The Spatio-Temporal Autoregressive Distributed Lag Modelling Approach to an Analysis of Dynamic Networks *

Yongcheol Shin and Michael Thornton
University of York

Revised, November 2019

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 (2019), the only paper in considering heterogeneous spatial parameters. To deal with the simultaneity arising from spatial-lagged dependent variables, we develop both QML-based and 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 diffusion multipliers that can capture adjustments as well as dynamic 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 Network, Iraqi War Casualties
JEL CODES: C23, C55.

*We gratefully acknowledge many helpful interactions with Badi Baltagi, Francesco Bravo, Jia Chen, In Choi, Matthew Greenwood-Nimmo, David Kang, Uih Ran Lee, Rui Lin, Roger Moon, Viet Nguyen, Paulo Santos Monteiro, Peter Wagner, Weining Wang, Yoon Jae Whang, Chaowen Zheng, and the comments of seminar participants at Seoul National University, Sogang University, Yonsei University and University of York as well as delegates at the 11th York Econometrics Symposium: Big Data Analysis and its Applications in Economics and Finance (York, April 2018), the 24th International Panel Data Conference (Seoul, June 2018), the Joint Symposium on Statistical Penalisation Methods and Dimension Reduction Methods for Economic and Financial Analysis (York, September 2018), the KAEA Session III, jointly with American Economic Association (Atlanta, January 2019), the 11th French Econometrics Conference (Marseille, November 2019). 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, economic geographers 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 popular spatial autoregressive model imposes the restriction that spatial spill-overs flow only through the dependent variable but more general forms allowing the explanatory variables of one unit to impact the dependent variable of another directly, such as the spatial Durbin model, are also widely used. 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.

Much of the early work on spatial models was done with large cross-section dimensions in mind but, given the growing availability of spatial datasets with a large time dimension, it is of interest to explore time dynamics in greater detail. Imposing a static model on data with a time dimension is, in effect, assuming that the spatial system is only observed on (or relatively close to) its equilibrium path and thereby misses the opportunity to study of the transition between equilibria that might be afforded by regular data collection. In this respect our work mirrors recent interest in dynamic spatial models, 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. 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 to differentiate short-run and long-run behaviour, 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, and researchers must decide how and whether to normalise. 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, hereafter ABP), 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 develop a quasi-maximum likelihood (QML) estimator and a control function estimator 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 computationally difficult for large cross-sections and the problem is exacerbated by heterogeneity across the spatial parameters. 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 in terms of bias and root mean square error. 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, heterogeneity across a large number of units can make the meaningful interpretation of results difficult. In contrast to homogeneous parameter models, where interest often centres on the single spatial parameter, here we have numerous parameters performing complementary roles within the network. Another contribution of this paper is in developing a comprehensible and widely applicable format for the presentation of estimation results alongside the tools to analyse the evolving importance of particular nodes. We avoid the use of potentially misleading mean group estimators and instead provide two quantities, based on work by Shin et al. (2014) and Greenwood-Nimmo, Nguyen and Shin (2015): the individual spatio-temporal dynamic multipliers; and, the system diffusion multipliers, leading to our network connectedness matrix, each of which exists for a range of time horizons. The network connectedness matrices form a sequence of output network matrices resulting from the input network matrix, $\boldsymbol{W}$, and the STARDL coefficients. They also give rise to two intuitive measures of the developing role of each

node within the network: their external motivation, reflecting the extent to, and direction in, which they are steered by the network; and, their systemic influence, reflecting their relative importance within it.

We then demonstrate the usefulness of our approach in an empirical application considering the effect of enemy casualties on civilian deaths across the 18 governorates of Iraq in the period following the 2003 invasion. The period in question was characterised by an armed insurgency in which mobile militias, provoked by conditions across the country, were drawn to particular hotspots but then able to relocate to supportive conditions when threatened by coalition forces. The time and spatial contamination we find is consistent with the use of violence against civilians as a means to maintain influence in the nascent political institutions as well as in retribution for past acts. Our network connectedness measures highlight the importance of the capital, Baghdad, as a net propagator of violence against civilians, in particular to the Shia centre of Basrah and to the governorate of Diyala. At the same time, many other governorates had strong institutions throughout the period and were able to maintain relatively high levels of security, even when geographically positioned between two hotspots. Our use of heterogenous coefficients has a considerable advantage in modelling the range of effects the network had on different provinces over different time horizons.

The structure of the paper is as follows. Section 2 presents the basic specification of the model and discusses in detail its underlying properties. Two estimators and their asymptotic distributions are discussed in section 3, which then considers some extensions to the basic framework. Section 4 develops the spatio-temporal dynamic and diffusion multipliers and discusses their use in measures of network connectedness. 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 aftermath of the 2003 Iraq war. Section 7 concludes. Aall proofs and further details of the estimators are relegated to the Appendix.

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):

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

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, $\boldsymbol{x}_{it} = (x^1_{it}, ..., x^K_{it})'$ is a $K \times 1$ vector of exogenous regressors with a $K \times 1$ vector of parameters, $\boldsymbol{\pi}_0 = (\pi^1_0, ..., \pi^K_0)'$. 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

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

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

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 information on both time and spatial dependence. If $\phi_{i\ell}$'s and $\boldsymbol{\pi}_{i\ell}$'s are statistically significant, this points to the usual temporal dynamics. In addition, if $\phi^*_{i\ell}$'s (the spatial interaction 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:

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

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

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NN} \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1' \\ \vdots \\ \boldsymbol{w}_N' \end{bmatrix}, \tag{5}$$

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

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

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

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

$$\boldsymbol{y}_t = \phi \boldsymbol{y}_{t-1} + \phi^*_0 \boldsymbol{W} \boldsymbol{y}_t + \phi^*_1 \boldsymbol{W} \boldsymbol{y}_{t-1} + \pi_0 \boldsymbol{x}_t + \pi_1 \boldsymbol{x}_{t-1} + \pi^*_0 \left(\boldsymbol{W} \otimes \boldsymbol{I}_K\right) \boldsymbol{x}_t + \pi^*_1 \left(\boldsymbol{W} \otimes \boldsymbol{I}_K\right) \boldsymbol{x}_{t-1} + \boldsymbol{u}_t. \tag{6}$$

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.[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 further discussion of the issues involved in the identification and estimation of this model.

[1] Somewhat misleading interpretations of spillover effects will be drawn if homogeneity is imposed falsely.

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

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

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

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

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

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 $\sigma_i^2 > 0$ 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 value of $\left(\boldsymbol{\phi}_0^{*\prime}, \boldsymbol{\theta}', \boldsymbol{\sigma}'\right)'$ lies in the interior of a compact set, where $\boldsymbol{\phi}_0^* = (\phi_{10}^*, \ldots, \phi_{N0}^*)'$, $\boldsymbol{\theta} = \left(\boldsymbol{\phi}_1', \ldots, \boldsymbol{\phi}_p', \boldsymbol{\phi}_1^{*\prime}, \ldots, \boldsymbol{\phi}_p^{*\prime}, \boldsymbol{\pi}_0', \ldots, \boldsymbol{\pi}_q', \boldsymbol{\pi}_0^{*\prime}, \ldots, \boldsymbol{\pi}_q^{*\prime}, \boldsymbol{\alpha}'\right)'$, and $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \ldots, \sigma_N^2\right)'$,

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 $\boldsymbol{S}\left(\boldsymbol{\phi}_0^*\right)) \equiv \boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W}$ 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 6: The explanatory variables $\boldsymbol{x}_{it}$ are random variables such that $E(\|\boldsymbol{x}_{it}\|^4) \leq C$ for all i and t, and they are independent of the idiosyncratic errors u_{js} for all (i, j, t, s).

Assumption 1 both limits the probability of extreme values of the disturbance and rules out the possibility that u_{it} is degenerate for any $i = 1, \ldots, N$. Here we model the time dependence through time lags of the dependent and explanatory variables such that u_{it} are serially uncorrelated with $E(\boldsymbol{u}_t \boldsymbol{u}_s') \equiv \boldsymbol{\Sigma}_u = diag\left(\boldsymbol{\sigma}^2\right)$ for $t = s$, and is null otherwise. The more common requirement that u_{it} be *i.i.d.* has been weakened to allow cross-section heteroskedasticty. Assumption 2 is standard in the literature on extremum estimators. Assumption 3 is common within the spatial literature, containing the normalising convention that no unit acts as its own neighbour and limiting spatial diffusion so that y_{it}^* and $\boldsymbol{x}_{it}^*$ remain bounded if y_{jt} and $\boldsymbol{x}_{jt}$ are bounded for all $i, j = 1, \ldots, N$ and $t = 1, \ldots, T$.

Assumption 4 is needed to limit 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 (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. The conditions on $\boldsymbol{\Phi}_0^* \boldsymbol{W}$ under 4(a) are sufficient for those given under 4(b), since we may write $\left(\boldsymbol{I}_N - \boldsymbol{\Phi}_0^* \boldsymbol{W}\right)^{-1} = \boldsymbol{I}_N + \boldsymbol{\Phi}_0^* \boldsymbol{W} + \left(\boldsymbol{\Phi}_0^* \boldsymbol{W}\right)^2 + \ldots$, 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, for which we make Assumption 5, which is a familiar condition from the literature on dynamic systems (e.g. Hamilton, 1994). Assumption 5 is the necessary condition for a stable relationship existing between $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$.[2] Assumption 5 generalises Assumption 4 in Mutl (2009) and 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 $\tilde{\boldsymbol{B}}\left(L\right) \left(= \sum_{\ell=0}^{\infty} \tilde{\boldsymbol{B}}_\ell L^\ell \right) = \tilde{\boldsymbol{\Phi}}\left(L\right)^{-1} \tilde{\boldsymbol{\Pi}}_\ell\left(L\right)$. 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). Assumption 6 together with Assumptions 1 and 3 is needed to ensure $E(\|\boldsymbol{\chi}_{it}\|^4) \leq C$ for all i and t, see the Appendix.

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). 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 using instrumental variables (IV), see by Kelejian and Prucha (1998, 1999), and the generalised method of moments (GMM), see Elhorst

[2] 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 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.

(2010) and Lee and Yu (2014). The growing availability of datasets with both spatial and time dimension has sparked interest in dynamic spatial panel data models, see Lee and Yu (2014). In what follows we define $r = \max\{p, q\}$ and $\bar{T} = T - r$ as the sample size after allowing for required lags.

3.1 Quasi-Maximum Likelihood Estimation

We denote the parameters in (1), including the error variance as $\boldsymbol{\xi} = \left(\boldsymbol{\phi}_0^{*\prime}, \boldsymbol{\theta}', \boldsymbol{\sigma}^{2\prime}\right)'$ where $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_N'\right)'$ and $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, ..., \sigma_N^2\right)'$. We use $\tilde{}$ to denote true value of these parameters, $\tilde{\boldsymbol{\xi}} = \left(\tilde{\boldsymbol{\phi}}_0^{*\prime}, \tilde{\boldsymbol{\theta}}', \tilde{\boldsymbol{\sigma}}^{2\prime}\right)'$.

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

$$\mathcal{L}_{\bar{T}}(\boldsymbol{\xi}) = -\frac{N\bar{T}}{2}\ln(2\pi) - \frac{\bar{T}}{2}\ln|\boldsymbol{\Sigma}_u| + \bar{T}\ln|\boldsymbol{S}(\boldsymbol{\phi}_0^*)| - \frac{1}{2}\sum_{t=r+1}^{T}\boldsymbol{u}_t'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{u}_t. \tag{12}$$

Since the quasi-likelihood reverts to standard if $\boldsymbol{\phi}_0^*$ is known, the closed form solutions for $\boldsymbol{\theta}$ and $\boldsymbol{\sigma^2}$ given $\boldsymbol{\phi}_0^*$ are (see the Appendix for details):

$$\hat{\boldsymbol{\theta}}_i(\phi_{0i}^*) = \left(\frac{1}{\bar{T}}\sum_{t=r+1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=r+1}^{T}\boldsymbol{\chi}_{it}\left[y_{it} - \phi_{0i}^* y_{it}^*\right]\right), \tag{13}$$

$$\begin{aligned}\hat{\sigma}_i^2(\phi_{0i}^*) &= \frac{1}{\bar{T}}\sum_{t=r+1}^{T}\left(y_{it} - \phi_{0i}^* y_{it}^* - \hat{\boldsymbol{\theta}}_i(\phi_{0i}^*)'\boldsymbol{\chi}_{it}\right)^2 \\ &= \frac{1}{\bar{T}}\sum_{t=r+1}^{T}\left(y_{it} - \phi_{0i}^* y_{it}^*\right)^2 \\ &\quad -\frac{1}{\bar{T}}\sum_{t=r+1}^{T}\left(y_{it} - \phi_{0i}^* y_{it}^*\right)\boldsymbol{\chi}_{it}'\left[\left(\frac{1}{\bar{T}}\sum_{t=r+1}^{T}\boldsymbol{\chi}_{it}\boldsymbol{\chi}_{it}'\right)\right]^{-1}\frac{1}{\bar{T}}\sum_{t=r+1}^{T}\boldsymbol{\chi}_{it}\left(y_{it} - \phi_{0i}^* y_{it}^*\right),\end{aligned} \tag{14}$$

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

$$\mathcal{L}_{\bar{T}}^c(\boldsymbol{\phi}_0^*) = -\frac{N\bar{T}}{2}\ln(2\pi + 1) - \frac{\bar{T}}{2}\sum_{i=1}^{N}\ln\hat{\sigma}_i^2(\phi_{0i}^*) + \bar{T}\ln|\boldsymbol{S}(\boldsymbol{\phi}_0^*)|, \tag{15}$$

and the heterogeneous spatial parameters $\boldsymbol{\phi}_0^*$ can be estimated more conveniently by

$$\hat{\boldsymbol{\phi}}_0^* = \arg\max_{\boldsymbol{\phi}_0^* \in \Theta_{\phi_0^*}} \mathcal{L}_{\bar{T}}^c(\boldsymbol{\phi}_0^*).$$

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

$$Q_{\bar{T}}(\boldsymbol{\xi}) = E\left[\mathcal{L}_{\bar{T}}(\boldsymbol{\phi}_0^*, \boldsymbol{\theta}, \boldsymbol{\Sigma}_u)\right]. \tag{16}$$

In order to evaluate this expression, we express $\boldsymbol{S}(\boldsymbol{\phi}_0^*)\boldsymbol{y}_t$ in terms of $\boldsymbol{\chi}_t$ and $\boldsymbol{u}_t$. Define $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{S}^{-1} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_N]'$, with rows $\boldsymbol{g}_i' = [g_{i1}, \ldots, g_{iN}]$, and $\boldsymbol{S} = \boldsymbol{S}\left(\tilde{\phi}_0^*\right)$. Note that $\boldsymbol{I} + \tilde{\boldsymbol{\Phi}}_0^* \boldsymbol{G} = \boldsymbol{S}^{-1}$ so that

$$\boldsymbol{S}\left(\boldsymbol{\phi}_0^*\right)\boldsymbol{S}^{-1} = \boldsymbol{S}^{-1} - \boldsymbol{\Phi}_0^* \boldsymbol{G} = \boldsymbol{I} + \left(\tilde{\boldsymbol{\Phi}}_0^* - \boldsymbol{\Phi}_0^*\right)\boldsymbol{G},$$

and hence

$$\boldsymbol{S}(\boldsymbol{\phi}_0^*)\boldsymbol{y}_t = \boldsymbol{S}\left(\boldsymbol{\phi}_0^*\right)\boldsymbol{S}^{-1}\left[\boldsymbol{\Theta}\boldsymbol{\chi}_t + \boldsymbol{u}_t\right] = \left[\boldsymbol{I} + \left(\tilde{\boldsymbol{\Phi}}_0^* - \boldsymbol{\Phi}_0^*\right)\boldsymbol{G}\right]\boldsymbol{\Theta}\boldsymbol{\chi}_t + \boldsymbol{S}\left(\boldsymbol{\phi}_0^*\right)\boldsymbol{S}^{-1}\boldsymbol{u}_t.$$

Let $\boldsymbol{s}_i^{*\prime}(\phi_{0i}^*)$ denote the i'th row of $\boldsymbol{S}\left(\boldsymbol{\phi}_0^*\right)$. Note that the scalar argument reflects independence from any ϕ_{0j}^* for $j \neq i$, which is the result of modelling cross-section dependence through observable spatial contamination rather than unobserved factors as in Song (2013).

The i'th row of the above relationship can then be written

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

where $\kappa_{it}(\phi_{i0}^*)$ is reflects the part of $(y_{it} - \phi_{i0}^* y_{it}^*)$ driven by the pre-determined variables in the system, $\boldsymbol{\chi}_t$, as opposed to the disturbance, $\boldsymbol{s}_i'(\phi_{0i}^*)\boldsymbol{S}^{-1}\boldsymbol{u}_t$.

For given $\boldsymbol{\phi}_0^*$, the values of $\boldsymbol{\theta}$ that maximises $Q_{\bar{T}}(\boldsymbol{\xi})$ is given by

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

and the value of σ_i^2 is given by

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

Substituting in we have the concentrated expected quasi-likelihood function

$$Q_{\bar{T}}^c(\boldsymbol{\phi}_0^*) \equiv \max_{\boldsymbol{\theta},\boldsymbol{\sigma}^2} E\left[\mathcal{L}_{\bar{T}}\left(\boldsymbol{\phi}_0^*, \boldsymbol{\theta}, \boldsymbol{\sigma}^2\right)\right] = -\frac{N\bar{T}}{2}\ln(2\pi + 1) - \frac{\bar{T}}{2}\sum_{i=1}^{N}\ln\bar{\sigma}_i^2(\phi_{0i}^*) + \bar{T}\ln\left|\boldsymbol{S}(\boldsymbol{\phi}_0^*)\right|. \quad (17)$$

It follows from the definition of $\boldsymbol{S}$ that $\boldsymbol{s}_i'(\tilde{\phi}_{0i}^*)\boldsymbol{S}^{-1} = \boldsymbol{e}_i'$, the i'th row of the $N \times N$ identity matrix. Hence, at the true parameter value, $\kappa_{it}(\phi_{i0}^*) = \boldsymbol{\chi}_{it}'\tilde{\boldsymbol{\theta}}_i$, which ensures that $\bar{\boldsymbol{\theta}}_i(\tilde{\phi}_{0i}^*) = \tilde{\boldsymbol{\theta}}_i$ and $\bar{\sigma}_i^2(\tilde{\phi}_{0i}^*) = \tilde{\sigma}_i^2$. Away from the true parameter value, however, $\kappa_{it}(\phi_{i0}^*)$ will be a weighted sum of $\boldsymbol{\chi}_{jt}$, $j = 1, \ldots, N$. In order to identify ϕ_{i0}^*, we require that $\kappa_{it}(\phi_{i0}^*)$ is not entirely driven by $\boldsymbol{\chi}_{it}$ alone, but depends on $\boldsymbol{\chi}_{jt}$ for some $j \neq i$ such that the difference between $\kappa_{it}(\phi_{i0}^*)$ and $\kappa_{it}(\tilde{\phi}_{i0}^*)$ cannot be bridged by a different choice of $\boldsymbol{\theta}_i$. This is encapsulated in the following assumption.

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

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

Assumption 7 mirrors the local identification Assumption 8 in Lee (2004) and Assumption G2 in Li (2017). It asserts 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 identify locally the parameter vector $\boldsymbol{\phi}_0^*$, given $\boldsymbol{\chi}_t$, and hence $\boldsymbol{\theta}_i$ and σ_i^2, using the closed form expressions in (13) and (14). Arguments similar to Lee (2004) may be deployed to establish the following Theorem (see the Appendix).

Theorem 1 (**The QML estimator**) Consider the STARDL model (8) and suppose that Assumptions 1-7 hold. Then, as $T \to \infty$,

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

where

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

Detailed expressions are given in the Appendix.

Even after $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ have been concentrated out of (12), the remaining optimisation over $\boldsymbol{\phi}^*$, requires repeated evaluation of the determinant $N \times N$ matrix, $\boldsymbol{I}_N - \boldsymbol{\Phi}_0^*\boldsymbol{W}$, making the maximisation of (12) or (15) numerically burdensome for large N. The technique proposed by Ord (1975), which greatly simplifies the calculation once the eigenvalues of $\boldsymbol{W}$ have been found, has proved popular in the estimation of models with a homogeneous spatial autoregressive parameter despite some potential to become numerically unstable, when N is large and $\boldsymbol{W}$ asymmetric, both of which are typically true in applied work.[3] This technique is not applicable, however, when the spatial autoregressive parameters are heterogeneous. With this in mind, we consider a computationally simpler method that exploits the naturally available instruments to control the endogeneity within (8).

[3]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 $\boldsymbol{W}$.

3.2 Control Function Estimation

We adopt a control function (CF) approach to instrumenting for the endogeneity in (8).[4] Let $\boldsymbol{z}_{it}$ be the $L \times 1$ vector of exogenous/pre-determined variables:

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

where the $L_1 \times 1$ vector, $\boldsymbol{\chi}_{it}$ contains all all exogenous and pre-determined variables included in (8). The $L_2 \times 1$ vector $\boldsymbol{z}^2_{it}$ 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 $\boldsymbol{\chi}_{it}$ (see Section 3.2.1).

Assumption 7': For each $i = 1, \ldots, N$ there exist a vector, $\boldsymbol{z}^{2\prime}_{it}$ such that $(\boldsymbol{z}^2_{it} u_{it})$ is a stationary and ergodic martingale difference sequence, $\lim_{T\to\infty} \left\{ E\left(\frac{1}{\overline{T}} \sum_{t=r+1}^{T} \boldsymbol{z}^2_{it} y^*_{it} \right) \right\} \neq \mathbf{0}$ and for all L_2 vectors $\boldsymbol{\gamma} \neq 0$

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

Assumption 7' carries the usual condition that $\boldsymbol{z}^2_{it}$ is asymptotically correlated with y^*_{it} but is not asymptotically linearly dependent on $\boldsymbol{\chi}_{it}$. Whereas Assumption 7 asserted that $\kappa(\phi^*_{i0})$, a linear combination (depending on ϕ^*_{i0}) of $\boldsymbol{\chi}_t$ was not asymptotically linearly dependent on $\boldsymbol{\chi}_{it}$, Assumption 7' allows for a wider choice of variables to be used, including higher orders of time lag which do not feature in $\kappa(\phi^*_{i0})$.

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

$$y^*_{it} = \boldsymbol{\varphi}'_i \boldsymbol{z}_{it} + v_{it} \text{ with } E\left(\boldsymbol{z}'_{it} v_{it}\right) = \mathbf{0} \tag{18}$$

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

$$u_{it} = \rho_i v_{it} + e_{it} \tag{19}$$

where $\rho_i = E\left(v_{it} u_{it}\right) / E\left(v^2_{it}\right)$. 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 (18) 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 (19), we obtain the following transformation of (8):

$$y_{it} = \phi^*_{i0} y^*_{it} + \boldsymbol{\theta}'_i \boldsymbol{\chi}_{it} + \rho_i v_{it} + e_{it} \tag{20}$$

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 (20).

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

We propose the two-step procedure: (i) obtain the reduced form residuals, $\hat{v}_{it} = y^*_{it} - \hat{\boldsymbol{\varphi}}_i' \boldsymbol{z}_{it}$ from (18) and (ii) run the following regression:

$$y_{it} = \phi^*_{i0} y^*_{it} + \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it} + \rho_i \hat{v}_{it} + e^*_{it} \tag{21}$$

where $e^*_{it} = e_{it} + \rho_i \left(\boldsymbol{\varphi}_i - \hat{\boldsymbol{\varphi}}_i\right)' \boldsymbol{z}_{it}$ depends on the sampling error in $\hat{\boldsymbol{\varphi}}_i$ unless $\rho_i = 0$ (exogeneity test). Then, the OLS estimator of the parameters in (21) will be consistent. OLS standard errors will not be consistent, however, due to the presence of an estimated regressor in $\hat{v}_{it}$. 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 distribution of the control function estimator is given in the following Theorem.

Theorem 2 (The CF estimator) Under Assumptions 1-6 and 7', as $T \to \infty$, the OLS estimator of $\boldsymbol{\beta}_i = \left(\phi^*_{i0}, \boldsymbol{\theta}_i'\right)'$ in (21) is consistent and asymptotically normally distributed as

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

$$AVar\left(\hat{\boldsymbol{\beta}}_i\right) \to_p \hat{\sigma}_i^2 \left(\sum_{t=r+1}^{T} \tilde{\boldsymbol{X}}_{it} \tilde{\boldsymbol{X}}_{it}'\right)^{-1},$$

where $\tilde{\boldsymbol{X}}_{it}' = \left(y^*_{i,t} - \hat{v}_{it}, \boldsymbol{\chi}_{it}'\right)$ and $\hat{\sigma}_i^2 = \bar{T}^{-1} \sum_{t=r+1}^{T} \hat{u}_{it}^2$, where $\hat{u}_{it} = \hat{e}_{it} + \hat{v}_{it}\hat{\rho}$ and $\hat{\rho}$ is the OLS estimate of ρ from (21).

It is not too surprising that the same $\sqrt{T}$ rate of consistency also applies to the control function estimator. The expression for the variance follows from the cross-section heteroskedaticity assumption on the disturbances made in Assumption 1. The estimator would remain consistent if time heteroskedasticity were allowed, in which case

$$AVar\left(\hat{\boldsymbol{\beta}}_i\right) \to_p \left(\sum_{t=r+1}^{T} \tilde{\boldsymbol{X}}_{it} \tilde{\boldsymbol{X}}_{it}'\right)^{-1} \left(\sum_{t=r+1}^{T} \hat{u}_{it}^2 \tilde{\boldsymbol{X}}_{it} \tilde{\boldsymbol{X}}_{it}'\right) \left(\sum_{t=r+1}^{T} \tilde{\boldsymbol{X}}_{it} \tilde{\boldsymbol{X}}_{it}'\right)^{-1}.$$

- **YC: which $AVar$ do we use in the simulation and application? The first one or the 2nd? As mentioned by Francesco, what's the comparison between the asymptotic distributions of QML and CF estimators?**

Conversely, Assumption 1 could be strengthened to cross-section homoskedasticity, $E(u_{it}^2) = \sigma^2$, $\forall i$, with σ^2 estimated by $\hat{\sigma}^2 = (N\bar{T})^{-1} \sum_{i=1}^{N} \sum_{t=r+1}^{T} \hat{u}_{it}^2$. {YC: do we need this?}

3.2.1 The selection of instrumental variables

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

$$\boldsymbol{y}_t^* = \boldsymbol{G}\left[\sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \boldsymbol{y}_{t-\ell} + \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell^* \boldsymbol{W} \boldsymbol{y}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_\ell \boldsymbol{x}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_\ell^* \left(\boldsymbol{W} \otimes \boldsymbol{I}_K\right) \boldsymbol{x}_{t-\ell} + \boldsymbol{\alpha} + \boldsymbol{u}_t\right]. \tag{22}$$

This suggests that

$$\left[\sum_{\ell=1}^{p}\boldsymbol{W}^{2}\boldsymbol{y}_{t-\ell},\sum_{\ell=1}^{p}\boldsymbol{W}^{3}\boldsymbol{y}_{t-\ell},\ldots,\sum_{\ell=0}^{q}\boldsymbol{W}^{2}\boldsymbol{x}_{t-\ell},\sum_{\ell=0}^{q}\boldsymbol{W}^{3}\boldsymbol{x}_{t-\ell},\ldots\right] \tag{23}$$

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):

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

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

$$\boldsymbol{\Phi}\left(L\right)\boldsymbol{y}_t=\boldsymbol{\Phi}^{*}\left(L\right)\boldsymbol{W}\boldsymbol{y}_t+\boldsymbol{\Pi}\left(L\right)\boldsymbol{x}_t+\boldsymbol{\Pi}^{*}\left(L\right)\left(\boldsymbol{W}\otimes\boldsymbol{I}_K\right)\boldsymbol{x}_t+\boldsymbol{\alpha}+\boldsymbol{u}_t \tag{24}$$

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

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

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

$$\boldsymbol{y}_t^{*}=\boldsymbol{G}\left\{\boldsymbol{\Psi}_1\left(L\right)\boldsymbol{y}_t+\boldsymbol{\Xi}\left(L\right)\boldsymbol{x}_t+\left[\boldsymbol{\Phi}\left(L\right)\right]^{-1}\left[\boldsymbol{\alpha}+\boldsymbol{u}_t\right]\right\} \tag{25}$$

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

$$\left[\boldsymbol{W}^{2}\boldsymbol{y}_{t-p-1},\boldsymbol{W}^{2}\boldsymbol{y}_{t-p-2},...,\boldsymbol{W}\boldsymbol{x}_{t-p-1},\boldsymbol{W}\boldsymbol{x}_{t-p-2},...\right] \tag{26}$$

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.

3.3 The STARDL model with both individual and time effects

Consider the STARDL model with both individual and time effects, α_i and τ_t:

$$y_{it}=\sum_{\ell=1}^{p}\phi_{i\ell}y_{i,t-\ell}+\sum_{\ell=0}^{p}\phi_{i\ell}^{*}y_{i,t-\ell}^{*}+\sum_{\ell=0}^{q}\boldsymbol{\pi}_{i\ell}^{\prime}\boldsymbol{x}_{i,t-\ell}+\sum_{\ell=0}^{q}\boldsymbol{\pi}_{i\ell}^{\prime *}\boldsymbol{x}_{i,t-\ell}^{*}+\alpha_i+\tau_t+u_{it}, \tag{27}$$

which can be written as

$$\mathring{y}_{it}=\sum_{\ell=1}^{p}\phi_{i\ell}\mathring{y}_{i,t-\ell}+\sum_{\ell=0}^{p}\phi_{i\ell}^{*}\mathring{y}_{i,t-\ell}^{*}+\sum_{\ell=0}^{q}\boldsymbol{\pi}_{i\ell}^{\prime}\mathring{\boldsymbol{x}}_{i,t-\ell}+\sum_{\ell=0}^{q}\boldsymbol{\pi}_{i\ell}^{\prime *}\mathring{\boldsymbol{x}}_{i,t-\ell}^{*}+\mathring{u}_{it},$$

[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.

where $\mathring{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$, and similarly for $\mathring{\boldsymbol{x}}_{it}$ and $\mathring{u}_{it}$. In this case, we need to account for the correlations between lagged y_{it} and $\mathring{u}_{it}$, between lagged y^*_{it} and $\mathring{u}_{it}$, as well as between current y^*_{it} and $\mathring{u}_{it}$. We first consider for $\ell = 1, 2, ...$:

$$Cov\left(y_{i,t-\ell}, \mathring{u}_{it}\right) = O\left(\frac{1}{T}\right) + O\left(\frac{1}{NT}\right); Cov\left(y^*_{i,t-\ell}, \mathring{u}_{it}\right) = O\left(\frac{1}{NT}\right).$$

These terms would disappear asymptotically as $T \to \infty$. Next,we have:

$$Cov\left(y^*_{it}, \mathring{u}_{it}\right) = Cov\left(y^*_{it}, u_{it}\right) + O\left(\frac{1}{N}\right)$$

Fortunately, the CF approach already controls for this correlation, provided instruments are chosen so that Assumption 7' holds for the process $\mathring{u}_{it}$. This rules out the combinations of $\boldsymbol{y}_t$ (for example, $\boldsymbol{W}^2\boldsymbol{y}_{t-p-1}$) as valid instruments but otherwise, given normalisation conditions sufficient to identify $\alpha_i + \tau_t$, such as $\sum_{i=1}^{N} \alpha_i = 0$ and $\sum_{t=1}^{T} \tau_t = 0$, we expect that the CF estimator of (27) will be consistent and follow the asymptotic normal distributions. This problem has been addressed in the context of the homogeneous parameter spatial dynamic panel data model by Lee and Yu (2010b) who show that the QML estimator, applied to a modified version of the above transformed data, is $\sqrt{T}$ consistent and asymptotically normally distributed.

4 Spatio-Temporal Network Analysis

Many of the tools of network analysis, such as popular centrality statistics and clustering algorithms, can be applied to spatial models to facilitate their interpretation, with the graphical visualisation and interpretation of network models growing rapidly in recent years. The STARDL model captures both spatial interactions and diffusion dependence within a dynamic network. In contrast to homogeneous parameter models, where interest often centres on the single spatial parameter, here we have numerous parameters performing complementary roles. Another contribution of this paper is in developing a comprehensible and widely applicable format for the presentation of the estimation results alongside the tools to analyse the evolving importance of particular nodes within the network. To do so we provide two quantities, based on work by Shin et al. (2014) and Greenwood-Nimmo, Nguyen and Shin (2015): the individual spatio-temporal dynamic multipliers; and, the system diffusion multipliers each of which exists for a range of time horizons. The latter are collected in a sequence of network connectedness matrices that can be interpreted as output network matrices resulting from the input network matrix, $\boldsymbol{W}$, and the STARDL coefficients. They also give rise to two intuitive measures of the developing role of each node within the network: their external motivation, reflecting the extent to, and direction in, which they are steered by the network; and, their systemic influence, reflecting their relative importance within it.

For this practical network-oriented approach, we need only $\sqrt{T}$-consistent estimators of the heterogeneous parameters. Our approach contrasts with the pooled or mean-group versions of the estimators, which have been routinely proposed in the main panel data literature. In many applications, including that in the following section, there is no economic reason to expect the coefficients of the model to be non-negative, or to share a common sign. Regions may effectively be competitors for the location of a particular industrial hub which, once established, sucks activity from its immediate surroundings. A pooled or mean group estimator would then be subject to an element of netting off that has the potential to produce a misleading global picture. A single

pooled estimator is also unable to discern the relative importance of individual nodes beyond that pre-supposed by the input matrix, $\boldsymbol{W}$. Moreover, the asymptotic distribution of the pooled or the mean-group estimators may be different, and their convergence rates depend on whether the parameters are homogeneous or heterogeneous.

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 $h = 0, 1, 2...$ Rewrite the STARDL(p, q) model, (1) as[6]

$$\phi_i(L)\, y_{it} = \phi_i^*(L)\, y_{it}^* + \boldsymbol{\pi}_i(L)\, \boldsymbol{x}_{it} + \boldsymbol{\pi}_i^*(L)\, \boldsymbol{x}_{it}^* + u_{it} \tag{28}$$

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

$$y_{it} = \tilde{\phi}_i^*(L)\, y_{it}^* + \tilde{\boldsymbol{\pi}}_i(L)\, \boldsymbol{x}_{it} + \tilde{\boldsymbol{\pi}}_i^*(L)\, \boldsymbol{x}_{it}^* + \tilde{u}_{it} \tag{29}$$

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 relationship:

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

where $\phi_{ij} = 0$ for $j < 1$ and $\tilde{\phi}_{i0}^* = \phi_{i0}^*$, $\tilde{\phi}_{ij}^* = 0$ for $j < 0$ by construction. Similarly for $\tilde{\boldsymbol{\pi}}_{ij}'$ and $\tilde{\boldsymbol{\pi}}_{ij}^{*\prime}$ for $j = 0, 1, ...$

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:[7]

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

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.

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 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.

[6] To construct the dynamic multipliers, we should use the structural parameters in (1) which are consistently estimated by the STARDL estimator from (21). Without loss of generality we drop the intercept, α_i.

[7] 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 $\left[\tilde{\boldsymbol{\Phi}}(L)\right]^{-1}$, we have:

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

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 elsewhere. 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 or 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 propose an average of the (cumulative) off diagonal elements over all rows to produce a summary that corresponds to the indirect 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. 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.

4.3 Dynamic 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 apply the network approach to an analysis of the $N \times N$ matrix of the diffusion multipliers of each regressor in $\boldsymbol{x}_t$. Here we follow Diebold and Yilmaz (2014) and Greenwood-Nimmo, Nguyen and Shin (2019, GNS) and apply the generalised connectedness measures.

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, $\boldsymbol{y}_t$. Rewrite the $N \times N$ matrices, $\boldsymbol{d}_{x^k}^{H}$ in (36) in terms of the following $N \times N$ connectedness matrix (we suppress the horizon index to avoid cluttering our notation.):

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

The main diagonal elements of $\mathbb{C}$ represent (cumulative) own-region impacts that arise from both time and spatial dependence. The off-diagonal elements 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 spillover effect.

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

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

which lie on the prime diagonal of $\mathbb{C}$. Next, we define the cross-from or *spill-in* 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. 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 aggregate impact of the regressor in country j attributable to all sources. Notice that the underlying diffusion multipliers in (37) are not normalised and may be positive or negative, which are different from the (normalised) connectedness matrix constructed using the positive FEVDs as in Diebold and Yilmaz (2014). In this situation $TOT_{j\leftarrow \bullet}$ may understate the total magnitude of the impacts on region j. To measure this we use the customary absolute row sum of $\mathbb{C}$, which we denote

$$ATOT_{j\leftarrow \bullet} = \sum_{i=1}^{N} |\phi_{j\leftarrow i}|. \tag{41}$$

Similarly, we define the total contributions *to* all other countries (or *spill-out* contributions) as

$$T_{\bullet\leftarrow j} = \sum_{i=1, i\neq j}^{N} \phi_{i\leftarrow j} \tag{42}$$

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

$$N_j = T_{\bullet \leftarrow j} - F_{j \leftarrow \bullet}. \tag{43}$$

Finally, we define a pair of indices to address succinctly 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 within the global economy. In response to the first question, we propose the following intuitive index of *external motivation* :

$$EM_j = \frac{F_{j \leftarrow \bullet}}{ATOT_{j \leftarrow \bullet}},\ j = 1, ..., N$$

where $-1 \leq EM_j \leq 1$ expresses the relative importance and direction of spill-ins in determining the conditions in the j-th country. As $EM_j \to 1(-1)$, then conditions in group j are dominated by positive (negative) spill-ins, as opposed to direct effects. If region j receives contradictory spill-ins or if their magnitude is small in comparison to direct effects then $EM_j \to 0$. In response to the second question we follow Alter and Bayer (2014) in noting that by definition, $\sum_{j=1}^{N} N_j = 0$, and that total net connectedness across the network can be measured by

$$TNP = \sum_{j:N_j>0} N_j = \frac{1}{2}\sum_{j=1}^{N} |N_j|. \tag{44}$$

We then develop the systemic influence index by

$$SI_j = \frac{N_j}{TNP},\ j = 1, ..., N,$$

where $-1 \leq SI_j \leq 1$. For any horizon h, if the j-th region is broadly neutral within the network, with a tendency to spill-outs to roughly match spill-ins, then SI_j will be close to zero. Region j will positive (negative) net shock propagator if $0 \leq SI_j \leq 1$ ($-1 \leq SI_j \leq 0$) with a tendency to transmit positive (negative) spill-outs and/or to receive negative (positive) spill-ins. The measures of systemic influence balance across the network with the sum of the positive (negative) values equalling one (minus one). A simple network with two connected nodes would see them located at ± 1; two disconnected nodes would locate at zero. With further nodes these influences become diluted but a dominant node transmitting a positive spillover without significant feedback from other smaller nodes would see its SI_j close to one, while the other nodes clustered below zero and those falling within its sphere of influence appearing more negative.

When studying connectedness among regions, the coordinate pair (EM_j, SI_j) in dependence-influence space provides an elegant representation of regions i's role in the system. There is the tendency for points to cluster on a north-west to south east access, since positive (negative) spill-ins contribute negatively (positively) to a region's net connectedness. A region in the south-west (north-east) quadrant would be one for which negative (positive) spill-ins were outweighed by larger negative (positive) spill-outs, leading to a negative (positive) net connectedness measure.

5 Monte Carlo Simulations

We investigate the small sample properties of the 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:

$$y_{it} = \phi_i y_{i,t-1} + \pi_{i0} x_{it} + \pi_{i1} x_{i,t-1} + \phi^*_{i0} y^*_{it} + \phi^*_{i1} y^*_{i,t-1} + \pi^*_{i0} x^*_{it} + \pi^*_{i1} x^*_{i,t-1} + u_{it} \tag{45}$$

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), 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 $N = (25, 50, 75, 100)$ over $T = (50, 100, 200)$. 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 case, parameter values must satisfy Assumption 5 (the stationarity of $\boldsymbol{y}_t$) with 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.[9] Each specification is explored over $R = 1,000$ repetitions.

We consider two 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 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

$$x_{i,t} = \rho_i x_{i,t-1} + v_{i,t}, \; v_{i,t} \sim N\left(0, 1 - \rho_i^2\right), \tag{46}$$

where $\rho_i \sim U\left[0.4, 0.6\right]$, alongside heteroskedastic disturbances with $u_{it} \sim N(0, \sigma_i^2)$, with $\sigma_i^2 = 0.5 + 0.25 \times \eta_i$ 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 $= N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{j=1}^{R} (\hat{\alpha}_{ij} - \alpha_{i0})$, in Tables 1 and 2;

Average RMSE $= N^{-1} \sum_{i=1}^{N} \sqrt{R^{-1} \sum_{j=1}^{R} (\hat{\alpha}_{ij} - \alpha_{i0})^2}$, 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 instrument set of time and spatial lags of exogenous variables, $y^{**}_{i,t-1}$ and x^{**}_{it}, where

[8] The cases of $b = 4$ and 20 were also considered but produced qualitatively similar results. These results will be available upon request.

[9] 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.

$y^{**}_{i,t-1} = \boldsymbol{w}_i'\boldsymbol{W}\boldsymbol{y}_{t-1}$ and $x^{**}_{it} = \boldsymbol{w}_i'\boldsymbol{W}\boldsymbol{x}_t$ are the second spatial lags.[10] 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 the next b neighbours' neighbours as our instruments.

The initial values for each iteration were provided by the (inconsistent) ordinary least squares estimates of (45). The exogenous variables were then concentrated out and leaving an iteration over the N vector $\boldsymbol{\phi}^*$ as in (15), with estimates of the other parameters recovered by least squares regression conditional on $\hat{\boldsymbol{\phi}}^*$.[11]

Table 1 show that both the CF and the QML estimates perform reasonably well. This is very reassuring given the range of time and spatial dependence possible in (45). 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 are comparable with the control function estimator 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 (45) open more channels through which a time lagged variable may potentially impact on a y_{it}. Interestingly, there is no noticeable deterioration in the bias of either estimator even as b rises.

Table 2 indicates that both methods are reasonably robust to heteroskedasticity in the disturbances and to time dependence in the exogenous variables. Indeed time dependence in the exogenous variables improves CF estimates of contemporaneous spatial parameters, ϕ^*, due to the stronger correlation between y^{*}_{it} and the instrument x^{**}_{it}, at the expense of parameters on lagged exogenous variables, π_1, due to increased correlation between the exogenous regressors.

Comparing Tables 3 and 4 it is clear that QML is the more efficient estimator. 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 interest.

A number of interesting patters appear in Tables 5 and 6. Firstly it is clear that the CF 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 CF 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 $T = 200$.

[10] This is not the only possible instrument set, with higher spatial and/or time lags also valid and generated internally,

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

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.

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

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 due to armed violence in the aftermath of the 2003 Iraq war. The data cover monthly deaths between 2004 and 2009 (with two missing months May 2004 and March 2009) across the 18 provincial governorates of Iraq shown in Figure 1, collected by the Pentagon and released notoriously by WikiLeaks in October 2010. Deaths are classified into those of enemy or anti-coaltion insurgents, including foreign and Iraqi fighters, and civilians, the vast majority of whom were Iraqi but a minority were foreign security contractors. These data present a rare opportunity to analyse the intensity of armed violence and its spatio-temporal diffusion through the different regions of the country, if any, notwithstanding some concern over their lack of transparency in the attribution of deaths to armed violence and in distinguishing civilians from combatants.

The period is dominated by an insurgent uprising involving relatively mobile militia groups motivated by conditions across, and able to relocate within, the country. The first part of the period, beginning in the spring of 2004, is characterised by militia groups seizing effective control of cities and their surrounds and holding them until dislodged by coalition forces, whereupon they would melt away into the local population before regrouping in another sympathetic area, often in a different governorate. The first such outbreak involved members of the former ruling Ba'ath party, who had been summarily ejected from powerful roles in the state by the new coalition authority, and was centred on the city of Fallujah, in Anbar, a Sunni-dominated area, which they took and held in the face of the relatively weak institutions that prevailed in the country. Once the uprising was put-down many of these forces relocated elsewhere within the so-called Sunni triangle, whose other vertices are Baghdad and Tikrit in Saladin. This group was not alone in taking up arms, however. Mounting unrest among poorer Shia Iraqi's led to the eruption of a second branch of the insurgency, concentrated initially on the ancient city and neighbouring governorate of Najaf, before spreading to other Shia areas, and settling particularly heavily on the southern region of Basrah. During the later part of the period there was a explosion of sectarian violence, initiated by a bomb attack in Samarra, in Saladin in February 2006. This led to rival militias attacking civilians of the other group, most often in Baghdad but spreading to neighbouring regions, particularly Diyala, which had been the centre for Al Qaeda in Iraq and was becoming the centre for the group now known as Islamic State. It is important to note that violence tended to concentrate on particular cities and regions, with many others, in particular the Kurdish provinces in the north of the country which maintained strong local institutions, left relatively untouched.

Table 7 presents the number of civilian deaths across each governorate and per 1,000 inhabitants, and the number of enemy deaths in descending order.

Table 7: Civilian and Enemy casualties across governorates (2004-2009)

	Civilian				Enemy		
Rank	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	Babil	2,251	3.41	0.086	Najaf	1,064	4.44
7	Anbar	2,191	3.32	0.084	Basrah	467	1.95
8	Tamim	1,780	2.69	0.068	Babil	417	1.74
9	Wassit	887	1.34	0.034	Tamim	394	1.64
10	Kerbala	819	1.24	0.031	Wassit	265	1.10
11	Qadisiyah	468	0.71	0.018	Qadisiyah	160	0.67
12	Najaf	335	0.51	0.013	Dhi-qar	142	0.59
13	Dhi-qar	280	0.42	0.011	Karbala	120	0.50
14	Arbil	235	0.36	0.009	Maysan	51	0.21
15	Maysan	135	0.20	0.005	Arbil	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	Dohuk	40	0.06	0.002	Dohuk	5	0.02
	Others	490	0.74		Others	135	0.56
Sum		66,081	100			23,984	100
Spearman's rank correlation coefficient				0.897			
p-value				0.000			

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

There is a strong and clear correlation between the ranking of governorates by civilian and enemy deaths, with a Spearman's rank correlation coefficient close to 0.9, but it is equally clear that the proportions of each type of casualty do not match for many governorates. Baghdad, the capital governorate where roughly a quarter of population reside[12], was affected disproportionately, suffering over half of civilian deaths and experiencing civilian death toll per 1,000 inhabitants more than five times that in any other governorate, largely the result of sectarian attacks. Despite this concentration, 73% of insurgent deaths occurred outside of Baghdad, with Anbar governorate providing the highest toll of insurgent casualties but experiencing a small fraction of the civilian casualties suffered in Baghdad. It is worth noting that many provinces were relatively unaffected, particularly the predominantly Kurdish areas of Arbil and Sulaymaniyah, where the local institutions were able to maintain order virtually throughout the period, and some areas of the south, such as Maysan and Dhi-Qar, even though many of these areas share a border with governorates that saw significant casualties.

We estimate the STARDL(1, 1) model with civilian casualties as the dependent variable and enemy casualties as the independent variable across $N = 18$ governorates and $T = 70$ months of data using quasi-maximum likelihood. Modelling civilian casualties in terms of civilian and enemy casualties with both time and spatial lags has numerous advantages in light of the history of the conflict: it looks to explain casualties both in terms of coalition action against the insurgency and sectarian violence; it models the security situation as evolving through time, capturing the dynamic nature of the conflict; and, it explicitly models contagion within the country. To do so we deploy a row standardised inverse distance matrix as the weight matrix $\boldsymbol{W}$. While this choice uses distance to act as a proxy for the ease of travel between areas for armed militias it is relatively naive in ignoring the ethnic and religious partitions in the country. This makes the heterogeneity in the susceptibility of each governorate to the spillover effects from neighbours, and indeed in the direction of those effects, another attractive property of our specification.

Due to the large number of individual equations and the complicated, inter-linked and dynamic system, the STARDL coefficients themselves are not the clearest way to understand the full effects of enemy and civilian deaths in fuelling the insurgency. Instead in Figure 2 we present the dynamic multipliers for each province with respect to neighbouring civilian casualties (y^*) and insurgent deaths within (x) and in neighbouring provinces (x^*). There is a consistent pattern across all governorates that civilian casualties are increased by civilian casualties in neighbouring regions and that these are noticeably more important than insurgent casualties either within the region or its neighbours. Magnitudes vary greatly between governorates, however, with Basrah and Diyala standing out as the two regions that are particularly open to influence by civilian casualties elsewhere. This is not surprising given that later uprisings by Shia and Sunni groups centred on these two governorates. On the other hand, Arbil, a relatively peaceful predominantly Kurdish area in the north, appears well insulated after the initial shock. The dynamic multipliers attempt to match the spread and persistence of the insurgency and so it is not suprising that, in the main, the heterogeneous coefficient ϕ_{i1}, ϕ^*_{io} and ϕ^*_{i1} relating to time, spatial and time-spatial lags are positive. Not every region was affected by the insurgency, however, with some relatively peaceful areas, such as Dhi-Qar and Sulaymaniyah, sharing borders with more violent neighbours, respectively Basrah and Diyala. Whereas a homogeneous parameter model tends to predict contagion on the basis of contaminated neighbours, our heterogeneous model is sufficiently flexible to capture a range of differing time and spatial dynamics, in these cases by allowing negative estimates of ϕ^*_{io}

[12]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.

and positive estimates of ϕ^*_{i1} to cancel over the long term. Since the STARDL(1,1) contains only one time lag, the multipliers generally proceed in a predictable way to their long run values. The jagged reversions displayed by Maysan, Dihok and Karbala are the result of an estimated negative value for ϕ_{i1} (not statistically significant), reflecting the relatively low number of civilian deaths in those governorates and the sporadic nature of the violence.

The provinces display a wider pattern of response to x and x^*, measuring the effects of insurgent deaths from within and from neighbouring provinces, produced partly by the estimates of π_{i0}, π_{i1}, π^*_{io} and π^*_{i1}. For some governorates, such as An-Najaf and Basrah, the effect of insurgent deaths within the province is to reduce civilian deaths, perhaps through reduced capacity, while insurgent deaths in neighbouring regions increase them. This type of behaviour is consistent with a story of militia relocation to these provinces having suffered heavy casualties and possibly military defeat elsewhere, as was indeed the case. For other governorates, for example Baghdad, the opposite picture emerges, with insurgent deaths within the province increasing civilian deaths, while those in neighbouring provinces reduce them. This type of behaviour is consistent with a cycle of retribution fuelled by an influx of insurgents from surrounding provinces that was typical of the sectarian violence that came to dominate the later conflict, particularly in Baghdad. The long-run effects of x on y are are positive for only 13 of the 18 provinces, so that 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 of insurgent capabilities; alongside, the strategic decision to target resources or to enact retribution at other provinces where that is thought to be more effective. This latter reason shows up as the long-run impact on y of x^*, which is positive for nine provinces. Of those, perhaps the most striking is Basrah, which alongside Sala ad-Din, appears particularly open to blowback from insurgent deaths outside its locality. Together with the capital, Baghdad, the Basrah was an important military posting during the period of the war, with British forces stationed there from the initiation of the Iraq war. This oil-rich region became the focus of a Shia uprising by the so-called Mahdi Army under the leadership of Muqtada al-Sadr until control of Basrah International Airport was handed to Iraqis in January 2009. In contrast 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.

The cumulative diffusion multipliers in Figure 3 support this analysis, taking account of network as well as time effects. It is immediately noticeable that direct effects are relatively small: all are under 0.5 apart from Baghdad and the relatively peaceful Sulaymaniyah. Spill-in effects are broadly positive and can exceed direct effects over the long term, such as in An-Najaf and Al Qadisiyah. Negative spill-outs are noted in Diyala, Basrah, Dhi-Qar and Karbala, with the largest positive spill-out coming from Baghdad. 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, with relatively little influence from outside as shown by a slightly negative spill-in diffusion multiplier. Nor is it surprising that the breakdown spread out from the capital, as shown by the large and positive spill-out diffusion multiplier. The areas most open to substantially positive spill-in effects are Basrah and Diyala, suggesting that this is where the impact from Baghdad was felt. Both areas have negative, if comparatively small, spill-out diffusion multipliers, suggesting that they were good areas to engage in counter-insurgency. In 2006 Diyala was designated the capital of the Islamic caliphate that the group Islamic State hoped to establish in Iraq and soon became seen as a safe haven for Sunni insurgents. At the same time, its religious mix left it particularly open to

sectarian violence, partly in retribution to events in Baghdad.

It is of significant interest to consider the estimated network in more detail by looking at the impact of an enemy death in one governorate on civilian deaths over time in the others. Figures 4 and 5 illustrate the $N \times N$ matrices of the cumulative diffusion multipliers, $\boldsymbol{d}_x^H$ for $h = 0$ and $h = 40$ respectively, with the recipient shown on the vertical axis. Not surprisingly, the strongest instantaneous effects come from within, although spatial spillovers are not uncommon. Some provinces, such as Sulaymaniyah, Arbil and Babil, appear to react particularly quickly, but there were relatively few actual outbreaks of violence in these provinces. After 40 months, the cumulative effects of an insurgent death in a neighbouring state sometimes exceed those of a death within the state, such as in Al-Qadisiyah and Wasif. In the case of Basrah and Diyala, the cumulative effects of insurgent deaths within one province is not just to reduce civilian deaths in that province, as discussed earlier, but to reduce civilian deaths in the other. These are not adjacent provinces and the effect contrasts with that from most other provinces, particularly Baghdad and Sulaymaniyah. The transmission of violence from Baghdad is also relatively clear from Figure 5, and it is worth noting that this transmission is able to affect Basrah in the far south without having much impact on governorates such as Dhi-Qar, Maysan and Muthannia that lie in between. This pattern is broadly consistent with the precarious power balance in the country at the time, with rival groups using violence to assert relevance and influence over the nascent political institutions. Insurgent casualties in areas dominated by one militia removes the need for a rival group to carry out attacks in order to maintain relative importance. If, however, those casualties are inflicted by Kurdish forces, who are closely allied to the coalition regime, then maintaining that relative importance requires an increase in violence in other areas.

Figures 6, 7 and 8 show the external motivation / systemic influence based on the cumulative diffusion multipliers at horizons 0, 1 and 40, respectively. It is noticeable that in the short run the external motivation is initially mixed but in the long run it is universally positive, reflecting the upward trend in all spill-in diffusion multipliers in Figure 3. The predominently Kurdish areas of Arbil, Sualymaniuah and At-Ta'mim are among the least motivated from outside, reflecting their relatively strong local security institutions. They are also among the governorates with the largest positive systemic influence to increase violence elsewhere, alongside Baghdad and Anbar. Baghdad grows in systemic importance with the time horizon, overtaking Anbar, Babil and At-Ta'mim at longer time horizons, while remaining one of the least externally motivated governorates. Baghdad's positive systemic influence reflects the significance of spill-outs, in terms of increased civilian casualties, from insurgent casualties in Baghdad. Basrah has the largest negative systemic influence at all three horizons and among the most positive external motivation, reflecting the tendency for positive spill-ins and the speed to transmission to this province. Transmission to Diyala, where many attacks on civilians were carried out in retribution appears slower. The initial motivation for Diyala is negative and the province exerts little systemic influence at horizon 0 but begins to move south east from horizon 1 and appears next to Basrah in the long run. This marks out Basrah and Diayala as among the regions most easily motivated by casualties elsewhere in the country and with a large negative systemic influence suggesting a tendency for violence to spill-in to these regions in particular.

Summing up our method has had success in identifying spillovers of violence between governorates using a relatively naive weighting matrix and has sufficient flexibility to model regions with very different experience of armed violence over the period within the same framework. Our measures of network connectedness have highlighted both the most significant influencers and the most heavily influenced and the magnitude and direction of their motivation. Our analysis also

highlights the significant differences in the experience of coalition members in the aftermath of the war. Areas in the north under the control of the Iraqi Kurds were able to maintain viable security institutions and were spared significant inflows of violence against civilians from elsewhere in the country. They appear to have had significant influence over the network of violence, however, and our results suggest that civilian deaths would have risen across the country had there been more insurgent deaths in these regions. In contrast, the experience of UK forces in Basrah is one of very strong spatial spill-ins of violence from across the country and particularly from the capital, Baghdad. Baghdad itself was under the control of US forces, acting in support of fledgling Iraqi security forces and was particularly prone to time persistent violence enacted in retribution.

7 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 aftermath of the 2003 Iraq war.

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, which remains a topic of ongoing research. Eventually, this project aims to develop the general econometric 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 big datasets.

8 References

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

Alter A. and A. Beyer (2014):The dynamics of spillover effects during the European sovereign debt turmoil. *Journal of Banking & Finance* 42: 134-153.

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.

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., 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.

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 Unified 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.

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.

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.

Song, M., (2013): Asymptotic Theory for Dynamic Heterogeneous Panels with Cross-Sectional Dependence and its Applications. Columbia University 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 Monographs 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.

9 Appendix

9.1 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

$$\boldsymbol{y}_t = \sum_{\ell=0}^{\infty} \bar{\boldsymbol{B}}_\ell \bar{\boldsymbol{x}}_{t-\ell}, \tag{47}$$

where $\bar{\boldsymbol{B}}_\ell = \left[\tilde{\boldsymbol{B}}_\ell,\, \boldsymbol{B}_\ell\right]$, with typical element $\bar{b}_{\ell,ij}$, and $\bar{\boldsymbol{x}}_{t-\ell} = \left[\boldsymbol{x}'_{t-\ell},\, \tilde{\boldsymbol{u}}_{t-\ell'}\right]'$ 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 coefficients of the lag polynomial in $\bar{\boldsymbol{B}}_\ell$ are absolutely summable. Thus,

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

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. □

The following results are corollaries of the above and Assumptions 3 and 6.

Corollary 1 The process $\boldsymbol{\chi}_{it}$ has finite fourth moments for all i.

Corollary 2 The processes that are candidates for $\boldsymbol{z}_{it}$ discussed in 3.2.1 have finite fourth moments for all i.

9.2 Derivatives of (12)

The first derivatives of (12) are

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

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

The individual components of the Hessian are then

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

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

9.3 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 $\frac{1}{\bar{T}} \mathcal{L}_{\bar{T}}^c(\boldsymbol{\phi}_0^*)$; (ii) $\frac{1}{\bar{T}} \mathcal{L}_{\bar{T}}^c(\boldsymbol{\phi}_0^*) \rightarrow_p$

$\frac{1}{\bar{T}}Q^c(\boldsymbol{\phi}_0^*)$, uniformly in $\boldsymbol{\phi}_0^*$; and (iii) $\frac{1}{\bar{T}}Q^c(\boldsymbol{\phi}_0^*)$ is uniquely maximised at the true value $\tilde{\boldsymbol{\phi}}_0^*$. Point (i) is easily established by considering the two terms of $\mathcal{L}_{\bar{T}}^c(\boldsymbol{\phi}_0^*)$ that depend on $\boldsymbol{\phi}_0^*$, $\ln|S(\boldsymbol{\Phi}_0^*)|$ and $\sum_{i=1}^N \ln \frac{1}{\bar{T}} \left(\boldsymbol{y}_i - \phi_{i0}^* \boldsymbol{y}_i^*\right)' \boldsymbol{M}_i \left(\boldsymbol{y}_i - \phi_{i0}^* \boldsymbol{y}_i^*\right)$. 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

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

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

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

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 $\bar{\sigma}_i^2$ while the remainder are $o_p(1)$.

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

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

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

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

The function, $Q_{\bar{T}}^{c1}(\boldsymbol{\phi}_0^*)$ 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, $\frac{1}{\bar{T}}\left[Q_{\bar{T}}^{c1}(\boldsymbol{\phi}_0^*) - Q_{\bar{T}}^{c1}(\tilde{\boldsymbol{\phi}}_0^*)\right] \leq 0$. Assumption 7 ensures that $\bar{\sigma}_i^2(\phi_{0i}^*) > \varsigma_i^2(\phi_{0i}^*)$ for all $i = 1, \ldots, N$ and thus $\frac{1}{\bar{T}}Q_{\bar{T}}^{2c}(\boldsymbol{\phi}_0^*) < 0$.

Asymptotic Normality Having shown identifiability and consistency, we define $\boldsymbol{\xi} \equiv \left[\boldsymbol{\phi}_0^{*\prime}, \boldsymbol{\theta}', \boldsymbol{\sigma}^{2\prime}\right]'$, with $\tilde{\boldsymbol{\xi}}$ denoting the true parameter values. Since the vector function, $\frac{\partial \mathcal{L}_{\bar{T}}}{\partial \boldsymbol{\xi}}$ is 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 QML estimator then follows from normalising and rearranging:

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

Under Assumption 1, $\boldsymbol{u}_t$ is stationary with finite fourth moments so is $\boldsymbol{\chi}_t$ under Assumptions 4, 5 and 6. 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 $E\zeta_{it}^3 = \mu_i^3$ and $E\zeta_{it}^4 = \mu_i^4$. 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 $\boldsymbol{\aleph} = E\frac{1}{\bar{T}}\sum_{t=r+1}^{T}\boldsymbol{\Xi}_t\boldsymbol{\Xi}_t'$ partitioned in such a way that $\boldsymbol{\aleph}_{ij} = E\frac{1}{\bar{T}}\sum_{t=r+1}^{T}\boldsymbol{\Xi}_{it}\boldsymbol{\Xi}_{jt}'$ and with block columns $\boldsymbol{\aleph}_i = E\frac{1}{\bar{T}}\sum_{t=r+1}^{T}[\boldsymbol{\Xi}_{1t}', \ldots \boldsymbol{\Xi}_{Nt}']'\boldsymbol{\Xi}_{it}'$ Evaluating at the true parameter value, $\tilde{\boldsymbol{\xi}}$ and using 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:

$$\begin{aligned}
\frac{1}{\sqrt{\bar{T}}}\frac{\partial \mathcal{L}_{\bar{T}}(\tilde{\boldsymbol{\xi}})}{\partial \phi_{i0}^*} &= \frac{1}{\sqrt{\bar{T}}}\sum_{t=r+1}^{T}\left\{\boldsymbol{g}_i[\boldsymbol{\Theta}\boldsymbol{\Xi}_t + \boldsymbol{\zeta}_t]\zeta_{it} - g_{ii}\right\}, \text{ so that} \\
\frac{1}{\sqrt{\bar{T}}}\frac{\partial \mathcal{L}_{\bar{T}}(\tilde{\boldsymbol{\xi}})}{\partial \boldsymbol{\phi}_0^*} &= \frac{1}{\sqrt{\bar{T}}}\sum_{t=r+1}^{T}\left\{\boldsymbol{G}_0[\boldsymbol{\Theta}\boldsymbol{\Xi}_t + \boldsymbol{\zeta}_t] \odot \boldsymbol{\zeta}_t - \text{vecd}(\boldsymbol{G}_0)\right\} \\
\frac{1}{\sqrt{\bar{T}}}\frac{\partial \mathcal{L}_{\bar{T}}}{\partial \boldsymbol{\theta}_i} &= -\frac{1}{\sqrt{\bar{T}}}\sum_{t=r+1}^{T}\boldsymbol{\Xi}_{it}\zeta_{it}, \\
\frac{1}{\sqrt{\bar{T}}}\frac{\partial \mathcal{L}_{\bar{T}}}{\partial \sigma_i^2} &= \frac{1}{\sqrt{\bar{T}}}\frac{1}{2\sigma_i^2}\sum_{t=r+1}^{T}(\zeta_{it}^2 - 1).
\end{aligned}$$

The assumptions relating to $\boldsymbol{X_t}$ and the absence of serial correlation in $\boldsymbol{u}_t$ ensure that $E\boldsymbol{\Xi}_t\boldsymbol{\zeta}_t' = 0$ such 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 expectations (see Davidson and Mackinnon (1993, p292)), we have: $E\left\{\sum_{t=r+1}^{T}\frac{\partial \mathcal{L}_t}{\partial \boldsymbol{\xi}}\sum_{s=r+1}^{T}\frac{\partial \mathcal{L}_s}{\partial \boldsymbol{\xi}'}\right\} = 0$ 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 $r = 1, 2, 3$.

$$
\begin{aligned}
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^3 g_{ii}\boldsymbol{g}_i'\boldsymbol{\Theta}E\left\{\frac{1}{\bar{T}}\sum_{t=r+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}_{j0}^{*\prime}}\right\} &= g_{ij}g_{ji},\ i \neq j,\\
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^3 g_{ii}E\left\{\frac{1}{\bar{T}}\sum_{t=r+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=r+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=r+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).
\end{aligned}
$$

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

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

with all other elements null.

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

9.4 Proof of Theorem 2

We make use of the following expression for the inverse of a partitioned matrix, e.g. Horn and Johnson (1985, p18),[13]

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} & -\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} A_{12}A_{22}^{-1} \\ -\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} A_{21}A_{11}^{-1} & \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \end{bmatrix}.$$

When applied to the least square formula for equation (21) we have that

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_i \\ \hat{\rho}_i \end{pmatrix} = \begin{bmatrix} \left(\boldsymbol{K}_i'\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}\boldsymbol{K}_i\right)^{-1} & -\left(\boldsymbol{K}_i'\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}\boldsymbol{K}_i\right)^{-1}\boldsymbol{K}_i'\hat{\boldsymbol{v}}_i\left(\hat{\boldsymbol{v}}_i'\hat{\boldsymbol{v}}_i\right)^{-1} \\ -\left(\hat{\boldsymbol{v}}_i'\boldsymbol{M}_{\boldsymbol{K}_i}\hat{\boldsymbol{v}}_i\right)^{-1}\hat{\boldsymbol{v}}_i'\boldsymbol{K}_i\left(\boldsymbol{K}_i'\boldsymbol{K}_i\right)^{-1} & \left(\hat{\boldsymbol{v}}_i'\boldsymbol{M}_{\boldsymbol{K}_i}\hat{\boldsymbol{v}}_i\right)^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{K}_i'\boldsymbol{y}_i \\ \hat{\boldsymbol{v}}_i'\boldsymbol{y}_i \end{bmatrix},$$

with the result that

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_i \\ \hat{\rho}_i \end{pmatrix} = \begin{bmatrix} \left(\boldsymbol{K}_i'\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}\boldsymbol{K}_i\right)^{-1}\boldsymbol{K}_i'\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}\boldsymbol{y}_i \\ \left(\hat{\boldsymbol{v}}_i'\boldsymbol{M}_{\boldsymbol{K}_i}\hat{\boldsymbol{v}}_i\right)^{-1}\hat{\boldsymbol{v}}_i'\boldsymbol{M}_{\boldsymbol{K}_i}\boldsymbol{y}_i \end{bmatrix}, \tag{48}$$

where $\boldsymbol{K}_i = (\boldsymbol{y}_i^*, \boldsymbol{\chi}_i)$, $\boldsymbol{y}_i^* = \left(y_{i,r+1}^*, y_{i,r+1}^*, \ldots, y_{i,T}^*\right)'$, $\boldsymbol{\chi}_i' = \left(\boldsymbol{\chi}_{i,r+1}', \boldsymbol{\chi}_{i,r+1}', \ldots, \boldsymbol{\chi}_{i,T}'\right)'$, $\hat{\boldsymbol{v}}_i^* = \left(\hat{v}_{i,r+1}^*, \hat{v}_{i,r+1}^*, \ldots, \hat{v}_{i,T}^*\right)'$ and 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 $\boldsymbol{A}$. denotes the $(t-r)$'th row of the matrix $\tilde{\boldsymbol{X}}_i = \left(\tilde{\boldsymbol{X}}_{i,r+1}', ..., \tilde{\boldsymbol{X}}_{i,T}'\right)'$, Note that the annihilation matrix for $\hat{\boldsymbol{v}}_i$, $[\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}] = [\boldsymbol{I} - \boldsymbol{M}_{\boldsymbol{z}i}\boldsymbol{y}_i^*(\boldsymbol{y}_i^{*\prime}\boldsymbol{M}_{\boldsymbol{z}i}\boldsymbol{y}_i^*)^{-1}\boldsymbol{y}_i^{*\prime}\boldsymbol{M}_{\boldsymbol{z}i}]$. Since $\boldsymbol{\chi}_i$ lies in the column space of $\boldsymbol{z}_i$, $[\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}]\boldsymbol{\chi}_i = \boldsymbol{\chi}_i$, whereas

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

the fitted values from regression (18). Hence $\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}\boldsymbol{K}_i = \tilde{\boldsymbol{X}}_i$. Substituting equation (1) into the top of (48) and rearranging

$$\sqrt{\bar{T}}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\right) = \left(\frac{1}{\bar{T}}\boldsymbol{K}_i'\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}\boldsymbol{K}_i\right)^{-1}\frac{1}{\sqrt{\bar{T}}}\boldsymbol{K}_i'\boldsymbol{M}_{\hat{\boldsymbol{v}}_i}\boldsymbol{u}_i = \left(\frac{1}{\bar{T}}\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right)^{-1}\frac{1}{\sqrt{\bar{T}}}\tilde{\boldsymbol{X}}_i'\boldsymbol{u}_i. \tag{49}$$

Following Assumptions 1, 6 and 7' and the corollaries to Lemma 1,

$$\text{plim } \bar{T}^{-1}\sum_{t=r+1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}' = \boldsymbol{C}_1, \text{plim } \bar{T}^{-1}\sum_{t=r+1}^{T}\boldsymbol{z}_{it}y_{it}^* = \boldsymbol{c}_2, \text{plim } \bar{T}^{-1}\sum_{t=r+1}^{T}y_{it}^{*2} = c_3,$$

where $\boldsymbol{C}_1$ is a finite $L \times L$ positive definite matrix, $\boldsymbol{c}_2$ a finite L vector and c_3 is a positive constant, in the case that $\boldsymbol{z}_{it}$ consists of time and or spatial lags of $\boldsymbol{\chi}_{it}$. If external variables are used in the control regression, however, then further assumptions are needed on their moments in place of Corollary 2. The sequence $\tilde{\boldsymbol{x}}_{i,t}u_{i,t}$ is a stationary, ergodic martingale difference sequence with finite fourth moments and denoting

$$\text{plim } \bar{T}^{-1}\sum_{t=r+1}^{\bar{T}}\tilde{\boldsymbol{x}}_{i,t}\tilde{\boldsymbol{x}}_{i,t}' = \boldsymbol{C}_4,$$

[13]The top line here is a translation of the bottom given in the text.

where $\boldsymbol{C}_4$ is a finite $(L_1+1)\times(L_1+1)$. Using Chebyshev's inequality it follows that

$$\text{plim } \bar{T}^{-1}\sum_{t=r+1}^{T}\tilde{\boldsymbol{x}}_{i,t}\tilde{\boldsymbol{x}}_{i,t}'u_{it}^2=\boldsymbol{C}_4 Eu_{it}^2=\sigma_i^2\boldsymbol{C}_4.$$

Then following White (1984) Corollary 5.26

$$\bar{T}^{-1/2}\sum_{t=r+1}^{\bar{T}}\tilde{\boldsymbol{x}}_{i,t}u_{it}\rightarrow_d N\left(0,\sigma_i^2\boldsymbol{C}_4\right).$$

Since $\hat{\boldsymbol{\beta}}_i\rightarrow_p\boldsymbol{\beta}_i$, and $\hat{v}_{it}\rightarrow_p v_{it}$ the pseudo-residuals from (21), $\hat{u}_{it}=\hat{e}_{it}+\hat{v}_{it}\hat{\rho}$ can be used to provide a consistent estimator of σ_i^2, $\hat{\sigma}_i^2=\bar{T}^{-1}\sum_{t=r+1}^{T}\hat{u}_{it}^2$. The consistency and asymptotic normality of the instrumental variables estimator then follow from White (1984, 5.27). □

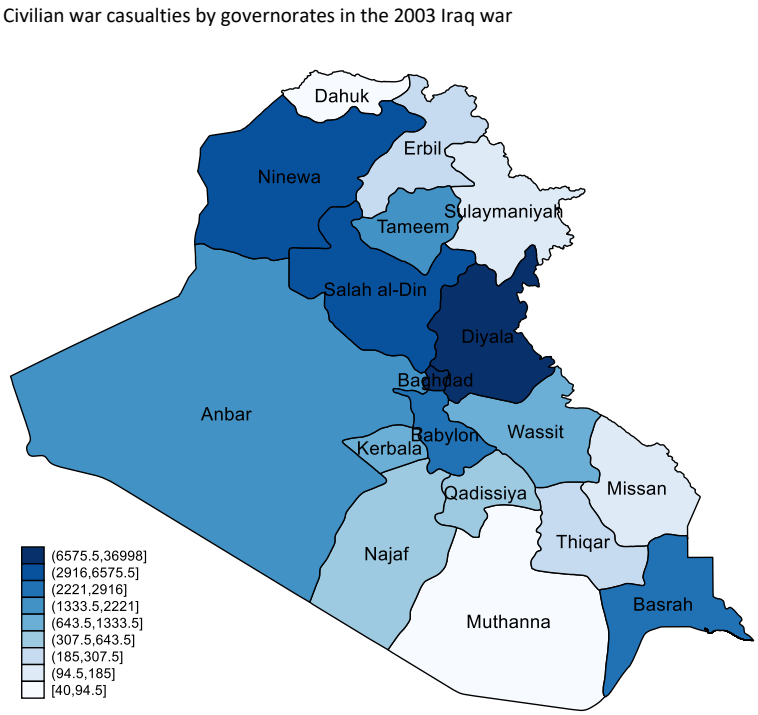

Figure 1: The governorates of Iraq.

Figure 2: Dynamic Multipliers: y^* (dotted); x (solid) and x^* (dashed).

Figure 3: Connectedness measures: direct (solid); spill-in (dashed); spill-out (dash/dotted); net (dotted).

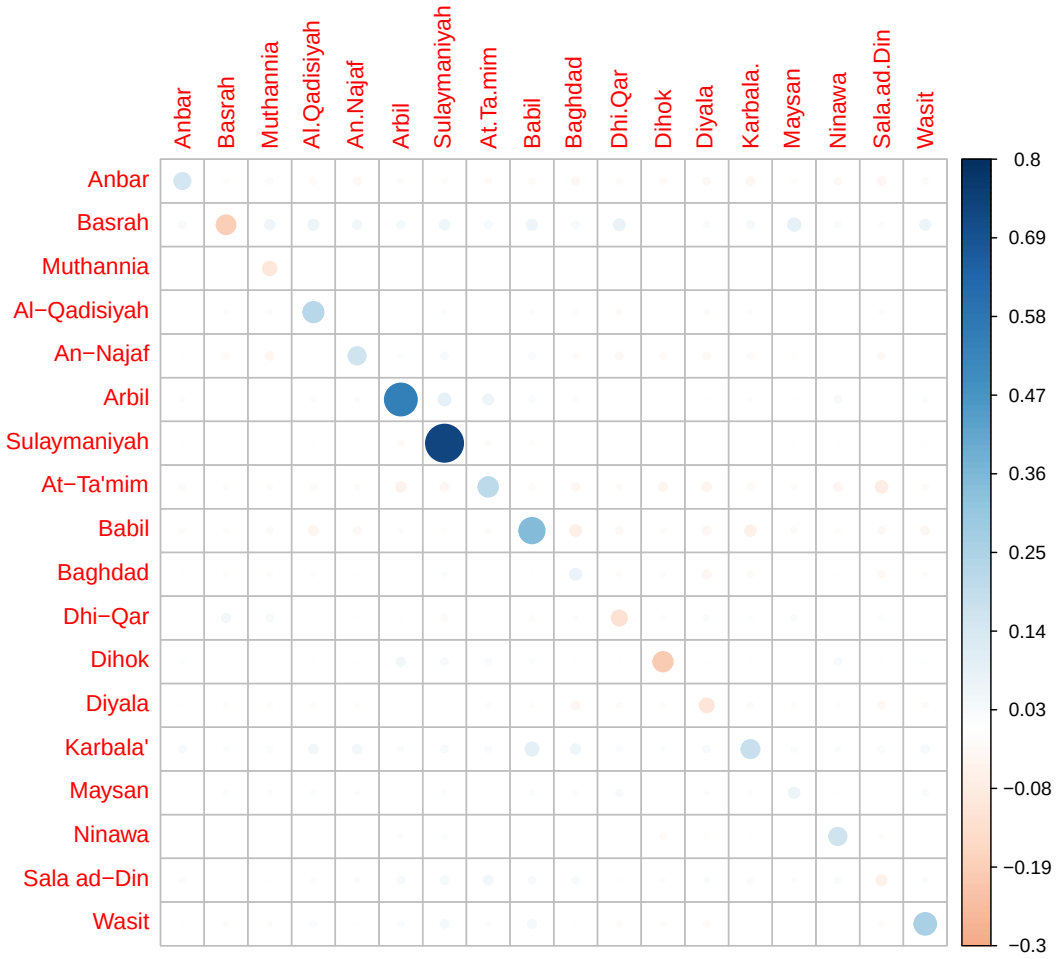

Figure 4: Heatmap at horizon 0.

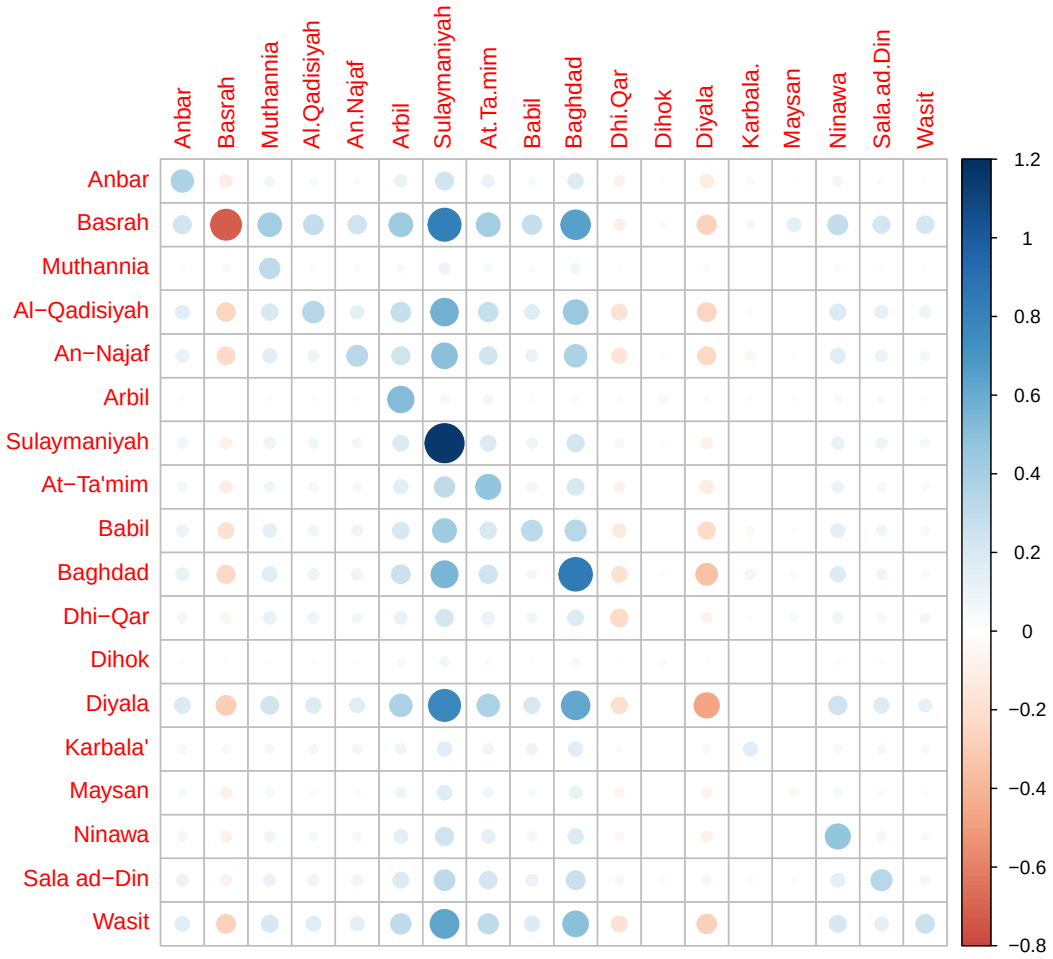

Figure 5: Cumulative heatmap at horizon 40.

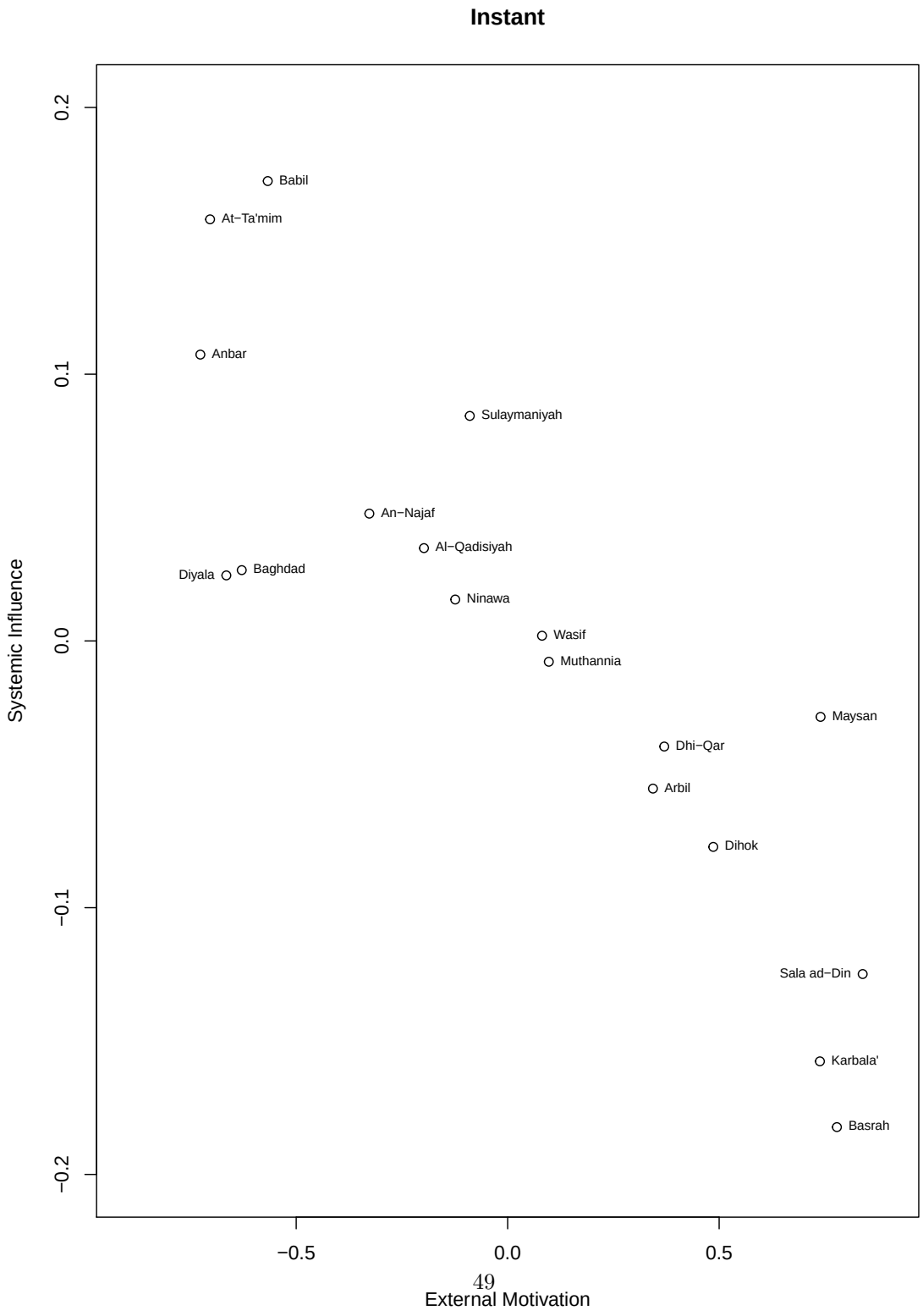

Figure 6: Network analysis at horizon 0.

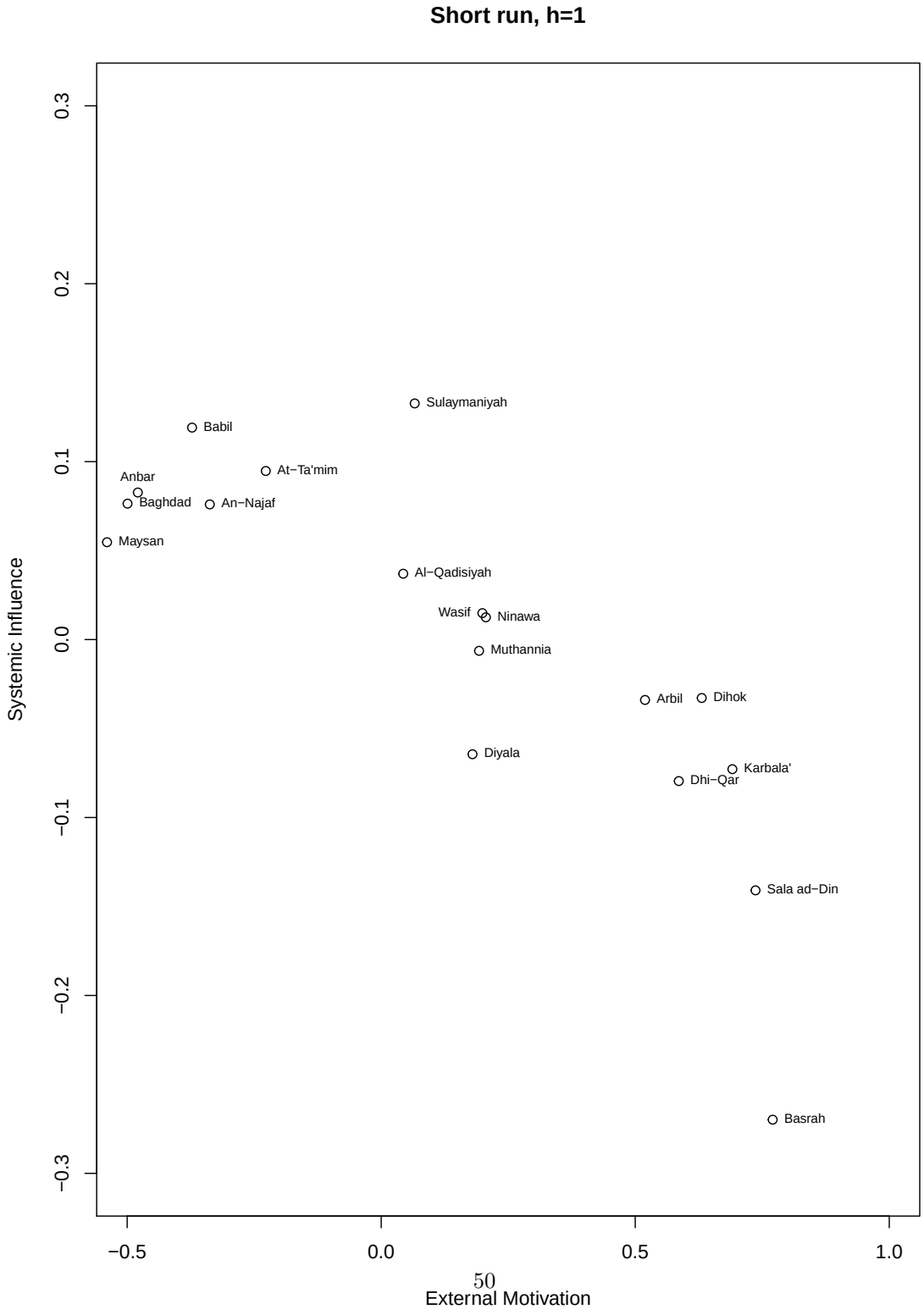

Figure 7: Network analysis at horizon 1.

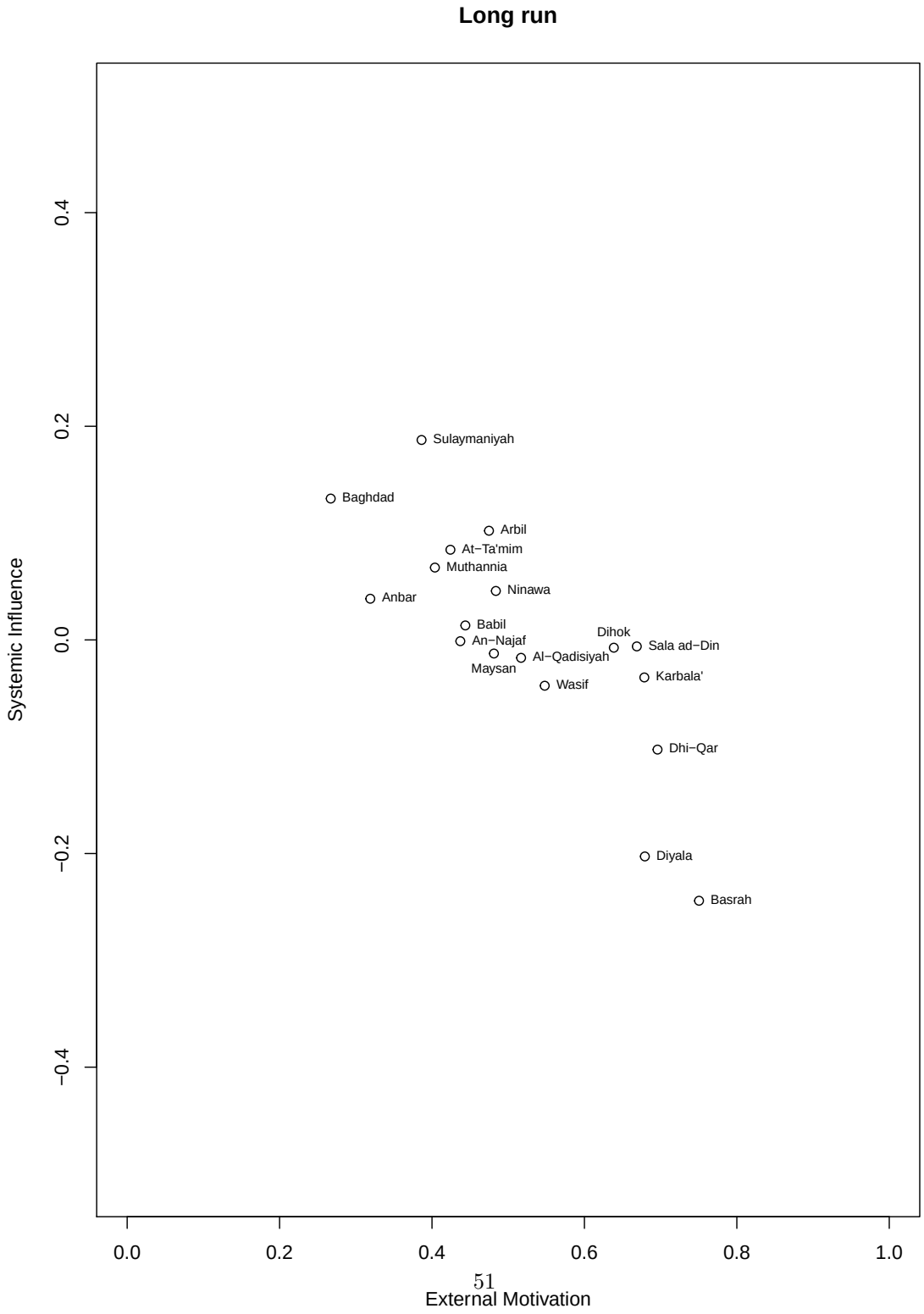

Figure 8: Network analysis at horizon 40.