	0.00	2	0.106	-0.004	0.105	-0.005	0.104	-0.004	0.103	-0.003
0.25	0.00	2	0.105	-0.005	0.105	-0.005	0.107	-0.004	0.103	-0.003
0.75	0.00	2	0.139	0.004	0.107	-0.005	0.108	-0.005	0.103	-0.003
0.90	0.00	2	0.116	-0.006	0.106	-0.006	0.108	-0.005	0.103	-0.003
-0.90	0.25	2	0.089	-0.006	0.090	-0.006	0.090	-0.005	0.087	-0.002
-0.75	0.25	2	0.088	-0.006	0.090	-0.005	0.089	-0.005	0.087	-0.002
-0.25	0.25	2	0.089	-0.006	0.090	-0.007	0.088	-0.006	0.087	-0.002
0.00	0.25	2	0.089	-0.006	0.088	-0.006	0.087	-0.006	0.087	-0.002
0.25	0.25	2	0.090	-0.008	0.088	-0.007	0.089	-0.008	0.087	-0.002
0.75	0.25	2	0.130	-0.037	0.090	-0.007	0.095	-0.010	0.087	-0.002
0.90	0.25	2	0.099	-0.013	0.089	-0.006	0.093	-0.009	0.087	-0.002
-0.90	0.50	2	0.068	-0.007	0.068	-0.006	0.069	-0.006	0.068	-0.002
-0.75	0.50	2	0.068	-0.006	0.068	-0.007	0.068	-0.006	0.068	-0.002
-0.25	0.50	2	0.069	-0.007	0.069	-0.007	0.069	-0.007	0.068	-0.002
0.00	0.50	2	0.069	-0.008	0.068	-0.007	0.070	-0.009	0.068	-0.002
0.25	0.50	2	0.067	-0.009	0.069	-0.007	0.071	-0.010	0.068	-0.002
0.75	0.50	2	0.124	-0.058	0.070	-0.009	0.077	-0.016	0.068	-0.002
0.90	0.50	2	0.080	-0.020	0.070	-0.009	0.073	-0.013	0.068	-0.002
-0.90	0.90	2	0.028	-0.007	0.027	-0.007	0.027	-0.008	0.025	0.000
-0.75	0.90	2	0.028	-0.008	0.027	-0.007	0.027	-0.008	0.025	0.000
-0.25	0.90	2	0.031	-0.012	0.028	-0.009	0.030	-0.011	0.025	0.000
0.00	0.90	2	0.033	-0.016	0.029	-0.010	0.033	-0.014	0.025	0.000
0.25	0.90	2	0.040	-0.022	0.030	-0.012	0.039	-0.020	0.025	0.000
0.75	0.90	2	0.056	-0.031	0.030	-0.011	0.050	-0.027	0.025	0.000
0.90	0.90	2	0.041	-0.018	0.029	-0.010	0.044	-0.022	0.025	0.000
-0.90	-0.90	3	0.194	0.003	0.193	0.001	0.189	0.001	0.187	-0.011
-0.75	-0.90	3	0.193	0.003	0.193	0.000	0.187	0.002	0.187	-0.011
-0.25	-0.90	3	0.189	0.006	0.189	0.000	0.185	0.005	0.187	-0.011
0.00	-0.90	3	0.187	0.015	0.187	0.000	0.183	0.007	0.187	-0.011
0.25	-0.90	3	0.191	0.034	0.185	0.001	0.179	0.013	0.187	-0.011
0.75	-0.90	3	0.395	0.255	0.183	0.004	0.189	0.039	0.187	-0.011
0.90	-0.90	3	0.204	0.047	0.188	0.002	0.187	0.023	0.187	-0.011

Table D3 cont.

Unweighted Spatial GM Estimators of ρ

Initial Estimator / True Values			AH1		AH2		AB		True	
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	-0.50	3	0.170	-0.005	0.171	-0.005	0.170	-0.004	0.168	-0.010
-0.75	-0.50	3	0.171	-0.004	0.171	-0.005	0.168	-0.002	0.168	-0.010
-0.25	-0.50	3	0.169	-0.002	0.170	-0.006	0.166	-0.003	0.168	-0.010
0.00	-0.50	3	0.167	0.002	0.170	-0.006	0.165	-0.003	0.168	-0.010
0.25	-0.50	3	0.167	0.009	0.167	-0.007	0.164	-0.002	0.168	-0.010
0.75	-0.50	3	0.261	0.127	0.167	-0.003	0.170	0.011	0.168	-0.010
0.90	-0.50	3	0.172	0.020	0.171	-0.006	0.167	0.006	0.168	-0.010
-0.90	-0.25	3	0.154	-0.006	0.155	-0.007	0.154	-0.006	0.152	-0.009
-0.75	-0.25	3	0.155	-0.006	0.155	-0.006	0.152	-0.006	0.152	-0.009
-0.25	-0.25	3	0.154	-0.005	0.155	-0.007	0.150	-0.007	0.152	-0.009
0.00	-0.25	3	0.152	-0.005	0.154	-0.008	0.149	-0.007	0.152	-0.009
0.25	-0.25	3	0.150	-0.001	0.152	-0.009	0.150	-0.006	0.152	-0.009
0.75	-0.25	3	0.199	0.054	0.150	-0.007	0.151	0.001	0.152	-0.009
0.90	-0.25	3	0.152	0.001	0.153	-0.007	0.151	0.001	0.152	-0.009
-0.90	0.00	3	0.136	-0.007	0.135	-0.008	0.133	-0.009	0.132	-0.009
-0.75	0.00	3	0.136	-0.007	0.136	-0.007	0.133	-0.007	0.132	-0.009
-0.25	0.00	3	0.135	-0.007	0.134	-0.009	0.131	-0.008	0.132	-0.009
0.00	0.00	3	0.132	-0.008	0.134	-0.009	0.131	-0.009	0.132	-0.009
0.25	0.00	3	0.134	-0.007	0.133	-0.010	0.132	-0.008	0.132	-0.009
0.75	0.00	3	0.166	0.001	0.134	-0.009	0.132	-0.005	0.132	-0.009
0.90	0.00	3	0.136	-0.009	0.132	-0.010	0.132	-0.005	0.132	-0.009
-0.90	0.25	3	0.111	-0.007	0.112	-0.008	0.109	-0.008	0.109	-0.007
-0.75	0.25	3	0.111	-0.008	0.113	-0.008	0.109	-0.008	0.109	-0.007
-0.25	0.25	3	0.112	-0.009	0.111	-0.009	0.109	-0.010	0.109	-0.007
0.00	0.25	3	0.111	-0.008	0.111	-0.009	0.110	-0.011	0.109	-0.007
0.25	0.25	3	0.111	-0.011	0.110	-0.009	0.111	-0.011	0.109	-0.007
0.75	0.25	3	0.148	-0.040	0.111	-0.011	0.111	-0.013	0.109	-0.007
0.90	0.25	3	0.115	-0.019	0.112	-0.012	0.109	-0.011	0.109	-0.007
-0.90	0.50	3	0.083	-0.007	0.084	-0.007	0.083	-0.007	0.082	-0.006
-0.75	0.50	3	0.083	-0.007	0.085	-0.007	0.083	-0.008	0.082	-0.006
-0.25	0.50	3	0.085	-0.009	0.084	-0.009	0.084	-0.010	0.082	-0.006
0.00	0.50	3	0.085	-0.010	0.085	-0.009	0.085	-0.011	0.082	-0.006
0.25	0.50	3	0.084	-0.013	0.084	-0.009	0.085	-0.011	0.082	-0.006
0.75	0.50	3	0.139	-0.059	0.087	-0.012	0.088	-0.017	0.082	-0.006
0.90	0.50	3	0.094	-0.022	0.087	-0.012	0.086	-0.016	0.082	-0.006

Table D3 cont.

Unweighted Spatial GM Estimators of ρ

Initial Estimator True Values			AH1		AH2		AB		True	
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.90	3	0.033	-0.007	0.033	-0.006	0.032	-0.006	0.031	0.001
-0.75	0.90	3	0.034	-0.007	0.033	-0.007	0.032	-0.007	0.031	0.001
-0.25	0.90	3	0.035	-0.012	0.033	-0.009	0.035	-0.010	0.031	0.001
0.00	0.90	3	0.038	-0.016	0.035	-0.011	0.037	-0.015	0.031	0.001
0.25	0.90	3	0.044	-0.022	0.036	-0.012	0.044	-0.020	0.031	0.001
0.75	0.90	3	0.066	-0.035	0.037	-0.012	0.054	-0.029	0.031	0.001
0.90	0.90	3	0.052	-0.020	0.037	-0.010	0.053	-0.023	0.031	0.001

Table D4

Weighted Spatial GM Estimators of ρ

Initial Estimator / True Values			AH1		AH2		AB		True	
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	-0.90	1	0.036	0.017	0.036	0.017	0.041	0.021	0.023	0.000
-0.75	-0.90	1	0.037	0.018	0.036	0.017	0.041	0.022	0.023	0.000
-0.25	-0.90	1	0.044	0.026	0.039	0.020	0.047	0.027	0.023	0.000
0.00	-0.90	1	0.051	0.032	0.039	0.022	0.053	0.032	0.023	0.000
0.25	-0.90	1	0.069	0.044	0.041	0.024	0.063	0.041	0.023	0.000
0.75	-0.90	1	0.087	0.060	0.039	0.021	0.073	0.051	0.023	0.000
0.90	-0.90	1	0.052	0.032	0.037	0.019	0.059	0.039	0.023	0.000
-0.90	-0.50	1	0.048	0.002	0.049	0.003	0.048	0.004	0.047	-0.001
-0.75	-0.50	1	0.048	0.003	0.049	0.003	0.049	0.004	0.047	-0.001
-0.25	-0.50	1	0.049	0.006	0.049	0.004	0.049	0.006	0.047	-0.001
0.00	-0.50	1	0.049	0.009	0.048	0.004	0.048	0.007	0.047	-0.001
0.25	-0.50	1	0.052	0.015	0.048	0.006	0.050	0.010	0.047	-0.001
0.75	-0.50	1	0.117	0.070	0.049	0.006	0.054	0.018	0.047	-0.001
0.90	-0.50	1	0.058	0.020	0.048	0.004	0.051	0.013	0.047	-0.001
-0.90	-0.25	1	0.054	0.000	0.054	0.000	0.054	0.000	0.054	-0.002
-0.75	-0.25	1	0.054	0.000	0.055	0.000	0.053	0.000	0.054	-0.002
-0.25	-0.25	1	0.054	0.001	0.054	0.000	0.053	0.000	0.054	-0.002
0.00	-0.25	1	0.053	0.002	0.053	0.001	0.053	0.001	0.054	-0.002
0.25	-0.25	1	0.054	0.006	0.053	0.001	0.053	0.002	0.054	-0.002
0.75	-0.25	1	0.089	0.042	0.054	0.002	0.055	0.007	0.054	-0.002
0.90	-0.25	1	0.058	0.012	0.055	0.000	0.052	0.004	0.054	-0.002
-0.90	0.00	1	0.057	-0.002	0.057	-0.002	0.057	-0.002	0.056	-0.003
-0.75	0.00	1	0.057	-0.001	0.057	-0.002	0.056	-0.002	0.056	-0.003
-0.25	0.00	1	0.057	-0.002	0.057	-0.002	0.056	-0.003	0.056	-0.003
0.00	0.00	1	0.055	-0.003	0.056	-0.003	0.056	-0.003	0.056	-0.003
0.25	0.00	1	0.056	-0.002	0.055	-0.002	0.055	-0.003	0.056	-0.003
0.75	0.00	1	0.074	-0.001	0.057	-0.002	0.057	-0.002	0.056	-0.003
0.90	0.00	1	0.062	-0.001	0.057	-0.001	0.056	-0.001	0.056	-0.003
-0.90	0.25	1	0.056	-0.003	0.056	-0.003	0.055	-0.003	0.054	-0.002
-0.75	0.25	1	0.056	-0.003	0.056	-0.004	0.056	-0.004	0.054	-0.002
-0.25	0.25	1	0.056	-0.004	0.055	-0.004	0.055	-0.005	0.054	-0.002
0.00	0.25	1	0.055	-0.006	0.056	-0.005	0.055	-0.006	0.054	-0.002
0.25	0.25	1	0.057	-0.008	0.055	-0.005	0.056	-0.007	0.054	-0.002
0.75	0.25	1	0.095	-0.050	0.056	-0.006	0.057	-0.010	0.054	-0.002
0.90	0.25	1	0.061	-0.011	0.056	-0.004	0.057	-0.007	0.054	-0.002

Table D4 cont.

Weighted Spatial GM Estimators of ρ

Initial Estimator / True Values			AH1		AH2		AB		True	
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.50	1	0.049	-0.006	0.049	-0.006	0.050	-0.007	0.047	-0.002
-0.75	0.50	1	0.050	-0.006	0.049	-0.006	0.050	-0.007	0.047	-0.002
-0.25	0.50	1	0.050	-0.008	0.050	-0.007	0.050	-0.009	0.047	-0.002
0.00	0.50	1	0.053	-0.011	0.051	-0.008	0.051	-0.010	0.047	-0.002
0.25	0.50	1	0.056	-0.015	0.051	-0.008	0.053	-0.013	0.047	-0.002
0.75	0.50	1	0.120	-0.076	0.050	-0.010	0.057	-0.021	0.047	-0.002
0.90	0.50	1	0.059	-0.018	0.050	-0.007	0.053	-0.015	0.047	-0.002
-0.90	0.90	1	0.037	-0.018	0.037	-0.017	0.041	-0.020	0.022	-0.001
-0.75	0.90	1	0.038	-0.019	0.038	-0.018	0.042	-0.021	0.022	-0.001
-0.25	0.90	1	0.044	-0.025	0.039	-0.021	0.047	-0.026	0.022	-0.001
0.00	0.90	1	0.053	-0.032	0.041	-0.023	0.052	-0.031	0.022	-0.001
0.25	0.90	1	0.071	-0.047	0.041	-0.024	0.065	-0.041	0.022	-0.001
0.75	0.90	1	0.094	-0.066	0.040	-0.022	0.078	-0.054	0.022	-0.001
0.90	0.90	1	0.055	-0.035	0.037	-0.020	0.061	-0.041	0.022	-0.001
-0.90	-0.90	2	0.115	0.005	0.115	0.005	0.118	0.007	0.118	-0.001
-0.75	-0.90	2	0.117	0.007	0.114	0.006	0.116	0.009	0.118	-0.001
-0.25	-0.90	2	0.116	0.011	0.115	0.009	0.118	0.014	0.118	-0.001
0.00	-0.90	2	0.120	0.019	0.116	0.009	0.121	0.018	0.118	-0.001
0.25	-0.90	2	0.126	0.034	0.117	0.012	0.128	0.025	0.118	-0.001
0.75	-0.90	2	0.307	0.210	0.119	0.016	0.134	0.041	0.118	-0.001
0.90	-0.90	2	0.140	0.046	0.117	0.011	0.125	0.026	0.118	-0.001
-0.90	-0.50	2	0.111	0.002	0.110	0.001	0.110	0.002	0.110	-0.001
-0.75	-0.50	2	0.110	0.002	0.110	0.001	0.110	0.002	0.110	-0.001
-0.25	-0.50	2	0.109	0.003	0.108	0.003	0.109	0.004	0.110	-0.001
0.00	-0.50	2	0.110	0.006	0.110	0.003	0.110	0.006	0.110	-0.001
0.25	-0.50	2	0.111	0.014	0.110	0.006	0.114	0.008	0.110	-0.001
0.75	-0.50	2	0.191	0.108	0.112	0.005	0.115	0.013	0.110	-0.001
0.90	-0.50	2	0.123	0.021	0.110	0.005	0.113	0.010	0.110	-0.001
-0.90	-0.25	2	0.102	-0.001	0.103	-0.001	0.103	0.001	0.102	-0.002
-0.75	-0.25	2	0.102	-0.001	0.102	-0.001	0.102	0.001	0.102	-0.002
-0.25	-0.25	2	0.101	0.000	0.102	-0.001	0.101	0.000	0.102	-0.002
0.00	-0.25	2	0.100	0.001	0.102	0.001	0.102	0.000	0.102	-0.002
0.25	-0.25	2	0.101	0.005	0.104	0.002	0.103	0.002	0.102	-0.002
0.75	-0.25	2	0.145	0.051	0.103	0.001	0.105	0.006	0.102	-0.002
0.90	-0.25	2	0.111	0.006	0.103	0.001	0.105	0.003	0.102	-0.002

Table D4 cont.

Weighted Spatial GM Estimators of ρ

Initial Estimator			AH1		AH2		AB		True	
True Values										
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.00	2	0.092	-0.002	0.092	-0.002	0.093	-0.002	0.091	-0.002
-0.75	0.00	2	0.091	-0.002	0.092	-0.002	0.093	-0.002	0.091	-0.002
-0.25	0.00	2	0.091	-0.002	0.092	-0.001	0.094	-0.002	0.091	-0.002
0.00	0.00	2	0.091	-0.002	0.092	-0.001	0.094	-0.001	0.091	-0.002
0.25	0.00	2	0.090	0.000	0.092	-0.001	0.093	-0.001	0.091	-0.002
0.75	0.00	2	0.129	0.005	0.091	-0.003	0.094	-0.003	0.091	-0.002
0.90	0.00	2	0.104	-0.003	0.093	-0.003	0.096	-0.003	0.091	-0.002
-0.90	0.25	2	0.080	-0.003	0.080	-0.003	0.082	-0.003	0.079	-0.002
-0.75	0.25	2	0.080	-0.003	0.080	-0.002	0.081	-0.004	0.079	-0.002
-0.25	0.25	2	0.078	-0.005	0.080	-0.003	0.080	-0.004	0.079	-0.002
0.00	0.25	2	0.078	-0.004	0.080	-0.003	0.081	-0.004	0.079	-0.002
0.25	0.25	2	0.079	-0.004	0.080	-0.004	0.080	-0.005	0.079	-0.002
0.75	0.25	2	0.122	-0.033	0.079	-0.005	0.083	-0.009	0.079	-0.002
0.90	0.25	2	0.086	-0.010	0.081	-0.005	0.084	-0.006	0.079	-0.002
-0.90	0.50	2	0.065	-0.005	0.064	-0.005	0.066	-0.004	0.065	-0.001
-0.75	0.50	2	0.065	-0.005	0.064	-0.004	0.065	-0.004	0.065	-0.001
-0.25	0.50	2	0.064	-0.005	0.066	-0.005	0.066	-0.007	0.065	-0.001
0.00	0.50	2	0.064	-0.006	0.065	-0.005	0.067	-0.008	0.065	-0.001
0.25	0.50	2	0.064	-0.009	0.065	-0.007	0.067	-0.009	0.065	-0.001
0.75	0.50	2	0.123	-0.060	0.065	-0.008	0.071	-0.014	0.065	-0.001
0.90	0.50	2	0.072	-0.016	0.067	-0.007	0.069	-0.010	0.065	-0.001
-0.90	0.90	2	0.033	-0.009	0.032	-0.009	0.033	-0.010	0.029	0.000
-0.75	0.90	2	0.034	-0.011	0.033	-0.010	0.033	-0.010	0.029	0.000
-0.25	0.90	2	0.037	-0.016	0.034	-0.012	0.037	-0.014	0.029	0.000
0.00	0.90	2	0.040	-0.020	0.035	-0.013	0.040	-0.017	0.029	0.000
0.25	0.90	2	0.049	-0.028	0.037	-0.015	0.048	-0.025	0.029	0.000
0.75	0.90	2	0.068	-0.040	0.039	-0.014	0.059	-0.035	0.029	0.000
0.90	0.90	2	0.048	-0.023	0.036	-0.013	0.053	-0.028	0.029	0.000
-0.90	-0.90	3	0.166	0.013	0.165	0.014	0.167	0.016	0.165	0.003
-0.75	-0.90	3	0.167	0.013	0.164	0.014	0.167	0.015	0.165	0.003
-0.25	-0.90	3	0.164	0.018	0.160	0.012	0.163	0.016	0.165	0.003
0.00	-0.90	3	0.165	0.021	0.162	0.014	0.161	0.019	0.165	0.003
0.25	-0.90	3	0.173	0.033	0.162	0.016	0.167	0.027	0.165	0.003
0.75	-0.90	3	0.366	0.239	0.167	0.018	0.177	0.043	0.165	0.003
0.90	-0.90	3	0.182	0.049	0.167	0.010	0.164	0.027	0.165	0.003

Table D4 cont.

Weighted Spatial GM Estimators of ρ

Initial Estimator / True Values			AH1		AH2		AB		True	
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	-0.50	3	0.148	0.005	0.148	0.007	0.146	0.007	0.147	0.002
-0.75	-0.50	3	0.147	0.006	0.145	0.006	0.146	0.006	0.147	0.002
-0.25	-0.50	3	0.148	0.009	0.144	0.006	0.146	0.007	0.147	0.002
0.00	-0.50	3	0.146	0.012	0.142	0.007	0.145	0.007	0.147	0.002
0.25	-0.50	3	0.144	0.017	0.143	0.008	0.144	0.013	0.147	0.002
0.75	-0.50	3	0.240	0.122	0.147	0.006	0.144	0.018	0.147	0.002
0.90	-0.50	3	0.157	0.022	0.145	0.002	0.144	0.012	0.147	0.002
-0.90	-0.25	3	0.135	0.004	0.133	0.004	0.133	0.004	0.134	0.002
-0.75	-0.25	3	0.135	0.003	0.134	0.004	0.132	0.003	0.134	0.002
-0.25	-0.25	3	0.133	0.004	0.130	0.003	0.131	0.003	0.134	0.002
0.00	-0.25	3	0.132	0.005	0.129	0.004	0.128	0.004	0.134	0.002
0.25	-0.25	3	0.131	0.011	0.128	0.003	0.130	0.007	0.134	0.002
0.75	-0.25	3	0.177	0.058	0.132	0.002	0.131	0.005	0.134	0.002
0.90	-0.25	3	0.140	0.011	0.133	-0.001	0.132	0.005	0.134	0.002
-0.90	0.00	3	0.118	0.002	0.117	0.002	0.117	0.001	0.117	0.000
-0.75	0.00	3	0.118	0.001	0.117	0.001	0.116	0.001	0.117	0.000
-0.25	0.00	3	0.116	0.001	0.114	0.002	0.114	0.001	0.117	0.000
0.00	0.00	3	0.115	0.001	0.115	0.001	0.114	0.001	0.117	0.000
0.25	0.00	3	0.113	0.002	0.113	0.000	0.114	0.001	0.117	0.000
0.75	0.00	3	0.145	0.006	0.115	-0.002	0.117	-0.001	0.117	0.000
0.90	0.00	3	0.123	-0.003	0.119	-0.004	0.119	-0.002	0.117	0.000
-0.90	0.25	3	0.099	-0.001	0.098	-0.002	0.098	-0.003	0.098	-0.001
-0.75	0.25	3	0.098	-0.001	0.098	-0.001	0.098	-0.003	0.098	-0.001
-0.25	0.25	3	0.095	-0.002	0.096	-0.002	0.097	-0.003	0.098	-0.001
0.00	0.25	3	0.097	-0.002	0.096	-0.002	0.096	-0.004	0.098	-0.001
0.25	0.25	3	0.097	-0.005	0.097	-0.003	0.097	-0.003	0.098	-0.001
0.75	0.25	3	0.133	-0.038	0.097	-0.006	0.099	-0.006	0.098	-0.001
0.90	0.25	3	0.105	-0.011	0.103	-0.006	0.100	-0.008	0.098	-0.001
-0.90	0.50	3	0.076	-0.004	0.075	-0.004	0.075	-0.004	0.076	-0.002
-0.75	0.50	3	0.075	-0.004	0.075	-0.004	0.076	-0.005	0.076	-0.002
-0.25	0.50	3	0.075	-0.005	0.075	-0.004	0.076	-0.006	0.076	-0.002
0.00	0.50	3	0.075	-0.006	0.076	-0.004	0.077	-0.007	0.076	-0.002
0.25	0.50	3	0.078	-0.009	0.076	-0.005	0.078	-0.008	0.076	-0.002
0.75	0.50	3	0.137	-0.064	0.078	-0.008	0.081	-0.012	0.076	-0.002
0.90	0.50	3	0.088	-0.016	0.078	-0.008	0.079	-0.010	0.076	-0.002

Table D4 cont.

Weighted Spatial GM Estimators of ρ

Initial Estimator			AH1		AH2		AB		True	
True Values										
Φ	ρ	W	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
-0.90	0.90	3	0.038	-0.011	0.038	-0.010	0.037	-0.011	0.035	-0.001
-0.75	0.90	3	0.038	-0.011	0.039	-0.010	0.037	-0.011	0.035	-0.001
-0.25	0.90	3	0.042	-0.015	0.039	-0.012	0.040	-0.014	0.035	-0.001
0.00	0.90	3	0.046	-0.021	0.040	-0.014	0.044	-0.019	0.035	-0.001
0.25	0.90	3	0.055	-0.029	0.042	-0.016	0.053	-0.027	0.035	-0.001
0.75	0.90	3	0.077	-0.045	0.041	-0.016	0.066	-0.036	0.035	-0.001
0.90	0.90	3	0.057	-0.027	0.041	-0.014	0.059	-0.030	0.035	-0.001

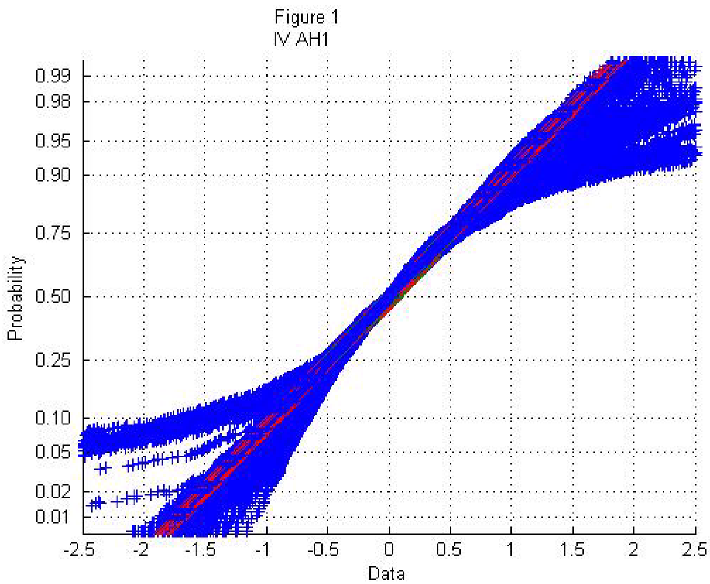

Figure1: QQ Plot of IV Estimator AH1

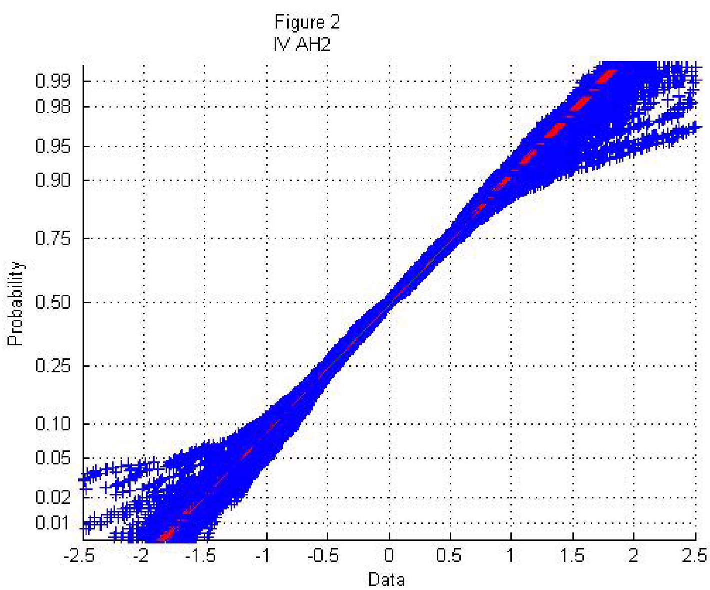

Figure 2: QQ Plot of IV Estimator AH2

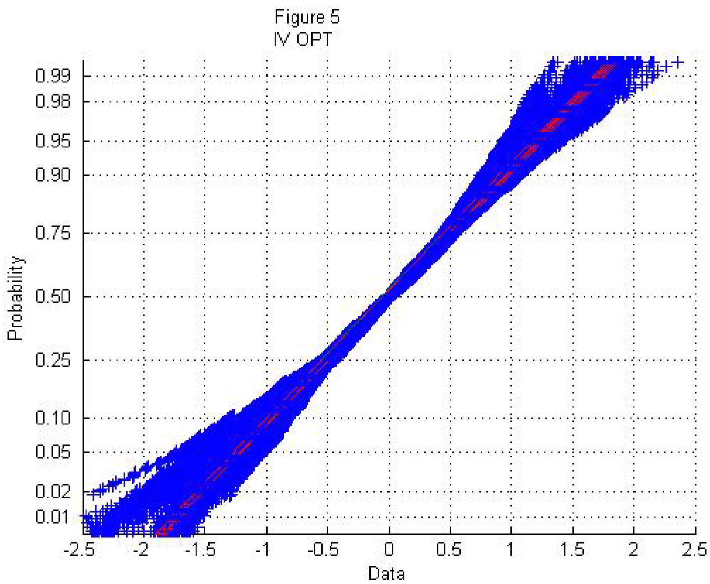

Figure 3: QQ Plot of IV Estimator AB

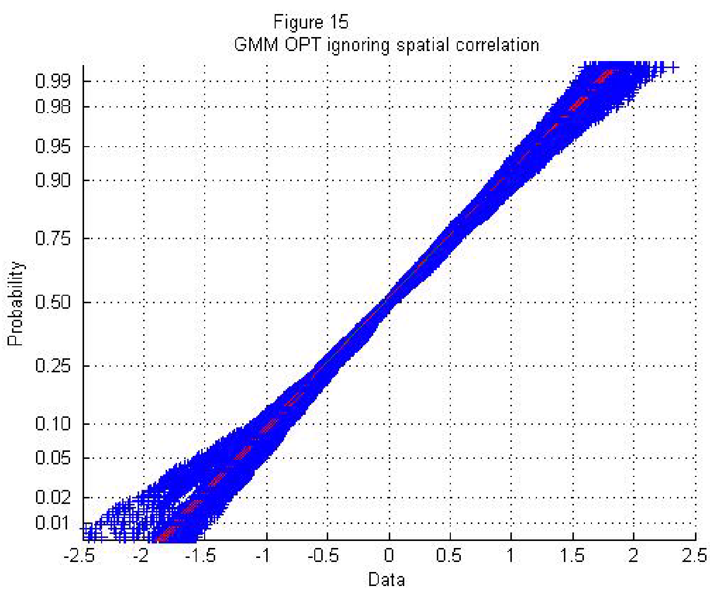

Figure 4: QQ Plot of GMM Estimator AB Ignoring Spatial Correlation

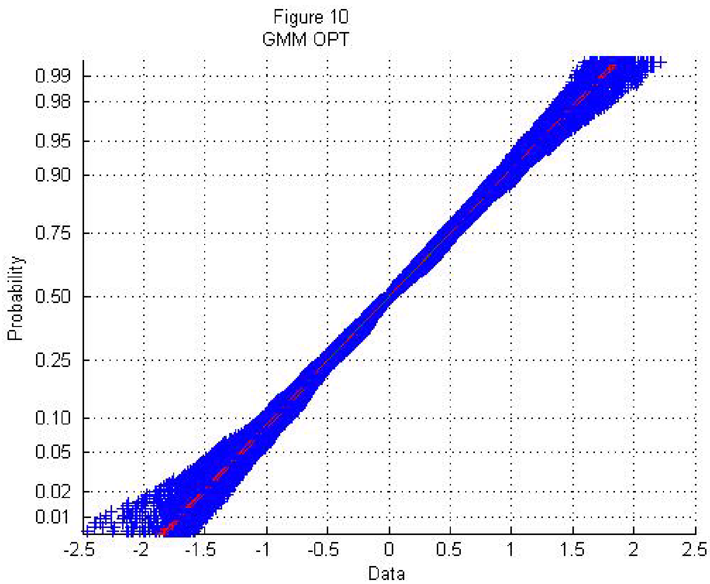

Figure 5: QQ Plot of GMM Estimator AB based on $\widehat{\mathbf{V}}^{mix}$

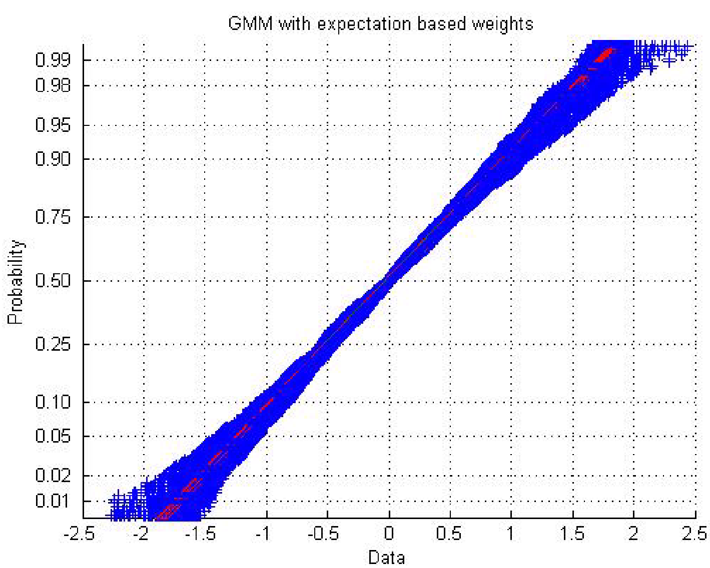

Figure 6: QQ Plot of GMM Estimator AB based on $\widehat{\mathbf{V}}^{E}$

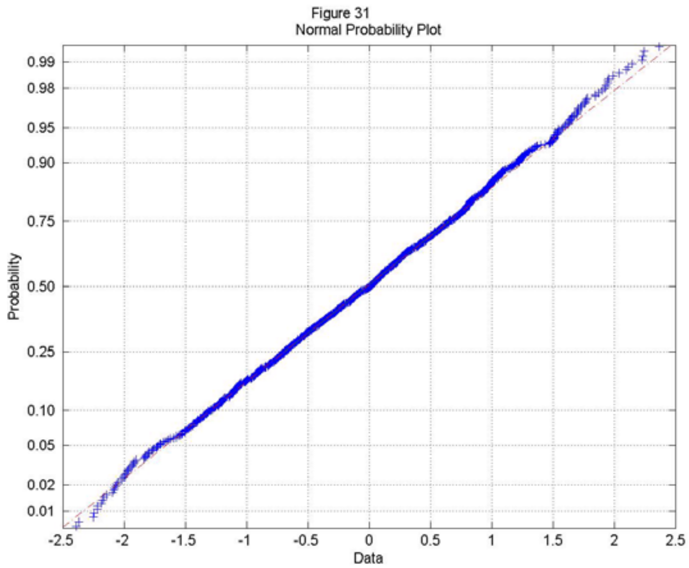

Figure 7: Normal Probability QQ Plot

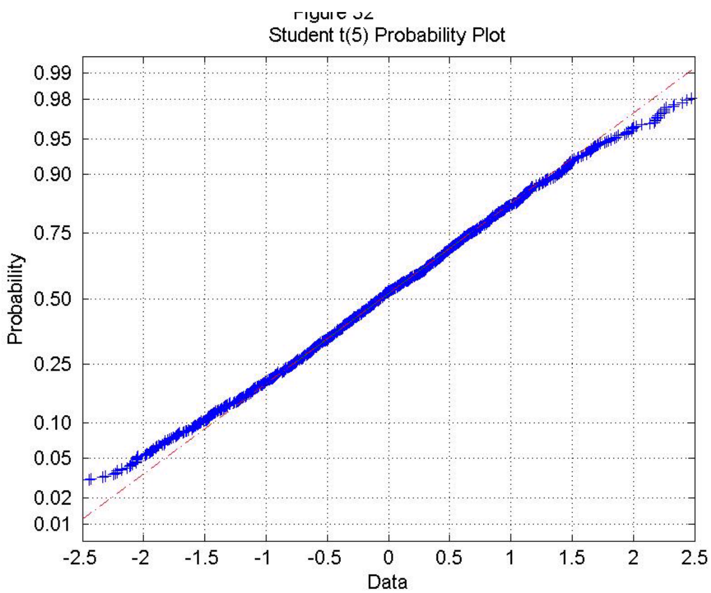

Figure 8: Student t Probability QQ Plot

E Appendix: Symbols and Notation Used

In this Appendix, I provide a brief explanation of the different (standard) symbols used throughout the thesis.

N	cross-sectional dimension of the data under consideration
T	time dimension of the data under consideration

$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbf{e}_T$	$T \times 1$ vector of ones
$\mathbf{J}_T$	$T \times T$ matrix of ones
$\mathbf{Q}_0$	transformation matrix that subtracts location specific sample means
$\mathbf{Q}_1$	transformation matrix that calculates location specific sample means
Δ	first difference operator (in time dimension)
$\mathbf{D}$	first difference transformation matrix

$\forall$	for all (logical predicate)
$\exists$	exists (logical predicate)
$\in$	relation operator 'belongs to a set'
∞	infinity
$\mathbb{R}$	set of real numbers
$\mathbb{N}$	set of natural numbers

$\sigma(x)$ neighborhood of a real number x

$\sup$ supremum

$\inf$ infimum

$\min$ minimum

$\arg\min_{\delta\in\Theta}\{\}$ argument that maximizes a maximization problem in brackets with parameters δ restricted to a set Θ

$\overline{\lim}_{n\to\infty} a_n$ limes superior of the sequence a_n

$\otimes$ Kronecker product operator

$\|\mathbf{M}\|$ matrix norm $[tr(\mathbf{M}'\mathbf{M})]^{1/2}$

$\lambda_{\min}(\mathbf{\Omega})$ smallest eigenvalue of a matrix $\mathbf{\Omega}$

$diag(d_1, ..., d_N)$ diagonal matrix with $d_1, ..., d_N$ on the main diagonal

$E(\mathbf{y})$ expected value of a vector/scalar $\mathbf{y}$

$VC(\mathbf{y})$ variance covariance matrix of a vector $\mathbf{y}$

$Cov(z_1, z_2)$ covariance of a two scalar random variables

$\xrightarrow{d}$ convergence in distribution

$\xrightarrow{p}$ convergence in probability

$\xrightarrow{r}$ convergence in r-th mean

$N(\mathbf{x}, \mathbf{\Omega})$ multivariate normal distribution with mean $\mathbf{x}$ and variance covariance matrix $\mathbf{\Omega}$

L^p space of random variables with finite p-th absolute moments

$|x|$ absolute value of a number/random variable

$\|\chi\|_r$ $[E(\chi^r)]^{1/r}$

$O_p(k)$ sequence random variables is of order in probability of at most N^k

$O(k)$ deterministic sequence is of order of at most N^k

2SLS two stage least squares

3SLS three stage least squares

CV covariance (estimator)

GLS generalized least squares

GM generalized moments

GMM generalized method of moments

HAC heteroscedasticity and autocorrelation consistent

IV instrumental variable

LIML limited information maximum likelihood

LSDV least-squares dummy variable (estimator)

MD minimum distance

ML maximum likelihood

OLS ordinary least squares

SAR spatial autoregressive

STAR space-time autoregressive

STARMA space-time autoregressive moving average

SUR seemingly unrelated regressions

VAR vector autoregressive

VARMA vector autoregressive moving average

WG within group

F Appendix: Inequalities

In this Appendix, I provide a list of inequalities used throughout the thesis. The following is based on, e.g. Bierens (1994), Section 1.4.

F.1 Deterministic Inequalities

(**Bernoulli**) Let $x \in \mathbb{R}, x > 1$ and $n \in \mathbb{N}$. Then

$$(1+x)^n \geq 1 + nx, \tag{C.1.1}$$

with the inequality being sharp for $x \neq 0$ and $n > 1$.

(**Triangle**) Let $x, y \in \mathbb{C}$. Then

$$|x| - |y| \leq |x \pm y| \leq |x| + |y| \,. \tag{C.1.2}$$

F.2 Stochastic Inequalities

(**Chebyshev**) Let X be a non-negative random variable with a finite mean μ_X and finite variance σ_X^2. Then for any $\varepsilon \in \mathbb{R}, \varepsilon > 0$

$$P\left(|X - \mu_X| > \sqrt{\frac{\sigma_X^2}{\varepsilon}}\right) \leq \varepsilon. \tag{C.2.3}$$

(**Holder**) Let X and Y be random variables and let $p, q \in \mathbb{R}, p > 1, \frac{1}{p} + \frac{1}{q} = 1$.

Then

$$E\left(|XY|\right) \leq \left[E\left(|X|^{p}\right)\right]^{\frac{1}{p}}\left[E\left(|Y|^{q}\right)\right]^{\frac{1}{q}}. \tag{C.2.4}$$

(**Cauchy-Schwartz**) For $p = q = 2$, we have

$$E\left(|XY|\right) \leq \sqrt{E\left(|X|^{2}\right)}\sqrt{E\left(|Y|^{2}\right)}. \tag{C.2.5}$$

(**Lyapunov**) For $Y = 1$ we have for $p > 1$

$$E\left(|X|\right) \leq \left[E\left(|X|^{p}\right)\right]^{\frac{1}{p}}. \tag{C.2.6}$$

(**Minkowski**) If for some $p \geq 1, E\left(|X|^{p}\right) < \infty$ and $E\left(|Y|^{p}\right) < \infty$, then

$$E\left(|X+Y|\right) \leq \left[E\left(|X|^{p}\right)\right]^{\frac{1}{p}}\left[E\left(|Y|^{p}\right)\right]^{\frac{1}{p}}. \tag{C.2.7}$$

(**Jensen**) Let X be a random variable and $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be a convex real function. Then

$$f\left[E\left(X\right)\right] \leq E\left[f\left(X\right)\right]. \tag{C.2.8}$$

Observe that by selecting the random variables to be constants, the above inequalities can be applied in the deterministic case as well.

Since the mean of a finite number of non-random variables in $\mathbb{R}$ may be considered as mathematical expectations, it follows from Hölder's inequality that for real numbers x_i, y_i, $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$:

$$\left|\sum_{i=1}^{m} x_i y_i\right| \leq \left(\sum_{i=1}^{m} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{m} |y_i|^q\right)^{\frac{1}{q}}. \tag{C.2.9}$$

Similarly from Lyapunov's inequality (or by selecting $y_i = 1$ in the above):

$$\left|\sum_{i=1}^{m} x_i\right|^p \leq m^{p-1} \sum_{i=1}^{m} |x_i|^p, \; p \geq 1. \tag{C.2.10}$$

Finally, by Minkowski's inequality

$$\left|\sum_{i=1}^{m} x_i + y_i\right|^{\frac{1}{p}} \leq \left(\sum_{i=1}^{m} |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{m} |y_i|^q\right)^{\frac{1}{q}}. \tag{C.2.11}$$

Note if x_i and y_i are random variables, then the last three inequalities hold for all their realizations. As a result, we can apply these inequalities also in cases where x_i and y_i are stochastic. The same holds for the triangle inequality.

References

[1] Abraham, B., 1983, The Exact Likelihood for a Space Time Model, *Metrika*, 30, 239-243.

[2] Ahn, S.C., and P. Schmidt, 1995, Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-27.

[3] Alonso-Borrego, C. and M. Arellano, 1999, Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data, *Journal of Business & Economic Statistics*, 17, 36-49.

[4] Alvarez, J. and M. Arellano, 2003, The Time-Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators, *Econometrica*, 71(4), 1121-59.

[5] Anderson, T.W., and C. Hsiao, 1981, Estimation of Dynamic Models with Error Components, *Journal of American Statistical Association*, 76, 598-606.

[6] Anderson, T.W., and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, *Journal of Econometrics*, 18, 47-82.

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

[8] Anselin L. and S. Hudak, 1992, Spatial Econometrics in Practice; A Review of Software Options, *Regional Science and Urban Economics*, 22, 509-536.

[9] Arellano, M., 1989, A Note on the Anderson-Hsiao Estimator for Panel Data, *Economics Letters*, 31, 337-341.

[10] Arellano, M., and S.R. Bond, 1991, Some Tests of Specification for Panel Data: Monte Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297.

[11] Arellano, M., and O. Bover, 1995, Another Look at the Instrumental Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29-51.

[12] Audretsch, D.B., Feldmann, M.P., 1996. R&D spillovers and the geography of innovation and production, *American Economic Review*, 86, 630-640.

[13] Baltagi, B.H., 1995 and 2002, *Econometric Analysis of Panel Data* Wiley, New York.

[14] Baltagi, B.H., Li, D., 2001a. Double length artificial regressions for testing spatial dependence, *Econometric Reviews*, 20, 31-40.

[15] Baltagi, B.H., Li, D., 2001b. LM test for functional form and spatial error correlation, *International Regional Science Review*, 24, 194-225.

[16] Baltagi, B.H., S.H. Song and W. Koh, 2003, Testing Panel Data Regression Models with Spatial Error Correlation, *Journal of Econometrics*,117(1),123-50.

[17] Bartsch, H.J., 1987, *Mathematische Formeln* (VEB Fachbuchverlag, Leipzig, DDR).

[18] Bernat Jr., G., 1996, Does Manufacturing matter? A spatial econometric view of Kaldor's laws, *Journal of Regional Science*, 36, 463-477.

[19] Besley, T., Case, A., 1995. Incumbent behavior: Vote-seeking, taxsetting, and yardstick competition, *American Economic Review*, 85, 25-45.

[20] Bhargava, A. and J.D. Sargan, 1983, Estimating Dynamic Random Effects Models from Panel Models Covering Short Time Periods, *Econometrica*, 51, 1635-1659.

[21] Bierens, H.J., 1994, *Topics in Advanced Econometrics* (Cambridge University Press, Cambridge, U.K.).

[22] Binder, M., C. Hsiao, and M.H. Pesaran, 2002, Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration, Working Paper, University of Maryland, College Park.

[23] Binder, M., C. Hsiao, J. Mutl, and M.H. Pesaran, 2002, Computational Issues in the Estimation of Higher-Order Panel Vector Autoregressions, *mimeo*, University of Maryland, College Park.

[24] Blundell, R.W., and S.R. Bond, 1998, Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 82, 135-156.

[25] Blundell, R.W., Smith, R.J., 1991, Initial conditions and efficient estimation in dynamic panel data models, *Annales d'Economie et de Statistique*, 20/21, 109–123.

[26] Bollinger, C., 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.

[27] Bronnenberg, B.J. and V. Mahajan, 2001, Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables, *Marketing Science*, 20, 284-299.

[28] Brown, P.E., K.F. Karesen, G.O. Roberts and S. Tonellato, 2000, Blur-Generated Non-Separable Space-Time Models, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 62(4), 847-860.

[29] 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.

[30] Case, A., 1991. Spatial patterns in household demand, *Econometrica*, 59, 953-966.

[31] Case, A., Hines Jr., J., Rosen, H., 1993. Budget spillovers and fiscal policy independence: Evidence from the States, *Journal of Public Economics*, 52, 285-307.

[32] Chamberlain, G., 1982, Multivariate Regression Models for Panel Data, *Journal of Econometrics*, 18, 5-46.

[33] Chamberlain, G., 1984, Panel data, Ch. 22 in: Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Vol. II (North-Holland, Amsterdam).

[34] Chang, Y., 2002, Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency, *Journal of Econometrics*, 110(2), 261-92.

[35] Chen, X. and T. Conley, 2001, A New Semiparametric Spatial Model for Panel Time Series, *Journal of Econometrics*, 105(1), 59-83.

[36] Cliff, A. and J. Ord, 1973, *Spatial Autocorrelation* (Pion, London).

[37] Cliff, A. and J. Ord, 1981, *Spatial Processes,Models and Applications* (Pion, London).

[38] Conley, T., 1999, GMM estimation with cross sectional dependence, *Journal of Econometrics* 92, 1-45.

[39] Cressie, N., 1993, *Statistics of Spatial Data* (Wiley, New York).

[40] Das, D., Kelejian, H.H., Prucha, I.R., 2003. Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances, *Papers in Regional Science*, 82, 1-26.

[41] Dhrymes, P.J., 1984, *Mathematics for Econometrics* (Springer-Verlag, New York).

[42] Došlá, Z. and O. Došlý, 1999, *Differenciální počet funkcí více promněnných* (Masarykova univerzita v Brně, Czech Republic).

[43] Dowd, M. R., LeSage, J. P., 1997. Analysis of spatial contiguity influences on state price level formation, *International Journal of Forecasting*, 13, 245-253.

[44] Driscoll, J. and A. Kraay, 1995, Spatial Correlations on Panel Data, *Policy Research Working Paper*, 1553, The World Bank.

[45] Driscoll, J. and A. Kraay, 1998, Consistent covariance matrix estimation with spatially dependent panel data, *The Review of Economics and Statistics,*80, 549-560.

[46] Elhorst, J.P., 2001, Dynamic models in space and time, *Geographical Analysis*, 33, 119–140.

[47] Gänsler, P. and W. Slute, *Wahrscheinlichkeitstheorie* (Springer Verlag, New York).

[48] Giacomini, R. and C.W.J. Granger, 2004, Aggregation of space-time processes, *Journal of Econometrics*, 118, 7-26.

[49] Hahn, J., 1999, How informative is the initial condition in the dynamic panel model with fixed effects, *Journal of Econometrics*, 93, 309–326.

[50] Hahn, J. and G. Kuersteiner, 2002, Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when both n, and T are Large, *Econometrica*, 70(4), 1639-1657.

[51] Haining, R., 1990, *Spatial data Analysis in the Social and Environmental Sciences* (Cambridge University Press, Cambridge).

[52] Hansen, L.P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029-1054.

[53] Hansen, L.P., J. Heaton and A. Yaron, 1996, Finite Sample Properties of Some Alternative GMM Estimators, *Journal of Business & Economic Statistics*, 14, 262-280.

[54] Harris, R.D.F and E. Tzavalis, 1999, Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed, *Journal of Econometrics*, 91, 201-226.

[55] Horn, R.H. and C.H. Johnson, 1985, *Matrix Analysis* (Cambridge University Press, Cambridge).

[56] Horn, R.H. and C.H. Johnson, 1991, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge).

[57] Holtz-Eakin, D., 1994. Public sector capital and the productivity puzzle, *Review of Economics and Statistics*, 76, 12-21.

[58] Holtz-Eakin, D., W. Newey and H. Rosen, 1988, Estimating Vector Autoregressions with Panel Data, *Econometrica*, 56, 1371-1395.

[59] Hordijk, L., 1979, Problems in estimating econometric relations in space, *Papers of the Regional Science Association*, 42, 99-115.

[60] Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu, 2002, Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods, *Journal of Econometrics*, 109(1), 107-50.

[61] Judson, R.A., Owen, A.L., 1999, Estimating dynamic panel data models: a guide for macroeconomists, *Economic Letters*, 65, 9–15.

[62] Karr, Alan F., 1993, *Probability* (Springer-Verlag, New York).

[63] Kapoor, M., H.H. Kelejian and I.R. Prucha, 2005, Panel Data Models with Spatially Correlated Error Components, *Journal of Econometrics*, forthcoming.

[64] Keane, M.P. and D.E. Runkle, 1992, On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous, *Journal of Business & Economic Statistics*, 10(1), 1-9.

[65] Kelejian, H.H., 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.

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

[67] Kelejian, H.H. and I.R. Prucha, 1999, A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* 40, 509-533.

[68] Kelejian, H.H. and I.R. Prucha, 2001, On the asymptotic distribution of the Moran I test statistic with applications, *Journal of Econometrics*, 104, 219-257.

[69] Kelejian, H.H. and I.R. Prucha, 2005, HAC Estimation in a Spatial Framework, Department of Economics, University of Maryland, College Park, mimeo.

[70] Kelejian, H.H. and I.R. Prucha, 2004, Estimation of simultaneous systems of spatially interelated cross sectional equations, *Journal of Econometrics*, 118, 27-50.

[71] Kelejian, H.H., I.R. Prucha and Y. Yuzefovich, 2004, Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results, Department of Economics, University of Maryland, forthcoming in J. LeSage and R.K. Pace, *Advances in Econometrics*, New York: Elsevier.

[72] Kelejian, H.H., Robinson, D., 1997. Infrastructure productivity estimation and its underlying econometric specifications, *Papers in Regional Science*, 76, 115-131.

[73] Kelejian, H.H., Robinson, D., 2000. Returns to investment in navigation infrastructure: An equilibrium approach, *Annals of Regional Science*, 34, 83-108.

[74] Kiviet, J.F., 1995, On bias, inconsistency and efficiency in various estimators of dynamic panel data models, *Journal of Econometrics*, 68, 53–78.

[75] Korniotis, G.M.,2005, A Dynamic Panel Estimator with Both Fixed and Spatial Effects, *mimeo*, University of Notre Dame.

[76] Kyriakidis, P.C and A.G. Journel, 1999, Geostatistical Space-Time Models: A Review, *Mathematical Geology*, 31(6), 651-84.

[77] Lee, L.-F., 2001a, Generalized method of moments estimation of spatial autoregressive processes, Department of Economics, Ohio State University, mimeo.

[78] Lee, L.-F., 2001b, GMM and 2SLS estimation of mixed regressive, spatial autoregressive models, Department of Economics, Ohio State University, mimeo.

[79] Lee, L.F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregessive models, *Econometric Theory*, 18, 252-277.

[80] Lee, L.F., 2003. Best spatial Two-Stage Least Squares estimators for a spatial autoregressive model with autoregressive disturbances, *Econometric Reviews*, 22, 307-335.

[81] Lee, L.-F., 2004, Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models, *Econometrica*, 72(6), 1899-1925.

[82] LeSage, J. P., 1997. Bayesian estimation of spatial autoregressive models, *International Regional Science Review*, 20, 113-129.

[83] LeSage, J. P., 1999. A spatial econometric analysis of China's economic growth, *Journal of Geographic Information Sciences*, 5, 143-153.

[84] LeSage, J. P., 2000. Bayesian estimation of limited dependent variable spatial autoregressive models, *Geographic Analysis*, 32, 19-35.

[85] LeSage, J.P and A. Krivelyova, 1999, A spatial prior for Bayesian vector autoregressive models, *Journal of Regional Science*, 39(2), 297-317.

[86] LeSage,J.P. and R.K. Pace, 2004, Models for Spatially Dependent Missing Data, *Journal of Real Estate Finance and Economics*, 29(2), 233-254.

[87] Mutl, J., 2006, Misspecification of Space: An Illustration Using Growth Convergence Regressions, *Workshop on Spatial Econometrics and Statistics*, Rome, Italy.

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

[89] Nerlove, M., 1967, Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections, *Econometrica*, 39, 359-387.

[90] Newey, W.K. and K.D. West, 1987, A Simple, Positive Semi-Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55(3), 703-708.

[91] Nickell, S., 1981, Biases in Dynamic Models with Fixed Effects, *Econometrica*, 16, 1-32.

[92] Pace, R., Barry, R., 1997. Sparse spatial autoregressions, *Statistics and Probability Letters*, 33, 291-297.

[93] Pace, R.K., R. Barry, J. Clapp and M. Rodriquez, 1998, Spatiotemporal Autoregressive Models of Neighborhood Effects, *Journal of Real Estate Finance and Economics*,17, 15-33.

[94] Pfeifer, P.E., Deutsch, S.J., 1980, A three-stage iterative procedure for space-time modeling, *Technometrics*, 22, 35–47.

[95] Pinkse, J., Slade, M.E., 1998. Contracting in space: An application of spatial statistics to discrete-choice models, *Journal of Econometrics*, 85, 125-154.

[96] Pinkse, J., Slade, M. E., Brett, C., 2002. Spatial price competition: A semiparametric approach, *Econometrica*, 70, 1111-53.

[97] Pötscher, B.M. and I.R. Prucha, 1994, Generic uniform convergence and equicontinuity concepts for random functions: An exploration of the basic structure, *Journal of Econometrics*, 60, 23-63.

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

[99] Pötscher, B.M. and I.R. Prucha, 2001, Basic elements of asymptotic theory. In B.H. Baltagi, ed., *A Companion to Theoretical Econometrics* (Blackwell, New York), 201-229.

[100] Prucha, I.R., 1985, Maximum likelihood and instrumental variable estimation in simultaneous equation systems with error components, *International Economic Review* 26, 491-506.

[101] Prucha, I.R., 2004, Econ 721 Handout: Lag Operators and Difference Equations, *mimeo*, University of Maryland.

[102] Prucha, I.R., 2005, Econ 624 Handout: Classical Nonlinear Econometrics Models, *mimeo*, University of Maryland.

[103] Rao, C.R., 1973, *Linear Statistical Inference and Its Applications* (Wiley, New York).

[104] Rey, S. and M. Boarnet, 1999, A taxonomy of spatial econometric models for simultaneous systems. In L. Anselin and R. Florax, eds., *Advances in Spatial Econometrics* (Springer Verlag, New York).

[105] Sargan, J.D., 1958, The Estimation of Economic Relationships Using Instrumental Variables, *Econometrica*, 26, 393-415.

[106] Schmidt, P., 1976, *Econometrics* (Marcel Dekker, New York).

[107] Sevestre, P. and A. Trognon, 1985, A note on autoregressive error-component models, *Journal of Econometrics*, 29, 231-245.

[108] 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.

[109] Stoffer, D.S., 1986, Estimation and Identification of Space-Time ARMAX Models in the Presence of Missing Data, *Journal of the American Statistical Association*, 81(395), 762-62.

[110] Trognon, A., 1978, Miscellaneous Asymptotic Properties of Ordinary Least Squares and Maximum Likelihood Estimators in Dynamic Error Component Models, *Annales de l'INSEE*, 30, 632-657.

[111] Vigil, R., 1998. Interactions among municipalities in the provision of police services: A spatial econometric approach, University of Maryland, Ph.D. Thesis.

[112] Whittle, P., 1954, On stationary processes in the plane, *Biometrica* 41, 434-449.

[113] Yang, Z., 2005, Quasi-Maximum Likelihood Estimation for Spatial Panel Data Regressions, *mimeo*, Singapore Management University.

[114] Zellner, A., 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Society*, 57, 348-368.