The Spatio-Temporal Autoregressive Distributed Lag Modelling Approach to an Analysis of the Spatial Heterogeneity and Diffusion Dependence*

Yongcheol Shin and Michael Thornton
University of York
Preliminary and incomplete

Abstract

Given the growing availability of large datasets, we propose the spatio-temporal autoregressive distributed lag (STARDL) model which allows spatial and temporal coefficients to differ jointly across the spatial units. Our model encompasses the widely used spatial dynamic panel data models as well as the heterogeneous spatial autoregressive model recently proposed by Aquaro, Bailey and Pesaran (2015), the only paper in considering heterogeneous spatial parameters but without any dynamics. To deal with a degree of simultaneity associated with the spatial-lagged dependent variables, we develop the QML-based and the control function-based STARDL estimators, which are shown to be consistent and asymptotically normally distributed when the time dimension is large, irrespective of whether the number of the spatial units is large or not. Furthermore, by deriving the system dynamic spatial panel data representation, we can develop the dynamic and the diffusion multipliers that can capture dynamic adjustments as well as network connectedness from initial to new equilibrium following an economic perturbation in a flexible manner. The utility of our proposed STARDL models is demonstrated by Monte Carlo studies as well as the empirical application to the Iraqi war casualties during 2003-2010.

Key Words: Spatial Heterogeneity, Diffusion Dependence, STARDL, Control Function, Dynamic and Diffusion Multipliers, Iraqi War Casualties
JEL CODES: C23, C55.

*We are grateful to seminar/conference participants at the University of York and York Econometrics Symposium (26-27 April 2018), and especially Jia Chen, In Choi, Matthew Greenwood-Nimmo, Uih Ran Lee, Rui Lin, Viet Nguyen, Weining Wang, for helpful comments and suggestions. The usual disclaimer applies.

1 Introduction

The ability of spatial econometric models to capture co-dependencies across a known terrain or network at relatively low parametric cost has proved highly attractive to economists and regional scientists. Following early work by Cliff and Ord (1973), Upton and Fingleton (1985), Anselin (1988), Cressie (1993), Kelejian and Robinson (1993) these models have been used in a wide range of applications. The spatial lag or spatial autoregressive model, which captures spatial correlation within the system through the dependent variable has proved the most popular but can be restrictive. A more general form, the spatial Durbin model, is also widely used and allows the explanatory variables of one unit to impact the dependent variable of another directly as well as indirectly through their impact on the original dependent variable. For example, an improvement in school quality in one area will directly improve house prices in neighbouring areas, whose residents may access those schools, as well as indirectly through the transmission of increased house prices from one area to its neighbours. Investigating identification in spatial Durbin models under both instrumental variable and maximum likelihood estimation, Lee and Yu (2016) show that significant biases can arise if relevant Durbin terms are omitted while their unnecessary inclusion causes no material loss of efficiency. Our model is in the spirit of the spatial Durbin model but generalises it to include time as well as spatial lags.

As the availability of spatial datasets with a large time dimension grows it is of interest to explore time dynamics in greater detail. Much of the early work on spatial models was done with large cross-section dimensions in mind. In data with a time dimension, a static model is, in effect, assuming that the spatial system is only observed on (or relatively close to) its equilibrium path. Because many systems take some time to adjust fully to shocks, interest in dynamic spatial models has grown recently, see Elhorst (2014) for an overview, with the spatial panel data model, see Anselin et al. (2009), Baltagi et al. (2003), Lee and Yu (2011), among others, now well established. Yu et al. (2008) study the stable spatial dynamic panel data model, featuring individual time lags, spatial time lags and contemporaneous spatial lags. Maximum likelihood estimation requires bias correction when the time dimension is small but alternative approaches, using time lags as instrumental variables, are available. Elhorst (2010) uses Monte Carlo studies to investigate small sample performances of various estimators. While these endow the system with some temporal memory they are incapable of capturing the dynamics seen in many economic series. We therefore propose to generalise the spatial panel model to higher-order temporal dynamics through the spatial-temporal autoregressive distributed lag (STARDL) model. In time series econometrics, the autoregressive distributed lag (ARDL) model has proved an extremely effective tool for both the estimation of dynamic parameters and for understanding the interaction of variables over time, becoming widely used in particular to differentiate short-run and long-run behaviour and the process of adjustment to new equilibria, see Pesaran and Shin (1998), Pesaran et al. (2001) and Shin et al. (2014). We adopt this approach in the novel context of spatially correlated data.

A significant drawback to spatial models is that the spatial weighting matrix must be known *a priori*. A range of methods have been used to construct weighting matrices in applied work, including contiguity, inverse distance or measures of similarity, with the only restriction that no unit can be its own neighbour. The researcher may also choose to normalise the rows of this matrix, for example to sum to unity. When, as in the overwhelming majority of cases, a common spatial parameter is used this matrix determines not only which units are to be considered as neighbours and their relative importance but also the relative intensity with which each unit is influenced by its neighbours. Only the general intensity of transmission within the system is left to be estimated; the relative degree of openness to transmission or susceptibility to external influence of each unit is assumed known.

We follow a recent paper by Aquaro, Bailey and Pesaran (2015), which is unusual in allowing heterogeneous spatial lag parameters, although our model is more general in allowing for time dynamics and also encompasses the widely used spatial Durbin model as a special case. This relaxation offers a number of advantages. Firstly, the predictions of the model are invariant to any row-normalisation of the spatial weighting matrix, in the sense that the residuals are not affected by multiplying the spatial weights of the neighbours and dividing the spatial coefficient of unit i by an arbitrary constant. Secondly it enables the researcher to estimate the relative openness of each spatial unit. As is common in spatial models, our set up assumes that the relative importance of different neighbours is known while leaving the importance of the spatial effect relative to other influences and relative to other spatial units to be estimated.

Allowing for parameter heterogeneity raises significant complications for estimation and interpretation that this paper addresses. As some simultaneity is inevitable in spatial models, we propose to develop a quasi-maximum likelihood (QML) estimator and one utilising internal instruments. QML techniques, based around a transformation of the data, were developed by Ord (1975), Cliff and Ord (1981) and Anselin (1980) among others and the asymptotic properties studied rigorously in Lee (2004). The evaluation of the quasi-likelihood requires calculation of a Jacobian of a matrix which grows with the cross-section dimension, typically by calculating the eigenvalues of the weighting matrix. As pointed out by Kelejian and Prucha (1998), this becomes compuationally difficult for large cross-sections and the problem is exacerbated by heterogeneity across the spatial parameters, which prohibits the typical approach. In the light of these difficulties we also develop an alternative control function approach, following the instrumental variables/ method of moments approach that has been developed by Anselin (1980), Kelejian and Robinson (1993), Kelejian and Prucha (1998, 1999, 2004), Lee and Liu (2010), Lee and Yu (2014) and Kuersteiner and Prucha (2018). The choice of the optimal set of instruments are also discussed in Anselin (1980), Land and Deane (1992) and Lee (2003). Importantly, the presence of time and spatial correlation equips our model with a greater variety of instruments, which we exploit in developing the estimator of the parameters. We undertake an asymptotic analysis of both estimators, establishing conditions for the stability of the model and the identification of the parameters and show that, under certain conditions they are consistent and normally distributed asymptotically.

We explore the properties of both the QML and control function estimators in a Monte Carlo simulation. Our results indicate that both methods provide good estimates in finite samples, with both bias and root mean square error falling as the time dimension increases. The results are largely unaffected by increases in the cross-sections dimension, with the control function providing a far easier algorithm to implement. The control function estimator is also shown to be robust to heteroskedasticity and to time dependence in the regressor, while the QML estimator has the lower root mean square error. We investigate both methods using different row-normalised weighting matrices and find that performance is maintained regardless of its sparsity.

While offering useful flexibility, allowing heterogeneity across a large number of units can make it difficult to interpret results in a meaningful way, once estimated, and to analyse the influence of both spatial interactions and diffusion dependence. Another contribution of the paper is to provide two such measures, the individual spatio-temporal dynamic multipliers and the system diffusion multipliers, based on work by Shin et al. (2014). We then illustrate their usefulness in an empirical application looking at casualty data from the 2003 Iraq war and its aftermath. Although the conduct of the war itself was initially quick and decisive, subsequent attempts by the coalition forces to maintain the safety of the Iraqi population during the subsequent insurgency required considerable efforts over a longer time horizon. Using novel (and famously unofficial) casualty data, we analyse the effect of insurgent deaths on those of civilians in each of the 18 provinces of Iraq, modelling both time and spatial contaminations. Our results highlight the contrasting experiences of the most populous, the capital Baghdad, where the majority of civilian casualties occurred but which, nevertheless, was largely a net propagator of civilian violence, and the Shia centre of Basrah, where the number of civilian casualties were particularly sensitive to events in neighbouring provinces.

Spatial model are not the only way that cross-section dependence can be modelled and interest in factor models has also grown, with recent efforts to combine the two undertaken by Shi and Lee (2017), Bai and Li (2015) and Kuersteiner and Prucha (2018). As an extension we incorporate observed and then unobserved common factors within our framework, providing algorithms based on the control function and QML approaches.

The structure of the paper is as follows. Section 2 presents the basic specification of the model and discusses in detail its underlying properties. Section 3 presents a number of extensions and discusses their properties. Section 4 develops the spatio-temporal dynamic and diffusion multipliers. Section 5 presents Monte Carlo simulation evidence of the control function and the QML estimators. Section 6 demonstrates the utility of our proposed models, providing an empirical illustration analysing the time and cross sectional dependence between civilian and military casualties during the 2003 Iraq war and its aftermath. Section 7 provides an important extension to the joint modelling of spatial dependence and unobserved factors. Section 8 concludes.

2 The STARDL Model

Consider the spatio-temporal autoregressive distributed lag model of order p and q with the heterogeneous parameters (STARDL(p,q) for short):

$$y_{it} = \sum_{\ell=1}^{p} \phi_{i\ell} y_{i,t-\ell} + \sum_{\ell=0}^{p} \phi_{i\ell}^{*} y_{i,t-\ell}^{*} + \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}' \boldsymbol{x}_{i,t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}'^{*} \boldsymbol{x}_{i,t-\ell}^{*} + \alpha_i + u_{it}, \quad (1)$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where y_{it} is the scalar dependent variable of the ith spatial unit at time t, $\boldsymbol{x}_{it} = (x_{it}^1, ..., x_{it}^K)'$ is a $K \times 1$ vector of exogenous regressors with a $K \times 1$ vector of parameters, $\boldsymbol{\pi}_0 = (\pi_0^1, ..., \pi_0^K)'$. Similarly for $y_{i,t-\ell}$ and $\boldsymbol{x}_{i,t-\ell}$. Spatial interactions between units, both contemporaneously and with lags, are captured via the spatial variables, y_{it}^* and $\boldsymbol{x}_{it}^*$, defined by

$$y_{it}^* \equiv \sum_{j=1}^{N} w_{ij} y_{jt} = \boldsymbol{w}_i \boldsymbol{y}_t \text{ with } \underset{N\times 1}{\boldsymbol{y}_t} = (y_{1t}, ..., y_{Nt})', \quad (2)$$

$$\underset{K\times 1}{\boldsymbol{x}_{it}^*} = \left(x_{it}^{1*}, ..., x_{it}^{K*}\right)' \equiv \left(\sum_{j=1}^{N} w_{ij} x_{jt}^1, ..., \sum_{j=1}^{N} w_{ij} x_{jt}^K\right)' = \left(\boldsymbol{w}_i \otimes \boldsymbol{I}_K\right) \boldsymbol{x}_t; \ \underset{NK\times 1}{\boldsymbol{x}_t} = \begin{pmatrix} \boldsymbol{x}_{1t} \\ \vdots \\ \boldsymbol{x}_{Nt} \end{pmatrix} \quad (3)$$

where $\boldsymbol{w}_i = (w_{i1}, ..., w_{iN})$ denotes a $1 \times N$ vector of (non-stochastic) spatial weights determined *a priori* with $w_{ii} = 0$. Notice that the specification in (1) is sufficiently general by controlling for fixed effects through individual-specific intercepts, α_i.

The STARDL(p,q) specification in (1) reveals some information. If $\phi_{i\ell}$'s and $\boldsymbol{\pi}_{i\ell}$'s are statistically significant, this points to the usual temporal dynamics (e.g. the ARDL approach by Pesaran et al. (2001)). In addition, if $\phi_{i\ell}^*$'s (the endogenous effect) and $\boldsymbol{\pi}_{i\ell}^*$'s (the contextual effect in terms of Manski (1993)) are statistically significant, this indicates an importance of spatial dependence as well as spatio-temporal or diffusion dynamics.

Stacking the N individual STARDL(p,q) regressions (1), we have the following system spatial representation:

$$\boldsymbol{y}_t = \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \boldsymbol{y}_{t-\ell} + \sum_{\ell=0}^{p} \boldsymbol{\Phi}_\ell^* \boldsymbol{W} \boldsymbol{y}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_\ell \boldsymbol{x}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_\ell^* \left(\boldsymbol{W} \otimes \boldsymbol{I}_K\right) \boldsymbol{x}_{t-\ell} + \boldsymbol{\alpha} + \boldsymbol{u}_t \quad (4)$$

where $\boldsymbol{W}$ is the $N \times N$ (non-stochastic) spatial weight or network matrix that characterises all the connections given by

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NN} \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_N \end{bmatrix} \text{ with } w_{ii} = 0, \quad (5)$$

$\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$ and $\boldsymbol{\Phi}_\ell$, $\boldsymbol{\Phi}_\ell^*$, $\boldsymbol{\Pi}_\ell$, $\boldsymbol{\Pi}_\ell^*$ are diagonal matrices consisting of the heterogeneous parameters,

$$\underset{N\times N}{\boldsymbol{\Phi}_\ell} = \begin{bmatrix} \phi_{1\ell} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_{N\ell} \end{bmatrix}, \ell = 1, ...p; \underset{N\times N}{\boldsymbol{\Phi}_\ell^*} = \begin{bmatrix} \phi_{1\ell}^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_{N\ell}^* \end{bmatrix}, \ell = 0, 1, ..., p$$

$$\underset{N\times NK}{\boldsymbol{\Pi}_\ell} = \begin{bmatrix} \boldsymbol{\pi}_{1\ell}' & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\pi}_{N\ell}' \end{bmatrix}, \quad \underset{N\times NK}{\boldsymbol{\Pi}_\ell^*} = \begin{bmatrix} \boldsymbol{\pi}_{1\ell}'^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\pi}_{N\ell}'^* \end{bmatrix} \text{ for } \ell = 0, 1, ..., q.$$

The representation in (4) is more general and nests a range of popular models seen in the literature. Consider the special case with homogeneous parameters and with $p = q = 1$, in which case we obtain the following dynamic spatial Durbin model analysed by Lee and Yu (2009) and Elhorst (2012):

$$\boldsymbol{y}_t = \phi \boldsymbol{y}_{t-1} + \phi_0^* \boldsymbol{W} \boldsymbol{y}_t + \phi_1^* \boldsymbol{W} \boldsymbol{y}_{t-1} + \pi_0 \boldsymbol{x}_t + \pi_1 \boldsymbol{x}_{t-1} + \pi_0^* (\boldsymbol{W} \otimes \boldsymbol{I}_K) \boldsymbol{x}_t + \pi_1^* (\boldsymbol{W} \otimes \boldsymbol{I}_K) \boldsymbol{x}_{t-1} + \boldsymbol{u}_t. \tag{6}$$

In practice it is difficult to provide meaningful interpretation on the homogeneous spatial parameters, especially ϕ_0^* and ϕ_1^* as well as π_0^* and π_1^*. But our proposed approach can deliver much more flexible and sensible interpretations by allowing these parameters to be heterogeneous across spatial units. For example, we can allow direct spillovers from neighbouring exogenous variables, e.g. improved amenities increase house prices in neighbouring areas directly (close to good amenities) and indirectly (close to an area of rising prices). See also Elhorst (2012) for more discussions on the identification and estimation issues.

To date, only one paper, proposed by Aquaro, Bailey and Pesaran (2015, hereafter ABP), examines the heterogeneous spatial autoregressive (HSAR) panel data model where the spatial coefficients are allowed to be heterogeneous. ABP consider the following model:

$$y_{it} = \phi_0^* \sum_{j=1}^{N} w_{ij} y_{jt} + \boldsymbol{\pi}_i' \boldsymbol{x}_{it} + \alpha_i + u_{it} \tag{7}$$

They derive conditions under which the heterogeneous spatial coefficients are identified and develop a quasi maximum likelihood (QML) estimation procedure when both the time and cross section dimensions are large. The STARDL model encompasses the HSAR model, that does not consider temporal dynamics and diffusion dependence explicitly.

2.1 Stability conditions and assumptions

We rewrite (1) compactly as

$$y_{it} = \phi_{i0}^* y_{it}^* + \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it} + u_{it} \tag{8}$$

where $\boldsymbol{\chi}_{it} = \left(y_{i,t-1}, ..., y_{i,t-p}, y^*_{i,t-1}, ..., y^*_{i,t-p}, \boldsymbol{x}'_{it}, ..., \boldsymbol{x}'_{i,t-q}, \boldsymbol{x}^{*\prime}_{it}, ..., \boldsymbol{x}^{*\prime}_{i,t-q}, 1\right)'$ and $\boldsymbol{\theta}_i = \left(\boldsymbol{\phi}'_i, \boldsymbol{\phi}^{*\prime}_i, \boldsymbol{\pi}'_i, \boldsymbol{\pi}'^*_i, \alpha_i\right)'$ with $\boldsymbol{\phi}_i = \left(\phi_{i1}, ..., \phi_{ip}\right)'$, $\boldsymbol{\phi}^*_i = \left(\phi^*_{i1}, ..., \phi^*_{ip}\right)'$, $\boldsymbol{\pi}_i = \left(\boldsymbol{\pi}'_{i0}, ..., \boldsymbol{\pi}'_{iq}\right)'$, $\boldsymbol{\pi}^*_i = \left(\boldsymbol{\pi}^{*\prime}_{i0}, ..., \boldsymbol{\pi}^{*\prime}_{iq}\right)'$. Stacking (8), we have

$$\boldsymbol{y}_t = \boldsymbol{\Phi}^*_0 \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{\Theta} \boldsymbol{\chi}_t + \boldsymbol{u}_t \tag{9}$$

where $\boldsymbol{\Phi}^*_0 = diag\left(\phi^*_{10}, ..., \phi^*_{N0}\right)$, $\boldsymbol{\Theta} = diag\left(\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_N\right)$, and $\boldsymbol{\chi}_t = \left(\boldsymbol{\chi}'_{1t}, ...\boldsymbol{\chi}'_{Nt}\right)'$.

We begin with the following assumptions:

Assumption 1: The disturbances $\{u_{it}\}$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$, are independent across i and t with zero mean, heterogeneous variance $\sigma^2_i > 0$ but without time dependence, $E(u_{it}u_{is}) = 0 \; \forall t \neq s$. In addition, $E\left|u_{it}\right|^{4+\epsilon} < \infty$ for some $\epsilon > 0$.

Assumption 2: The true parameter vector $\left(\boldsymbol{\phi}^{*\prime}_0, \boldsymbol{\theta}', \boldsymbol{\sigma}'\right)'$, where $\boldsymbol{\phi}^*_0 = \left(\phi^*_{10}, \ldots, \phi^*_{N0}\right)'$, $\boldsymbol{\theta} = \left(\boldsymbol{\phi}'_1, \ldots, \boldsymbol{\phi}'_p, \boldsymbol{\phi}^{*\prime}_1, \ldots, \boldsymbol{\phi}^{*\prime}_p, \boldsymbol{\pi}'_0, \ldots, \boldsymbol{\pi}'_q, \boldsymbol{\pi}^{*\prime}_0, \ldots, \boldsymbol{\pi}^{*\prime}_q, \boldsymbol{\alpha}'\right)'$, and $\boldsymbol{\sigma}^2 = \left(\sigma^2_1, \ldots, \sigma^2_N\right)'$, are in a compact set.

Assumption 3: The spatial weights matrix $\boldsymbol{W}$ is non-stochastic with zero diagonals and uniformly bounded for all N with absolute row and column sums.

Assumption 4: Either : (a) as $N \to \infty$, $\left(\boldsymbol{I}_N - \boldsymbol{\Phi}^*_0 \boldsymbol{W}\right)^{-1}$ exists and is uniformly bounded for all N with uniformly bounded absolute row and column sums; or (b) for bounded N, the eigenvalues of $\boldsymbol{\Phi}^*_0 \boldsymbol{W}$ lie inside the unit circle such that the matrix $\boldsymbol{S}\left(\boldsymbol{\Phi}^*_0\right)) = \boldsymbol{I}_N - \boldsymbol{\Phi}^*_0 \boldsymbol{W}$ is invertible for all $\boldsymbol{\Phi}^*_0 \in \Theta_{\Phi^*_0}$, where $\Theta_{\Phi^*_0}$ is the compact parameter space.

As a result of Assumption 4, we rewrite equation (4) as

$$\tilde{\boldsymbol{\Phi}}\left(L\right) \boldsymbol{y}_t = \tilde{\boldsymbol{\Pi}}_\ell\left(L\right) \boldsymbol{x}_{t-\ell} + \tilde{\boldsymbol{u}}_t, \tag{10}$$

where L denotes the lag operator, $\tilde{\boldsymbol{\Phi}}\left(z\right) = \boldsymbol{I} - \sum_{\ell=1}^{p} \tilde{\boldsymbol{\Phi}}_\ell z^\ell$, and $\tilde{\boldsymbol{\Pi}}_\ell\left(z\right) = \sum_{\ell=0}^{q} \tilde{\boldsymbol{\Pi}}_\ell z^\ell$ are $N \times N$ a matrix polynomials of order p and q respectively with $\tilde{\boldsymbol{\Phi}}_\ell = \left(\boldsymbol{I}_N - \boldsymbol{\Phi}^*_0 \boldsymbol{W}\right)^{-1}\left(\boldsymbol{\Phi}_\ell + \boldsymbol{\Phi}^*_\ell \boldsymbol{W}\right)$, $\tilde{\boldsymbol{\Pi}}_\ell = \left(\boldsymbol{I}_N - \boldsymbol{\Phi}^*_0 \boldsymbol{W}\right)^{-1}\left[\boldsymbol{\Pi}_\ell + \boldsymbol{\Pi}^*_\ell\left(\boldsymbol{W} \otimes \boldsymbol{I}_K\right)\right]$, and $\tilde{\boldsymbol{u}}_t = \left(\boldsymbol{I}_N - \boldsymbol{\Phi}^*_0 \boldsymbol{W}\right)^{-1} \boldsymbol{u}_t$.

Assumption 5 (Time stability): The roots of the characteristic equation $\left|\tilde{\boldsymbol{\Phi}}\left(z\right)\right| = \left|\boldsymbol{I}_N - \sum_{\ell=1}^{p} \tilde{\boldsymbol{\Phi}}_\ell z^\ell\right|$ lie outside the unit circle.

Assumption 5 implies that we can rewrite (10) as an infinite order moving average:

$$\boldsymbol{y}_t = \tilde{\boldsymbol{\Phi}}\left(L\right)^{-1}\left(\sum_{\ell=0}^{q} \tilde{\boldsymbol{\Pi}}_\ell \boldsymbol{x}_{t-\ell} + \tilde{\boldsymbol{u}}_t\right) \equiv \sum_{\ell=0}^{\infty} \tilde{\boldsymbol{B}}_\ell \boldsymbol{x}_{t-\ell} + \sum_{\ell=0}^{\infty} \boldsymbol{B}_\ell \tilde{\boldsymbol{u}}_{t-\ell}, \tag{11}$$

where $\tilde{\boldsymbol{B}}\left(L\right)\left(= \sum_{\ell=0}^{\infty} \tilde{\boldsymbol{B}}_\ell\right) = \tilde{\boldsymbol{\Phi}}\left(L\right)^{-1} \tilde{\boldsymbol{\Pi}}_\ell\left(L\right)$. Then, it follows that the sums $\sum_{\ell=0}^{\infty} \left\|\boldsymbol{B}_\ell\right\|_1$ and $\sum_{\ell=0}^{\infty} \left\|\boldsymbol{B}_\ell\right\|_\infty$ are bounded by some constant C, see also Li (2017) for more discussions.

Assumption 6: The explanatory variables $\boldsymbol{x}_{it}$ are random variables such that $E(\left\|\boldsymbol{x}_{it}\right\|^4) \leq C$ for all i and t, and they are independent of the idiosyncratic errors u_{js} for all (i, j, t, s).

Under Assumptions 5 and 6, it is easily seen that $E(\|\boldsymbol{\chi}_{it}\|^4) \leq C$ for all i and t, and they are uncorrelated with the idiosyncratic errors u_{js} for all (i,j,t,s). For (local) identification we require the following assumption:

Assumption 7: $p\lim_{N,T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left(\boldsymbol{\chi}'_{it}\boldsymbol{\chi}_{it}\right)$ is strictly positive definite with the largest eigenvalue bounded by some constant C for $i = 1, ..., N$. At the true parameter values of $(\boldsymbol{\phi}_0^*, \boldsymbol{\theta}, \boldsymbol{\Sigma}_u)$, the following matrix

$$\lim_{N,T\to\infty} \left\{ \frac{1}{N(T-q)} \left[\sum_{t=q}^{T} E\left(\boldsymbol{V}'_t\boldsymbol{V}_t\right) - \sum_{t=q}^{T} E\left(\boldsymbol{V}'_t\boldsymbol{\chi}_t\right) \left(\sum_{t=q}^{T} E\left(\boldsymbol{\chi}'_t\boldsymbol{\chi}_t\right) \right)^{-1} \sum_{t=q}^{T} E\left(\boldsymbol{\chi}'_t\boldsymbol{V}_t\right) \right] \right\}$$

is positive definite, where $\boldsymbol{V}_t = \boldsymbol{G}\boldsymbol{\chi}_t\boldsymbol{\theta}$ with $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{S}^{-1}\left(\boldsymbol{\Phi}_0^*\right)$.

Assumption 1 both limits the probability of extreme values of the disturbance and rules out the possibility that u_{it} is degenerate for any $i = 1, \ldots, N$. Time dependence is ruled out implicitly in the literature; here we model it through time lags of the dependent and explanatory variables. The more common requirement that u_{it} be *i.i.d.* has been weakened to allow heteroskedasticty. Assumption 2 is standard in econometrics literature. Assumption 3 is common within the spatial literature and contains the normalising convention that no unit acts as its own neighbour, thereby ensuring that y_{it}^* remains bounded for all $i = 1, \ldots, N$ if y_{jt} and $\boldsymbol{x}_{jt}$ are bounded for all $j = 1, \ldots, N$ and $t = 1, \ldots, T$. The bounded absolute row sums of $\boldsymbol{W}$ follow naturally when the rows of $\boldsymbol{W}$ are normalised to sum to unity, whereupon bounded absolute column sums amounts to a restriction on the degree of influence of individual observations within the system, which is only needed to accommodate the case with $N \to \infty$; it is satisfied automatically when N is bounded.

Assumption 4 limits the degree of contemporaneous spatial feedback within the system, without which it would be possible that the elements of $\boldsymbol{y}_t$ and the variance of $\boldsymbol{u}_t$ would not be finite. It ensures that the variance of the disturbance in the equation (10), $Var\left(\tilde{\boldsymbol{u}}_t\right) = \left(\boldsymbol{I} - \boldsymbol{\Phi}_0^*\boldsymbol{W}\right)^{-1}\boldsymbol{\Sigma}_u\left(\boldsymbol{I} - \boldsymbol{W}'\boldsymbol{\Phi}_0'^*\right)^{-1}$ is bounded where $\boldsymbol{\Sigma}_u = Var\left(\boldsymbol{u}_t\right) = diag\left(\sigma_1^2, \ldots, \sigma_N^2\right)$. The conditions on $\boldsymbol{\Phi}_0^*\boldsymbol{W}$ given under 4(a) are sufficient for those given under 4(b), since under 4(a) we may write $\left(\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}\right)^{-1} = \boldsymbol{I}_N + \boldsymbol{\Phi}_0^*\boldsymbol{W} + (\boldsymbol{\Phi}_0^*\boldsymbol{W})^2 + \ldots$, and by the properties that all norms are sub-additive and bounded below by the absolute value of the largest eigenvalue, see Horn and Johnson (1985, 5.6). If N is bounded (while $T \to \infty$), then the expression in 4(b) is equivalent to 4(a). This is not the same as assuming that the (unconditional) variance of $\boldsymbol{y}_t$, is finite and does not depend upon t, for which we make Assumption 5, which is a familiar condition from the literature on dynamic systems (e.g. Hamilton, 1994). Assumption 5 generalises Assumption 4 in Mutl (2009), who assumes that the largest eigenvalue of $\tilde{\boldsymbol{\Phi}}_1$ is strictly less than one when $p = 1$, and is the necessary condition for the stable relationship existing between $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$.[1]

Assumption 7 is a (local) identification condition for $\boldsymbol{\phi}_0^*$ and $\boldsymbol{\Sigma}_u$. Notice that when we use Assumption 7 to identify $\boldsymbol{\phi}_0^*$, we implicitly assume that $\boldsymbol{\theta} \neq 0$. See

[1]There has been a great deal of interest in stationarity of spatial dynamic panel data models (e.g. Lee and Yu, 2010). Let ω_i denote an eigenvalue of $\boldsymbol{W}$, then, for the homogeneous

Bai and Li (2015) for the global identification condition.

3 Estimation and Inference

Following the early work of Cliff and Ord (1973), it is well-known that the endogeneity caused by contemporaneous spillovers across spatial units makes estimation by ordinary least squares inconsistent. Quasi-Maximum Likelihood (QML) techniques, based upon a data transformation removing the endogeneity, have proved popular, see Anselin (1988) and Lee (2004) establishing their asymptotic properties. For applications in which the number of spatial units is large, however, the computational cost to evaluating the effect of this transformation on the log likelihood can be prohibitive. An alternative approach based on the use of moment conditions have been developed by Kelejian and Prucha (1998, 1999) using instrumental variables (IV), then the generalised method of moments (GMM) respectively, see Elhorst (2010) and Lee and Yu (2014) among others. The growing availability of datasets with both spatial and time dimension has sparked interest in models for dynamic spatial panel data. Working in a static framework, Baltagi et al. (2007) consider combined tests of spatial correlation, serial correlation and random effects. Kapoor et al. (2007) estimate panel data models with spatially correlated error components using GMM while Yu et al. (2008) suggest QML methods for the estimation of dynamic spatial panel models. Lee and Yu (2014) consider GMM methods on dynamic spatial panel models with multiple spatial lags.

3.1 Quasi-Maximum Likelihood Estimation

Let $\bar{T} = T - q$ denote the sample size after allowing for lags and let $\boldsymbol{\xi} = \left(\boldsymbol{\phi}_0^{*\prime}, \boldsymbol{\theta}', \boldsymbol{\sigma}^{2\prime}\right)'$ with $\boldsymbol{\sigma}^2 = \left(\sigma_i^2, ..., \sigma_i^2\right)'$ denote the parameters. The true value of the parameters are denoted $\tilde{\boldsymbol{\xi}} = \left(\tilde{\boldsymbol{\phi}}_0^{*\prime}, \tilde{\boldsymbol{\theta}}', \tilde{\boldsymbol{\sigma}}^{2\prime}\right)'$ and with, for example, $\tilde{\boldsymbol{\sigma}}^2 = \left(\tilde{\sigma}_1^2, ..., \tilde{\sigma}_N^2\right)'$. Define $\boldsymbol{\chi} = [\boldsymbol{\chi}_{q+1}', \ldots, \boldsymbol{\chi}_T']'$ and $\boldsymbol{\chi}_i = [\boldsymbol{\chi}_{i(q+1)}', \ldots, \boldsymbol{\chi}_{iT}']'$.

Under the specification in equation (9) alongside Assumptions 1 and 6, the density of $\boldsymbol{y}_t$ may be conditioned recursively for $t = q+1, \ldots, T$ on $\boldsymbol{\chi}_t$, which is to say the independent and pre-determined regressors. Following Lee (2004), the QML estimator can be constructed using equation (9) as the optimiser of

parameter case with $p = 1$, Assumption 5 reduces to

$$\left|\frac{\phi_1 + \phi_1^*\omega_i}{1 - \phi_0^*\omega_i}\right| < 1, \forall i.$$

Lee and Yu (2009) consider row normalised weight matrix, for which $\max \omega_i = 1$. When other parameters are in the non-negative but stable region, this simplifies further to $\left|\phi_1 + \phi_0^* + \phi_1^*\right| < 1$. See also Elhorst (2014) for a wider set of stability conditions in the homogeneous spatial dynamic panel data models.

the function:

$$\mathcal{L}_{\bar{T}}\left(\boldsymbol{\xi}\right)=-\frac{n\bar{T}}{2}\ln\left(2\pi\right)-\frac{\bar{T}}{2}\ln\left|\boldsymbol{\Sigma}_{u}\right|+\bar{T}\ln\left|S(\boldsymbol{\Phi}_{0}^{*})\right|-\frac{1}{2}\sum_{t=q+1}^{T}\boldsymbol{u}_{t}^{\prime}\boldsymbol{\Sigma}_{u}^{-1}\boldsymbol{u}_{t},\qquad(12)$$

where $\boldsymbol{\phi}_{0}^{*}=(\phi_{01}^{*},\ldots,\phi_{0N}^{*})^{\prime}$ and $\boldsymbol{\theta}=\left(\boldsymbol{\theta}_{1}^{\prime},...,\boldsymbol{\theta}_{N}^{\prime}\right)^{\prime}$.

It is easy to show, see the appendix, that the closed form solution for $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ given $\boldsymbol{\Phi}^{*}$ are

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{i}(\phi_{0i}^{*}) &= \left(\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}^{\prime}\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\left[y_{it}-\phi_{0i}^{*}y_{it}^{*}\right]\right),\\
\hat{\sigma}_{i}^{2}(\phi_{0i}^{*}) &= \frac{1}{\bar{T}}\sum_{t=q+1}^{T}\left(y_{it}-\phi_{0i}^{*}y_{it}^{*}-\hat{\boldsymbol{\theta}}_{i}^{\prime}\boldsymbol{\chi}_{it}\right)^{2}\\
&= \frac{1}{\bar{T}}\sum_{t=q+1}^{T}\left(y_{it}-\phi_{0i}^{*}y_{it}^{*}\right)^{2}\\
&\quad -\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\left(y_{it}-\phi_{0i}^{*}y_{it}^{*}\right)\boldsymbol{\chi}_{it}^{\prime}\left[\left(\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}^{\prime}\right)\right]^{-1}\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\left(y_{it}-\phi_{0i}^{*}y_{it}^{*}\right).
\end{aligned}$$

After substituting in, maximising (12) is equivalent to maximising the following concentrated log-likelihood function:

$$\mathcal{L}_{\bar{T}}^{c}\left(\boldsymbol{\phi}_{0}^{*}\right)=-\frac{n\bar{T}}{2}\ln\left(2\pi+1\right)+\ln\left|S(\boldsymbol{\Phi}_{0}^{*})\right|-\frac{1}{2}\sum_{i=1}^{N}\ln\hat{\sigma}_{i}^{2}(\phi_{0i}^{*}).\qquad(13)$$

The heterogeneous spatial parameters $\boldsymbol{\phi}_{0}^{*}=(\phi_{i0}^{*},...,\phi_{N0}^{*})^{\prime}$ can be estimated more conveniently by

$$\hat{\boldsymbol{\phi}}_{0}^{*}=\arg\max_{\boldsymbol{\phi}_{0}^{*}\in\Theta_{\phi_{0}^{*}}}L_{C}\left(\boldsymbol{\phi}_{0}^{*}\right).$$

Consider the following non-stochastic functions of the parameters, based on the (conditional) expectation of the log-likelihood function.

$$Q_{\bar{T}}(\boldsymbol{\xi})=E\left[\mathcal{L}_{\bar{T}}\left(\boldsymbol{\phi}_{0}^{*},\boldsymbol{\theta},\boldsymbol{\Sigma}_{u}\right)\right].\qquad(14)$$

In order to evaluate this expression, it will be very valuable to express $\boldsymbol{S}(\boldsymbol{\Phi}_{0}^{*})\boldsymbol{y}_{t}$ in terms of $\boldsymbol{\chi}_{t}$ and $\boldsymbol{u}_{t}$. Define $\boldsymbol{G}=\boldsymbol{W}\boldsymbol{S}^{-1}=[\boldsymbol{g}_{1},\ldots,\boldsymbol{g}_{N}]^{\prime}$, with rows $\boldsymbol{g}_{i}^{\prime}=[g_{i1},\ldots,g_{iN}]$, and with $\boldsymbol{S}=\boldsymbol{S}\left(\tilde{\boldsymbol{\Phi}}_{0}^{*}\right)$. Note that $\boldsymbol{I}+\tilde{\boldsymbol{\Phi}}_{0}^{*}\boldsymbol{G}=\boldsymbol{S}^{-1}$ so that the relationship

$$\boldsymbol{S}\left(\boldsymbol{\Phi}_{0}^{*}\right)\boldsymbol{S}^{-1}=\boldsymbol{S}^{-1}-\boldsymbol{\Phi}_{0}^{*}\boldsymbol{G}=\boldsymbol{I}+\left(\tilde{\boldsymbol{\Phi}}_{0}^{*}-\boldsymbol{\Phi}_{0}^{*}\right)\boldsymbol{G}.$$

Basic manipulation produces

$$\boldsymbol{S}(\boldsymbol{\Phi}_{0}^{*})\boldsymbol{y}_{t}=\boldsymbol{S}\left(\boldsymbol{\Phi}_{0}^{*}\right)\boldsymbol{S}^{-1}\left[\boldsymbol{\Theta}\boldsymbol{\chi}_{t}+\boldsymbol{u}_{t}\right]=\left[\boldsymbol{I}+\left(\tilde{\boldsymbol{\Phi}}_{0}^{*}-\boldsymbol{\Phi}_{0}^{*}\right)\boldsymbol{G}\right]\boldsymbol{\Theta}\boldsymbol{\chi}_{t}+\boldsymbol{S}\left(\boldsymbol{\Phi}_{0}^{*}\right)\boldsymbol{S}^{-1}\boldsymbol{u}_{t}.$$

Using $\boldsymbol{s}_i^{*\prime}(\phi_{0i}^*)$ to denote the i'th row of $\boldsymbol{S}(\boldsymbol{\Phi}_0^*)$, with its scalar argument reflecting that this does not depend on any ϕ_{0j}^* for $j \neq i$, the i'th row of this relationship is

$$y_{it} - \phi_{i0}^* y_{it}^* = \boldsymbol{s}_i'(\phi_{0i}^*)\boldsymbol{S}^{-1}\left[\boldsymbol{\Theta}\boldsymbol{\chi}_t + \boldsymbol{u}_t\right] \equiv \kappa_{it}(\phi_{i0}^*) + \boldsymbol{s}_i'(\phi_{0i}^*)\boldsymbol{S}^{-1}\boldsymbol{u}_t,$$

where $\kappa_{it}(\phi_{i0}^*)$ is defined implicitly.

Then, for given $\boldsymbol{\Phi}^*$, the values of $\boldsymbol{\theta}$ that maximises $Q_{\bar{T}}(\boldsymbol{\xi})$ is given by

$$\begin{aligned}
\bar{\boldsymbol{\theta}}_i(\phi_{0i}^*) &= \left(E\sum_{t=q+1}^{T}\frac{1}{\bar{T}}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right)^{-1}\left(E\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\left[y_{it} - \phi_{0i}^* y_{it}^*\right]\right) \\
&= \left(E\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right)^{-1}\left(E\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\kappa_{it}(\phi_{i0}^*)\right).
\end{aligned}$$

It then follows that and the value of σ_i^2 is given by

$$\begin{aligned}
\bar{\sigma}_i^2(\phi_{0i}^*) &= E\sum_{t=q+1}^{T}\frac{1}{\bar{T}}\left\{\left[y_{it} - \phi_{i0}^* y_{it}^* - \bar{\boldsymbol{\theta}}_i'(\phi_{0i}^*)\boldsymbol{\chi}_{it}\right]'\left[y_{it} - \phi_{i0}^* y_{it}^* - \bar{\boldsymbol{\theta}}_i'(\phi_{0i}^*)\boldsymbol{\chi}_{it}\right]\right\} \\
&= E\sum_{t=q+1}^{T}\frac{1}{\bar{T}}\left\{\left[\kappa_{it}(\phi_{i0}^*) + \boldsymbol{s}_i'(\phi_{i0}^*)\boldsymbol{S}^{-1}\boldsymbol{u}_t - \left(E\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\kappa_{it}(\phi_{i0}^*)\right)'\left(E\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right)^{-1}\boldsymbol{\chi}_{it}\right]^2\right\} \\
&= E\left(\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\kappa_{it}(\phi_{i0}^*)^2\right) \\
&\quad - E\left(\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\kappa_{it}(\phi_{i0}^*)\boldsymbol{\chi}_{it}'\right)\left[E\left(\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right)\right]^{-1}E\left(\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\kappa_{it}(\phi_{i0}^*)\right) \\
&\quad + tr\left\{\boldsymbol{S}^{-1\prime}\boldsymbol{s}_i(\phi_{0i}^*)\boldsymbol{s}_i'(\phi_{0i}^*)\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\right\}.
\end{aligned}$$

Then we have

$$Q_{\bar{T}}^c(\boldsymbol{\phi}_0^*) \equiv \max_{\boldsymbol{\theta},\boldsymbol{\sigma}^2} E\left[\mathcal{L}_{\bar{T}}\left(\boldsymbol{\phi}_0^*, \boldsymbol{\theta}, \boldsymbol{\sigma}^2\right)\right] = -\frac{n\bar{T}}{2}\ln(2\pi + 1) + \ln|S(\boldsymbol{\Phi}_0^*)| - \frac{1}{2}\sum_{i=1}^{N}\ln\bar{\sigma}_i^2(\phi_{0i}^*), \tag{15}$$

The variable, $\kappa_{it}(\phi_{i0}^*)$ represents the part of $(y_{it} - \phi_{i0}^* y_{it}^*)$ that can be 'explained' by the pre-determined variables in the system. Note that as $\boldsymbol{s}_i'(\tilde{\phi}_{0i}^*)\boldsymbol{S}^{-1} = \boldsymbol{e}_i'$, the i'th row of the $N \times N$ identity matrix, at the true parameter value $\kappa_{it}(\tilde{\phi}_{i0}^*) = \boldsymbol{\chi}_{it}'\tilde{\boldsymbol{\theta}}_i$ ensuring that $\bar{\boldsymbol{\theta}}_i(\tilde{\phi}_{0i}^*) = \tilde{\boldsymbol{\theta}}_i$ and $\bar{\sigma}_i^2(\tilde{\phi}_{0i}^*) = \tilde{\sigma}_i^2$, as expected. Away from the true parameter value, however, $\kappa_{it}(\tilde{\phi}_{i0}^*)$ will be a weighted sum of $\boldsymbol{\chi}_{jt}$. In order to identify ϕ_{i0}^*, we require that some of the explanation comes from $\boldsymbol{\chi}_{jt}$, for some $j \neq i$, that is to say which cannot equally be explained with a different choice of $\boldsymbol{\theta}_i$, if $\phi_{i0}^* \neq \tilde{\phi}_{i0}^*$, which is encapsulated in the following assumption.

Assumption 7: For all $\phi^*_{i0} \neq \tilde{\phi}^*_{i0}$ and for all i

$$\lim_{T\to\infty} \left\{ E\left(\frac{1}{\bar{T}} \sum_{t=q+1}^{T} v_{it}(\phi^*_{0i})^2 \right) \right.$$
$$\left. -E\left(\sum_{t=q+1}^{T} \frac{1}{\bar{T}} v(\phi^*_{0i})_{it} \boldsymbol{\chi}'_{it} \right) \left[E\left(\frac{1}{\bar{T}} \sum_{t=q+1}^{T} \boldsymbol{\chi}_{it}\boldsymbol{\chi}'_{it} \right) \right]^{-1} E\left(\sum_{t=q+1}^{T} \frac{1}{\bar{T}} \boldsymbol{\chi}_{it} v(\phi^*_{0i})_{it} \right) \right\} > 0.$$

Assumption 7 mirrors Assumption 8 in Lee (2004) and Assumption G2 in Li (2017). It states that $\kappa_{it}(\phi^*_{i0})$ cannot be co-linear with $\boldsymbol{\chi}_{it}$ for any value of ϕ^*_{0i} other than the true value. This is enough to locally identify the parameter vector $\boldsymbol{\phi}^*_0$, given $\boldsymbol{\chi}_t$, and hence, given that other parameters follow as closed form expressions, $\boldsymbol{\xi}$. Arguments similar to Lee (2004), see the appendix, may be deployed to establish the following Theorem.

Theorem 1 (**The QML estimator**) Consider the STARDL model (8) and suppose that (i) Assumptions 1-7 hold, andThen, as $T \to \infty$, we have:

$$\sqrt{\bar{T}}\left(\hat{\boldsymbol{\xi}} - \tilde{\boldsymbol{\xi}}\right) \to_d N\left(0, AVar\left(\hat{\boldsymbol{\xi}}\right)\right) \text{ with } AVar\left(\hat{\boldsymbol{\xi}}\right) = \boldsymbol{H}^{-1}_{\bar{T}}\left(\hat{\boldsymbol{\xi}}\right) \boldsymbol{J}_{\bar{T}}\left(\hat{\boldsymbol{\xi}}\right) \boldsymbol{H}^{-1}_{\bar{T}}\left(\hat{\boldsymbol{\xi}}\right)$$

where

$$\boldsymbol{J}_{\bar{T}}\left(\boldsymbol{\xi}\right) = \frac{-1}{\bar{T}} \left(\frac{\partial \mathcal{L}\left(\boldsymbol{\xi}\right)}{\partial \boldsymbol{\xi}} \right) \left(\frac{\partial \mathcal{L}\left(\boldsymbol{\xi}\right)}{\partial \boldsymbol{\xi}} \right)'; \boldsymbol{H}_{\bar{T}}\left(\boldsymbol{\xi}\right) = \frac{-1}{\bar{T}} \frac{\partial^2 \mathcal{L}\left(\boldsymbol{\xi}\right)}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'},$$

the non-zero elements of which are given in the appendix.

Although, the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ may be concentrated out of (12), following Lee (2004), leaving an optimisation over $\boldsymbol{\Phi}^*$, the need for repeated evaluation of the determinant $N \times N$ matrix, $\boldsymbol{I}_N - \boldsymbol{\Phi}^*_0 \boldsymbol{W}$, can make the maximisation of (12) or (13) numerically burdensome for large N. In the homogenous case, this is often done with using the technique proposed by Ord (1975), which is straightforward once the eigenvalues of $\boldsymbol{W}$ have been found. [2] This may be challenging if N is large and $\boldsymbol{W}$ asymmetric, both of which are typically true in applied work, because the computation of the eigenvalues becomes numerically unstable. This technique is not applicable, however, when parameters are heterogeneous and other methods must be found. With this in mind, we next consider a computationally simpler method that exploits the naturally available instruments to control the endogeneity within equation (8).

3.2 Control Function Estimation

The alternative approach to dealing with the endogeneity in (8) is to use instruments. Here we do so via a control function (CF). [3] Let $\boldsymbol{z}_{it}$ be the $L \times 1$ vector

[2]In the homogeneous case with $\phi^*_{i0} = \phi^*_0$, for all i $|S(\boldsymbol{\Phi}^*_0)| = \prod_{i=1}^{n}(1 - \phi^*_0 \omega_i)$, where the ω_i are eigenvalues of $\boldsymbol{W}$.

[3]Most models that are linear in parameters are estimated using standard IV methods – two stage least squares (2SLS). The CF approach relies on the same kinds of identification conditions. However, in models with nonlinearities or random coefficients, the form of exogeneity is stronger and more restrictions are imposed on the reduced forms.

of exogenous/ pre-determined variables:

$$\boldsymbol{z}_{it}=\left(\boldsymbol{z}_{it}^{1\prime},\boldsymbol{z}_{it}^{2\prime}\right)'$$

where the $L_1\times 1$ vector, $\boldsymbol{z}_{it}^{1}=\boldsymbol{\chi}_{it}$ contains all all exogenous and pre-determined variables included in (8). The $L_2\times 1$ vector $\boldsymbol{z}_{it}^{2}$ contains additional pre-determined variables related to $\boldsymbol{y}_{it}^{*}$ but not to u_{it} with $L_2\geq 1$, such as higher orders of spatial and time lagged variables in $\boldsymbol{z}_{it}^{1}$ (see Section 3.2.1 below).

Assumption 7': For each $i=1,\ldots,N$ there exist a vector, $\boldsymbol{z}_{it}^{2\prime}$ such that $(\boldsymbol{z}_{it}^{2}u_{it})$ is a stationary and ergodic martingale difference sequence, $\lim_{T\to\infty}\left\{E\left(\frac{1}{T}\sum_{t=q+1}^{T}\boldsymbol{z}_{it}^{2}y_{it}^{*}\right)\right\}\neq\mathbf{0}$ and for all L_1 vectors $\boldsymbol{\gamma}\neq 0$

$$\begin{aligned}&\lim_{T\to\infty}\left\{E\left(\frac{1}{T}\sum_{t=q+1}^{T}\boldsymbol{\gamma}'\boldsymbol{z}_{it}^{2}\boldsymbol{z}_{it}^{2\prime}\boldsymbol{\gamma}\right)\right.\\&\left.-E\left(\sum_{t=q+1}^{T}\frac{1}{T}\boldsymbol{\gamma}'\boldsymbol{z}_{it}^{2}\boldsymbol{\chi}_{it}'\right)\left[E\left(\frac{1}{T}\sum_{t=q+1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right)\right]^{-1}E\left(\sum_{t=q+1}^{T}\frac{1}{T}\boldsymbol{\chi}_{it}\boldsymbol{z}_{it}^{2\prime}\boldsymbol{\gamma}\right)\right\}>0.\end{aligned}$$

Assumption 7' carries the usual condition that $\boldsymbol{z}_{it}^{2}$ is not asymptotically linearly dependent on $\boldsymbol{\chi}_{it}$ but is asymptotically correlated with y_{it}^{*}. The above presentation makes clear that the identification does not rely on $\kappa(\phi_{i0}^{*})$ but may utilise other relevant variables provided they are uncorrelated with the disturbance process.

We now run the reduced form regression of y_{it}^{*} on $\boldsymbol{z}_{it}$

$$y_{it}^{*}=\boldsymbol{\varphi}_i'\boldsymbol{z}_{it}+v_{it}\text{ with }E\left(\boldsymbol{z}_{it}'v_{it}\right)=\mathbf{0}\tag{16}$$

Then, apply the linear projection of u_{it} on v_{it} as follows:

$$u_{it}=\rho_i v_{it}+e_{it}\tag{17}$$

where $\rho_i=E\left(v_{it}u_{it}\right)/E\left(v_{it}^{2}\right)$. By construction, $E\left(\boldsymbol{z}_{it}'e_{it}\right)=\mathbf{0}$ and $E\left(v_{it}e_{it}\right)=0$. The endogeneity is now fully reflected in $E\left(v_{it}u_{it}\right)$ since from (16) we have:

$$Cov\left(y_{it}^{*},u_{it}\right)=Cov\left(\boldsymbol{\varphi}_i'\boldsymbol{z}_{it},u_{it}\right)+Cov\left(v_{it},u_{it}\right)=Cov\left(v_{it},u_{it}\right).$$

Replacing u_{it} by (17), we obtain the following transformation of (1):

$$y_{it}=\sum_{h=1}^{p}\phi_{ih}y_{i,t-h}+\sum_{h=0}^{q}\boldsymbol{\pi}_{ih}'\boldsymbol{x}_{i,t-h}+\sum_{h=0}^{p}\phi_{ih}^{*}y_{i,t-h}^{*}+\sum_{h=0}^{q}\boldsymbol{\pi}_{ih}'^{*}\boldsymbol{x}_{i,t-h}^{*}+\alpha_i+\rho_i v_{it}+e_{it}\tag{18}$$

where v_{it} is the control variable, rendering the new error terms, e_{it} uncorrelated with y_{it}^{*} as well as with v_{it} and other regressors in (18). We rewrite (18) compactly as

$$y_{it}=\boldsymbol{\beta}_i'\boldsymbol{q}_{it}+e_{it},\ t=1,...,T,\tag{19}$$

where $\boldsymbol{q}_{it} = \left(y_{it}^{*}, \boldsymbol{z}_{it}^{1\prime}, 1, v_{it}\right)^{\prime}$ denotes the tth row of the matrix of the regressors, $\boldsymbol{q}_i = \left(\boldsymbol{q}_{i1}^{\prime}, ..., \boldsymbol{q}_{iT}^{\prime}\right)^{\prime}$, and $\boldsymbol{\beta}_i = \left(\phi_{i0}^{*}, \boldsymbol{\theta}_i^{\prime}, \rho_i\right)^{\prime}$.

Assumption 7' ensures that $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{q}_{it}e_{it} \rightarrow_d N\left(0, \sigma_{ei}^2\left(\boldsymbol{q}_i^{\prime}\boldsymbol{q}_i\right)^{-1}\right)$ and that $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{z}_{it}v_{it} \rightarrow_d N\left(0, \sigma_{vi}^2\left(\boldsymbol{z}_i^{\prime}\boldsymbol{z}_i\right)^{-1}\right)$, where $\boldsymbol{z}_i = \left(\boldsymbol{z}_{i1}^{\prime}, ..., \boldsymbol{z}_{iT}^{\prime}\right)^{\prime}$. We propose the two-step procedure: (i) obtain the reduced form residuals, $\hat{v}_{it} = y_{it}^{*} - \hat{\boldsymbol{\varphi}}_i^{\prime}\boldsymbol{z}_{it}$ from (16) and (ii) run the following regression:

$$y_{it} = \sum_{h=1}^{p}\phi_{ih}y_{i,t-h} + \sum_{h=0}^{q}\boldsymbol{\pi}_{ih}^{\prime}\boldsymbol{x}_{i,t-h} + \sum_{h=0}^{p}\phi_{ih}^{*}y_{i,t-h}^{*} + \sum_{h=0}^{q}\boldsymbol{\pi}_{ih}^{\prime *}\boldsymbol{x}_{i,t-h}^{*} + \alpha_i + \rho_i\hat{v}_{it} + e_{it}^{*} \tag{20}$$

where $e_{it}^{*} = e_{it} + \rho_i\left(\boldsymbol{\varphi}_i - \hat{\boldsymbol{\varphi}}_i\right)^{\prime}\boldsymbol{z}_{it}$ depends on the sampling error in $\hat{\boldsymbol{\varphi}}_i$ unless $\rho_i = 0$ (exogeneity test). Then, the OLS estimator from (20) will be consistent. We refer this estimator to as the STARDL-CF estimator. We note in passing that the practical advantage of the CF approach mainly lies in preserving the structural parameters in (1), which will be used for conducting the dynamic counterfactual analysis below.

The following Theorem shows that the STARDL estimator of the parameter vector $\boldsymbol{\beta}_i = \left(\phi_{i0}^{*}, \boldsymbol{\theta}_i^{\prime}\right)^{\prime}$ in (19) is $\sqrt{T}$-consistent and follows the asymptotic normal distribution.

Theorem 2 Under Assumptions 1-6 and 7', as $T \rightarrow \infty$, the OLS estimator from (20) is consistent and asymptotically normally distributed as

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\beta}}_i\right)) \rightarrow_d N\left(0, AVar\left(\hat{\boldsymbol{\beta}}_i\right)\right),$$

where

$$AVar\left(\hat{\boldsymbol{\beta}}_i\right) \rightarrow_p \hat{\sigma}_i^2\left(\tilde{\boldsymbol{X}}_i^{\prime}\tilde{\boldsymbol{X}}_i\right)^{-1},$$

where $\tilde{\boldsymbol{X}}_{it}^{\prime} = \left(y_{i,t}^{*} - \hat{v}_{it}, \boldsymbol{\chi}_{it}^{\prime}\right)$ denotes the t'th row of the matrix $\tilde{\boldsymbol{X}}_i = \left(\tilde{\boldsymbol{X}}_{i1}^{\prime}, ..., \tilde{\boldsymbol{X}}_{iT}^{\prime}\right)^{\prime}$, and $\hat{\sigma}_i^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_{it}^2$, where $\hat{u}_{it} = \hat{e}_{it} + \hat{v}_{it}\hat{\rho}$.[4]

3.2.1 The selection of instrumental variables

By modelling the spatial and dynamic effects jointly, we can obtain the valid IVs internally as follows: Under Assumption 4 we can express $\boldsymbol{y}_t^{*}$ as

$$\boldsymbol{y}_t^{*} = \boldsymbol{G}\left[\sum_{\ell=1}^{p}\boldsymbol{\Phi}_{\ell}\boldsymbol{y}_{t-\ell} + \sum_{\ell=1}^{p}\boldsymbol{\Phi}_{\ell}^{*}\boldsymbol{W}\boldsymbol{y}_{t-\ell} + \sum_{\ell=0}^{q}\boldsymbol{\Pi}_{\ell}\boldsymbol{x}_{t-\ell} + \sum_{\ell=0}^{q}\boldsymbol{\Pi}_{\ell}^{*}\left(\boldsymbol{W}\otimes\boldsymbol{I}_K\right)\boldsymbol{x}_{t-\ell} + \boldsymbol{\alpha} + \boldsymbol{u}_t\right]. \tag{21}$$

[4] The above allows for heterogeneity in the parameters and depends only on large T. Homogeneity (or indeed clustering) are nested within the model, but these are not treated asymptotically. With homogenous parameters it would be possible to get a faster $\sqrt{NT}$ rate of convergence.

This suggests that

$$\left[\sum_{\ell=1}^{p}\boldsymbol{W}^2\boldsymbol{y}_{t-\ell},\sum_{\ell=1}^{p}\boldsymbol{W}^3\boldsymbol{y}_{t-\ell},...,\sum_{\ell=0}^{q}\boldsymbol{W}^2\boldsymbol{x}_{t-\ell},\sum_{\ell=0}^{q}\boldsymbol{W}^3\boldsymbol{x}_{t-\ell},...\right] \tag{22}$$

can be used as the IV for $\boldsymbol{y}_t^*$. [5] Hence, we employ the following set of IVs for y_{it}^* in the individual STARDL regression, (1):

$$\boldsymbol{z}_{it}=\left(\sum_{\ell=1}^{p}y_{i,t-\ell}^{**},\sum_{\ell=1}^{p}y_{i,t-\ell}^{***},...,\sum_{\ell=0}^{q}x_{i,t-\ell}^{**},\sum_{\ell=0}^{q}x_{i,t-\ell}^{***},...\right)$$

where $y_{i,t-\ell}^{**}=\sum_{j=1}^{N}w_{ij}^{(2)}y_{j,t-\ell}$, $y_{i,t-\ell}^{***}=\sum_{j=1}^{N}w_{ij}^{(3)}y_{j,t-\ell}$, $x_{i,t-\ell}^{**}=\sum_{j=1}^{N}w_{ij}^{(2)}x_{j,t-\ell}$ and $x_{i,t-\ell}^{***}=\sum_{j=1}^{N}w_{ij}^{(3)}x_{j,t-\ell}$ with $w_{ij}^{(2)}$ and $w_{ij}^{(3)}$ being the (i,j)th element of $\boldsymbol{W}^2$ and $\boldsymbol{W}^3$, respectively.

Next, we can derive the IVs from the higher time lags by rewriting (4) as

$$\boldsymbol{\Phi}(L)\boldsymbol{y}_t=\boldsymbol{\Phi}^*(L)\boldsymbol{W}\boldsymbol{y}_t+\boldsymbol{\Pi}(L)\boldsymbol{x}_t+\boldsymbol{\Pi}^*(L)(\boldsymbol{W}\otimes\boldsymbol{I}_K)\boldsymbol{x}_t+\boldsymbol{\alpha}+\boldsymbol{u}_t \tag{23}$$

where $\boldsymbol{\Phi}(L)=\boldsymbol{I}_N-\sum_{\ell=1}^{p}\boldsymbol{\Phi}_\ell L^\ell$, $\boldsymbol{\Phi}^*(L)=\sum_{\ell=0}^{p}\boldsymbol{\Phi}_\ell^* L^\ell$, $\boldsymbol{\Pi}(L)=\sum_{\ell=0}^{q}\boldsymbol{\Pi}_\ell L^\ell$, $\boldsymbol{\Pi}^*(L)=\sum_{\ell=0}^{q}\boldsymbol{\Pi}_\ell^* L^\ell$, and

$$\boldsymbol{y}_t=\boldsymbol{\Psi}(L)\boldsymbol{y}_t+\boldsymbol{\Xi}(L)\boldsymbol{x}_t+[\boldsymbol{\Phi}(L)]^{-1}[\boldsymbol{\alpha}+\boldsymbol{u}_t]$$

where $\boldsymbol{\Psi}(L)=[\boldsymbol{\Phi}(L)]^{-1}\boldsymbol{\Phi}^*(L)\boldsymbol{W}$ and $\boldsymbol{\Xi}(L)=[\boldsymbol{\Phi}(L)]^{-1}[\boldsymbol{\Pi}(L)+\boldsymbol{\Pi}^*(L)(\boldsymbol{W}\otimes\boldsymbol{I}_K)]$. As $\boldsymbol{\Psi}_0=\boldsymbol{\Phi}_0^*\boldsymbol{W}$, we have:

$$\boldsymbol{y}_t^*=\boldsymbol{G}\left\{\boldsymbol{\Psi}_1(L)\boldsymbol{y}_t+\boldsymbol{\Xi}(L)\boldsymbol{x}_t+[\boldsymbol{\Phi}(L)]^{-1}[\boldsymbol{\alpha}+\boldsymbol{u}_t]\right\} \tag{24}$$

where $\boldsymbol{\Psi}_1(L)=\sum_{\ell=1}^{\infty}\boldsymbol{\Psi}_\ell L^\ell$. This suggests that the following additional IVs

$$\left[\boldsymbol{W}^2\boldsymbol{y}_{t-p-1},\boldsymbol{W}^2\boldsymbol{y}_{t-p-2},...,\boldsymbol{W}\boldsymbol{x}_{t-p-1},\boldsymbol{W}\boldsymbol{x}_{t-p-2},...\right] \tag{25}$$

could be used for $\boldsymbol{y}_t^*$. See also Kelejian and Prucha (1999), and Lee and Yu (2014) for discussion of an optimal set of instruments.

4 The Spatio-Temporal Multipliers

Another important issue is how best to present the STARDL estimation results as these involve the large dimensional spatial interactions and diffusion dependence. In this Section we provide two such measures, the individual spatio-temporal dynamic multipliers and the system diffusion multipliers, respectively. Given the availability of large spatial datasets with a large time dimension these counterfactural analyses will provide the applied researchers the valuable tools.

[5] In practice we can apply the different weight matrices to construct $\boldsymbol{y}_t^*=\boldsymbol{W}\boldsymbol{y}_t$ and $\boldsymbol{x}_t^*=\boldsymbol{Q}\boldsymbol{x}_t$. In this case we can also use $\sum_{\ell=0}^{q}\boldsymbol{W}\boldsymbol{x}_{t-\ell}$ as an Internal IVs.

4.1 The Spatio-Temporal Dynamic Multipliers

Following Shin et al. (2014), it is straightforward to derive dynamic multipliers associated with unit changes in y_{it}^*, $\boldsymbol{x}_{it}$, $\boldsymbol{x}_{it}^*$ and $\boldsymbol{g}_t$ respectively, on $y_{i,t+h}$ for $h = 0, 1, 2...$ We rewrite the STARDL(p, q) model, (1) as[6]

$$\phi_i(L)\, y_{it} = \phi_i^*(L)\, y_{it}^* + \boldsymbol{\pi}_i(L)\, \boldsymbol{x}_{it} + \boldsymbol{\pi}_i^*(L)\, \boldsymbol{x}_{it}^* + u_{it} \tag{26}$$

where

$$\phi_i(L) = 1 - \sum_{\ell=1}^{p} \phi_{i\ell} L^\ell; \phi_i^*(L) = 1 - \sum_{\ell=0}^{p} \phi_{i\ell}^* L^\ell; \boldsymbol{\pi}_i(L) = \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}' L^\ell; \boldsymbol{\pi}_i^*(L) = \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}^{*\prime} L^\ell.$$

Premultiplying (26) by the inverse of $\phi_i(L)$, we obtain:

$$y_{it} = \tilde{\phi}_i^*(L)\, y_{it}^* + \tilde{\boldsymbol{\pi}}_i(L)\, \boldsymbol{x}_{it} + \tilde{\boldsymbol{\pi}}_i^*(L)\, \boldsymbol{x}_{it}^* + \tilde{u}_{it} \tag{27}$$

where $\tilde{\phi}_i^*(L) \left(= \sum_{j=0}^{\infty} \tilde{\phi}_{ij}^* L^j\right) = [\phi_i(L)]^{-1} \phi_i^*(L)$, $\tilde{\boldsymbol{\pi}}_i(L) \left(= \sum_{j=0}^{\infty} \tilde{\boldsymbol{\pi}}_{ij}' L^j\right) = [\phi_i(L)]^{-1} \boldsymbol{\pi}_i(L)$, $\tilde{\boldsymbol{\pi}}_i^*(L) \left(= \sum_{j=0}^{\infty} \tilde{\boldsymbol{\pi}}_{ij}^{*\prime} L^j\right) = [\phi_i(L)]^{-1} \boldsymbol{\pi}_i^*(L)$ and $\tilde{u}_{it} = [\phi_i(L)]^{-1} u_{it}$. The $\tilde{\phi}_{ij}^*$, $\tilde{\boldsymbol{\pi}}_{ij}'$ and $\tilde{\boldsymbol{\pi}}_{ij}^{*\prime}$ for $j = 0, 1, ...,$ can be evaluated using the following recursive relationships:

$$\tilde{\phi}_{ij}^* = \phi_{i1} \tilde{\phi}_{i,j-1}^* + \phi_{i2} \tilde{\phi}_{i,j-2}^* + \cdots + \phi_{i,j-1} \tilde{\phi}_{i1}^* + \phi_{ij} \tilde{\phi}_{i0}^* + \phi_{ij}^*, \ j = 1, 2, ... \tag{28}$$

where $\phi_{ij} = 0$ for $j < 1$ and $\tilde{\phi}_{i0}^* = \phi_{i0}^*$, $\tilde{\phi}_{ij}^* = 0$ for $j < 0$ by construction,

$$\tilde{\boldsymbol{\pi}}_{ij}' = \phi_{i1} \tilde{\boldsymbol{\pi}}_{i,j-1}' + \phi_{i2} \tilde{\boldsymbol{\pi}}_{i,j-2}' + \cdots + \phi_{i,j-1} \tilde{\boldsymbol{\pi}}_{i,1}' + \phi_{ij} \tilde{\boldsymbol{\pi}}_{i0}' + \boldsymbol{\pi}_{ij}', \ j = 1, 2, ... \tag{29}$$

where $\tilde{\boldsymbol{\pi}}_{i0}' = \boldsymbol{\pi}_{i0}'$, $\tilde{\boldsymbol{\pi}}_{ij}' = 0$ for $j < 0$, and

$$\tilde{\boldsymbol{\pi}}_{ij}^{*\prime} = \phi_{i1} \tilde{\boldsymbol{\pi}}_{i,j-1}^{*\prime} + \phi_{i2} \tilde{\boldsymbol{\pi}}_{i,j-2}^{*\prime} + \cdots + \phi_{i,j-1} \tilde{\boldsymbol{\pi}}_{i,1}^{*\prime} + \phi_{ij} \tilde{\boldsymbol{\pi}}_{i0}^{*\prime} + \boldsymbol{\pi}_{ij}', \ j = 1, 2, ... \tag{30}$$

where $\tilde{\boldsymbol{\pi}}_{i0}^{*\prime} = \boldsymbol{\pi}_{i0}^{*\prime}$, $\tilde{\boldsymbol{\pi}}_{ij}^{*\prime} = 0$ for $j < 0$.

The cumulative dynamic multipliers of y_{it}^*, $\boldsymbol{x}_{it}$ and $\boldsymbol{x}_{it}$ on $y_{i,t+h}$ for $h = 0, ..., H$, can be evaluated as follows:

$$m_{y_i^*}^H = \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial y_{it}^*} = \sum_{h=0}^{H} \tilde{\phi}_{ih}^*; \boldsymbol{m}_{x_i}^H = \sum_{h=0}^{H} \tilde{\boldsymbol{\pi}}_{ih}'; \boldsymbol{m}_{x_i^*}^H = \sum_{h=0}^{H} \tilde{\boldsymbol{\pi}}_{ih}^{*\prime}. \tag{31}$$

By construction, as $H \to \infty$, $m_{y_i^*}^H \to \beta_{y_i^*}; \boldsymbol{m}_{x_i}^H \to \boldsymbol{\beta}_{x_i}'; \boldsymbol{m}_{x_i^*}^H \to \boldsymbol{\beta}_{x_i^*}'$, where $\beta_{y_i^*}$, $\boldsymbol{\beta}_{x_i}$ and $\boldsymbol{\beta}_{x_i^*}$ are the associated long-run multipliers. In particular, the cumulative dynamic multiplier effects of the k-th regressors, x_{it}^k and x_{it}^{*k} on $y_{i,t+H}$ are the kth element of the $1 \times K$ vectors, $\boldsymbol{m}_{x_i}^H$ and $\boldsymbol{m}_{x_i^*}^H$:

$$m_{x_i^k}^H = \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial x_{it}^k} = \sum_{h=0}^{H} \tilde{\pi}_{ih}^k; m_{x_i^{*k}}^H = \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial x_{it}^{*k}} = \sum_{h=0}^{H} \tilde{\pi}_{ih}^{*k} \text{ for } k = 1, ..., K.$$

[6] To construct the dynamic multipliers, we should use the structural parameters in (1) which are consistently estimated by the STARDL estimator from (20). Without loss of generality we drop the intercept, α_i.

The STARDL model can be treated as an extended ARDL model for each spatial unit. Suppose that y_{it} is the domestic policy variable. An important feature of the STARDL model is to capture three different forms of dynamic adjustment from initial equilibrium to the new equilibrium following an economic perturbation with respect to domestic conditions (x_{it}), overseas conditions (x_{it}^*) and the overseas policy decisions (y_{it}^*). A careful investigation of the dynamic multipliers enables us to categorise the group of countries, say countries that focus on domestic conditions only (e.g. the US), and those that pay attention to both domestic and overseas conditions (e.g. the small open economies), in the short-run and the long-run.

We may apply the mean group estimation of the dynamic multipliers to investigate the overall average pattern of $m_{y_i^*}^H$, $\boldsymbol{m}_{x_i}^H$ and $\boldsymbol{m}_{x_i}^H$ provided with the bootstrap-based confidence intervals.

4.2 The System Diffusion Multipliers

We now develop the system diffusion multipliers which measure the joint impacts of $\boldsymbol{x}_t$ on $\boldsymbol{y}_{t+h}$ in space and time for $h = 0, 1, 2...$ We rewrite (10) as

$$\tilde{\boldsymbol{\Phi}}(L)\,\boldsymbol{y}_t = \tilde{\boldsymbol{\Pi}}(L)\,\boldsymbol{x}_t + \tilde{\boldsymbol{u}}_t, \tag{32}$$

where $\tilde{\boldsymbol{\Phi}}(L) = \boldsymbol{I}_N - \sum_{\ell=1}^{p} \tilde{\boldsymbol{\Phi}}_\ell L^\ell$ and $\tilde{\boldsymbol{\Pi}}(L) = \sum_{\ell=0}^{q} \tilde{\boldsymbol{\Pi}}_\ell L^\ell$. Premultiplying (32) by the inverse of $\tilde{\boldsymbol{\Phi}}(L)$, we obtain:

$$\boldsymbol{y}_t = \boldsymbol{B}(L)\,\boldsymbol{x}_t + \left[\tilde{\boldsymbol{\Phi}}(L)\right]^{-1} \tilde{\boldsymbol{u}}_t,\ \boldsymbol{B}(L)\left(= \sum_{j=0}^{\infty} \boldsymbol{B}_j L^j\right) = \left[\tilde{\boldsymbol{\Phi}}(L)\right]^{-1} \tilde{\boldsymbol{\Pi}}(L) \tag{33}$$

The diffusion multipliers, $\boldsymbol{B}_j$ for $j = 0, 1, ...$, can be evaluated as follows:

$$\boldsymbol{B}_j = \tilde{\boldsymbol{\Phi}}_1 \boldsymbol{B}_{j-1} + \tilde{\boldsymbol{\Phi}}_2 \boldsymbol{B}_{j-2} + \cdots + \tilde{\boldsymbol{\Phi}}_{j-1} \boldsymbol{B}_1 + \tilde{\boldsymbol{\Phi}}_j \boldsymbol{B}_0 + \tilde{\boldsymbol{\Pi}}_j,\ j = 1, 2, ... \tag{34}$$

where $\boldsymbol{B}_0 = \tilde{\boldsymbol{\Pi}}_0$ and $\boldsymbol{B}_j = 0$ for $j < 0$ by construction.

Then, the $N \times NK$ matrix of the cumulative diffusion multipliers can be evaluated as follows:

$$\boldsymbol{d}_x^H = \sum_{h=0}^{H} \frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{x}_t'} = \sum_{h=0}^{H} \boldsymbol{B}_h,\ H = 0, 1, 2, ... \tag{35}$$

The cumulative diffusion multipliers of x_{jt}^h on $y_{i,t+h}$ are given by the $(i, (j-1)\,k + h)$th element of $\boldsymbol{d}_x^H$. Let $\boldsymbol{S}^k$ be the $NK \times N$ selection matrix given by

$$\boldsymbol{S}^k = \left[\boldsymbol{i}_k, \boldsymbol{i}_{K+k}, ..., \boldsymbol{i}_{(N-1)K+k}\right]$$

where $\boldsymbol{i}_j$ is the $NK \times 1$ selection vector with unity on jth row and zeros otherwise. Then, the $N \times N$ matrix of total diffusion multiplier effects with respect to the kth regressor, $\boldsymbol{x}_t^k = \left(x_{1t}^k, x_{2t}^k, ..., x_{Nt}^k\right)'$ is obtained by

$$\boldsymbol{d}_{x^k}^H = \sum_{h=0}^{H} \frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{x}_t^{k\prime}} = \sum_{h=0}^{H} \boldsymbol{B}_h \boldsymbol{S}^k, \quad k = 1, ..., K \tag{36}$$

In the case of homogeneous spatial panel models (e.g. (6)), LeSage and Pace (2009) propose using the average of the main diagonal elements of the $N \times N$ matrix as a summary measure of the own-partial derivatives that they label a direct effect. The direct effect for region i includes some feedback loop effects that arise as a result of impacts passing through neighboring regions j and back to region i. They also propose an average of the (cumulative) off diagonal elements over all rows to produce a summary that corresponds to the indirect (other-region) effect associated with changes in the explanatory variable. Debarsy et al. (2012) extend it to the case of dynamic space-time panel data. This allows us to compute own- and cross-partial derivatives that trace the effects through time and space. Space-time dynamic models produce a situation where a change in the ith observation of the explanatory variable at time t will produce contemporaneous and future responses in all regions' dependent variables as well as other-region future responses. This is due to the presence of an individual time lag (time dependence), a spatial lag (spatial dependence) and a cross-product term reflecting the space-time diffusion. The main diagonal elements of the $N \times N$ matrix sums represent (cumulative) own-region impacts that arise from both time and spatial dependence. The sum of off-diagonal elements reflects both spillovers measuring contemporaneous cross-partial derivatives and diffusion measuring cross-partial derivatives that involve different time periods. We note that it is not possible to separate out the time dependence from spillover and diffusion effects. In the case with heterogeneous spatial coefficients, LeSage and Chin (2016) propose use of the N diagonal elements to produce observation-level direct effects for each of the N regions. As estimates of region specific indirect spill-in and spill-out effects, they propose use of the sum of off-diagonal elements in each row and column. Indeed, this approach is qualitatively similar to the Diebold and Yilmaz (2014) network measures such as in-degree or out-degree effects.

4.3 Network Analysis

To provide the informative summary output measure of the impacts of $\{x_{jt}\}_{j=1}^{N}$ on $\{y_{it}\}_{i=1}^{N}$, we now apply the network approach to an analysis of the $N \times N$ matrix of the diffusion multiplier effects of each regressor in $\boldsymbol{x}_t$. Here we follow Diebold and Yilmaz (2014) and Greenwood-Nimmo, Nguyen and Shin (2015, GNS) and apply the generalised connectedness measures to the $N \times N$ matrix of the diffusion multipliers.

At any horizon, h, one cross-tabulates the impacts of the single regressor, x_{jt}^{k} on the $N \times 1$ vector of endogenous variables, $\boldsymbol{y}_t$. We now rewrite the $N \times N$ matrices, $\boldsymbol{d}_{x^k}^{H}$ in (36) in terms of the following $N \times N$ connectedness matrix across the spatial units:

$$\mathbb{C} = \begin{bmatrix} \phi_{1\leftarrow 1} & \phi_{1\leftarrow 2} & \cdots & \phi_{1\leftarrow N} \\ \phi_{2\leftarrow 1} & \phi_{2\leftarrow 2} & \cdots & \phi_{2_1\leftarrow N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N\leftarrow 1} & \phi_{N\leftarrow 2} & \cdots & \phi_{N\leftarrow N} \end{bmatrix} \tag{37}$$

We suppress the horizon index to avoid cluttering our notation.

The main diagonal elements of the $N\times\ N$ matrix of dynamic multipliers, $\mathbb{C}$ represent (cumulative) own-region impacts that arise from both time and spatial dependence. The off-diagonal elements of this matrix reflect both spillovers measuring contemporaneous cross-partial derivatives and diffusion measuring cross-partial derivatives that involve different time periods (and also through $\boldsymbol{WX}$). The use of an arrow indicates the direction of the feedback or spillover effect. We also note that it is not possible to separate out the time dependence and spatial spillover from diffusion effects.

Example. Suppose that y is the policy rate and x is the inflation or output gap, and consider the ith row of $\mathbb{C}$. $\sum_{j=1}^{N}\phi_{i\leftarrow j}$ is the total impact of global economic activity on domestic policy rate, in which $\phi_{i\leftarrow i}$ represents the own direct contribution to the total impact, and $\phi_{i\leftarrow j}$ the indirect spillover contribution with respect to economic activities in other country, j.

We start with the (cumulative) own-region (price) impacts that arise from both time and spatial dependence $\left(H^V_{j\leftarrow j}\right)$, defined as

$$H_{j\leftarrow j} = \phi_{j\leftarrow j} \tag{38}$$

which lie on the prime diagonal of $\mathbb{C}$. These are defined as *from* or *spill-in* contributions. We may write the cross-from contribution as

$$F_{j\leftarrow\bullet} = \sum_{i=1, i\neq j}^{N} \phi_{j\leftarrow i} \tag{39}$$

where the subscript $j\leftarrow\bullet$ indicates that the directional effect is from all other countries to country j. Diebold and Yilmaz (2014) refer to this measure as the *spillover index* in the context of individual returns or volatilities across multiple stock markets. The following is true by construction:

$$H_{j\leftarrow j} + F_{j\leftarrow\bullet} = TOT_{j\leftarrow\bullet} = \sum_{i=1}^{N}\phi_{j\leftarrow i}. \tag{40}$$

where $TOT_{j\leftarrow\bullet}$ denotes the total impacts of the regressor in country j attributable to all sources. Similarly, we define the total contributions *to* all other countries (or *spill-out* contributions) as the j-th column sum minus the own-region contribution, $\phi_{j\leftarrow j}$, yielding:

$$T_{\bullet\leftarrow j} = \sum_{i=1, i\neq j}^{N} \phi_{i\leftarrow j} \tag{41}$$

which measures the total directional connectedness from country j *to* the other countries in the system. The net directional connectedness is then defined simply as

$$N_{\bullet\leftarrow j} = T_{\bullet\leftarrow j} - F_{j\leftarrow\bullet}. \tag{42}$$

It is straightforward to develop the following aggregate (non-directional) connectedness measures for the $N \times 1$ vector of endogenous variables, $\boldsymbol{y}$:

$$H = \sum_{j=1}^{N} H_{j\leftarrow j} \tag{43}$$

$$S = \sum_{j=1}^{N} F_{j\leftarrow\bullet} = \sum_{j=1}^{N} T_{\bullet\leftarrow j} \tag{44}$$

$$H + S = TOT_{\bullet\leftarrow\bullet} = \sum_{j=1}^{N} TOT_{j\leftarrow\bullet} \tag{45}$$

$$\sum_{j=1}^{N} N_{\bullet\leftarrow j} = 0 \tag{46}$$

We refer to H and S as the aggregate heatwave index (direct own-region impacts) and the aggregate cross-spillover contribution (indirect cross-partial and diffusion impacts). (45) states that the sum of the aggregate heatwave and spillover measures accounts for all of the (price) impacts in the entire system at any given horizon, denoted $TOT_{\bullet\leftarrow\bullet}$. Similarly, (46) notes that the aggregate net connectedness among all the N countries in the system is equal to zero.[7]

Finally, we define a pair of indices to succinctly address two questions of particular interest when measuring connectedness: (i) 'how dependent is the j-th country on external conditions?' and (ii) 'to what extent does the j-th country influence/is the j-th country influenced by the system as a whole?'. These measures are especially relevant when evaluating connectedness among geo-political units such as countries and economic blocs within the global economy. In response to the first question, we propose the following dependence index :

$$O_j = \frac{F_{j\leftarrow\bullet}}{W_{j\leftarrow\bullet} + F_{j\leftarrow\bullet}},\ j = 1, ..., N$$

where $0 \leq O_j \leq 1$ expresses the relative importance of external shocks for the j-th country. Specifically, as $O_j \to 1$, then conditions in group j are dominated by external shocks while group j is unaffected by external shocks if $O_j \to 0$. In a similar vein, we develop the influence index:

$$I_j = \frac{N_{j\leftarrow\bullet}}{T_{j\leftarrow\bullet} + F_{j\leftarrow\bullet}},\ j = 1, ..., N$$

[7]Following GNS, we can develop the more parsimonious block-based connectedness table.

where $-1 \leq I_j \leq 1$. For any horizon h, the j-th group is a net shock recipient if $-1 \leq I_j \leq 0$, a net shock transmitter if $0 \leq I_j \leq 1$. As such, the influence index measures the extent to which the j-th group influences or is influenced by conditions in the system.

When studying connectedness among countries, the coordinate pair (O_j, I_j) in dependence-influence space provides an elegant representation of country i's role in the global system. A classic small open economy would be located close to the point $(1, -1)$ while, by contrast, an overwhelmingly dominant economy would exist in the locale of $(0, 1)$. In this way, we are able to measure the extent to which the different economies of the world correspond to these stylised concepts.

5 Monte Carlo Simulations

We investigate the small sample properties of the proposed STARDL estimator via a Monte Carlo simulation study. We use the following data generating process based on the heterogeneous STARDL(1,1) model with one exogenous variable:

$$y_{it} = \phi_i y_{i,t-1} + \pi_{i0} x_{it} + \pi_{i1} x_{i,t-1} + \phi_{i0}^* y_{it}^* + \phi_{i1}^* y_{i,t-1}^* + \pi_{i0}^* x_{it}^* + \pi_{i1}^* x_{i,t-1}^* + u_{it} \quad (47)$$

where y_{it} is the scalar dependent variable and x_{it} is a single exogenous regressor related to the of the ith spatial unit at time t. Their spatially lagged values, are given by $y_{it}^* = \sum_{j=1}^N w_{ij} y_{jt}$ and $x_{it}^* = \sum_{j=1}^N w_{ij} x_{jt}$.

The row-normalised spatial weights matrix, $\boldsymbol{W}$ is based on a b-nearest neighbours specification, with null elements apart from the $b/2$ either side of the principle diagonal (with neighbours wrapping round to the start or end of the row as necessary), which are $1/b$. The symmetry of this matrix means that the column sums are also normalised to unity. We explore differing levels of spatial dependence by allowing $b = (2, 10)$, [8] within a system of $N = (25, 50, 75, 100)$ cross-section units over $T = (50, 100, 200)$ time periods. The individual scalar parameters involving time and spatial lags of $y_{i,t}$ $(\phi_i, \phi_{i0}^*, \phi_{i1}^*)$ are independent draws from a $U(0, 0.4)$ distribution while the parameters for time and spatial lags of $x_{i,t}$ $(\pi_{i0}, \pi_{i1}, \pi_{i0}^*, \pi_{i1}^*)$ are independent draws from a $U(0, 1)$ distribution.

For this first order case, Assumption 5 determining the stationarity of $\boldsymbol{y}_t$ depends upon the eigenvalues of

$$\left[\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W}\right]^{-1} \left[\boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_1^* \boldsymbol{W}\right]$$

lying within the unit circle. In our specification the largest of these eigenvalues ranges between 0.5 and 0.65 depending on N and on b, which affects the relative importance of each unit within the system. Each specification is explored over $R = 1,000$ repetitions.

We consider two particular experiments, exploring the effects of dependence in the exogenous variables and heteroskedasticity in the disturbances. Experiment 1 uses a set of independent exogenous variables as draws from a standard

[8] The cases of $b = 4$ and $b = 20$ were also considered but are not reported.

normal distribution and disturbances, u_{it}, is made up of draws from an independent standard normal distribution. The more general experiment 2 uses a set of serially correlated exogenous variables, generated according to

$$x_{i,t} = \rho_i x_{i,t-1} + v_{i,t},\ v_{i,t} \sim N\left(0, 1-\rho_i^2\right), \tag{48}$$

where $\rho_i \sim U[0.4, 0.6]$, alongside heteroskedastic disturbances with $u_{it} \sim N(0, \sigma_i^2)$, where $\sigma_i^2 = 0.5 + 0.25 \times \eta_i$ and $\eta_i \sim \chi_2^2$.

Let α_{i0} denote the value of a parameter α_i used to simulate the data and let $\hat{\alpha}_{ij}$ denote its estimate in the jth repetition, then we report the following statistics:

Average bias $= N^{-1}\sum_{i=1}^{N} R^{-1}\sum_{j=1}^{R}\left(\hat{\alpha}_{ij} - \alpha_{i0}\right)$, in Tables 1 and 2;

Average RMSE $= N^{-1}\sum_{i=1}^{N}\sqrt{R^{-1}\sum_{j=1}^{R}\left(\hat{\alpha}_{ij} - \alpha_{i0}\right)^2}$, in Tables 3 and 4;

Average Size $= N^{-1}\sum_{i=1}^{N} R^{-1}\sum_{j=1}^{R}\mathbb{I}\left(\left|\frac{\hat{\alpha}_{ij}-\alpha_{i0}}{\sigma_\alpha}\right| > t_{0.975}\right)$, where $\mathbb{I}(.)$ denotes the indicator function and σ_α is an estimate of the standard deviation for the parameter, in Tables 5 and 6.

We consider both control function and QML estimation methods. Our control function estimates are based on the an instrument set of time and spatial lags of exogenous variables, $y_{i,t-1}^{**}$ and x_{it}^{**}, where $y_{i,t-1}^{**} = \boldsymbol{w}_i \boldsymbol{W} \boldsymbol{y}_{t-1}$ and $x_{it}^{**} = \boldsymbol{w}_i \boldsymbol{W} \boldsymbol{x}_t$ are the second spatial lags. The b-nearest neighbour weighting matrix guarantees that these will not be collinear with $y_{i,t-1}^{*}$ or x_{it}^{*}. In fact this choice of instrument set has the intuitive interpretation that we are using that the next b neighbours' neighbours as our instruments. This is not the only possible instrument set, with higher spatial and/or time lags also valid and generated internally,

$$IV = \left[\boldsymbol{W}^2\boldsymbol{y}_{t-1}, \boldsymbol{W}^3\boldsymbol{y}_{t-1}, ..., \boldsymbol{W}^2\boldsymbol{x}_t, \boldsymbol{W}^2\boldsymbol{x}_{t-1}, ..., \boldsymbol{W}^2\boldsymbol{y}_{t-2}, \boldsymbol{W}\boldsymbol{x}_{t-2}, \boldsymbol{W}^2\boldsymbol{x}_{t-2}, ...\right] \tag{49}$$

The price for including extra instruments is potential multi-collinearity and we found that two instruments was often the best choice in this particular set-up.

The initial values for each iteration were provided by the (inconsistent) ordinary least squares estimates of (47). The exogenous variables were then concentrated out and leaving an iteration over the N vector $\boldsymbol{\phi}^{\ast}$ as in (13), with estimates of the other parameters recovered by least squares regression conditional on $\hat{\boldsymbol{\phi}}^{\ast}$. [9]

Table 1 show that both the CF and the QML estimates of the STARDL model perform reasonably well. This is very reassuring given the range of time and spatial dependence possible in (47). In both cases bias falls as T increases and is not greatly affected by N, supporting our theoretical prediction. The repeated uptick in bias between the cases $N = 25, 50$ and $N = 75, 100$ is likely to be due to the values appended into the parameter vectors in the latter cases. For small T the QML estimator has noticeably lower bias, but as T becomes large the results of the two estimators a comparable with the control function estimator

[9]Despite this concentration procedure this was a far more computationally intensive estimation procedure even for moderately sized N.

having a strong computational advantage. The estimates of all parameters have biases of similar magnitude with the coefficients on the contemporaneous terms, ϕ^*, π and π^*, exhibiting the lower biases than their equivalents on lagged terms and ϕ_1. This is not too surprising as the time dynamics in (47) open more channels through which a time lagged variable may potentially impact on a $y_{i,t}$. Interestingly, there is no noticeable deterioration in the bias of either estimator improves as b rises.

Table 2 indicates that both methods are reasonably robust to both to heteroskedasticity in the disturbances and to time dependence in the exogenous variables. Indeed the estimation of contemporaneous spatial effects, ϕ^*, shows some improvement in performance at the expense of parameters on lagged exogenous variables, π_1 for control function estimation. This is probably due to the fact that correlation between x_{it} and $x_{i,t-1}$ causes our chosen instrument set approaches the optimal. At the same time, this rise in regressor collinearity causes the latter.

Comparing Tables 3 and 4 it is clear that QML is by far the more efficient estimator. This is not surprising. It should be remembered, however, that these experiments are being played on MLE's home pitch and a comparison of the estimators under different distributional assumptions would be of considerable interest.

A number of interesting patters appear in Tables 5 and 6. Firstly it is clear that, the control function estimates tend to be under-sized, particularly for ϕ^* and ϕ_1^*, although this improves towards 5 per cent as T increases. The size for the QML estimates of these parameters is much closer to 5 per cent but the others are slightly over-sized. Secondly performance deteriorates with the number of connections, with the control function becoming more under-sized while QML becomes over-sized. The QML estimator recovers more successfully as T increases and is always within a percentage point for the longest time span of $T = 200$

Table 1: Average Bias - time independent X, homoskedastic errors

Control Function

2 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0164	0.0023	−0.0001	0.0182	0.0045	0.0001	0.0291	0.0024	0.0038	0.0269	0.0062	0.0018
ϕ_1^*	0.0072	0.0050	0.0026	0.0040	0.0033	0.0021	0.0020	0.0036	0.0014	0.0008	0.0029	0.0019
ϕ_1	−0.0187	−0.0094	−0.0039	−0.0184	−0.0087	−0.0045	−0.0205	−0.0102	−0.0053	−0.0204	−0.0094	−0.0049
π	−0.0034	−0.0017	0.0008	−0.0022	−0.0017	0.0001	−0.0080	−0.0004	−0.0004	−0.0055	−0.0012	−0.0003
π_1	−0.0002	0.0020	0.0018	0.0015	0.0013	0.0022	−0.0007	0.0035	0.0014	−0.0003	0.0027	0.0018
π^*	−0.0062	0.0005	−0.0004	−0.0091	−0.0018	0.0001	−0.0187	−0.0003	−0.0029	−0.0124	−0.0032	−0.0006
π_1^*	−0.0070	0.0010	0.0015	−0.0036	0.0002	0.0012	−0.0157	0.0019	−0.0015	−0.0137	−0.0019	0.0008

4 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0142	0.0025	−0.0005	0.0193	0.0030	−0.0028	0.0261	−0.0050	0.0043	0.0316	0.0008	0.0011
ϕ_1^*	0.0005	0.0049	0.0032	0.0031	0.0050	0.0039	0.0016	0.0079	0.0010	−0.0019	0.0045	0.0034
ϕ_1	−0.0160	−0.0107	−0.0047	−0.0207	−0.0101	−0.0052	−0.0214	−0.0110	−0.0054	−0.0209	−0.0102	−0.0055
π	−0.0009	−0.0006	0.0004	−0.0026	−0.0005	0.0003	−0.0025	0.0012	−0.0004	−0.0033	−0.0005	−0.0003
π_1	0.0029	0.0034	0.0021	0.0046	0.0048	0.0025	0.0060	0.0057	0.0016	0.0025	0.0040	0.0020
π^*	−0.0116	−0.0006	0.0013	−0.0147	−0.0046	0.0018	−0.0202	0.0029	−0.0037	−0.0232	−0.0012	−0.0010
π_1^*	−0.0082	−0.0019	−0.0006	−0.0099	−0.0003	0.0032	−0.0167	0.0036	−0.0026	−0.0192	0.0011	−0.0012

QML

2 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	−0.0037	−0.0020	−0.0014	−0.0029	−0.0010	−0.0008	−0.0025	−0.0015	−0.0004	−0.0030	−0.0013	−0.0005
ϕ_1^*	0.0075	0.0042	0.0023	0.0068	0.0031	0.0019	0.0076	0.0032	0.0013	0.0072	0.0028	0.0019
ϕ_1	−0.0122	−0.0057	−0.0030	−0.0116	−0.0055	−0.0029	−0.0125	−0.0054	−0.0031	−0.0118	−0.0055	−0.0030
π	0.0026	−0.0003	0.0005	0.0011	0.0000	0.0011	0.0008	0.0007	−0.0002	0.0016	0.0001	0.0002
π_1	0.0060	0.0029	0.0012	0.0043	0.0024	0.0014	0.0054	0.0023	0.0015	0.0053	0.0032	0.0009
π^*	0.0067	0.0024	0.0015	0.0053	0.0019	0.0009	0.0051	0.0021	0.0009	0.0055	0.0029	0.0018
π_1^*	0.0067	0.0030	0.0024	0.0094	0.0036	0.0006	0.0067	0.0045	0.0015	0.0087	0.0036	0.0015

10 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0046	0.0045	0.0004	0.0068	0.0027	0.0011	0.0046	0.0015	0.0021	0.0050	0.0032	0.0014
ϕ_1^*	0.0046	0.0014	0.0009	0.0055	0.0012	0.0013	0.0055	0.0024	0.0007	0.0021	0.0028	0.0011
ϕ_1	−0.0109	−0.0052	−0.0026	−0.0123	−0.0059	−0.0028	−0.0131	−0.0058	−0.0036	−0.0122	−0.0059	−0.0032
π	0.0012	0.0003	−0.0003	−0.0020	−0.0003	0.0002	−0.0001	0.0003	0.0002	0.0003	−0.0001	0.0000
π_1	0.0046	0.0018	0.0013	0.0050	0.0025	0.0015	0.0063	0.0031	0.0018	0.0058	0.0026	0.0008
π^*	−0.0022	−0.0029	−0.0029	−0.0048	−0.0004	−0.0008	−0.0001	−0.0007	−0.0017	−0.0037	−0.0022	−0.0006
π_1^*	−0.0078	−0.0043	0.0016	−0.0063	−0.0020	−0.0008	−0.0011	−0.0003	−0.0006	0.0021	−0.0027	−0.0022

Table 2: Average Bias - time dependent X, heteroskedastic errors

Control Function

2 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0110	−0.0002	−0.0016	0.0060	−0.0021	−0.0008	0.0081	0.0048	0.0004	0.0067	0.0033	−0.0003
ϕ_1^*	0.0161	0.0094	0.0064	0.0173	0.0099	0.0053	0.0188	0.0081	0.0049	0.0183	0.0090	0.0050
ϕ_1	−0.0361	−0.0180	−0.0094	−0.0396	−0.0172	−0.0087	−0.0376	−0.0188	−0.0091	−0.0374	−0.0183	−0.0090
π	−0.0034	−0.0010	0.0011	−0.0039	0.0004	0.0009	−0.0010	−0.0007	0.0001	−0.0013	−0.0013	0.0002
π_1	0.0209	0.0131	0.0067	0.0236	0.0131	0.0058	0.0230	0.0123	0.0062	0.0236	0.0116	0.0069
π^*	−0.0124	0.0011	−0.0004	−0.0039	0.0016	−0.0008	−0.0053	−0.0025	−0.0002	−0.0047	−0.0004	0.0002
π_1^*	0.0100	0.0079	0.0065	0.0164	0.0106	0.0053	0.0125	0.0060	0.0040	0.0130	0.0061	0.0042

10 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	−0.0151	0.0004	0.0026	−0.0056	−0.0001	0.0019	0.0057	−0.0031	0.0001	0.0279	0.0077	−0.0014
ϕ_1^*	0.0306	0.0117	0.0051	0.0272	0.0114	0.0048	0.0240	0.0148	0.0068	0.0085	0.0068	0.0068
ϕ_1	−0.0397	−0.0205	−0.0099	−0.0408	−0.0203	−0.0104	−0.0419	−0.0215	−0.0110	−0.0407	−0.0204	−0.0105
π	0.0005	−0.0011	0.0003	−0.0008	−0.0002	−0.0006	−0.0002	−0.0011	0.0004	−0.0010	−0.0006	−0.0009
π_1	0.0316	0.0158	0.0073	0.0308	0.0156	0.0082	0.0309	0.0169	0.0079	0.0293	0.0149	0.0088
π^*	0.0155	−0.0035	−0.0020	0.0013	−0.0035	−0.0001	−0.0002	0.0022	0.0002	−0.0188	−0.0071	0.0031
π_1^*	0.0206	0.0021	−0.0026	0.0125	0.0097	0.0036	−0.0008	0.0057	0.0029	−0.0225	0.0011	0.0020

QML

2 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	−0.0037	−0.0011	−0.0004	−0.0049	−0.0020	−0.0003	−0.0045	−0.0023	−0.0010	−0.0039	−0.0015	−0.0008
ϕ_1^*	0.0171	0.0077	0.0036	0.0162	0.0082	0.0041	0.0189	0.0091	0.0043	0.0177	0.0089	0.0040
ϕ_1	−0.0289	−0.0146	−0.0071	−0.0293	−0.0146	−0.0069	−0.0303	−0.0149	−0.0069	−0.0295	−0.0146	−0.0073
π	0.0001	0.0004	0.0003	0.0005	0.0000	0.0000	0.0001	−0.0004	0.0004	0.0004	0.0001	0.0004
π_1	0.0207	0.0096	0.0054	0.0225	0.0099	0.0052	0.0205	0.0106	0.0049	0.0201	0.0110	0.0049
π^*	0.0023	0.0016	0.0003	0.0032	0.0020	0.0005	0.0039	0.0023	0.0018	0.0041	0.0014	0.0013
π_1^*	0.0157	0.0073	0.0047	0.0213	0.0106	0.0050	0.0181	0.0097	0.0041	0.0188	0.0081	0.0042

10 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0039	0.0004	0.0012	0.0037	0.0017	0.0015	0.0018	0.0020	0.0014	0.0053	0.0019	0.0008
ϕ_1^*	0.0167	0.0062	0.0036	0.0150	0.0075	0.0031	0.0164	0.0080	0.0043	0.0156	0.0077	0.0038
ϕ_1	−0.0326	−0.0135	−0.0076	−0.0298	−0.0144	−0.0073	−0.0310	−0.0156	−0.0078	−0.0308	−0.0149	−0.0074
π	0.0018	0.0014	0.0007	−0.0017	0.0000	−0.0007	−0.0006	−0.0006	−0.0009	−0.0004	0.0009	0.0002
π_1	0.0228	0.0106	0.0055	0.0231	0.0109	0.0055	0.0234	0.0112	0.0064	0.0211	0.0104	0.0055
π^*	0.0007	−0.0025	−0.0017	−0.0072	−0.0011	0.0005	−0.0045	−0.0025	0.0003	−0.0045	−0.0006	−0.0006
π_1^*	0.0044	0.0034	0.0008	0.0047	0.0067	−0.0005	0.0094	0.0046	−0.0014	0.0029	0.0010	0.0021

Table 3: Average RMSE - time independent X, homoskedastic errors

Control Function

2 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	1.2446	0.6317	0.3436	1.3854	0.5339	0.3089	1.1537	0.6646	0.4148	1.5754	0.7015	0.3917
ϕ_1^*	0.6155	0.2147	0.1324	0.3967	0.1986	0.1239	0.4201	0.2380	0.1537	0.5384	0.2312	0.1445
ϕ_1	0.2158	0.1194	0.0669	0.2643	0.1026	0.0663	0.2018	0.1155	0.0736	0.2769	0.1201	0.0708
π	0.3705	0.1937	0.1150	0.4318	0.1763	0.1094	0.3708	0.1870	0.1219	0.4309	0.1896	0.1137
π_1	0.4002	0.2373	0.1351	0.5320	0.2019	0.1233	0.4086	0.2187	0.1303	0.6272	0.2222	0.1323
π^*	0.9958	0.4426	0.2495	0.7710	0.3975	0.2142	0.7038	0.3957	0.2402	0.9681	0.4143	0.2413
π_1^*	0.8734	0.4395	0.2536	0.9454	0.3759	0.2267	0.7971	0.4558	0.3077	1.0468	0.4788	0.2804

10 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	3.0142	1.4441	0.9624	2.9305	1.7517	0.8196	3.3261	1.7719	0.8246	3.8326	1.9139	0.8826
ϕ_1^*	1.1884	0.5812	0.3835	1.2450	0.6925	0.3487	1.3931	0.7154	0.3733	1.5438	0.8121	0.3836
ϕ_1	0.1983	0.1003	0.0657	0.2008	0.1147	0.0650	0.2549	0.1109	0.0661	0.2720	0.1174	0.0670
π	0.2785	0.1472	0.0927	0.2812	0.1612	0.0916	0.3357	0.1798	0.0919	0.4010	0.1651	0.0924
π_1	0.3403	0.1682	0.1095	0.3041	0.1948	0.1024	0.3411	0.2167	0.1017	0.3831	0.1877	0.1022
π^*	2.9307	1.3650	0.8985	2.7138	1.6716	0.7715	3.0674	1.7163	0.7984	3.5273	1.6812	0.7942
π_1^*	2.9973	1.4096	0.9646	2.8104	1.6205	0.8313	3.1273	1.8086	0.8125	3.3936	1.7584	0.8334

QML

2 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.3506	0.2242	0.1500	0.3431	0.2167	0.1478	0.3454	0.2178	0.1470	0.3510	0.2245	0.1506
ϕ_1^*	0.2529	0.1693	0.1158	0.2543	0.1694	0.1164	0.2577	0.1706	0.1177	0.2566	0.1714	0.1174
ϕ_1	0.1807	0.1217	0.0835	0.1806	0.1226	0.0846	0.1835	0.1236	0.0859	0.1826	0.1230	0.0850
π	0.2427	0.1617	0.1098	0.2420	0.1611	0.1108	0.2461	0.1615	0.1108	0.2445	0.1610	0.1105
π_1	0.2669	0.1767	0.1226	0.2634	0.1737	0.1206	0.2665	0.1743	0.1202	0.2638	0.1745	0.1198
π^*	0.3734	0.2484	0.1691	0.3722	0.2463	0.1687	0.3765	0.2499	0.1695	0.3775	0.2494	0.1711
π_1^*	0.3999	0.2642	0.1832	0.4059	0.2648	0.1827	0.4026	0.2627	0.1794	0.4052	0.2656	0.1820

10 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.5562	0.3724	0.2575	0.5446	0.3647	0.2506	0.5380	0.3598	0.2491	0.5504	0.3675	0.2514
ϕ_1^*	0.4352	0.2875	0.1976	0.4355	0.2879	0.1968	0.4334	0.2902	0.1983	0.4444	0.2961	0.2017
ϕ_1	0.1767	0.1189	0.0831	0.1782	0.1216	0.0841	0.1799	0.1222	0.0850	0.1787	0.1219	0.0850
π	0.2238	0.1504	0.1043	0.2226	0.1490	0.1040	0.2251	0.1499	0.1030	0.2261	0.1503	0.1034
π_1	0.2459	0.1647	0.1145	0.2415	0.1639	0.1138	0.2442	0.1655	0.1141	0.2427	0.1636	0.1132
π^*	0.8344	0.5691	0.3884	0.8597	0.5811	0.3939	0.8474	0.5734	0.3948	0.8444	0.5724	0.3899
π_1^*	0.8923	0.6116	0.4272	0.9173	0.6188	0.4306	0.9005	0.6050	0.4223	0.9092	0.6047	0.4187

Table 4: Average RMSE - time dependent X, heteroskedastic errors

Control Function

2 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.7353	0.3960	0.2524	0.9734	0.4078	0.2518	0.9932	0.4594	0.2686	0.9706	0.4740	0.2703
ϕ_1^*	0.2989	0.1863	0.1238	0.3648	0.1926	0.1276	0.3767	0.2084	0.1313	0.3708	0.2071	0.1285
ϕ_1	0.1562	0.0937	0.0625	0.1795	0.0961	0.0636	0.2868	0.0999	0.0650	0.2046	0.1018	0.0652
π	0.3238	0.1752	0.1151	0.3529	0.1854	0.1210	0.4021	0.1869	0.1200	0.3937	0.1857	0.1183
π_1	0.3306	0.1911	0.1275	0.4077	0.2021	0.1330	0.3957	0.2071	0.1333	0.3939	0.2056	0.1310
π^*	0.5939	0.3330	0.2118	0.6127	0.3184	0.2014	0.5847	0.3287	0.2059	0.6196	0.3324	0.2031
π_1^*	0.5823	0.3145	0.1992	0.8693	0.3277	0.2118	0.7870	0.3579	0.2164	0.7613	0.3596	0.2173

10 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	2.0820	1.1969	0.6574	1.9941	1.1746	0.6363	2.7413	1.1986	0.6176	7.4793	1.3384	0.6585
ϕ_1^*	1.0373	0.5560	0.3309	0.9541	0.5688	0.3171	1.3004	0.5899	0.3242	3.3241	0.6351	0.3365
ϕ_1	0.1777	0.0993	0.0634	0.1821	0.0992	0.0645	0.2415	0.1034	0.0650	0.2515	0.1075	0.0648
π	0.2907	0.1670	0.1056	0.3154	0.1776	0.1089	0.3633	0.1704	0.1075	0.3576	0.1734	0.1073
π_1	0.3282	0.1915	0.1202	0.3351	0.1876	0.1211	0.3914	0.2021	0.1209	0.3965	0.1912	0.1196
π^*	2.1892	1.2238	0.6787	2.0797	1.1830	0.6572	2.7036	1.1504	0.6386	5.4662	1.2393	0.6391
π_1^*	2.1857	1.2063	0.6748	2.0632	1.1846	0.6518	2.7402	1.1615	0.6225	7.9759	1.2665	0.6326

QML

2 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.2514	0.1599	0.1072	0.2586	0.1611	0.1091	0.2628	0.1659	0.1099	0.2677	0.1685	0.1122
ϕ_1^*	0.1850	0.1236	0.0851	0.1909	0.1258	0.0870	0.1947	0.1281	0.0883	0.1940	0.1279	0.0881
ϕ_1	0.1292	0.0869	0.0594	0.1314	0.0876	0.0599	0.1329	0.0885	0.0604	0.1333	0.0884	0.0608
π	0.2244	0.1463	0.0997	0.2280	0.1493	0.1013	0.2260	0.1486	0.1016	0.2251	0.1487	0.1016
π_1	0.2443	0.1623	0.1120	0.2473	0.1630	0.1117	0.2481	0.1640	0.1127	0.2488	0.1643	0.1126
π^*	0.3278	0.2215	0.1478	0.3292	0.2189	0.1485	0.3329	0.2191	0.1497	0.3341	0.2200	0.1501
π_1^*	0.3545	0.2406	0.1621	0.3615	0.2385	0.1623	0.3608	0.2374	0.1620	0.3618	0.2388	0.1637

10 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.3836	0.2661	0.1820	0.3888	0.2641	0.1811	0.3868	0.2603	0.1796	0.3961	0.2645	0.1834
ϕ_1^*	0.3022	0.2108	0.1450	0.3099	0.2124	0.1460	0.3128	0.2126	0.1462	0.3176	0.2151	0.1485
ϕ_1	0.1308	0.0875	0.0609	0.1309	0.0887	0.0611	0.1323	0.0887	0.0611	0.1315	0.0886	0.0614
π	0.2019	0.1387	0.0951	0.2068	0.1404	0.0965	0.2091	0.1398	0.0966	0.2074	0.1397	0.0965
π_1	0.2317	0.1583	0.1099	0.2316	0.1586	0.1094	0.2341	0.1575	0.1096	0.2323	0.1568	0.1096
π^*	0.7333	0.5144	0.3421	0.7411	0.5039	0.3384	0.7358	0.4973	0.3391	0.7262	0.4921	0.3350
π_1^*	0.8358	0.5857	0.3919	0.8349	0.5613	0.3822	0.8353	0.5609	0.3801	0.8229	0.5514	0.3759

Table 5: Average size - time independent X, homoskedastic errors

Control Function

2 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0272	0.0304	0.0360	0.0289	0.0321	0.0378	0.0304	0.0319	0.0364	0.0288	0.0308	0.0341
ϕ_1^*	0.0440	0.0410	0.0435	0.0448	0.0418	0.0424	0.0428	0.0406	0.0424	0.0424	0.0406	0.0411
ϕ_1	0.0565	0.0482	0.0486	0.0550	0.0490	0.0498	0.0548	0.0500	0.0488	0.0516	0.0479	0.0476
π	0.0443	0.0389	0.0432	0.0458	0.0441	0.0425	0.0475	0.0419	0.0449	0.0452	0.0429	0.0434
π_1	0.0454	0.0412	0.0432	0.0459	0.0427	0.0449	0.0458	0.0416	0.0441	0.0449	0.0410	0.0425
π^*	0.0388	0.0386	0.0394	0.0411	0.0405	0.0423	0.0434	0.0383	0.0409	0.0405	0.0383	0.0404
π_1^*	0.0409	0.0366	0.0415	0.0419	0.0387	0.0409	0.0436	0.0394	0.0424	0.0404	0.0388	0.0399

10 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0178	0.0206	0.0269	0.0192	0.0228	0.0268	0.0183	0.0204	0.0283	0.0168	0.0192	0.0257
ϕ_1^*	0.0275	0.0240	0.0318	0.0278	0.0274	0.0298	0.0271	0.0253	0.0312	0.0246	0.0241	0.0289
ϕ_1	0.0511	0.0475	0.0458	0.0485	0.0475	0.0470	0.0487	0.0457	0.0484	0.0446	0.0452	0.0451
π	0.0451	0.0421	0.0426	0.0441	0.0434	0.0440	0.0459	0.0423	0.0436	0.0418	0.0404	0.0428
π_1	0.0459	0.0423	0.0414	0.0450	0.0447	0.0433	0.0448	0.0410	0.0435	0.0428	0.0403	0.0421
π^*	0.0244	0.0248	0.0298	0.0251	0.0278	0.0299	0.0249	0.0250	0.0318	0.0225	0.0227	0.0296
π_1^*	0.0244	0.0254	0.0286	0.0273	0.0272	0.0309	0.0255	0.0257	0.0320	0.0244	0.0240	0.0289

QML

2 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0356	0.0368	0.0386	0.0376	0.0375	0.0409	0.0385	0.0383	0.0402	0.0368	0.0374	0.0381
ϕ_1^*	0.0656	0.0568	0.0540	0.0694	0.0582	0.0552	0.0680	0.0575	0.0540	0.0660	0.0583	0.0539
ϕ_1	0.0798	0.0670	0.0566	0.0767	0.0650	0.0579	0.0767	0.0641	0.0586	0.0770	0.0642	0.0569
π	0.0701	0.0639	0.0542	0.0715	0.0599	0.0542	0.0706	0.0607	0.0558	0.0716	0.0605	0.0551
π_1	0.0706	0.0627	0.0569	0.0723	0.0599	0.0552	0.0712	0.0594	0.0558	0.0707	0.0606	0.0543
π^*	0.0638	0.0569	0.0511	0.0670	0.0572	0.0530	0.0652	0.0586	0.0534	0.0655	0.0578	0.0544
π_1^*	0.0626	0.0574	0.0544	0.0662	0.0564	0.0524	0.0656	0.0572	0.0536	0.0645	0.0568	0.0517

10 connections

N		25			50			75			100	
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0816	0.0658	0.0556	0.0801	0.0646	0.0558	0.0795	0.0653	0.0577	0.0796	0.0651	0.0553
ϕ_1^*	0.0822	0.0645	0.0566	0.0820	0.0644	0.0558	0.0824	0.0671	0.0586	0.0827	0.0657	0.0561
ϕ_1	0.0826	0.0635	0.0562	0.0825	0.0668	0.0558	0.0849	0.0656	0.0588	0.0824	0.0651	0.0582
π	0.0835	0.0670	0.0596	0.0799	0.0638	0.0598	0.0826	0.0641	0.0577	0.0827	0.0660	0.0562
π_1	0.0818	0.0638	0.0596	0.0809	0.0651	0.0579	0.0812	0.0666	0.0582	0.0806	0.0664	0.0576
π^*	0.0804	0.0660	0.0586	0.0837	0.0667	0.0570	0.0821	0.0676	0.0579	0.0828	0.0675	0.0567
π_1^*	0.0812	0.0664	0.0588	0.0822	0.0651	0.0563	0.0825	0.0658	0.0570	0.0836	0.0666	0.0574

Table 6: Average size - time dependent X, heteroskedastic errors

Control Function

2 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0368	0.0386	0.0454	0.0422	0.0413	0.0447	0.0398	0.0415	0.0450	0.0381	0.0397	0.0436
ϕ_1^*	0.0486	0.0452	0.0479	0.0531	0.0472	0.0489	0.0489	0.0460	0.0484	0.0478	0.0458	0.0465
ϕ_1	0.0667	0.0592	0.0555	0.0679	0.0600	0.0567	0.0670	0.0586	0.0551	0.0646	0.0575	0.0543
π	0.0566	0.0530	0.0512	0.0581	0.0530	0.0509	0.0573	0.0514	0.0513	0.0557	0.0507	0.0498
π_1	0.0569	0.0504	0.0523	0.0568	0.0536	0.0523	0.0582	0.0515	0.0509	0.0559	0.0505	0.0511
π^*	0.0544	0.0486	0.0497	0.0557	0.0497	0.0512	0.0544	0.0491	0.0503	0.0523	0.0482	0.0479
π_1^*	0.0527	0.0488	0.0512	0.0566	0.0510	0.0508	0.0538	0.0505	0.0488	0.0524	0.0489	0.0494

10 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0296	0.0283	0.0368	0.0291	0.0296	0.0365	0.0283	0.0294	0.0375	0.0279	0.0284	0.0359
ϕ_1^*	0.0361	0.0341	0.0395	0.0351	0.0339	0.0388	0.0353	0.0343	0.0400	0.0343	0.0326	0.0380
ϕ_1	0.0616	0.0529	0.0526	0.0642	0.0542	0.0542	0.0619	0.0573	0.0542	0.0593	0.0531	0.0538
π	0.0583	0.0508	0.0493	0.0573	0.0492	0.0504	0.0561	0.0496	0.0499	0.0535	0.0493	0.0493
π_1	0.0528	0.0502	0.0490	0.0562	0.0513	0.0506	0.0559	0.0505	0.0498	0.0522	0.0483	0.0489
π^*	0.0418	0.0357	0.0407	0.0389	0.0385	0.0421	0.0390	0.0360	0.0403	0.0390	0.0360	0.0400
π_1^*	0.0434	0.0374	0.0433	0.0434	0.0390	0.0429	0.0416	0.0388	0.0429	0.0411	0.0375	0.0419

QML

2 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0451	0.0400	0.0431	0.0433	0.0406	0.0398	0.0419	0.0413	0.0421	0.0404	0.0379	0.0392
ϕ_1^*	0.0672	0.0574	0.0521	0.0676	0.0573	0.0518	0.0669	0.0557	0.0535	0.0655	0.0564	0.0522
ϕ_1	0.0832	0.0682	0.0579	0.0811	0.0673	0.0582	0.0808	0.0676	0.0573	0.0830	0.0666	0.0582
π	0.0787	0.0634	0.0570	0.0752	0.0633	0.0534	0.0735	0.0622	0.0565	0.0737	0.0625	0.0549
π_1	0.0772	0.0636	0.0562	0.0729	0.0631	0.0557	0.0754	0.0630	0.0566	0.0744	0.0624	0.0569
π^*	0.0722	0.0631	0.0532	0.0744	0.0627	0.0539	0.0697	0.0592	0.0541	0.0708	0.0602	0.0545
π_1^*	0.0706	0.0634	0.0553	0.0719	0.0588	0.0536	0.0710	0.0600	0.0542	0.0692	0.0591	0.0544

10 connections

N	25			50			75			100		
T	50	100	200	50	100	200	50	100	200	50	100	200
ϕ^*	0.0798	0.0664	0.0580	0.0822	0.0651	0.0579	0.0812	0.0654	0.0584	0.0818	0.0632	0.0573
ϕ_1^*	0.0794	0.0648	0.0563	0.0823	0.0678	0.0579	0.0804	0.0649	0.0585	0.0829	0.0657	0.0588
ϕ_1	0.0882	0.0674	0.0616	0.0874	0.0700	0.0599	0.0904	0.0679	0.0598	0.0868	0.0682	0.0602
π	0.0834	0.0653	0.0552	0.0845	0.0663	0.0586	0.0840	0.0669	0.0585	0.0839	0.0661	0.0560
π_1	0.0840	0.0678	0.0598	0.0831	0.0675	0.0582	0.0842	0.0639	0.0579	0.0825	0.0646	0.0580
π^*	0.0812	0.0628	0.0536	0.0835	0.0663	0.0576	0.0821	0.0655	0.0572	0.0836	0.0663	0.0564
π_1^*	0.0856	0.0642	0.0563	0.0833	0.0643	0.0587	0.0829	0.0669	0.0585	0.0828	0.0666	0.0573

6 Empirical Application to Spatio-temporal Diffusion of Armed Violence against Civilians in the Iraqi War

We demonstrate the usefulness of our model in an application to fatalities during the Iraq war. Since 2003, the Pentagon has compiled data on the deaths of civilians, insurgents and Iraqi security forces killed by armed violence with exact location and the time, which was released comparatively recently and notoriously. [10] Despite the lack of transparency surrounding classification of deaths due to armed violence and in distinguishing civilians from combatants, the data present a rare opportunity to infer the intensity of armed violence and its spatio-temporal diffusion through the different regions of the country, if any, during the war period.

The data cover monthly war deaths between 2004 and 2009 with two missing months (May 2004 and March 2009) across 18 governorates of Iraq. The period was characterised by an insurgency from March 2004, which begin in the cities of Fallujah, in Anbar governorate, and Najf, in Basrah governorate, before spreading more widely, and sectarian violence initiated by a bomb attack in Samarra, in Saladin governorate, in February 2006. Both types of conflict are typically fuelled by a cycle of atrocity, clampdown and reprisal and it is of interest to investigate the diffusion of violence across the country.

Table 7 presents the descriptive statistics of the Civilian and the Enemy categories. Although the Civilian category includes foreign security contractors, the vast majority of recorded deaths are of Iraqis. Enemy is the category for insurgents or anti-coalition forces.

Table 7: Descriptive statistics of war casualties (2004-2009)

Categories	Civilian	Enemy
Number of deaths	66,081	23,984
Number of deaths in Baghdad (% in total)	36,998 (55%)	6,526 (27%)
Number of incidents*	34,009	9,417
Deaths per incident	1.94	2.55
Monthly average deaths (95% confidence interval)	944.0 (711.3 to 1176.7)	342.6 (268.7 to 416.5)
Standard deviation	975.84	309.97
Median	483.5	283

*The incidents indicate the events involving at least 1 death from armed violence.

Table 8 presents the number of civilian deaths across each governorate, absolutely and per 1,000 inhabitants, and the number of enemy deaths. It is noteworthy that the majority of civilian deaths occurred in Baghdad, the

[10] The data were part of the Iraq War Logs, leaked by a whistle-blowing NGO WikiLeaks in October 2010.

capital governorate where a quarter of population reside. [11] The civilian death toll per 1,000 inhabitants was more than five times that in any other governorate. Despite this concentration, 73% of insurgent deaths occurred outside of Baghdad, with Anbar governorate providing the highest toll of insurgent casualties with less than 6 % of the civilian casualties suffered in Baghdad.

Table 8: Civilian and Insurgent casualties across governorates (2004-2009)

	Governorate	Deaths	%	Per 1000*	Governorate	Deaths	%
1	Baghdad	36,998	55.99	1.410	Anbar	6,602	27.53
2	Diyala	7,142	10.81	0.272	Baghdad	6,526	27.21
3	Ninewa	6,009	9.09	0.229	Diyala	3,211	13.39
4	Salah al-Din	3,197	4.84	0.122	Ninewa	2,615	10.90
5	*Basrah*	2,635	3.99	0.100	Salah al-Din	1,760	7.34
6	*Babylon*	2,251	3.41	0.086	Najaf	1,064	4.44
7	**Anbar**	2,191	3.32	0.084	Basrah	467	1.95
8	Kirkuk (*Tamim*)	1,780	2.69	0.068	Babylon	417	1.74
9	*Wassit*	887	1.34	0.034	Kirkuk (Tamim)	394	1.64
10	Kerbala	819	1.24	0.031	Wassit	265	1.10
11	Qadissiya	468	0.71	0.018	Qadissiya	160	0.67
12	**Najaf**	335	0.51	0.013	Thi-qar	142	0.59
13	Thi-qar	280	0.42	0.011	Kerbala	120	0.50
14	Erbil	235	0.36	0.009	Missan	51	0.21
15	Missan	135	0.20	0.005	Erbil	28	0.12
16	Sulaymaniyah	125	0.19	0.005	Muthanna	14	0.06
17	Muthanna	64	0.10	0.002	Sulaymaniyah	8	0.03
18	Dahuk	40	0.06	0.002	Dahuk	5	0.02
	Others	490	0.74		Others	135	0.56
Sum		66,081	100			23,984	100

*Deaths scaled by 1000 in the population based on the 2003 World Bank estimates (World Bank 2013).

Empirical results We estimate the STARDL(1,1) estimation, employing y = civilian casualties and x = enemy casualties with $N = 18$ (governorate level) and $T = 70$ (monthly data). A row standardised inverse distance has been employed to construct the weight matrix $\boldsymbol{W}$.[12]

The system is stable with the largest eigenvalue of time stability matrix being 0.86, indicating a high level of persistence, and the largest eigenvalue of spatial stability matrix being -0.086+0.252i. Values for the short-run dynamic

[11]The estimated number of population in Baghdad is 7,145,470 in 2007 (UN Joint Analysis Policy Unit, 2011). The estimated number of population in Iraq is 29,682,000 in 2007.

[12]We obtain the qualitatively similar results when using different $\boldsymbol{W}$ matrices (e.g. a contiguity matrix with and without row standardising).

and diffusion multipliers are reported in Table 9, while those for the long-run dynamic and diffusion multipliers are reported in Table 10. It is not surprising that the long-run spill-over effects (i.e. y^* on y) are positive for all 18 provinces: a rise in civilian casualties in neighbours leads, over time, to a cumulative increase in civilian deaths in any province. The long-run effects of x on y are more diverse, however, and are positive for only 13 of the 18 provinces. In the remaining five an increase in enemy combatant deaths serves to reduce, over time, civilian casualties in that province. Two reasons for this may be: the degradation over time of insurgent capabilities; alongside, the strategic decision to target resources/ retribution at other provinces where that is thought to be more effective, perhaps due to a lower level of security infrastructure or a lower density of patrolling. This latter reason shows up as the long-run impact of x^* on y, which is positive for 10 provinces.

Of those 10, the most striking is Basrah, which appears the most open to the impact of insurgent deaths outside its locality. A test of the significance of these affects cannot be rejected at the 5% level, strongly implying that armed violence against civilians in the neighbouring provinces lead to an increase in civilian casualties in Basrah. Together with Baghdad, the Basrah governorate was an important military posting during the period of the war, with British forces stationed there from the initiation of the Iraq war until control of Basrah International Airport was handed to Iraqis in January 2009. Our model highlights an interesting comparison between Basrah and the capital, Baghdad. Contrary to Basrah where spatial diffusion was prevalent, Baghdad shows strong and significant temporal diffusion (i.e., y_{t-1} on y_t) of armed violence against civilians. Baghdad was the major military post for the US forces in wartime, and was heavily armoured due to severe insurgency against military personnel, civilians and foreign contractors.

Cumulative dynamic and diffusion multipliers The cumulative dynamic multipliers (CDMs) for Basrah and Baghdad are shown alongside the mean group (MGE) for the country in Figure 4. The two provinces show strong opposite patterns with respect to y, x and x^*. The CDMs of y with respect to y^* in Basrah are strongly positive and reach the long-run elasticity, 2.5 within 2-3 months, reflecting an openness to the effects of casualties elsewhere, while those of Baghdad are initially negative, converging to the small positive value gradually. It is not too surprising that, as the largest city and the centre of the coalition government, the security situation in Baghdad was largely determined there. This is supported by the CDMs with respect to x^*, which show a similar pattern except the long-run CDM for Baghdad is negative: insurgent deaths in other provinces making Baghdad safer. The two provinces show contrasting pictures when insurgents are killed in that particular province. For Basrah the CDMs of y with respect to x are negative, but again reach the long-run elasticity, -0.4, quickly while those of Baghdad are positive, again converging to the long-run value of 0.6 gradually. MGE CDMs always tend to the intermediate figures, reaching the long-run estimates at 0.73, 0.18 and 0.04, respectively.

Alongside these CDMs, we are able to calculate Spill-in and Spill-out diffusion multipliers with respect to the explanatory variable x, figure 5. The spill-in diffusion multipliers of x in Basrah are substantially positive while those of Baghdad are slightly negative. Spill-out diffusion multipliers of x in Basrah are slightly negative while those of Baghdad are large and positive. It is noticeable that the net effects show nearly a mirror image (around the MGE line which is 0 by construction), displaying that net effects of Baghdad are large and positive while those of Basrah are large and negative .

The UK and the US had been the major political stakeholders which invaded Iraq in spite of strong disagreement with the UN and the international communities. The military forces of the two countries however appear to have faced substantially different challenges while they stationed in Iraq, with the UK forces being tested by spatial diffusion of armed violence, and the US forces challenged by self-generated temporal persistency in violence. Our analysis indicates that much of the climate for the insurgency was made in Baghdad but that its greatest effects were felt, with relatively little delay, in Basrah.

7 Extension to Joint Modelling of Spatial Dependence and Factors

Recently, a few studies have attempted to develop a combined approach that can accommodate both weak and strong CSD. Bailey et al. (2016) develop multi-step estimation procedure that can distinguish the relationship between spatial units that is purely spatial from that which is due to common factors. Mastromarco et al. (2015) propose the technique in modelling technical efficiency of stochastic frontier panels by combining the exogenously driven factor-based approach and an endogenous threshold regime selection advanced by Kapetanios et al. (2014). Shi and Lee (2017), Bai and Li (2015) and Kuersteiner and Prucha (2018) have developed the framework for jointly modelling spatial effects and unobserved factors. Following this trend we now propose two extensions, one with observed factors and another with unobserved factors.

7.1 STARDL models with observed common factors

Consider the STARDL(p, q) model with the $G \times 1$ vector of observed global factors, $\boldsymbol{g}_t = \left(g_t^1, ..., g_t^G\right)'$ (e.g. oil prices, commodity prices or the common currency such as the Euro):[13]

$$y_{it} = \sum_{\ell=1}^{p} \phi_{i\ell} y_{i,t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\pi}'_{i\ell} \boldsymbol{x}_{i,t-\ell} + \sum_{\ell=0}^{p} \phi^*_{i\ell} y^*_{i,t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\pi}'^*_{i\ell} \boldsymbol{x}^*_{i,t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\psi}'_{i\ell} \boldsymbol{g}_{t-\ell} + \alpha_i + u_{it} \tag{50}$$

where y_{it} is the scalar dependent variable, $\boldsymbol{x}_{it} = (x_{it}^1, ..., x_{it}^K)'$ is a $K \times 1$ vector of exogenous regressors, $y^*_{it} = \sum_{j=1}^N w_{ij} y_{jt}$ and $x^*_{it} = \sum_{j=1}^N w_{ij} x_{jt}$ are spatially lagged values, with their conformable parameters, $\phi_{i\ell}$, $\phi^*_{i\ell}$, $\boldsymbol{\pi}_{i\ell} = (\pi^1_{i\ell}, ..., \pi^K_{i\ell})'$, $\boldsymbol{\pi}^*_{i\ell} = (\pi^{*1}_{i\ell}, ..., \pi^{*K}_{i\ell})'$ and $\boldsymbol{\psi}_{i\ell} = \left(\psi^1_{i\ell}, ..., \psi^G_{i\ell}\right)'$.

[13] For notational simplicity we use the same lag order q for the global factors.

To deal with the endogeneity of y_{it}^* in (50) we can also apply the QML or the CF approach. Here, we run the following CF-augmented regression:

$$y_{it} = \sum_{h=1}^{p} \phi_{ih} y_{i,t-h} + \sum_{h=0}^{q} \boldsymbol{\pi}_{ih}' \boldsymbol{x}_{i,t-h} + \sum_{h=0}^{p} \phi_{ih}^* y_{i,t-h}^* + \sum_{h=0}^{q} \boldsymbol{\pi}_{ih}'^* \boldsymbol{x}_{i,t-h}^* + \sum_{h=0}^{q} \boldsymbol{\psi}_{ih}' \boldsymbol{g}_{t-h} + \alpha_i + \rho \hat{v}_{it} + e_{it}^* \tag{51}$$

where $\hat{v}_{it} = y_{it}^* - \hat{\boldsymbol{\varphi}}_i' \boldsymbol{z}_{it}$ and $e_{it}^* = e_{it}^* + \rho \left(\hat{\boldsymbol{\varphi}}_i - \boldsymbol{\varphi}_i\right)' \boldsymbol{z}_{it}$. As before, we employ the internal IVs as derived in (22) and (25).

Next, we rewrite the STARDL-F(p, q) model, (50) as

$$\phi_i(L) y_{it} = \phi_i^*(L) y_{it}^* + \boldsymbol{\pi}_i(L) \boldsymbol{x}_{it} + \boldsymbol{\pi}_i^*(L) \boldsymbol{x}_{it}^* + \boldsymbol{\psi}_i(L) \boldsymbol{g}_t + \alpha_i + u_{it} \tag{52}$$

where $\phi_i(L) = 1 - \sum_{\ell=1}^{p} \phi_{i\ell} L^\ell$, $\phi_i^*(L) = 1 - \sum_{\ell=0}^{p} \phi_{i\ell}^* L^\ell$, $\boldsymbol{\pi}_i(L) = \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}' L^\ell$, $\boldsymbol{\pi}_i^*(L) = \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}^{*\prime} L^\ell$, and $\boldsymbol{\psi}_i(L) = \sum_{\ell=0}^{q} \boldsymbol{\psi}_{i\ell}' L^\ell$. Premultiplying (52) by $[\phi_i(L)]^{-1}$, we obtain:

$$y_{it} = \tilde{\phi}_i^*(L) y_{it}^* + \tilde{\boldsymbol{\pi}}_i(L) \boldsymbol{x}_{it} + \tilde{\boldsymbol{\pi}}_i^*(L) \boldsymbol{x}_{it}^* + \tilde{\boldsymbol{\psi}}_i(L) \boldsymbol{g}_t + \tilde{u}_{it} \tag{53}$$

where $\tilde{u}_{it} = [\phi_i(L)]^{-1} u_{it}$, $\tilde{\phi}_i^*(L) = [\phi_i(L)]^{-1} \tilde{\phi}_i^*(L)$, $\tilde{\boldsymbol{\pi}}_i(L) = [\phi_i(L)]^{-1} \boldsymbol{\pi}_i(L)$, $\tilde{\boldsymbol{\pi}}_i^*(L) = [\phi_i(L)]^{-1} \boldsymbol{\pi}_i^*(L)$, and $\tilde{\boldsymbol{\psi}}_i(L) \left(= \sum_{h=0}^{\infty} \boldsymbol{\psi}_{ih}' L^h\right) = [\phi_i(L)]_i^{-1} \boldsymbol{\psi}_i(L)$.

The cumulative dynamic multiplier effects of y_{it}^*, $\boldsymbol{x}_{it}$, $\boldsymbol{x}_{it}^*$ and $\boldsymbol{g}_t$ on $y_{i,t+h}$ for $h = 0, 1, ..., H$, can be evaluated as follows:

$$m_{y_i^*}^H = \sum_{h=0}^{H} \tilde{\phi}_{ih}^*; \boldsymbol{m}_{x_i}^H = \sum_{h=0}^{H} \tilde{\boldsymbol{\pi}}_{ih}'; \boldsymbol{m}_{x_i^*}^H = \sum_{h=0}^{H} \tilde{\boldsymbol{\pi}}_{ih}^{*\prime}; \boldsymbol{m}_g^H = \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial \boldsymbol{g}_t'} = \sum_{h=0}^{H} \tilde{\boldsymbol{\psi}}_{ih}'$$

where $\tilde{\phi}_{ih}^*$, $\tilde{\boldsymbol{\pi}}_{ih}'$, and $\tilde{\boldsymbol{\pi}}_{ih}^{*\prime}$ for $h = 0, 1, ...,$ are evaluated using the recursive relationships, (28)-(30), and $\tilde{\boldsymbol{\psi}}_{ih}'$ for $h = 0, 1, ...,$ are evaluated as follows:

$$\tilde{\boldsymbol{\psi}}_{ih}' = \phi_{i1} \tilde{\boldsymbol{\psi}}_{i,h-1}' + \phi_{i2} \tilde{\boldsymbol{\psi}}_{i,h-2}' + \cdots + \phi_{i,h-1} \tilde{\boldsymbol{\psi}}_{i,1}' + \phi_{ih} \tilde{\boldsymbol{\psi}}_{i0}' + \boldsymbol{\pi}_{ih}', \; h = 1, 2, ...$$

where $\tilde{\boldsymbol{\psi}}_{i0}' = \boldsymbol{\pi}_{i0}'$, $\tilde{\boldsymbol{\psi}}_{ih}' = 0$ for $h < 0$. By construction, as $H \to \infty$, $m_{y_i^*}^H \to \beta_{y_i^*}; \boldsymbol{m}_{x_i}^H \to \boldsymbol{\beta}_{x_i}'; \boldsymbol{m}_{x_i^*}^H \to \boldsymbol{\beta}_{x_i^*}'; \boldsymbol{m}_{i,g}^H \to \boldsymbol{\beta}_{i,g}'$, where $\beta_{y_i^*}$, $\boldsymbol{\beta}_{x_i}$, $\boldsymbol{\beta}_{x_i^*}$ and $\boldsymbol{\beta}_{i,g}$ are the long-run multipliers. In particular, the cumulative dynamic multiplier effects of the k-th global factor, g_t^k on $y_{i,t+h}$ are the kth element of $\boldsymbol{m}_{i,g}^H$ given by

$$\boldsymbol{m}_{i,g^k}^H = \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial g_t^k} = \sum_{\ell=0}^{H} \psi_{i\ell}^k \text{ for } k = 1, ..., G$$

Next, to derive the diffusion multipliers of $\boldsymbol{y}_t$ with respect to $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$, we derive the following system spatial representation by stacking the individual STARDL-F(p, q) regressions, (50):

$$\boldsymbol{y}_t = \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \boldsymbol{y}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_\ell \boldsymbol{x}_{t-\ell} + \sum_{\ell=0}^{p} \boldsymbol{\Phi}_\ell^* \boldsymbol{W} \boldsymbol{y}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_\ell^* \left(\boldsymbol{W} \otimes \boldsymbol{I}_K\right) \boldsymbol{x}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Psi}_\ell \boldsymbol{g}_{t-\ell} + \boldsymbol{\alpha} + \boldsymbol{u}_t \tag{54}$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$, $\boldsymbol{\Phi}_\ell = diag\,(\phi_{1\ell}, ..., \phi_{N\ell})$ for $\ell = 1, ...p$, $\boldsymbol{\Phi}_\ell^* = diag\,(\phi_{1\ell}^*, ..., \phi_{N\ell}^*)$ for $\ell = 0, ...p$, and $\boldsymbol{\Pi}_\ell = diag\,(\boldsymbol{\pi}_{1\ell}', ..., \boldsymbol{\pi}_{N\ell}')$, $\boldsymbol{\Pi}_\ell^* = diag\,(\boldsymbol{\pi}_{1\ell}^{*\prime}, ..., \boldsymbol{\pi}_{N\ell}^{*\prime})$, $\boldsymbol{\Psi}_\ell = (\boldsymbol{\psi}_{1\ell}, ..., \boldsymbol{\psi}_{N\ell})'$ for $\ell = 0, ...q$. We then rewrite (54) as

$$\tilde{\boldsymbol{\Phi}}\,(L)\,\boldsymbol{y}_t = \tilde{\boldsymbol{\Pi}}\,(L)\,\boldsymbol{x}_t + \tilde{\boldsymbol{\Psi}}\,(L)\,\boldsymbol{g}_t + \tilde{\boldsymbol{\alpha}} + \tilde{\boldsymbol{u}}_t, \tag{55}$$

where $\tilde{\boldsymbol{\Phi}}\,(L) = \boldsymbol{I}_N - \sum_{\ell=1}^p \tilde{\boldsymbol{\Phi}}_\ell L^\ell$ with $\tilde{\boldsymbol{\Phi}}_\ell = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1} (\boldsymbol{\Phi}_\ell + \boldsymbol{\Phi}_\ell^* \boldsymbol{W})$, $\tilde{\boldsymbol{\Pi}}\,(L) = \sum_{\ell=0}^q \tilde{\boldsymbol{\Pi}}_\ell L^\ell$ with $\tilde{\boldsymbol{\Pi}}_\ell = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1} [\boldsymbol{\Pi}_\ell + \boldsymbol{\Pi}_\ell^* (\boldsymbol{W} \otimes \boldsymbol{I}_K)]$, $\tilde{\boldsymbol{\Psi}}\,(L) = \sum_{\ell=0}^q \tilde{\boldsymbol{\Psi}}_\ell L^\ell$ with $\tilde{\boldsymbol{\Psi}}_\ell = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1} \boldsymbol{\Psi}_\ell$, $\tilde{\boldsymbol{\alpha}} = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1} \boldsymbol{\alpha}$ and $\tilde{\boldsymbol{u}}_t = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W})^{-1} \boldsymbol{u}_t$. Premultiplying (55) by $\left[\tilde{\boldsymbol{\Phi}}\,(L)\right]^{-1}$, we obtain:

$$\boldsymbol{y}_t = \boldsymbol{B}\,(L)\,\boldsymbol{x}_t + \boldsymbol{C}\,(L)\,\boldsymbol{g}_t + \left[\tilde{\boldsymbol{\Phi}}\,(L)\right]^{-1} (\tilde{\boldsymbol{\alpha}} + \tilde{\boldsymbol{u}}_t) \tag{56}$$

where $\boldsymbol{B}\,(L) = \left[\tilde{\boldsymbol{\Phi}}\,(L)\right]^{-1} \tilde{\boldsymbol{\Pi}}\,(L)$ and $\boldsymbol{C}\,(L) = \left[\tilde{\boldsymbol{\Phi}}\,(L)\right]^{-1} \tilde{\boldsymbol{\Psi}}\,(L)$. The $\boldsymbol{B}_j$ and $\boldsymbol{C}_j$ for $j = 0, 1, ...$, are evaluated as follows:

$$\boldsymbol{B}_j = \tilde{\boldsymbol{\Phi}}_1 \boldsymbol{B}_{j-1} + \tilde{\boldsymbol{\Phi}}_2 \boldsymbol{B}_{j-2} + \cdots + \tilde{\boldsymbol{\Phi}}_{j-1} \boldsymbol{B}_1 + \tilde{\boldsymbol{\Phi}}_j \boldsymbol{B}_0 + \tilde{\boldsymbol{\Pi}}_j,\ j = 1, 2, ... \tag{57}$$

$$\boldsymbol{C}_j = \tilde{\boldsymbol{\Phi}}_1 \boldsymbol{C}_{j-1} + \tilde{\boldsymbol{\Phi}}_2 \boldsymbol{C}_{j-2} + \cdots + \tilde{\boldsymbol{\Phi}}_{j-1} \boldsymbol{C}_1 + \tilde{\boldsymbol{\Phi}}_j \boldsymbol{C}_0 + \tilde{\boldsymbol{\Psi}}_j,\ j = 1, 2, ... \tag{58}$$

where $\boldsymbol{B}_0 = \tilde{\boldsymbol{\Pi}}_0$ and $\boldsymbol{B}_j = 0$ for $j < 0$, and $\boldsymbol{C}_0 = \tilde{\boldsymbol{\Psi}}_0$ and $\boldsymbol{C}_j = 0$ for $j < 0$.

Then, the cumulative diffusion multiplier effects are evaluated as follows:

$$\boldsymbol{d}_{\boldsymbol{x}}^H = \sum_{h=0}^H \frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{x}_t'} = \sum_{h=0}^H \boldsymbol{B}_h; \boldsymbol{d}_{\boldsymbol{g}}^H = \sum_{h=0}^H \frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{g}_t'} = \sum_{h=0}^H \boldsymbol{C}_h,\ H = 0, 1, 2, ... \tag{59}$$

The cumulative diffusion multiplier effects of x_{jt}^k on $y_{i,t+h}$ are given by the $(i, (j-1)\,K + k)$th element of the $N \times NK$ matrix, $\boldsymbol{m}_{\boldsymbol{x}}^H$ while those of g_t^k on $y_{i,t+h}$ are given by the (i, k)th element of the $N \times G$ matrix, $\boldsymbol{m}_{\boldsymbol{g}}^H$.

$\boldsymbol{m}_g^H$ measures the individual dynamic multipliers while $\boldsymbol{d}_{\boldsymbol{g}}^H$ captures the system diffusion multiplier effects with respect to $\boldsymbol{g}_t$. Thus, the difference between $\boldsymbol{d}_{\boldsymbol{g}}^H$ and $\boldsymbol{m}_{\boldsymbol{g}}^H$ may indicate the additional spatial impacts of $\boldsymbol{g}_t$ at each forecast horizon, though it is not possible to separate out the spatial and time dependence from the diffusion multipliers with respect to $\boldsymbol{x}_t$.

7.2 STARDL Models with unobserved common factors

Consider the STARDL(p, q) model with an $r \times 1$ vector of unobserved common factors, $\boldsymbol{f}_t = (f_{1t}, ..., f_{rt})'$:

$$y_{it} = \sum_{\ell=1}^p \phi_{i\ell} y_{i,t-\ell} + \sum_{\ell=0}^p \phi_{i\ell}^* y_{i,t-\ell}^* + \sum_{\ell=0}^q \boldsymbol{\pi}_{i\ell}' \boldsymbol{x}_{i,t-\ell} + \sum_{\ell=0}^q \boldsymbol{\pi}_{i\ell}'^* \boldsymbol{x}_{i,t-\ell}^* + \alpha_i + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + u_{it}, \tag{60}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where y_{it} is the scalar dependent variable of the ith spatial unit at time t, $\boldsymbol{x}_{it} = (x_{it}^1, ..., x_{it}^K)'$ is a $K \times 1$ vector of exogenous regressors, $y_{it}^* = \sum_{j=1}^N w_{ij} y_{jt}$ and $x_{it}^* = \sum_{j=1}^N w_{ij} x_{jt}$ are spatially lagged values with conformable parameters, $\boldsymbol{\lambda}_i = (\lambda_{1,i}, ..., \lambda_{r,i})'$is the heterogenous loadings and u_{it} is the idiosyncratic errors. In the homogenous spatial model the QML and GMM approach have been developed together with unobserved factors (e.g. Bai and Li (2015), Shi and Lee (2017), and Kuersteiner and Prucha (2018)). By contrast no unified approach has been developed yet in the heterogenous case.

We develop the two estimation procedures, depending on whether we impose the specific data generating condition on $\boldsymbol{x}_{it}$'s or not. In the first approach we allow $\boldsymbol{x}_{it}$'s to be correlated arbitrarily with the common factors and/or the factor loadings. We then estimate $\boldsymbol{f}_t$ directly and develop an iterative principal components analysis. In the second approach, we assume that $\boldsymbol{x}_{it}$'s follow the VAR processes but also share the unobserved common factors. Then, we propose the QML-EM method in which case we only estimate the sample variance of the common factor $\boldsymbol{f}_t$, not $\boldsymbol{f}_t$.

For convenience we write (60) compactly as

$$y_{it} = \phi_{i0}^* y_{it}^* + \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it} + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + u_{it} \tag{61}$$

where $\boldsymbol{\theta}_i = \left(\boldsymbol{\phi}_i', \boldsymbol{\phi}_i^{*\prime}, \boldsymbol{\pi}_i', \boldsymbol{\pi}_i'^{*}\right)'$ and $\boldsymbol{\chi}_{it} = \left(\boldsymbol{y}_{i,-\ell}', \boldsymbol{y}_{i,-\ell}^{*\prime}, \boldsymbol{x}_{i,-\ell}', \boldsymbol{x}_{i,-\ell}^{*\prime}, 1\right)'$ with $\boldsymbol{\phi}_i = (\phi_{i1}, ..., \phi_{ip})'$, $\boldsymbol{\phi}_i^* = \left(\phi_{i1}^*, ..., \phi_{ip}^*\right)'$, $\boldsymbol{\pi}_i = \left(\boldsymbol{\pi}_{i0}', ..., \boldsymbol{\pi}_{iq}'\right)'$, $\boldsymbol{\pi}_i^* = \left(\boldsymbol{\pi}_{i0}^{*\prime}, ..., \boldsymbol{\pi}_{iq}^{*\prime}\right)'$, and $\boldsymbol{y}_{i,-p} = (y_{i,t-1}, ..., y_{i,t-p})'$, $\boldsymbol{y}_{i,-p}^* = \left(y_{i,t-1}^*, ..., y_{i,t-p}^*\right)'$, $\boldsymbol{x}_{i,-q} = \left(\boldsymbol{x}_{it}', ..., \boldsymbol{x}_{i,t-q}'\right)'$, $\boldsymbol{x}_{i,-q}^* = \left(\boldsymbol{x}_{it}^{*\prime}, ..., \boldsymbol{x}_{i,t-q}^{*\prime}\right)'$. Stacking (61), we have[14]

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_0^* \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{\Theta} \boldsymbol{\chi}_t + \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{u}_t \tag{62}$$

where $\boldsymbol{\Phi}_0^* = diag\left(\phi_{10}^*, ..., \phi_{N0}^*\right)$, $\boldsymbol{\Theta} = diag\left(\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_N'\right)$, $\boldsymbol{\chi}_t = \left(\boldsymbol{\chi}_{1t}', ...\boldsymbol{\chi}_{Nt}'\right)'$ and $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_N)'$.

The IPC-QML estimator We allow the explanatory variables $\boldsymbol{x}_{it}$ to be arbitrarily correlated with $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$, and estimate both $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$ together as parameters. We make the following additional assumptions.

Assumption F1: The $\boldsymbol{f}_t$ are random and independent of u_{is} for all t and s.

Assumption F2: The factor loadings $\boldsymbol{\lambda}_i$ are random such that $E\left(\|\Gamma_i\|^4\right) \leq C$ for all i and $N^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_u\boldsymbol{\Lambda} \to_p \boldsymbol{\Omega}_r$, where $\boldsymbol{\Sigma}_u = diag\left(\sigma_{u1}^2, ..., \sigma_{uN}^2\right)$, and $\boldsymbol{\Omega}_r$ is some positive definite matrix. Further, $\boldsymbol{\lambda}_i$s are independent of the idiosyncratic errors u_{jt} for all i and j.

[14]Notice that (62) can also be written as

$$\boldsymbol{y}_i = \phi_{i0}^* \boldsymbol{y}_i^* + \boldsymbol{\chi}_i \boldsymbol{\theta}_i + \boldsymbol{F} \boldsymbol{\lambda}_i + \boldsymbol{u}_i$$

where $\boldsymbol{y}_i = (y_{i1}, ..., y_{iT})'$, $\boldsymbol{y}_i^* = \left(y_{i1}^*, ..., y_{iT}^*\right)'$, $\boldsymbol{\chi}_i = \left(\boldsymbol{\chi}_{i1}', ..., \boldsymbol{\chi}_{iT}'\right)'$, $\boldsymbol{F} = (\boldsymbol{f}_1, ..., \boldsymbol{f}_T)'$, and $\boldsymbol{u}_i = (u_{i1}, ..., u_{iT})'$.

The (quasi) log-likelihood function of (62) can be derived as

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\Phi}_0^*,\boldsymbol{\Theta},\boldsymbol{F}\right) &= -\frac{NT}{2}\ln 2\pi-\frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2+T\ln\left|\boldsymbol{I}_N-\boldsymbol{\Phi}_0^*\boldsymbol{W}\right| \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\left[\left(\boldsymbol{I}_N-\boldsymbol{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t-\boldsymbol{\Theta}\boldsymbol{\chi}_t-\boldsymbol{\Lambda}\boldsymbol{f}_t\right]'\boldsymbol{\Sigma}_u^{-1}\left[\left(\boldsymbol{I}_N-\boldsymbol{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t-\boldsymbol{\Theta}\boldsymbol{\chi}_t-\boldsymbol{\Lambda}\boldsymbol{f}_t\right]\end{aligned} \tag{63}$$

Given $\boldsymbol{\Phi}_0^*$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Lambda}$, it is easily seen that $\boldsymbol{f}_t$ maximize $\mathcal{L}\left(\boldsymbol{\Phi}_0^*,\boldsymbol{\Theta},\boldsymbol{F}\right)$ at

$$\boldsymbol{f}_t=\left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_u^{-1}\left[\left(\boldsymbol{I}_N-\boldsymbol{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t-\boldsymbol{\Theta}\boldsymbol{\chi}_t\right]. \tag{64}$$

Substituting (64) in (63), we obtain the concentrated likelihood function:

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\Phi}_0^*,\boldsymbol{\Theta}\right) &= -\frac{NT}{2}\ln 2\pi-\frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2+T\ln\left|\boldsymbol{I}_N-\boldsymbol{\Phi}_0^*\boldsymbol{W}\right| \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\left[\left(\boldsymbol{I}_N-\boldsymbol{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t-\boldsymbol{\Theta}\boldsymbol{\chi}_t\right]'\boldsymbol{M}_u\left[\left(\boldsymbol{I}_N-\boldsymbol{\Phi}_0^*\boldsymbol{W}\right)\boldsymbol{y}_t-\boldsymbol{\Theta}\boldsymbol{\chi}_t\right]\end{aligned} \tag{65}$$

where

$$\boldsymbol{M}_u=\boldsymbol{\Sigma}_u^{-1}-\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_u^{-1}=\boldsymbol{\Sigma}_u^{-1}-\frac{1}{N}\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_u^{-1}$$

with $\frac{1}{N}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\Lambda}=\boldsymbol{I}_r$. The QML estimator of $\left(\boldsymbol{\Phi}_0^*,\boldsymbol{\Theta}\right)$ is defined by

$$\hat{\boldsymbol{\theta}}=\arg\max_{\boldsymbol{\theta}\in\Theta}\mathcal{L}\left(\boldsymbol{\Phi}_0^*,\boldsymbol{\Theta}\right)$$

where $\boldsymbol{\Theta}$ is the compact parameters space.

The IPC-QML estimator is updated via the following iterative procedure. Denote the estimates of $\boldsymbol{\Phi}_0^*$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}_u$ at the sth iteration by $\hat{\boldsymbol{\Phi}}_0^{*(s)}$, $\hat{\boldsymbol{\Theta}}^{(s)}$ and $\hat{\boldsymbol{\Sigma}}_u^{(s)}$, respectively.

Step 1: Given $\hat{\boldsymbol{\Phi}}_0^{*(s)}$, $\hat{\boldsymbol{\Theta}}^{(s)}$ and $\hat{\boldsymbol{\Sigma}}_u^{(s)}$, we estimate $\hat{\boldsymbol{\Lambda}}^{(s+1)}$ as the first r eigenvectors associated with the first r largest eigenvalues of the $N\times N$ matrix,

$$\hat{\boldsymbol{G}}=\left[\frac{1}{NT}\sum_{t=1}^{T}\left(\boldsymbol{y}_t-\hat{\boldsymbol{\Phi}}_0^{*(s)}\boldsymbol{W}\boldsymbol{y}_t-\hat{\boldsymbol{\Theta}}^{(s)}\boldsymbol{\chi}_t\right)\left(\boldsymbol{y}_t-\hat{\boldsymbol{\Phi}}_0^{*(s)}\boldsymbol{W}\boldsymbol{y}_t-\hat{\boldsymbol{\Theta}}^{(s)}\boldsymbol{\chi}_t\right)'\right]\left(\hat{\boldsymbol{\Sigma}}_u^{(s)}\right)^{-1}$$

and $\hat{\boldsymbol{f}}_t^{(s+1)}$ by

$$\hat{\boldsymbol{f}}_t^{(s+1)}=\frac{1}{N}\hat{\boldsymbol{\Lambda}}^{(s+1)\prime}\left(\hat{\boldsymbol{\Sigma}}_u^{(s)}\right)^{-1}\left(\boldsymbol{y}_t-\hat{\boldsymbol{\Phi}}_0^{*(s)}\boldsymbol{W}\boldsymbol{y}_t-\hat{\boldsymbol{\Theta}}^{(s)}\boldsymbol{\chi}_t\right),\ t=1,...,T.$$

We then construct

$$\hat{C}_{it}^{(s+1)}=\hat{\boldsymbol{\lambda}}_i^{(s+1)\prime}\hat{\boldsymbol{f}}_t^{(s+1)}\text{ and }\hat{\boldsymbol{C}}_i^{(s+1)}=\left(\hat{C}_{i1}^{(s+1)},...,\hat{C}_{iT}^{(s+1)}\right)'$$

Step 2: Given $\hat{C}_{it}^{(s+1)}$ and $\hat{\boldsymbol{C}}_i^{(s+1)}$, we update update $\hat{\boldsymbol{\Sigma}}_u^{(s+1)}$ by

$$\left(\hat{\sigma}_i^{2)}\right)^{(s+1)} = \frac{1}{T}\sum_{t=1}^{T}\left(\left(y_{it} - \hat{C}_{it}^{(s+1)}\right) - \hat{\phi}_{i0}^{*(s)} y_{it}^{*} - \hat{\boldsymbol{\theta}}_i^{(s)\prime}\boldsymbol{\chi}_{it}\right)^2, \ i = 1, ..., N$$

and $\hat{\boldsymbol{\theta}}_i^{(s+1)}$ by

$$\hat{\boldsymbol{\theta}}_i^{(s+1)} = \left(\boldsymbol{\chi}_i'\boldsymbol{\chi}_i\right)^{-1}\left(\boldsymbol{\chi}_i'\boldsymbol{\chi}_i\right)\left\{\left(\boldsymbol{y}_i - \hat{\boldsymbol{C}}_i^{(s+1)}\right) - \hat{\phi}_{i0}^{*(s)}\boldsymbol{y}_i^{*}\right\}, \ i = 1, ..., N.$$

Finally, we update $\hat{\boldsymbol{\Phi}}_0^{*(s+1)}$ by maximizing $\mathcal{L}\left(\boldsymbol{\Phi}_0^*, \boldsymbol{\Theta}\right)$ in (65) directly with respect to $\boldsymbol{\Phi}_0^*$ at $\boldsymbol{\Lambda} = \hat{\boldsymbol{\Lambda}}^{(s+1)}, \boldsymbol{\Sigma}_u = \hat{\boldsymbol{\Sigma}}_u^{(s+1)}$, and $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}^{(s+1)}$. We repeat Steps 1 and 2 until convergence at a preset tolerance. We may use the within-group estimator as the starting value for $\boldsymbol{\Phi}_0^*$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}_u^{(s)}$.

Notice that there is a bias term of order $O_p\left(N^{-1}\right)$, see also Lu (2017). This bias comes from the incidental parameters involved when estimating $\boldsymbol{f}_t$ directly as parameters. Similarly, the extra $O_p(N^{-2})$ term included in the average convergence rates for $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Lambda}_i$ occurs for the same reason. Because these extra terms depend only on N, the IPC-QML estimator is no longer consistent under fixed N.

Alternatively, it is often more convenient to write the above concentrated log-likelihood function $\mathcal{L}\left(\boldsymbol{\Phi}_0^*, \boldsymbol{\Theta}\right)$ in (65) as

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\Phi}_0^*, \boldsymbol{\Theta}\right) \quad = \quad & -\frac{NT}{2}\ln 2\pi - \frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2 + T\ln\left|\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}\right| \\ & -\frac{1}{2}\sum_{i=1}^{N}\frac{\left(\boldsymbol{y}_i - \phi_{i0}^*\boldsymbol{y}_i^* - \boldsymbol{\chi}_i\boldsymbol{\theta}_i\right)'\left(\boldsymbol{y}_i - \phi_{i0}^*\boldsymbol{y}_i^* - \boldsymbol{\chi}_i\boldsymbol{\theta}_i\right)}{\tilde{\sigma}_i^2}\end{aligned} \tag{66}$$

where $\tilde{\sigma}_i^2$ is the ith diagonal element of the $N \times N$ matrix, $\boldsymbol{M}_u^{-1}$. Then, following ABP, we update $\hat{\boldsymbol{\Phi}}_0^{*(s+1)}$ by maximising the following concentrated log-likelihood function:

$$\begin{aligned}\mathcal{L}_C\left(\boldsymbol{\Phi}_0^*\right) \quad \propto \quad & T\ln\left|\boldsymbol{I}_N - \hat{\boldsymbol{\Phi}}_0^{*(s+1)}\boldsymbol{W}\right| \\ & -\frac{1}{2T}\sum_{i=1}^{N}\left(\left(\boldsymbol{y}_i - \hat{\boldsymbol{C}}_i^{(s+1)}\right) - \hat{\phi}_{i0}^{*(s+1)}\boldsymbol{y}_i^*\right)'\boldsymbol{M}_{\chi_i}\left(\left(\boldsymbol{y}_i - \hat{\boldsymbol{C}}_i^{(s+1)}\right) - \hat{\phi}_{i0}^{*(s+1)}\boldsymbol{y}_i^*\right)\end{aligned} \tag{67}$$

where $\boldsymbol{M}_{\chi_i} = \boldsymbol{I}_T - \boldsymbol{\chi}_i\left(\boldsymbol{\chi}_i'\boldsymbol{\chi}_i\right)^{-1}\boldsymbol{\chi}_i'$ and $\boldsymbol{\chi}_i = \left(\boldsymbol{\chi}_{i1}', ...\boldsymbol{\chi}_{iT}'\right)'$. Then, we can update other parameters, $\hat{\boldsymbol{\theta}}_i^{(s+1)}$ by least squares applied to the individual equations (61) conditional on $\hat{\phi}_{i0}^{*(s+1)}$ and $\hat{\boldsymbol{C}}_i^{(s+1)}$, and similarly for $\left(\hat{\sigma}_i^{2)}\right)^{(s+1)}$, $i = 1, ..., N$.

We may apply the STARDL-CF estimator as follows: Given the consistent estimate of $\hat{C}_{it}$, we update all other parameters including ϕ_{i0}^* by running the following augmented regression:

$$\left(y_{it} - \hat{C}_{it}\right) = \sum_{\ell=1}^{p} \phi_{i\ell} y_{i,t-\ell} + \sum_{\ell=0}^{p} \phi_{i\ell}^{*} y_{i,t-\ell}^{*} + \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}' \boldsymbol{x}_{i,t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}'^{*} \boldsymbol{x}_{i,t-\ell}^{*} + \alpha_i + \rho \hat{v}_{it} + e_{it}^{*}, \tag{68}$$

where $\hat{v}_{it} = y_{it}^{*} - \hat{\boldsymbol{\varphi}}_i' \boldsymbol{z}_{it}$ and $e_{it}^{*} = e_{it} + \rho \left(\hat{\boldsymbol{\varphi}}_i - \boldsymbol{\varphi}_i\right)' \boldsymbol{z}_{it} + \left(C_{it} - \hat{C}_{it}\right)$.

The QML-EM estimator Next, we develop the QML-EM algorithms. To this end we assume that $\boldsymbol{x}_{it}$ follows the VAR(p) process and shares the same unobserved factors, $\boldsymbol{f}_t$:[15]

$$\boldsymbol{x}_{it} = \sum_{\ell=1}^{p} \Psi_i \boldsymbol{x}_{i,t-\ell} + \boldsymbol{b}_i + \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \boldsymbol{v}_{it} \tag{69}$$

Then, (60) and (69) can be written as

$$\begin{aligned} & \left[\begin{array}{c} y_{it} - \sum_{\ell=1}^{p} \phi_{i\ell} y_{i,t-\ell} - \sum_{\ell=0}^{p} \phi_{i\ell}^{*} y_{i,t-\ell}^{*} - \sum_{\ell=0}^{p} \boldsymbol{\pi}_{i\ell}' \boldsymbol{x}_{i,t-\ell} - \sum_{\ell=0}^{p} \boldsymbol{\pi}_{i\ell}'^{*} \boldsymbol{x}_{i,t-\ell}^{*} \\ \boldsymbol{x}_{it} - \sum_{\ell=1}^{p} \Psi_i \boldsymbol{x}_{i,t-\ell} \end{array} \right] \\ = \; & \boldsymbol{\mu}_i + \boldsymbol{\Phi}_i' \boldsymbol{f}_t + \boldsymbol{\epsilon}_{it} \end{aligned} \tag{70}$$

where

$$\boldsymbol{\mu}_i = \left[\begin{array}{c} \alpha_i \\ \boldsymbol{v}_i \end{array} \right]; \boldsymbol{\Phi}_i' = \left[\begin{array}{c} \boldsymbol{\lambda}_i' \\ \boldsymbol{\gamma}_i' \end{array} \right]; \boldsymbol{\epsilon}_{it} = \left[\begin{array}{c} u_{it} \\ \boldsymbol{v}_{it} \end{array} \right]$$

Let $\boldsymbol{z}_t = \left(\boldsymbol{z}_{1t}', \boldsymbol{z}_{2t}', ..., \boldsymbol{z}_{Nt}'\right)'$ with $\boldsymbol{z}_{it} = \left(y_{it}, \boldsymbol{x}_{it}'\right)'$. Then, we can write (70) compactly as

$$\boldsymbol{D}(L) \underset{N(k+1)\times 1}{\boldsymbol{z}_t} = \boldsymbol{\mu} + \underset{N(k+1)\times r}{\boldsymbol{\Phi}} \underset{r\times 1}{\boldsymbol{f}_t} + \boldsymbol{\epsilon}_t \tag{71}$$

where $\boldsymbol{\mu} = \left(\boldsymbol{\mu}_1', ..., \boldsymbol{\mu}_N'\right)'$, $\boldsymbol{\Phi} = \left(\boldsymbol{\Phi}_1, ..., \boldsymbol{\Phi}_N\right)'$, $\boldsymbol{\epsilon}_t = \left(\boldsymbol{\epsilon}_{1t}', ..., \boldsymbol{\epsilon}_{Nt}'\right)'$, and $\boldsymbol{D}(L) = \boldsymbol{D}_0 - \sum_{\ell=1}^{p} \boldsymbol{D}_\ell L^\ell$ with $\boldsymbol{D}_0$ and $\boldsymbol{D}_\ell$ being the $N(k+1) \times N(k+1)$ matrices with the (i,j) sub-blocks respectively given by

$$\underset{(k+1)\times(k+1)}{\boldsymbol{D}_{0,ij}} = \left\{ \begin{array}{c} \left[\begin{array}{cc} 1 & -\boldsymbol{\pi}_{i0}' \\ 0 & \boldsymbol{I}_k \end{array} \right], \; if \; i = j \\ \left[\begin{array}{cc} \phi_{i0}^{*} w_{ij} & \boldsymbol{\pi}_{i0}^{*\prime} w_{ij} \\ 0 & 0 \end{array} \right], \; if \; i \neq j \end{array} \right\}$$

$$\underset{(k+1)\times(k+1)}{\boldsymbol{D}_{\ell,ij}} = \left\{ \begin{array}{c} \left[\begin{array}{cc} \phi_{i\ell} & \boldsymbol{\pi}_{i\ell}' \\ 0 & \Psi_{i\ell} \end{array} \right], \; if \; i = j \\ \left[\begin{array}{cc} \phi_{i\ell}^{*} w_{ij} & \boldsymbol{\pi}_{i\ell}^{*\prime} w_{ij} \\ 0 & 0 \end{array} \right], \; if \; i \neq j \end{array} \right\}, \ell = 1, ..., p$$

To analyse model (71), we make the following assumptions (e.g. Bai and Li, 2104).

[15] For convenience we set $p = q$ in (60) without of loss of generality.

Assumption Q1: The idiosyncratic errors $\boldsymbol{\epsilon}_{it} = (u_{it}, \boldsymbol{v}_{it}')'$ are such that (i) u_{it} is independent and identically distributed over t and uncorrelated over i with $E(u_{it}) = 0$ and $E(u_{it}^4) \leq \infty$ for all $i = 1, ..., N$ and $t = 1, \ldots, T$. (ii) $\boldsymbol{v}_{it}$ is also independent and identically distributed over t and uncorrelated over i with $E(\boldsymbol{v}_{it}) = 0$ and $E(\|\boldsymbol{v}_{it}\|^4) \leq \infty$ for all $i = 1, ..., N$ and $t = 1, \ldots, T$. (iii) u_{it} is independent of $\boldsymbol{v}_{js}$ for all (i, j, t, s). Let $\boldsymbol{\Sigma}_{ii} = diag\left(\sigma_i^2, \boldsymbol{\Sigma}_{iiv}\right)$ denote the variance matrix $\boldsymbol{\epsilon}_{it}$, where σ_i^2 is the variance of e_{it} and $\boldsymbol{\Sigma}_{iiv}$ denotes the variance matrix of $\boldsymbol{v}_{it}$.

Assumption Q2: There exists a $C > 0$ sufficiently large such that (i) $\|\Phi_i\| \leq C$ for all $i = 1, ..., N$; (ii) $C^{-1} \leq \tau_{\min}(\boldsymbol{\Sigma}_{jj}) \leq \tau_{\max}(\boldsymbol{\Sigma}_{jj}) \leq C$ for all $j = 1, ..., N$, where $\tau_{\min}(\boldsymbol{\Sigma}_{jj})$ and $\tau_{\max}(\boldsymbol{\Sigma}_{jj})$ denote the smallest and largest eigenvalues of $\boldsymbol{\Sigma}_{jj}$; (iii) There exists an $r \times r$ positive matrix $\boldsymbol{Q}$ such that $\boldsymbol{Q} = \lim_{N\to\infty} N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\Phi}$, and $\boldsymbol{\Sigma}_{\epsilon\epsilon} = diag\left(\boldsymbol{\Sigma}_{11}, ..., \boldsymbol{\Sigma}_{NN}\right)$.

Assumption Q3: The variances $\boldsymbol{\Sigma}_{ii}$ for all i and $\boldsymbol{M}_{ff}$ are estimated in a compact set, *i.e.* all the eigenvalues of $\hat{\boldsymbol{\Sigma}}_{ii}$ and $\hat{\boldsymbol{M}}_{ff}$ are in an interval $\left[C^{-1}, C\right]$.

Assumption Q4: Identification conditions.[16] To remove the rotational indeterminacy, we impose the normalization restrictions: (i) $\bar{\boldsymbol{f}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{f}_t = 0$; (ii) $\boldsymbol{M}_{ff} = T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{f}_t - \bar{\boldsymbol{f}}\right)\left(\boldsymbol{f}_t - \bar{\boldsymbol{f}}\right)' = \boldsymbol{I}_r$; and (iii) $N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{\Phi}$ is diagonal with the diagonal elements being distinct and arranged in descending order.

The objective function for the model (71) is given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{\xi}) \quad = \quad & -\frac{1}{2N}\ln|\boldsymbol{\Sigma}_{zz}| + \frac{1}{N}\ln|\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}| \\ & -\frac{1}{2NT}\sum_{t=1}^{T}\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)'\boldsymbol{\Sigma}_{zz}^{-1}\left(\boldsymbol{D}_0\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)\end{aligned} \tag{72}$$

where $\boldsymbol{\xi} = (\boldsymbol{\phi}_0^*, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}_{\epsilon\epsilon})$ with $\boldsymbol{\phi}_0^* = (\phi_{10}^*, ..., \phi_{N0}^*)'$, $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_N'\right)'$ and $\boldsymbol{\Sigma}_{zz} = \Phi\Phi' + \boldsymbol{\Sigma}_{\varepsilon\varepsilon}$.

Following Bai and Li (2014) we develop the QML-EM algorithms as follows: The algorithm combines the usual maximisation procedure with the EM algorithm. Let $\boldsymbol{\xi}^{(s)} = \left(\boldsymbol{\phi}_0^{*(s)}, \boldsymbol{\theta}^{(s)}, \boldsymbol{\Phi}^{(s)}, \boldsymbol{\Sigma}_{\epsilon\epsilon}^{(s)}\right)$ denote the estimates at the sth iteration. Our updating procedures consist of two steps.

Step 1: We update $\boldsymbol{\Phi}^{(s)}, \boldsymbol{\Sigma}_{\epsilon\epsilon}^{(s)}$ and $\boldsymbol{\theta}^{(s)}$ according to the EM algorithm:

$$\Phi^{(s+1)} = \left[\frac{1}{T}\sum_{t=1}^{T}E\left(\left(\boldsymbol{D}_0^{(s)}\boldsymbol{z}_t - \sum_{\ell=1}^{p}\boldsymbol{D}_\ell^{(s)}\boldsymbol{z}_{t-\ell}\right)\boldsymbol{f}_t'|\theta_s\right)\right]\left[\frac{1}{T}\sum_{t=1}^{T}E\left(\boldsymbol{f}_t\boldsymbol{f}_t'|\theta_s\right)\right]^{-1}$$

$$\boldsymbol{\Sigma}_{\epsilon\epsilon}^{(s+1)} = Dg\left[\begin{array}{c}\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{D}_0^{(s)}\boldsymbol{z}_t - \sum\limits_{\ell=1}^{p}\boldsymbol{D}_\ell^{(s)}\boldsymbol{z}_{t-\ell}\right)\left(\boldsymbol{D}_0^{(s)}\boldsymbol{z}_t - \sum\limits_{\ell=1}^{p}\boldsymbol{D}_\ell^{(s)}\boldsymbol{z}_{t-\ell}\right)' \\ -\boldsymbol{\Phi}^{(s+1)}\boldsymbol{\Phi}^{(s)\prime}\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{D}_0^{(s)}\boldsymbol{z}_t - \sum\limits_{\ell=1}^{p}\boldsymbol{D}_\ell^{(s)}\boldsymbol{z}_{t-\ell}\right)\left(\boldsymbol{D}_0^{(s)}\boldsymbol{z}_t - \sum\limits_{\ell=1}^{p}\boldsymbol{D}_\ell^{(s)}\boldsymbol{z}_{t-\ell}\right)'\end{array}\right] \tag{73}$$

[16] The estimation of key parameters $\boldsymbol{\omega} = (\boldsymbol{\phi}_0^{*\prime}, \boldsymbol{\theta}')'$ and $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ is invariant to the different normalization restrictions.

and

$$\begin{aligned}\boldsymbol{\theta}_i^{(s+1)} &= \left[\sum_{t=1}^{T}\frac{1}{\left(\sigma_i^{(s+1)}\right)^2}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right]^{-1}\left[\sum_{t=1}^{T}\frac{1}{\left(\sigma_i^{(s+1)}\right)^2}\boldsymbol{\chi}_{it}\left(y_{it}-\phi_{i0}^{*(s)}y_{it}^{*}-\boldsymbol{\lambda}_i^{(s+1)\prime}\boldsymbol{f}_t^{(s)}\right)\right] \\ &= \left[\sum_{t=1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right]^{-1}\left[\sum_{t=1}^{T}\boldsymbol{\chi}_{it}\left(y_{it}-\phi_{i0}^{*(s)}y_{it}^{*}-\boldsymbol{\lambda}_i^{(s+1)\prime}\boldsymbol{f}_t^{(s)}\right)\right]\end{aligned}$$

for $i=1,...,N$, where Dg is the operator which sets the entries of its argument to zeros if the counterparts of $E\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right)$ are zeros; $\left(\sigma_i^{(s+1)}\right)^2$ is the $[(i-1)(k+1)+1]$th diagonal element of $\boldsymbol{\Sigma}_{\epsilon\epsilon}^{(s+1)}$ and $\boldsymbol{\lambda}_i^{(s+1)\prime}$ is the transpose of the $[(i-1)(k+1)+1]$th row of $\boldsymbol{\Phi}^{(s+1)}$. In addition,

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}E\left(\left(\boldsymbol{D}_0\boldsymbol{z}_t-\sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)\boldsymbol{f}_t'|\theta_s\right) \\ =\;& \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{D}_0\boldsymbol{z}_t-\sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)\left(\boldsymbol{D}_0\boldsymbol{z}_t-\sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)'\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)^{-1}\boldsymbol{\Phi}^{(s)}\end{aligned}$$

$$\frac{1}{T}\sum_{t=1}^{T}E\left(\boldsymbol{f}_t\boldsymbol{f}_t'|\theta_s\right)=\boldsymbol{I}_r-\boldsymbol{\Phi}^{(s)\prime}\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)^{-1}\boldsymbol{\Phi}^{(s)}$$

$$+\boldsymbol{\Phi}^{(s)\prime}\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{D}_0\boldsymbol{z}_t-\sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)\left(\boldsymbol{D}_0\boldsymbol{z}_t-\sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right)'\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)^{-1}\boldsymbol{\Phi}^{(s)}$$

and

$$\boldsymbol{f}_t^{(s)}=\boldsymbol{\Phi}^{(s)\prime}\left(\boldsymbol{\Sigma}_{zz}^{(s)}\right)^{-1}\left(\boldsymbol{D}_0\boldsymbol{z}_t-\sum_{\ell=1}^{p}\boldsymbol{D}_\ell\boldsymbol{z}_{t-\ell}\right).$$

Step 2: We update $\boldsymbol{\phi}_0^*$ by maximising (72) with respect to $\boldsymbol{\phi}_0^*$ at $\boldsymbol{\theta}_i=\hat{\boldsymbol{\theta}}_i^{(s+1)}$, $\boldsymbol{\Phi}=\hat{\boldsymbol{\Phi}}^{(s+1)}$ and $\boldsymbol{\Sigma}_{\epsilon\epsilon}=\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon}^{(s+1)}$ with an initial value of $\boldsymbol{\phi}_0^*$ at $\hat{\boldsymbol{\phi}}_0^{*(s)}$.

This iterative procedure guarantees that the value of likelihood function does not decrease in each iteration. This is because

$$L\left(\rho^{(s)},\beta^{(s+1)},\Phi^{(s+1)},\Sigma_{\epsilon\epsilon}^{(s+1)}\right)\geq L\left(\rho^{(s)},\beta^{(s)},\Phi^{(s)},\Sigma_{\epsilon\epsilon}^{(s)}\right)$$

$$L\left(\rho^{(s+1)},\beta^{(s+1)},\Phi^{(s+1)},\Sigma_{\epsilon\epsilon}^{(s+1)}\right)\geq L\left(\rho^{(s)},\beta^{(s+1)},\Phi^{(s+1)},\Sigma_{\epsilon\epsilon}^{(s+1)}\right)$$

We show that the limit of the iterated solution satisfies the first order conditions and therefore possesses the local optimality property. For the initial estimates, $\boldsymbol{\xi}^{(0)}=\left(\boldsymbol{\phi}_0^{*(0)},\boldsymbol{\theta}^{(0)},\boldsymbol{\Phi}^{(0)},\boldsymbol{\Sigma}_{\epsilon\epsilon}^{(0)}\right)$, $\boldsymbol{\phi}_0^{*(0)}$ and $\boldsymbol{\theta}^{(0)}$ can be set to the within group estimator, ignoring the endogeneity issue. And $\boldsymbol{\Phi}^{(0)}$ and $\boldsymbol{\Sigma}_{\epsilon\epsilon}^{(0)}$ are then the maximizer of (72) at $\boldsymbol{\phi}_0^{*(0)}$ and $\boldsymbol{\theta}^{(0)}$.

Remark: Notice that $\Sigma_{\epsilon\epsilon}$ is the $N(k+1) \times N(k+1)$ block-diagonal matrix given by

$$\underset{N(k+1)\times N(k+1)}{\Sigma_{\epsilon\epsilon}} = \begin{bmatrix} \Sigma_{11} & 0 & \cdots & 0 \\ 0 & \Sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{NN} \end{bmatrix}, \quad \underset{(k+1)\times(k+1)}{\Sigma_{ii}} = \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & \Sigma_{iiv} \end{bmatrix}$$

Then, σ_i^2 is estimated by

$$\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{it}^2, \;\; i = 1, ..., N$$

where

$$\hat{u}_{it} = \left(y_{it} - \left(\hat{\boldsymbol{\lambda}}_i' \hat{\boldsymbol{f}}_t\right)\right) - \hat{\phi}_{i0}^* y_{it}^* - \hat{\boldsymbol{\theta}}_i' \boldsymbol{\chi}_{it},$$

Next, Σ_{iiv} can be estimated by

$$\hat{\Sigma}_{iiv} = \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{v}}_{it} \hat{\boldsymbol{v}}_{it}', \;\; i = 1, ..., N$$

where

$$\hat{\boldsymbol{v}}_{it} = \left(\boldsymbol{x}_{it} - \hat{\boldsymbol{\gamma}}_i' \hat{\boldsymbol{f}}_t\right) - \sum_{\ell=1}^{p} \hat{\Psi}_i \boldsymbol{x}_{i,t-\ell} - \hat{\boldsymbol{b}}_i$$

where $\hat{\Psi}_i$ and $\hat{\boldsymbol{b}}_i$ are the OLS estimator obtained from the following modified VAR model:

$$\left(\boldsymbol{x}_{it} - \hat{\boldsymbol{\gamma}}_i' \hat{\boldsymbol{f}}_t\right) = \sum_{\ell=1}^{p} \Psi_i \boldsymbol{x}_{i,t-\ell} + \boldsymbol{b}_i + \boldsymbol{v}_{it}.$$

Then, update $\Sigma_{\epsilon\epsilon}$ by

$$\underset{N(k+1)\times N(k+1)}{\hat{\Sigma}_{\epsilon\epsilon}} = \begin{bmatrix} \hat{\Sigma}_{11} & & 0 \\ & \ddots & \\ 0 & & \hat{\Sigma}_{NN} \end{bmatrix}, \quad \underset{(k+1)\times(k+1)}{\hat{\Sigma}_{ii}} = \begin{bmatrix} \hat{\sigma}_i^2 & 0 \\ 0 & \hat{\Sigma}_{iiv} \end{bmatrix} \tag{74}$$

The results in (73) and (74) are equivalent.

Remark: Alternatively, we may combine the (concentrated) QML estimator with E-algorithm: Suppose that we obtain the estimate of

$$\boldsymbol{\Phi}_i' \boldsymbol{f}_t = \begin{bmatrix} \boldsymbol{\lambda}_i' \boldsymbol{f}_t \\ \gamma_i' \boldsymbol{f}_t \end{bmatrix}$$

by E-algorithm. Then we construct

$$\hat{\boldsymbol{C}}_i^{(1)} = \hat{\boldsymbol{\lambda}}_i' \hat{\boldsymbol{f}}_t$$

Then, we may maximise the following concentrated LF to obtain the estimate of $\boldsymbol{\phi}_0^* = (\phi_{10}^*, ..., \phi_{N0}^*)'$:

$$\mathcal{L}_C\left(\boldsymbol{\phi}_0^*\right) \propto T \ln\left|S(\boldsymbol{\Phi}_0^*)\right| - \frac{1}{2T}\sum_{i=1}^{N}\left(\boldsymbol{y}_i - \hat{\boldsymbol{C}}_i^{(1)} - \phi_{i0}^*\boldsymbol{y}_i^*\right)' \boldsymbol{M}_i \left(\boldsymbol{y}_i - \hat{\boldsymbol{C}}_i^{(1)} - \phi_{i0}^*\boldsymbol{y}_i^*\right) \tag{75}$$

Then, we can estimate other parameters, $\hat{\boldsymbol{\theta}}_i^{(1)}$ by least squares applied to the individual equations (61) conditional on $\hat{\phi}_{i0}^{*(1)}$ and $\hat{\boldsymbol{C}}_i^{(1)}$.

Remark: Alternatively, we may combine the STARDL-CF estimator with the EM algorithm. Given the consistent estimate of $\hat{C}_{it}$, we update all other parameters including ϕ_{i0}^* by running the following augmented regression:

$$\left(y_{it} - \hat{C}_{it}\right) = \sum_{\ell=1}^{p}\phi_{i\ell}y_{i,t-\ell} + \sum_{\ell=0}^{p}\phi_{i\ell}^* y_{i,t-\ell}^* + \sum_{\ell=0}^{q}\boldsymbol{\pi}_{i\ell}'\boldsymbol{x}_{i,t-\ell} + \sum_{\ell=0}^{q}\boldsymbol{\pi}_{i\ell}'^*\boldsymbol{x}_{i,t-\ell}^* + \alpha_i + \rho\hat{v}_{it} + e_{it}^*, \tag{76}$$

where $\hat{v}_{it} = y_{it}^* - \hat{\boldsymbol{\varphi}}_i'\boldsymbol{z}_{it}$ and $e_{it}^* = e_{it} + \rho\left(\hat{\boldsymbol{\varphi}}_i - \boldsymbol{\varphi}_i\right)'\boldsymbol{z}_{it} + \left(C_{it} - \hat{C}_{it}\right)$.

8 Concluding Remarks

The issue of cross-sectional dependence is developing very rapidly, with increasing interest being taken in modelling growing number of datasets with both cross-section and time dimension. The STARDL model provides a simple way of capturing dependence along both dimensions, based on the popular ARDL model in time series. We adopt the convention of allowing parameters to be heterogeneous across cross-section units and discuss the conditions under which these models are stable. Under widely held conditions, we show that both the QML and the control function estimator are $\sqrt{T}$ consistent and asymptotically normally distributed. Monte Carlo evidence supports the validity of both methods in finite samples. The counter-weight to the degree of sophistication in any model is the subsequent difficultly in interpreting the results they give. We propose two methods for analysing the patterns produced, and illustrate their use in analysing casualty data for the 2003 Iraq war and its aftermath. As an extension of the basic model we consider estimation under the presence of observed and unobserved common factors.

There remains a number of interesting challenges to be addressed. We have, throughout, assumed our spatial weighting matrix to be not only known but determined exogenously, ruling out a number of exciting areas of research in social networks and team formation. Although our control function approach has the potential, under certain conditions, to control for this source of endogeneity further work is required to determine how it may be applied. Our current approach built upon internally generated instruments is unlikely to be valid. We have also restricted ourselves to linear effects, both in time and across space, and to modelling conditional means rather than other parts of the conditional distribution such as medians or quantiles. These remain topic of ongoing research. Eventually, this project aims to develop the general econometrics models that can accommodate spatial and factor dependence, spatial heterogeneity, endogenous spatial weight matrix as well as spatial nonlinearity in a unified framework by combining all the recent advances. These works will be of great applicability to a variety of the big dataset.

9 References

Ahn S. and A. Horenstein (2013): Eigenvalue ratio test for the number of factors. *Econometrica* 81: 1203-1227

Andrews, D.W.K. (2005): Cross-section regression with common shocks. *Econometrica* 73: 1551-1585.

Anselin, L. (1980) Estimation methods for spatial autoregressive structures. E *Regional Science Dissertation & Monograph Series, Program in Urban and Regional Studies*, Cornell University (8).

Anselin, L. (1988): *Spatial Econometrics: Methods and Models.* Kluwer Academic, Boston, MA.

Anselin, L. (2009) *Spatial regression. The SAGE handbook of spatial analysis.* 1, pp.255-276.

Aquaro, M., N. Bailey and M.H. Pesaran (2015): "Quasi Maximum Likelihood Estimation of Spatial Models with Heterogeneous Coefficients," *mimeo.*, University of Southern California.

Bai, J. (2003): Inferential theory for factor models of large dimensions. *Econometrica* 71: 135-171.

Bai, J. (2009): Panel data models with interactive fixed effects. *Econometrica* 77: 1229-1279.

Bai, J. (2013): "Likelihood Approach to Dynamic Panel Models with Interactive Effects," *mimeo.*, Columbia University.

Bai, J. and K. Li (2014): "Spatial Panel Data Models with Common Shocks," *mimeo.*, Columbia University.

Bai, J. and K. Li (2015): "Dynamic Spatial Panel Data Models with Common Shocks," *mimeo.*, Columbia University.

Bailey, N., S. Holly and M.H. Pesaran (2016): "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-sectional Dependence," forthcoming in *Journal of Applied Econometrics.*

Baltagi, B. H., S. H. Song, and W. Koh (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117: 123-150.

Baltagi B.H. (2005): *Econometric Analysis of Panel Data*, 3rd edition. Wiley: Chichester.

Baltagi, B. H. and A. Pirotte (2007). Panel unit root tests and spatial dependence. *Journal of Applied Econometrics*: 360: 339-360.

Baltagi, B., S.H. Song, B.C. Jung and W. Koh (2007): "Testing for serial correlation, spatial autocorrelation and random effects using panel data," *Journal of Econometrics* 140, 5-51.

Chudik, A., K. Mohaddes, M.H. Pesaran and M. Raissi (2017): "Is there a debt-thrshold effect on output growth?" *Review of Economics and Statistics* 99: 135-150.

Cliff, A.D. and J.K. Ord (1973): *Spatial Autocorrelation.* Pion, London.

Cliff, A.D. and J.K. Ord (1981): *Spatial processes: models & applications.* Taylor & Francis.

Conley, T.G. and B. Dupor (2003): A spatial analysis of sectoral complementarity. *Journal of Political Economy* 111: 311-352.

Cressie, N.A.C. (1993): *Statistics of Spatial Data.* Wiley, New York.

Davidson, R. MacKinnon, J.G. (1993): *Estimation and Inference in Econometrics.* Oxford University Press, Oxford.

Debarsy, N., C. Ertur and J.P. LeSage (2012): "Interpreting dynamic space–time panel data models," Statistical Methodology 9, 158-171.

Defever, F., B. Heid and M. Larch (2015): Spatial exporters. *Journal of International Economics* 95: 145-C156.

Diebold, F.X. and K. Yilmaz (2014): "On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms," *Journal of Econometrics* 182, 119-134.

Elhorst, J.P. (2005): Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis* 37: 85-106.

Elhorst, J.P. (2010): Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics* 40: 272-282.

Elhorst, J.P. Lacombe, D.J. and G.Piras (2012): On model specification and parameter space definitions in higher order spatial econometric models. *Regional Science and Urban Economics* 42:211-220.

Elhorst, J.P. (2014): *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels.* Springer, Berlin, Heidelberg.

Elliott, M., B. Golub and M.O. Jackson (2014): Financial Networks and Contagion. *American Economic Review* 104: 3115-53.

Fan, J., Y. Liao, Y and M. Mincheva (2011): High dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics* 39: 3320.

Greenwood-Nimmo, M., Nguyen, V.H. and Y. Shin (2015): Measuring the connectedness of the global economy. *Melbourne Institute Working Paper* No. 7/15.

Gunnella V., C. Mastromarco, L. Serlenga and Y. Shin (2015): "The Euro Effects on Intra-EU Trade Flows and Balances: Evidence from the Cross Sectionally Correlated Panel Gravity Models," *mimeo.*, University of York.

Hausman, J.A. and W.E. Taylor (1981): "Panel Data and Unobservable Individual Effect," *Econometrica* 49: 1377-1398.

Hahn, J. and G. Kuersteiner (2002): Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70: 1639-1657.

Hamilton J.D. (1994): *Time Series Analysis.* Princeton University Press, Princeton.

Horn, R.A. and C.R. Johnson (1985): *Matrix Analysis.* Cambridge University Press, Cambridge.

Kapetanios, G. and M.H. Pesaran (2005): "Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns,", CESifo Working Paper Series 1416, CESifo Group Munich.

Kapetanios, G., J. Mitchell and Y. Shin (2014): "A nonlinear panel data model of cross-sectional dependence," *Journal of Econometrics* 179, 134-157.

Kapetanios, G., L. Serlenga and Y. Shin (2017): "The Multi-dimensional Heterogeneous Panel Data with the Hierarchical Multi-factor Error Structure," *mimeo.*, University of York.

Kapoor, M., Kelejian, H.H. and I.R. Prucha (2007): "Panel data models with spatially correlated error components," *Journal of Econometrics* 140, 97–130.

Kelejian, H.H. and I.R. Prucha (1998): "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance," *Journal of Real Estate Finance Economics* 17, 99-121.

Kelejian, H.H. and I.R. Prucha (1999): "A generalized moments estimator for the autoregressive parameter in a spatial model," *International Economic Review* 40, 509-533.

Kelejian, H.H. and I.R. Prucha (2010): Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157: 53-67.

Kelejian, H.H. and D.P. Robinson (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. *Papers in regional science* 72, 297-312.

Kuersteiner, G.M. and I.R. Prucha (2018): "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity," *mimeo.*, University of Maryland.

Land, K.C. and G. Deane (1992) On the large-sample estimation of regression models with spatial-or network-effects terms: A two-stage least squares approach. *Sociological methodology* 22:221-248.

Lee, L.F. (2004): Asymptotic distributions of quasi-maximum likelihood estimator for spatial autoregressive models. *Econometrica* 72: 1899-1925.

Lee, L.F. and J. Yu (2010a): Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40: 255-271.

Lee, L.F. and J. Yu (2010b): A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* 26: 564-597.

Lee, L.f. and J. Yu (2011): A UnifiÂed Estimation Approach for Spatial Dynamic Panel Data Models: Stability, Spatial Cointegration and Explosive Roots in *Handbook of Empirical Economics and Finance* edited by Ullah, A. and D. E. A. Giles.

Lee, L.F. and J. Yu (2013): Identification of spatial durbin panel models. Working paper; forthcoming in *Journal of Applied Econometrics.*

Lee, L.F and J. Yu (2014) "Efficient GMM estimation of spatial dynamic panel data models with fixed effects," *Journal of Econometrics* 180, 174-197.

Lee, L.F. and J. Yu (2015): Spatial panel data models, chapter 12, in: Baltagi, B.H. (Ed.), *The Oxford Handbook of Panel Data*. Oxford University Press, New York, NY.

LeSage, J.P and Y.Y. Chin (2016): "Interpreting heterogeneous coefficient spatial autoregressive panel models," *Economics Letters* 142, 1-5.

LeSage, J.P. and R.K. Pace (2009): *Introduction to Spatial Econometrics*. CRC Press, Boca Raton.

Li, K. (2017): "Fixed-effects dynamic spatial panel data models and impulse response analysis," *Journal of Econometrics* 198, 102-121.

Lin, X. and L.F. Lee (2010): GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157: 34-52.

Liu, X. and L.F. Lee (2010): GMM estimation of social interaction models with centrality. *Journal of Econometrics* 159: 99-115.

Lu, L. (2017): "Simultaneous Spatial Panel Data Models with Common Shocks," *mimeo.*, Federal Reserve Bank of Boston.

Lu, X. and L. Su (2015): Shrinkage estimation of dynamic panel data models with interactive fixed effects. Working paper. Singapore Management School

Manski, C.F. (1993): Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60: 531-542.

Mastromarco, C., L. Serlenga and Y. Shin (2016): "Modelling Technical Inefficiency in Cross Sectionally Dependent Stochastic Frontier Panels," forthcoming in *Journal of Applied Econometrics* 31: 281-297.

Mastromarco, C., L. Serlenga and Y. Shin, (2015): "Multilateral Resistance and Euro Effects on Trade Flows," forthcoming in *Spatial Econometric Interaction Modelling* eds. by G. Arbia and R. Patuelli. Springer: Berlin.

Mutl, J. (2009): Consistent Estimation of Global VAR Models. *Economics Series*, Institute for Advanced Studies, Vienna.

Onatski A. (2009): A formal statistical test for the number of factors in the approximate factor models. *Econometrica* 77: 1447-1479

Ord, K. (1975): Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70: 120-297.

Pesaran, M.H. (2006): Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967-1012.

Pesaran, M.H. (2015): "Testing Weak Cross-sectional Dependence in Large Panels," *Econometric Reviews* 34: 1089-1117.

Pesaran, M.H. and E. Tosetti (2011): Large panels with common factors and spatial correlation. *Journal of Econometrics* 161: 182-202.

Shi, W. and L.F. Lee (2017): "Spatial Dynamic Panel Data Models with Interactive Fixed Effects," *Journal of Econometrics* 197: 323-347.

Shin, Y., Yu, B. and M. Greenwood-Nimmo (2014) Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. In *Festschrift in Honor of Peter Schmidt* (pp. 281-314). Springer, New York.

Sloboda, J., H. Dardagan, M. Spagat and M.H.R. Hicks (2011): Iraq Body Count: A case study in the uses of incident-based conflict casualty data. Working paper.

Stiglitz, J. and L. Bilmes (2008): *The Three Trillion Dollar War.* W.W. Norton.

Su L., S. Jin S. and Y. Zhang (2015): Specification test for panel data models with interactive fixed effects. *Journal of Econometrics* 186: 222-244.

Su, L. and Z. Yang (2015): QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185: 230-258.

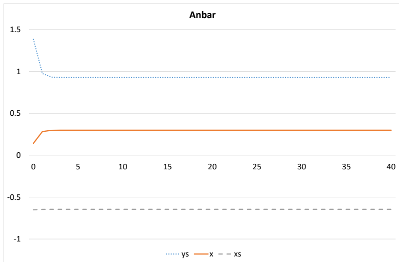
Upton, G. and B. Fingleton (1985): *Spatial data analysis by example. Volume 1: Point pattern and quantitative data.* Wiley, Chichester.

White, H (1984): *Asymptotic Theory for Econometricians.* Academic Press, San Diego.

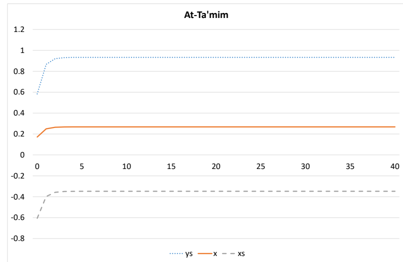
White, H. (1994): *Estimation, inference and specification analysis.* Econometric Society Mongraphs No 22, Cambridge University Press, Cambridge.

Yu, J., R. de Jong and L.F. Lee (2008): "Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large," *Journal of Econometrics* 146, 118-134.

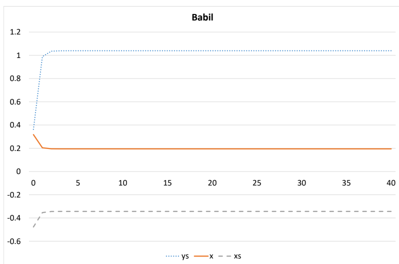
Yu, J. and L.F. Lee (2010): "Estimation of unit root spatial dynamic data models," *Econometric Theory* 26, 1332–1362.



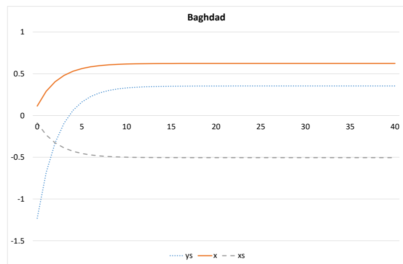
(a) Anbar



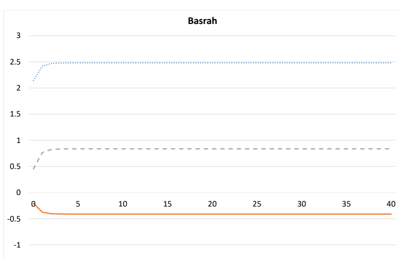
(b) At Ta'mim



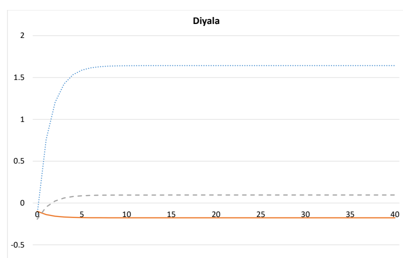
(c) Babil



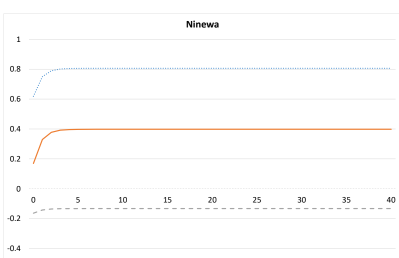
(d) Baghdad



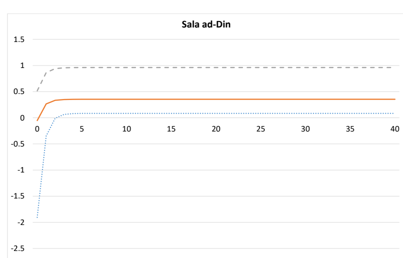
(e) Basrah



(f) Diyala

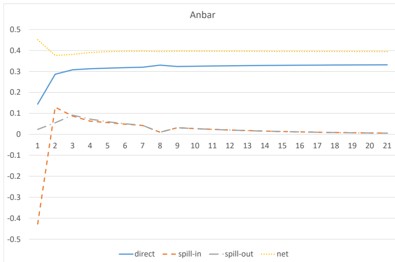


(g) Ninewa

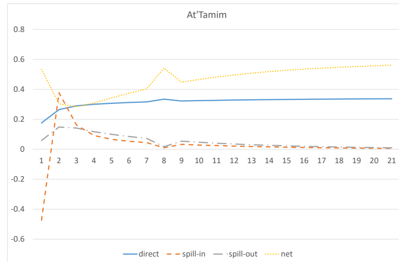


(h) Sala ad-Din

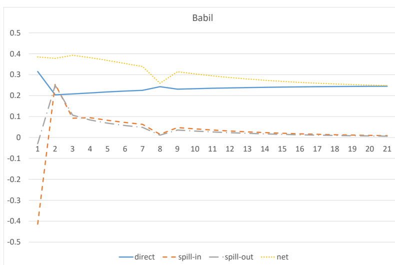
Figure 1: Dynamic Multipliers: y^* (dotted); x (solid) and x^* (dashed).



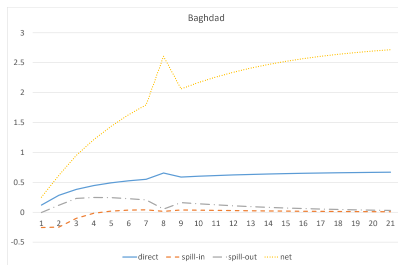
(a) Anbar



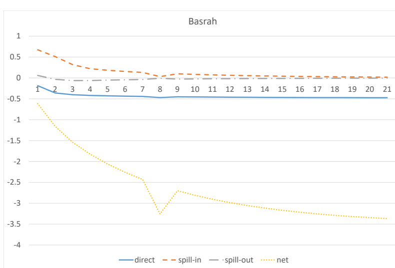
(b) At Ta'mim



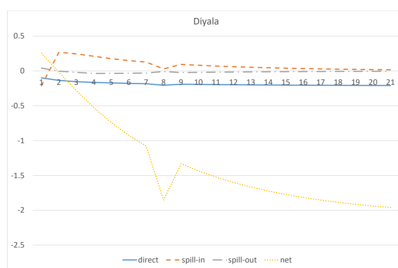
(c) Babil



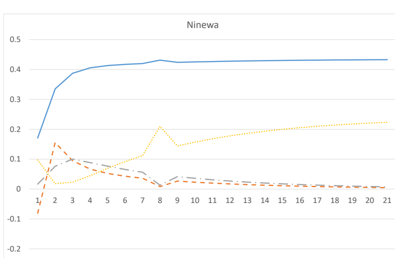
(d) Baghdad



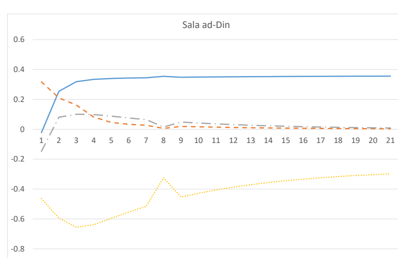
(e) Basrah



(f) Diyala

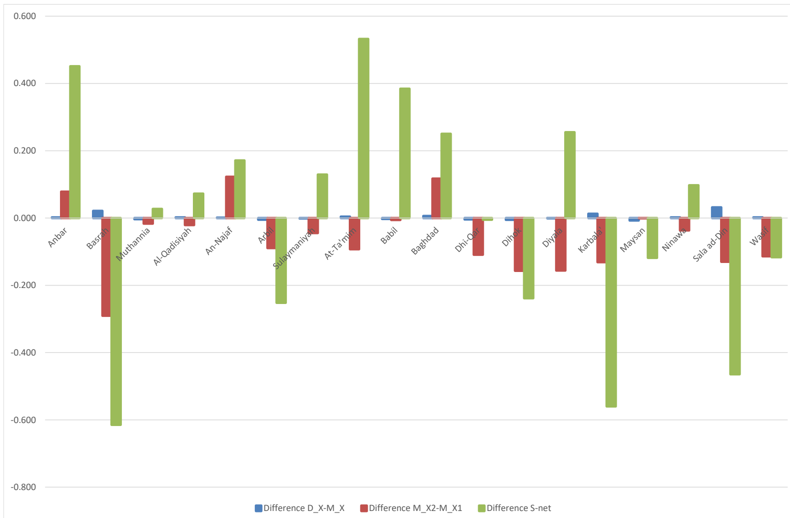


(g) Ninewa

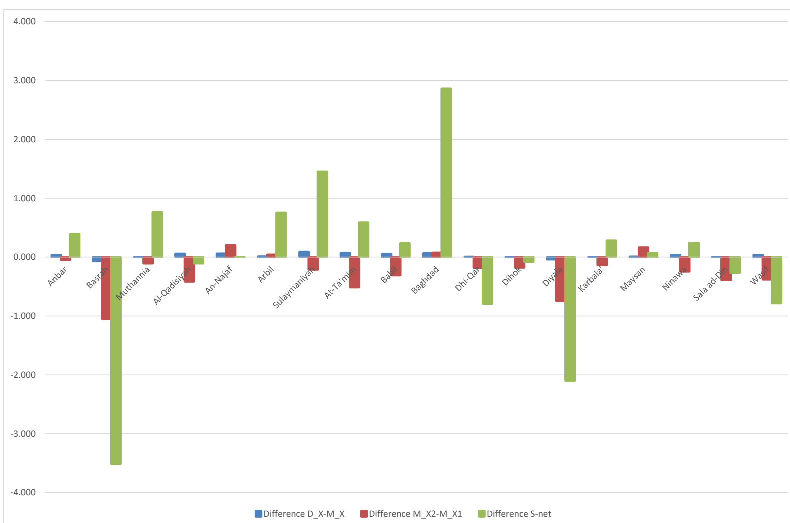


(h) Sala ad-Din

Figure 2: Connectedness measures: direct (solid); spill-in (dashed); spill-out (dash/dotted); net (dotted).



(a) Short-run



(b) Long-run

Figure 3: Difference between dynamic and diffusion multipliers.

Table 9: Summary of Dynamic Multipliers and Diffusion Multipliers for Short Run Horizon $h = 0$

	ARDL	STARDL			System			Difference		
	MX1	MYS	MX2	MXS	DX	SI	SO	DX-MX	MX2-MX1	S-net
Anbar	0.064	1.382	0.143	−0.651	0.144	−0.428	0.023	0.002	0.078	0.451
Basrah	0.084	2.142	−0.207	0.452	−0.185	0.674	0.059	0.021	−0.290	−0.615
Muthannia	−0.056	−0.894	−0.073	0.141	−0.076	0.005	0.032	−0.004	−0.017	0.027
Al-Qadisiyah	0.250	−1.014	0.229	0.075	0.231	−0.055	0.017	0.002	−0.021	0.072
An-Najaf	0.051	1.243	0.173	−0.340	0.174	−0.156	0.015	0.001	0.123	0.171
Arbil	0.598	1.231	0.509	0.110	0.503	0.222	−0.030	−0.005	−0.089	−0.252
Sulaymaniyah	0.799	−0.242	0.754	−0.050	0.753	−0.073	0.056	−0.001	−0.045	0.129
At-Ta'mim	0.264	0.584	0.172	−0.605	0.176	−0.474	0.059	0.004	−0.093	0.532
Babil	0.323	0.363	0.318	−0.477	0.315	−0.414	−0.030	−0.003	−0.006	0.385
Baghdad	−0.004	−1.231	0.113	−0.089	0.119	−0.256	−0.006	0.006	0.117	0.251
Dhi-Qar	0.009	−0.277	−0.100	0.155	−0.104	0.111	0.106	−0.004	−0.109	−0.005
Dihok	−0.061	0.492	−0.218	0.081	−0.223	0.183	−0.056	−0.005	−0.157	−0.239
Diyala	0.057	−0.132	−0.100	−0.193	−0.100	−0.214	0.041	0.000	−0.156	0.255
Karbala'	0.365	−0.914	0.234	0.525	0.247	0.471	−0.089	0.013	−0.131	−0.560
Maysan	0.090	−0.742	0.089	0.286	0.082	0.188	0.069	−0.007	0.000	−0.119
Ninawa	0.206	0.618	0.169	−0.164	0.171	−0.081	0.017	0.002	−0.037	0.097
Sala ad-Din	0.076	−1.906	−0.054	0.515	−0.022	0.317	−0.147	0.032	−0.130	−0.465
Wasif	0.333	−0.445	0.219	0.159	0.221	0.111	−0.006	0.002	−0.114	−0.117

Table 10: Summary of Dynamic Multipliers and Diffusion Multipliers for Long Run Horizon $h = 40$

	ARDL	STARDL			System			Difference		
	MX1	MYS	MX2	MXS	DX	SI	SO	DX-MX	MX2-MX1	S-net
Anbar	0.345	0.927	0.299	−0.645	0.333	0.273	0.667	0.035	−0.046	0.394
Basrah	0.636	2.480	−0.412	0.836	−0.479	3.038	−0.474	−0.067	−1.047	−3.512
Muthannia	0.407	0.006	0.300	0.212	0.301	0.218	0.978	0.000	−0.106	0.760
Al-Qadisiyah	0.679	0.955	0.265	0.157	0.321	1.044	0.938	0.056	−0.414	−0.106
An-Najaf	0.025	1.365	0.224	−0.578	0.283	0.721	0.723	0.059	0.199	0.002
Arbil	0.459	0.149	0.501	0.310	0.511	0.453	1.206	0.011	0.041	0.753
Sulaymaniyah	1.182	0.668	0.972	0.242	1.062	0.814	2.265	0.090	−0.210	1.450
At-Ta'mim	0.780	0.933	0.268	−0.348	0.340	0.602	1.188	0.072	−0.511	0.586
Babil	0.502	1.039	0.195	−0.343	0.248	0.654	0.887	0.053	−0.307	0.233
Baghdad	0.549	0.354	0.624	−0.505	0.687	−0.179	2.682	0.063	0.075	2.861
Dhi-Qar	0.062	0.596	−0.117	0.189	−0.112	0.826	0.035	0.004	−0.179	−0.791
Dihok	0.143	0.243	−0.033	−0.061	−0.031	0.200	0.121	0.002	−0.177	−0.079
Diyala	0.568	1.643	−0.177	0.096	−0.213	1.778	−0.320	−0.036	−0.744	−2.098
Karbala'	0.418	0.087	0.286	0.358	0.289	0.443	0.723	0.003	−0.132	0.280
Maysan	−0.214	0.246	−0.050	−0.038	−0.043	0.241	0.310	0.007	0.164	0.070
Ninawa	0.641	0.806	0.398	−0.133	0.435	0.599	0.839	0.037	−0.243	0.241
Sala ad-Din	0.743	0.084	0.353	0.960	0.357	1.041	0.777	0.004	−0.390	−0.264
Wasif	0.508	1.098	0.130	0.233	0.164	1.298	0.517	0.035	−0.378	−0.781

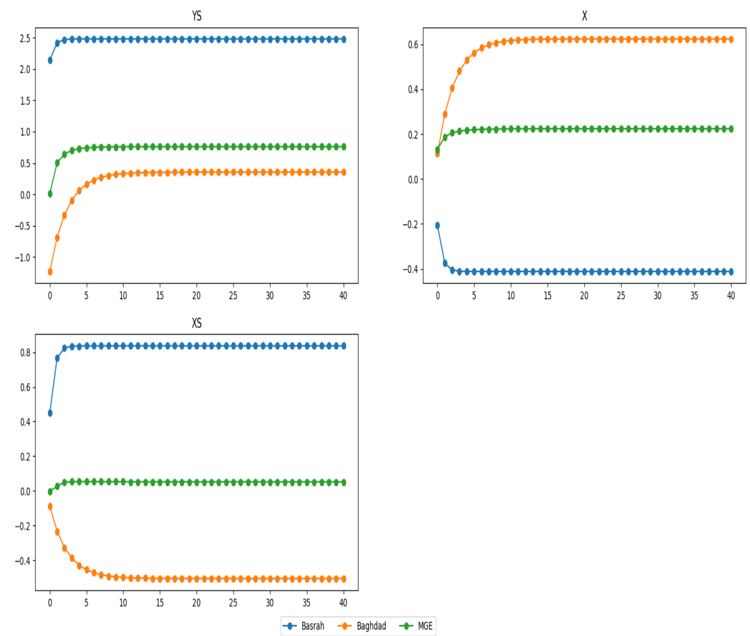


Figure 4: Dynamic Multipliers for Basrah and Baghdad

Appendix

Appendix 1: Lemmas

Lemma 1 The process $\boldsymbol{y}_t$ and $\boldsymbol{y}_t^*$ have finite fourth moments.

Proof In (11) the vector process $\boldsymbol{y}_t$ has been written using the lags of two independent processes with finite fourth moments. We can re-write (11) as

$$\boldsymbol{y}_t = \sum_{\ell=0}^{\infty} \bar{\boldsymbol{B}}_\ell \bar{\boldsymbol{x}}_{t-\ell}, \tag{77}$$

where $\bar{\boldsymbol{B}}_\ell = \left[\tilde{\boldsymbol{B}}_\ell, \boldsymbol{B}_\ell\right]$, with typical element $\bar{b}_{\ell,ij}$, and $\bar{\boldsymbol{x}}_{t-\ell} = \left[\boldsymbol{x}'_{t-\ell}, \tilde{\boldsymbol{u}}_{t-\ell}\boldsymbol{\prime}\right]'$ with typical element $\bar{x}_{t-\ell,j}$. By applying Hölders inequality twice,

$$E\left|\bar{x}_{i,r}\bar{x}_{h,s}\bar{x}_{k,t}\bar{x}_{l,u}\right| \leq \max_{i,t} E\bar{x}_{i,t}^4 \leq C_1 < \infty.$$

Under Assumption 5 the boundedness on q and of the elements of $\tilde{\boldsymbol{\Pi}}_\ell$ the coef-

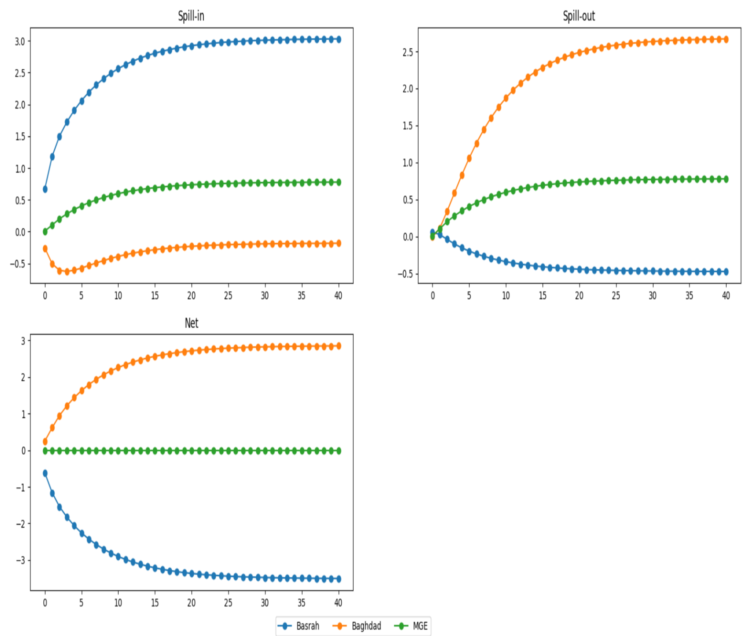


Figure 5: Spill-in and Spill-out for Basrah and Baghdad

ficients of the lag polynomial in $\bar{\boldsymbol{B}}_\ell$ are absolutely summable and so,

$$\begin{aligned} E\left|y_{i,r}y_{j,s}y_{k,t}y_{l,u}\right| &= E\left|\left(\sum_{\ell=0}^{\infty}\sum_{h}\bar{b}_{\ell,ih}\bar{x}_{r-\ell,h}\right)\left(\sum_{\ell=0}^{\infty}\sum_{h}\bar{b}_{\ell,jh}\bar{x}_{s-\ell,h}\right)\right. \\ &\quad \left.\times\left(\sum_{\ell=0}^{\infty}\sum_{h}\bar{b}_{\ell,kh}\bar{x}_{t-\ell,h}\right)\left(\sum_{\ell=0}^{\infty}\sum_{h}\bar{b}_{\ell,lh}\bar{x}_{u-\ell,h}\right)\right| \\ &\leq \sum_{\ell=0}^{\infty}\sum_{h}\left|\bar{b}_{\ell,ih}\right|\sum_{\ell=0}^{\infty}\sum_{h}\left|\bar{b}_{\ell,jh}\right|\sum_{\ell=0}^{\infty}\sum_{h}\left|\bar{b}_{\ell,kh}\right|\sum_{\ell=0}^{\infty}\sum_{h}\left|\bar{b}_{\ell,lh}\right|C_1 \leq C_2 < \infty. \end{aligned}$$

The above argument may then be repeated to show

$$\begin{aligned} E\left|y^*_{i,r}y^*_{j,s}y^*_{k,t}y^*_{l,u}\right| &= E\left|\sum_{h=1}^{n}w_{ih}y_{h,r}\sum_{h=1}^{n}w_{jh}y_{h,s}\sum_{h=1}^{n}w_{kh}y_{h,s}\sum_{h=1}^{n}w_{lh}y_{h,u}\right| \\ &\leq \left|\sum_{h=1}^{n}w_{ih}\right|\left|\sum_{h=1}^{n}w_{jh}\right|\left|\sum_{h=1}^{n}w_{kh}\right|\left|\sum_{h=1}^{n}w_{lh}\right|C_2 < \infty \end{aligned}$$

as the rows of $\boldsymbol{W}$ are bounded in absolute row sums under Assumption 3. $\square$

Lemma 2 The process $\boldsymbol{\chi}_{it}$ has finite fourth moments for all i.

Proof The proof results from the application of the arguments used in Lemma 1 to establish the boundedness of any four elements chosen from $\boldsymbol{y}_t$, $\boldsymbol{y}_t^*$ and $\boldsymbol{X}_t$ and $\boldsymbol{X}_t^*$ at any lead and lag. □

Appendix 2: Derrivatives of (12) The first derivatives of (12) are

$$
\begin{aligned}
\frac{\partial \mathcal{L}_{\bar{T}}}{\partial \phi_{0i}^*} &= -\bar{T} g_{ii} + \frac{1}{\sigma_i^2} \sum_{t=q+1}^{T} y_{it}^* \left(y_{it} - \phi_{0i}^* y_{it}^* - \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it} \right), \text{so that} \\
\frac{\partial \mathcal{L}_{\bar{T}}}{\partial \boldsymbol{\phi}_0^*} &= -\bar{T} \mathrm{vecd}(\boldsymbol{G}_0) + \boldsymbol{y}_y^* \odot \left(\boldsymbol{y}_t - \boldsymbol{\Phi}_0^* \boldsymbol{W} \boldsymbol{y}_t - \boldsymbol{\Theta} \boldsymbol{\chi}_t \right) \odot \boldsymbol{\sigma}^{-2}, \\
\frac{\partial \mathcal{L}_{\bar{T}}}{\partial \boldsymbol{\theta}_i} &= \frac{1}{\sigma_i^2} \sum_{t=q+1}^{T} \boldsymbol{\chi}_{it} \left([y_{it} - \phi_{0i}^* y_{it}^* - \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it} \right), \\
\frac{\partial \mathcal{L}_{\bar{T}}}{\partial \sigma_i^2} &= -\frac{\bar{T}}{2\sigma_i^2} + \frac{1}{2\sigma_i^4} \sum_{t=q+1}^{T} \left(y_{it} - \phi_{0i}^* y_{it}^* - \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it} \right)^2,
\end{aligned}
$$

where g_{ij} is i,j'th element of the matrix $\boldsymbol{G}$, where the operator $\mathrm{vecd}(A) \equiv [a_{11}, \dots a_{NN}]'$ extracts the principal diagonal from a square matrix as a column vector and $\boldsymbol{\sigma}^{-2} = [\sigma_1^{-2}, \dots, \sigma_n^{-2}]$.

The individual components of the Hessian are then

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \phi_{0i}^* \partial \phi_{0j}^*} &= -\bar{T} g_{ii}^2 - \frac{1}{\sigma_i^2} \sum_{t=q+1}^{T} (y_{it}^*)^2, \ i = j, \\
&= -\bar{T} g_{ij} g_{ji}, \ i \neq j, \ \text{so that} \\
\frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \boldsymbol{\phi}_0^* \partial \boldsymbol{\phi}_0^{*\prime}} &= -\bar{T} \boldsymbol{G} \odot \boldsymbol{G}' - \sum_{t=q+1}^{T} \mathrm{diag} \left\{ \boldsymbol{y}_t^* \odot \boldsymbol{y}_t^* \odot \boldsymbol{\sigma}^{-2} \right\}, \\
\frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \phi_{0i}^* \boldsymbol{\theta}_j'} &= -\frac{1}{\sigma_i^2} \sum_{t=q+1}^{T} y_{it}^* \boldsymbol{\chi}_{it}', \ i = j, \ 0 \text{ otherwise}, \\
\frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \phi_{0i}^* \partial \sigma_j^2} &= -\frac{1}{\sigma_i^4} \sum_{t=q+1}^{T} y_{it}^* \left(y_{it} - \phi_{0i}^* y_{it}^* - \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it} \right), \ i = j, \ 0 \text{ otherwise}, \\
\frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j'} &= -\frac{1}{\sigma_i^2} \sum_{t=q+1}^{T} \boldsymbol{\chi}_{it} \boldsymbol{\chi}_{it}', \ i = j, \ 0 \text{ otherwise}, \\
\frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \boldsymbol{\theta}_i \partial \sigma_j^2} &= -\frac{1}{\sigma_i^4} \sum_{t=q+1}^{T} \boldsymbol{\chi}_{it} [y_{it} - \phi_{0i}^* y_{it}^* - \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it}], \ i = j, \ 0 \text{ otherwise}, \\
\frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \sigma_i^2 \partial \sigma_j^2} &= \frac{\bar{T}}{2\sigma_i^4} - \frac{1}{\sigma_i^6} \sum_{t=q+1}^{T} [y_{it} - \phi_{0i}^* y_{it}^* - \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it}]^2 \ i = j, \ 0 \text{ otherwise},
\end{aligned}
$$

where the operator diag places its vector argument along the principal diagonal of a null, square matrix.

Appendix 3: Proof of Theorem 1

Consistency

Consistency of the QML estimator follows Lee (2004) and Theorem 3.4 of White (1994). It rests on establishing: (i) the stochastic equicontinuity of $\frac{1}{T}\mathcal{L}^c_T(\boldsymbol{\phi}^*_0)$ (ii) that $\frac{1}{T}\mathcal{L}^c_T(\boldsymbol{\phi}^*_0) \to_p \frac{1}{T}Q^c(\boldsymbol{\phi}^*_0)$, uniformly in $\boldsymbol{\phi}^*_0$; and, (iii) that $\frac{1}{T}Q^c(\boldsymbol{\phi}^*_0)$ is uniquely maximised at the true value $\tilde{\boldsymbol{\phi}}^*_0$. Point (i) is easily established by considering the two terms of $\mathcal{L}^c_T(\boldsymbol{\phi}^*_0)$ that depend on $\boldsymbol{\phi}^*_0$, $\ln|S(\boldsymbol{\Phi}^*_0)|$ and $\sum_{i=1}^N \ln \frac{1}{T}\left(\boldsymbol{y}_i - \phi^*_{i0}\boldsymbol{y}^*_i\right)' \boldsymbol{M}_i \left(\boldsymbol{y}_i - \phi^*_{i0}\boldsymbol{y}^*_i\right)$. The latter is clearly quadratic in ϕ^*_{i0} while the former is a continuous function of a sum of polynomials containing powers of minimum order 0 and maximum order 1 in any ϕ^*_{i0}.

Next consider that

$$\frac{1}{T}\mathcal{L}^c_T(\boldsymbol{\phi}^*_0) - \frac{1}{T}Q^c(\boldsymbol{\phi}^*_0) = \frac{1}{2}\sum_{i=1}^N \left[\ln\left(\bar{\sigma}^2_i(\phi^*_{i0})\right) - \ln\left(\hat{\sigma}^2_i(\phi^*_{i0})\right)\right],$$

and so (ii) follows if $\hat{\sigma}^2_i(\phi^*_{i0}) \to_p \bar{\sigma}^2_i(\phi^*_{i0})$, uniformly in ϕ^*_{i0}. Defining, for $i = 1, \ldots, N$, $w_{it}(\phi^*_{i0}) \equiv \boldsymbol{s}'_i(\phi^*_{0i})\boldsymbol{S}^{-1}\boldsymbol{u}_t$ we can write

$$\begin{aligned}
\hat{\sigma}^2_i(\phi^*_{i0}) &= \frac{1}{T}\sum_{t=q+1}^T \left(\kappa_{it}(\phi^*_{i0}) + w_{it}(\phi^*_{i0})\right)^2 \\
&\quad -\frac{1}{T}\sum_{t=q+1}^T \left(\kappa_{it}(\phi^*_{i0}) + w_{it}(\phi^*_{i0})\right)\boldsymbol{\chi}'_{it}\left[\sum_{t=q+1}^T \boldsymbol{\chi}_{it}\boldsymbol{\chi}'_{it}\right]^{-1}\sum_{t=q+1}^T \boldsymbol{\chi}_{it}\left(\kappa_{it}(\phi^*_{i0}) + w_{it}(\phi^*_{i0})\right) \\
&= \left\{\frac{1}{T}\sum_{t=q+1}^T \kappa_{it}(\phi^*_{i0})^2\right\} - \left\{\frac{1}{T}\sum_{t=q+1}^T \kappa_{it}(\phi^*_{i0})\boldsymbol{\chi}'_{it}\right\}\left[\left\{\frac{1}{T}\sum_{t=q+1}^T \boldsymbol{\chi}_{it}\boldsymbol{\chi}'_{it}\right\}\right]^{-1}\left\{\frac{1}{T}\sum_{t=q+1}^T \boldsymbol{\chi}_{it}\kappa_{it}(\phi^*_{i0})\right\} \\
&\quad +\left\{\frac{1}{T}\sum_{t=q+1}^T w_{it}(\phi^*_{i0})^2\right\} + 2\left\{\frac{1}{T}\sum_{t=q+1}^T \kappa_{it}(\phi^*_{i0})w_{it}(\phi^*_{i0})\right\} \\
&\quad -2\left\{\frac{1}{T}\sum_{t=q+1}^T \kappa_{it}(\phi^*_{i0})\boldsymbol{\chi}'_{it}\right\}\left[\left\{\frac{1}{T}\sum_{t=q+1}^T \boldsymbol{\chi}_{it}\boldsymbol{\chi}'_{it}\right\}\right]^{-1}\left\{\frac{1}{T}\sum_{t=q+1}^T \boldsymbol{\chi}_{it}w_{it}(\phi^*_{i0})\right\} \\
&\quad -\left\{\frac{1}{T}\sum_{t=q+1}^T w_{it}(\phi^*_{i0})\boldsymbol{\chi}'_{it}\right\}\left[\left\{\frac{1}{T}\sum_{t=q+1}^T \boldsymbol{\chi}_{it}\boldsymbol{\chi}'_{it}\right\}\right]^{-1}\left\{\frac{1}{T}\sum_{t=q+1}^T \boldsymbol{\chi}_{it}w_{it}(\phi^*_{i0})\right\}.
\end{aligned}$$

The sums of $\boldsymbol{\chi}_{it}w_{it}$ and $\boldsymbol{\chi}_{it}\kappa_{it}$ are weighted (linearly in $\boldsymbol{\phi}^*_0$) sums of sample estimates of cross-correlations between $\boldsymbol{\chi}_{it}\boldsymbol{u}_t$ and $\boldsymbol{\chi}_{it}\boldsymbol{\chi}_t$, respectively. As $T \to \infty$, under our assumptions these sample correlations converge in probability to their expectation and so the terms in $\{.\}$ converge in probability to their expectation uniformly in ϕ^*_{i0}, which means that the first three terms together converge in probability to $\bar{\sigma}^2_i$ while the remainder are $o_p(1)$.

Finally, to establish (iii) we express $Q^c_{\bar{T}}(\boldsymbol{\phi}^*_0) = Q^{c1}_{\bar{T}}(\boldsymbol{\phi}^*_0) + Q^{c2}_{\bar{T}(\boldsymbol{\phi}^*_0)}$, where

$$\begin{aligned} Q^{c1}_{\bar{T}}(\boldsymbol{\phi}^*_0) &\equiv -\frac{n\bar{T}}{2}\ln(2\pi+1) + \bar{T}\ln|S(\boldsymbol{\Phi}^*_0)| - \frac{\bar{T}}{2}\sum_{i=1}^{N}\ln\varsigma(\phi^*_{0i})^2_i, \\ Q^{c2}_{\bar{T}}(\boldsymbol{\phi}^*_0) &\equiv -\frac{\bar{T}}{2}\sum_{i=1}^{N}\left[\ln\bar{\sigma}^2_i(\phi^*_{0i}) - \ln\varsigma^2_i(\phi^*_{0i})\right], \end{aligned}$$

with $\varsigma^2_i(\phi^*_{0i}) \equiv tr\left\{\boldsymbol{S}^{-1\prime}\boldsymbol{s}_i(\phi^*_{0i})\boldsymbol{s}'_i(\phi^*_{0i})\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\right\}$. Noting that $Q^{c2}_{\bar{T}}(\tilde{\boldsymbol{\phi}}^*_0) = 0$, we may write

$$\frac{1}{\bar{T}}\left[Q^c_{\bar{T}}(\boldsymbol{\phi}^*_0) - Q^c_{\bar{T}}(\boldsymbol{\phi}^*_0)\right] = \frac{1}{\bar{T}}\left[Q^{c1}_{\bar{T}}(\boldsymbol{\phi}^*_0) - Q^{c1}_{\bar{T}}(\tilde{\boldsymbol{\phi}}^*_0)\right] + \frac{1}{\bar{T}}Q^{2c}_{\bar{T}}(\boldsymbol{\phi}^*_0).$$

The function $Q^{c1}_{\bar{T}}(\boldsymbol{\phi}^*_0)$ is the expected value of the exact log likelihood function for a heterogeneous spatial autoregressive model with normally distributed disturbances hence, by Jenson's inequality, $\frac{1}{\bar{T}}\left[Q^{c1}_{\bar{T}}(\boldsymbol{\phi}^*_0) - Q^{c1}_{\bar{T}}(\tilde{\boldsymbol{\phi}}^*_0)\right] \leq 0$. Assumption 7 ensures that $\bar{\sigma}^2_i(\phi^*_{0i}) > \varsigma^2_i(\phi^*_{0i})$ for all $i = 1,\ldots,N$ and hence $\frac{1}{\bar{T}}Q^{2c}_{\bar{T}}(\boldsymbol{\phi}^*_0) < 0$.

Asymptotic Normality

Having shown identifiability and consistency, we define $\boldsymbol{\xi} \equiv \left[\boldsymbol{\phi}^{*\prime}_0, \boldsymbol{\theta}', \boldsymbol{\sigma}^{2\prime}\right]'$, with $\tilde{\boldsymbol{\xi}}$ denoting the true parameter values. Since the vector function $\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\xi}}$ continuous and differentiable, we may write

$$0 = \frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}(\hat{\boldsymbol{\xi}})}{\partial\boldsymbol{\xi}} = \frac{\partial\mathcal{L}_{\bar{T}}(\tilde{\boldsymbol{\xi}})}{\partial\boldsymbol{\xi}} + \frac{1}{\sqrt{\bar{T}}}\frac{\partial^2\mathcal{L}_{\bar{T}}(\bar{\boldsymbol{\xi}})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'}\left(\hat{\boldsymbol{\xi}} - \tilde{\boldsymbol{\xi}}\right),$$

for some $\bar{\boldsymbol{\xi}} \in [\hat{\boldsymbol{\xi}}, \tilde{\boldsymbol{\xi}}]$ such that $\bar{\boldsymbol{\xi}} \to_p \tilde{\boldsymbol{\xi}}$. The asymptotic distribution of the QMLE estimator then follows from normalising and rearranging

$$\sqrt{\bar{T}}\left(\hat{\boldsymbol{\xi}} - \tilde{\boldsymbol{\xi}}\right) = \left[\frac{1}{\bar{T}}\frac{\partial^2\mathcal{L}_{\bar{T}}(\bar{\boldsymbol{\xi}})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'}\right]^{-1}\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}(\tilde{\boldsymbol{\xi}})}{\partial\boldsymbol{\xi}}.$$

As, following Assumption 1, $\boldsymbol{u}_t$ is stationary with finite fourth moments so is, under Assumptions 4, 5 and 6, $\boldsymbol{\chi}_t$ and hence a central limit theorem for stationary and ergodic processes can be applied to $\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}(\tilde{\boldsymbol{\xi}})}{\partial\boldsymbol{\xi}}$.

For ease of notation, we introduce two standardised variables. Define $\boldsymbol{\zeta}_t = [\zeta_{1t},\ldots\ \zeta_{Nt}]'$, such that $\boldsymbol{\zeta}_t \odot \boldsymbol{\sigma}^2 \equiv \boldsymbol{u}_t$, and denote $E\zeta^3_{it} = \mu^3_i$ and $E\zeta^4_{it} = \mu^4_i$. Define $\boldsymbol{\Xi}_t = [\boldsymbol{\Xi}'_{1t},\ldots,\boldsymbol{\Xi}'_{Nt}]'$ such that $\boldsymbol{\Xi}_t \odot \boldsymbol{\sigma}^2 \equiv \boldsymbol{\chi}_t$. We note that $E(\boldsymbol{\zeta}_t) = 0$ and $E(\boldsymbol{\zeta}_t\boldsymbol{\zeta}'_t) = \boldsymbol{I}$. Let $\boldsymbol{\aleph} = E\frac{1}{\bar{T}}\sum\limits_{t=q+1}^{T}\boldsymbol{\Xi}_t\boldsymbol{\Xi}'_t$ partitioned in such a way that $\boldsymbol{\aleph}_{ij} = E\frac{1}{\bar{T}}\sum\limits_{t=q+1}^{T}\boldsymbol{\Xi}_{it}\boldsymbol{\Xi}'_{jt}$ and with block columns $\boldsymbol{\aleph}_i = E\frac{1}{\bar{T}}\sum\limits_{t=q+1}^{T}[\boldsymbol{\Xi}_{1t}\prime,\ldots\boldsymbol{\Xi}_{Nt}\prime]'\boldsymbol{\Xi}'_{it}$ Evaluating at the true parameter value, $\tilde{\boldsymbol{\xi}}$ and making used of the relation $\boldsymbol{y}^*_t = \boldsymbol{G}\left(\boldsymbol{\Theta}\boldsymbol{\Xi}_t + \boldsymbol{\zeta}_t\right) \odot \boldsymbol{\sigma}^2$, produces the following expressions

$$\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}(\tilde{\boldsymbol{\xi}})}{\partial\phi^*_{i0}} = \frac{1}{\sqrt{\bar{T}}}\sum_{t=q+1}^{T}\{\boldsymbol{g}_i[\boldsymbol{\Theta}\boldsymbol{\Xi}_t+\boldsymbol{\zeta}_t]\zeta_{it}-g_{ii}\}\text{, so that}$$

$$\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}(\tilde{\boldsymbol{\xi}})}{\partial\boldsymbol{\phi}^*_{0}} = \frac{1}{\sqrt{\bar{T}}}\sum_{t=q+1}^{T}\{\boldsymbol{G}_0[\boldsymbol{\Theta}\boldsymbol{\Xi}_t+\boldsymbol{\zeta}_t]\odot\boldsymbol{\zeta}_t-\mathrm{vecd}(\boldsymbol{G}_0)\}$$

$$\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\theta}_i} = -\frac{1}{\sqrt{\bar{T}}}\sum_{t=q+1}^{T}\boldsymbol{\Xi}_{it}\zeta_{it},$$

$$\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\sigma_i^2} = \frac{1}{\sqrt{\bar{T}}}\frac{1}{2\sigma_i^2}\sum_{t=q+1}^{T}(\zeta_{it}^2-1).$$

The assumptions relating to $\boldsymbol{X_t}$ and the absence of serial correlation in $\boldsymbol{u}_t$ ensure that $E\boldsymbol{\Xi}_t\boldsymbol{\zeta}_t' = 0$ that, as expected, $E\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}(\tilde{\boldsymbol{\xi}})}{\partial\boldsymbol{\xi}} = 0$. By then deploying the law of iterated expecations, Davidson and Mackinnon (1993, p292) $E\left\{\sum_{t=q+1}^{T}\frac{\partial\mathcal{L}_t}{\partial\boldsymbol{\xi}}\sum_{s=q+1}^{T}\frac{\partial\mathcal{L}_s}{\partial\boldsymbol{\xi}'}\right\}=0$ for $t\neq s$.

The covariance matrix for $\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}(\tilde{\boldsymbol{\xi}})}{\partial\boldsymbol{\xi}}$ is based on taking the expectation of outer products, using the result that, for $i\neq j$, $E\zeta_{it}^r\zeta_{jt}=0$, for $r=1,2,3$.

$$E\left\{\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\phi}^*_{i0}}\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\phi}^*_{i0}}\right\} = g_{ii}^2+\boldsymbol{g}_i'\boldsymbol{g}_i+\boldsymbol{g}_i'\boldsymbol{\Theta}\boldsymbol{\aleph}\boldsymbol{\Theta}'\boldsymbol{g}_i+g_{ii}^2\mu_i^4+2\mu_i^3g_{ii}\boldsymbol{g}_i'\boldsymbol{\Theta}E\left\{\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\Xi}_{it}\right\},$$

$$E\left\{\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\phi}^*_{i0}}\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\phi}^{*\prime}_{j0}}\right\} = g_{ij}g_{ji},$$

$$E\left\{\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\phi}^*_{i0}}\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\theta}_i'}\right\} = \boldsymbol{g}_i'\boldsymbol{\Theta}\boldsymbol{\aleph}_i+\mu_i^3g_{ii}E\left\{\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\Xi}_{it}'\right\},$$

$$E\left\{\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\phi}^*_{i0}}\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\sigma_i^2}\right\} = \frac{1}{2\sigma_i^2}\left[\mu_i^3\boldsymbol{g}_i'\boldsymbol{\Theta}E\left\{\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\Xi}_t\right\}+g_{ii}(\mu_i^4-1)\right],$$

$$E\left\{\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\theta}_i}\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\theta}_i'}\right\} = \boldsymbol{\aleph}_{ii},$$

$$E\left\{\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\boldsymbol{\theta}_i}\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\sigma^2{}_i}\right\} = \frac{\mu_i^3}{2\sigma_i^2}E\left\{\frac{1}{\bar{T}}\sum_{t=q+1}^{T}\boldsymbol{\Xi}_{it}\right\}\text{, and finally}$$

$$E\left\{\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\sigma_i^2}\frac{1}{\sqrt{\bar{T}}}\frac{\partial\mathcal{L}_{\bar{T}}}{\partial\sigma^2{}_i}\right\} = \frac{1}{4\sigma_i^2}(\mu_i^4-1).$$

At the same time the non-zero elements of $\boldsymbol{H} \equiv -\frac{1}{\bar{T}} E \frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \xi \partial \xi'}$ are

$$\begin{aligned}
-\frac{1}{\bar{T}} E \frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \phi_{0i}^* \partial \phi_{0i}^*} &= g_{ij} g_{ji} + \boldsymbol{g}_i' \boldsymbol{g}_i + \boldsymbol{g}_i' \boldsymbol{\Theta} \boldsymbol{\aleph} \boldsymbol{\Theta}' \boldsymbol{g}_i, \ i = j, \\
-\frac{1}{\bar{T}} E \frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \phi_{0i}^* \partial \phi_{0j}^*} &= g_{ij} g_{ji}, \\
-\frac{1}{\bar{T}} E \frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \phi_{0i}^* \boldsymbol{\theta}_i'} &= \boldsymbol{g}_i' \boldsymbol{\Theta} \boldsymbol{\aleph}_i, \\
-\frac{1}{\bar{T}} E \frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \phi_{0i}^* \partial \sigma_i^2} &= \frac{g_{ii}}{\sigma_i^2}, \\
-\frac{1}{\bar{T}} E \frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i'} &= \boldsymbol{\aleph}_{ii}, \text{ and finally} \\
-\frac{1}{\bar{T}} E \frac{\partial^2 \mathcal{L}_{\bar{T}}}{\partial \sigma_i^2 \partial \sigma_i^2} &= \frac{1}{2\sigma_i^4}.
\end{aligned}$$

The result follows from It can be seen that when $\boldsymbol{u}_t$ is normally distributed so that $\mu_i^3 = 0$ and $E\mu_i^4 = 3$, $\forall i$ then QML estimation become exact maximum likelihood estimation and the information equality holds. □

Appendix 4: Proof of Theorem 2

In the linear case, it can be shown using the Frisch-Waugh Lovell theorem that control function estimation becomes two stage least squares estimation. This results from multiplying both sides of equation (20) by the annihilation matrix for $\hat{v}_{it}$, $[\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}] = [\boldsymbol{I} - \boldsymbol{M}_{\boldsymbol{z}i} \boldsymbol{y}_i^* (\boldsymbol{y}_i^{*\prime} \boldsymbol{M}_{\boldsymbol{z}i} \boldsymbol{y}_i^*)^{-1} \boldsymbol{y}_i^{*\prime} \boldsymbol{M}_{\boldsymbol{z}i}]$, where for any matrix $\boldsymbol{A}$ of full column rank we define $\boldsymbol{M}_{\boldsymbol{A}} \equiv \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}$ and $\boldsymbol{P}_{\boldsymbol{A}} \equiv \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}$ as the projection matrix into the column space of $\boldsymbol{A}$. Since $\boldsymbol{z}_i^1$ lies in the column space of $\boldsymbol{z}_i$, $[\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}]\boldsymbol{z}_i^1 = \boldsymbol{z}_i^1$, whereas

$$[\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}]\boldsymbol{y}_i^* = [\boldsymbol{I} - \boldsymbol{M}_{\boldsymbol{z}i} \boldsymbol{y}_i^* (\boldsymbol{y}_i^{*\prime} \boldsymbol{M}_{\boldsymbol{z}i} \boldsymbol{y}_i^*)^{-1} \boldsymbol{y}_i^{*\prime} \boldsymbol{M}_{\boldsymbol{z}i}]\boldsymbol{y}_i^* = \boldsymbol{P}_{\boldsymbol{z}i} \boldsymbol{y}_i^*,$$

the fitted values from regression (16), the t'the element of which is $y_{it}^* - \hat{v}_{it}$. Since is the projection of the endogenous variable y_{it}^* on the instruments $\boldsymbol{z}_{it}$, it follows that the two stage least squares residuals are given by $\hat{u}_{it} = \hat{e}_{it} + \hat{v}_{it}\hat{\rho}$, the part of y_{it} that is not explained by these exogenous regressors and fitted values.

Then, also following Assumptions 1, 6 and 7' and Lemma 2 the sequence $\boldsymbol{z}_{i,t} u_{i,t}$ is a stationary, ergodic martingale difference sequence with finite fourth moments. The consistency and asymptotic normality of the instrumental variables estimator then follow from White (1984, 5.27). □