UNIVERSITAT
ROVIRA I VIRGILI
DEPARTAMENT D'ECONOMIA

CENTRE DE RECERCA EN ECONOMIA INDUSTRIAL I ECONOMIA PÚBLICA

WORKING PAPERS

Col·lecció "DOCUMENTS DE TREBALL DEL DEPARTAMENT D'ECONOMIA - CREIP"

A correlated random effects spatial Durbin model

Karen Miranda
Óscar Martínez-Ibáñez
Miguel Manjón Antolín

Document de treball n.23- 2016

DEPARTAMENT D'ECONOMIA – CREIP
Facultat d'Economia i Empresa

UNIVERSITAT
ROVIRA I VIRGILI
DEPARTAMENT D'ECONOMIA

creip
CENTRE DE RECERCA EN ECONOMIA INDUSTRIAL I ECONOMIA PÚBLICA

Edita:

Departament d'Economia
www.fcee.urv.es/departaments/economia/public_html/index.html
Universitat Rovira i Virgili
Facultat d'Economia i Empresa
Av. de la Universitat, 1
43204 Reus
Tel.: +34 977 759 811
Fax: +34 977 758 907
Email: sde@urv.cat

CREIP
www.urv.cat/creip
Universitat Rovira i Virgili
Departament d'Economia
Av. de la Universitat, 1
43204 Reus
Tel.: +34 977 758 936
Email: creip@urv.cat

Adreçar comentaris al Departament d'Economia / CREIP

ISSN edició en paper: 1576 - 3382
ISSN edició electrònica: 1988 - 0820

DEPARTAMENT D'ECONOMIA – CREIP
Facultat d'Economia i Empresa

A correlated random effects spatial Durbin model*

K. Miranda, O. Martínez-Ibañez, and M. Manjón-Antolín

QURE-CREIP Department of Economics, Rovira i Virgili University

Abstract

We consider a correlated random effects specification of the spatial Durbin (dynamic) panel model with an error-term containing individual effects and their spatial spillovers. We derive the likelihood function of the model and the asymptotic properties of the quasi-maximum likelihood estimator. We also provide illustrative evidence from a growth-initial level equation and the country dataset analysed by Lee and Yu (2016). While largely replicating their estimates, our results indicate the existence of spatial contagion in the individual effects. In particular, estimated spill-in/out effects reveal the existence of groups of countries with common patterns in their spillovers.

Keywords: correlated random effects, Durbin model, spatial dynamic panel data

JEL Classification: C23

*This research was funded by grants ECO2017-88888-P and ECO2014-55553-P (Ministerio de Economía y Competitividad) and grant 2014FI_B00301 (Agaur, Generalitat de Catalunya). We thank participants at the 57^{th} ERSA Congress (University of Groningen, 2017), 11^{th} World Conference of the Spatial Econometrics Association (Singapore Management University, 2017), AQR PhD-seminar of the Faculty of Economics and Business (University of Barcelona, 2017), and 9^{th} Jean Paelinck Seminar (Technical University of Cartagena, 2017) for helpful comments. We are also grateful to B. Baltagi, P. Elhorst, Z. Yang and E. Lopez-Bazo for their comments on earlier versions of this paper. Usual caveats apply.

1 Introduction

The spatial Durbin model is a widely used specification in cross-section studies using georeferenced data (LeSage and Pace, 2009; LeSage, 2014). However, its use appears to be more limited with panel data. Although it has a certain appeal as a general framework to analyse spatial relations, concerns have been raised about its estimation and identification, particularly in its dynamic version (Elhorst et al., 2010; Elhorst, 2012). Despite these concerns, the spatial Durbin dynamic panel model (or, simply, dynamic spatial Durbin model) is expected to gain popularity in applied work, since identification conditions and Monte Carlo evidence for 2-Stage Least Squares (2SLS) and Quasi Maximum Likelihood (QML) estimators have recently been provided by Lee and Yu (2016). It is also interesting to note that Yu et al. (2008) and Su and Yang (2015) have analysed the asymptotic properties of QML estimators in restricted versions of the model specification analysed by Lee and Yu (2016).

In this paper we consider a correlated random effects specification (Mundlak, 1978; Chamberlain, 1982) of the spatial Durbin (dynamic) panel model and, following Yu et al. (2008) and Su and Yang (2015), derive the likelihood function of the model and proof that the QML estimator is consistent and asymptotically normal. To be precise, our model specification corresponds to a restricted version of the dynamic spatial Durbin model of Lee and Yu (2016), since we do not include the spatial lag of the lagged dependent variable among the regressors.[1] This means that, in terms of spatial dependence, our model specification lies somewhere in between that of Yu et al. (2008), who only consider the spatial lag of the dependent and lagged-dependent variables, and that of Su and Yang (2015, p. 231), in which "spatial dependence is present only in the error term". A major difference with respect to these papers is that while they consider a rather general variance-covariance matrix of the error term (which may contain individual and/or time effects), we consider an error-components structure with individual effects and their spatial spillovers (time effects can easily be incorporated), which results in a specific albeit involved variance-covariance matrix (see also Kapoor et al., 2007). Our proofs,

[1]See e.g. Elhorst (2012) for an overview of empirical studies using this model specification. Notice that the inclusion of the spatial lag of the lagged dependent variable would not make a substantial difference in proving the asymptotic properties of the QML estimator (other than complicate it).

however, are derived under rather standard assumptions in the spatial econometrics literature.[2]

Our model specification is inspired by the work of Beer and Riedl (2012), who advocate using an extension of the spatial Durbin model for panel data that controls for both the individual effects and the spatially weighted individual effects (see also Miranda et al., 2017). Ultimately, however, they argue that "it is (...) advisable to remove the spatial lag of the fixed effects from the equation as the inclusion of both, [the individual effects] and [their spatial spillovers], leads to perfect multicollinearity" (p. 302). Removing the spatial lag of the fixed effects does not generally preclude the consistent estimation of the parameters of the model. However, this practice rules out obtaining an estimate of the individual-specific effects (net of the spatially weighted effects), which can be critical in certain applications. This is the case, for example, in growth models, where a measure of the unobserved productivity of the geographical units under study can be obtained from the estimated individual effects (Islam, 1995). Distinguishing the individual effects from their spatial spillovers can thus provide interesting insights into how the unobserved characteristics of the neighbouring territories affect the output of a certain territory and, conversely, how the unobserved characteristics of a territory affect the output of the neighbouring territories.

To illustrate this point, we estimate a growth-initial level equation using OECD data from Lee and Yu (2016). Unlike previous studies (e.g., Yu and Lee, 2012; Ho et al., 2013), however, our model specification not only accounts for observable "technological interdependences" (à la Ertur and Koch 2007) but also for unobserved ones (through the spatial spillovers of the individual effects). Interestingly, our estimated coefficients and standard errors largely replicate those reported by Lee and Yu (2016). This means that, since the spatial autoregressive parameter is not statistically significant, "the role played by technological interdependence on the growth of [OECD] countries" may not be as important as previously thought (Ertur and Koch 2007, p. 1052; see also Elhorst et al. 2010). In contrast, our results point to the existence of "unobservable technological interdependences" (i.e., spatial contagion in the – weakly significant – individual effects). Following Islam (1995), this may be interpreted as evidence that the growth of some countries is partially explained by the impact that the (unobserved productivity)

[2]See e.g. Kelejian and Prucha (1998, 2001); Lee (2004); Yu et al. (2008) and Su and Yang (2015).

of the neighbouring countries have on their economies. Lastly, computation of the "spill-in" and "spill-out" effects of the individual effects indicate that countries that impact less/more on other countries tend to be those that are less/more affected by the spillovers from their neighbours (Debarsy et al., 2012; LeSage and Chih, 2016). Further, they tend to have larger/smaller individual effects.

The rest of the paper is organised as follows. In Section 2 we present the model. In Section 3 we discuss its estimation by QML and derive the asymptotic properties of the QML estimator. In Section 4 we provide illustrative evidence. Section 5 concludes.

2 Model specification

In this paper we are interested in the following dynamic spatial autoregressive model with spatially weighted regressors and spatially weighted fixed effects:

$$Y_{nt} = \rho_0 Y_{n,t-1} + \lambda_0 W_n Y_{nt} + X_{nt}\beta_{10} + W_n X_{nt}\beta_{20} + \mu_n + W_n \alpha_n + \varepsilon_{nt} \tag{2.1}$$

where the subindex 0 denotes the "true" parameters of the model (e.g, ρ_0, λ_0, β_{10} and β_{20}), $Y_{nt} = (y_{1t}, y_{2t}, \cdots, y_{nt})'$ is an $n-$dimensional vector of dependent variables at time t, W_n is the exogenous spatial weight matrix that describes the spatial arrangement of the units in the sample, $X_{nt} = (x'_{1t}, x'_{2t}, \cdots, x'_{nt})'$ is a $n \times K$ matrix of regressors (i.e., x_{it} is a row vector of $1 \times K$), and ε_{nt} is the $n-$dimensional vector of disturbances at time t, with $\varepsilon_{nt} \sim (0, \sigma^2_\varepsilon)$, whose stochastic properties are discussed below. We assume, without loss of generality, that data is available for $i = 1, \ldots, n$ spatial units and $t = 1, \ldots, T$ time periods.[3]

Notice that this model specification critically differs from alternative specifications of the spatial Durbin dynamic panel data model (see e.g. Elhorst 2012) in that it includes both the individual effects (μ_n) and their spatial counterparts (α_n). Although the inclusion of

[3] Dealing with a "complete panel" is just meant to simplify notation and the burden of some proofs. Our results can easily be extended to incomplete panels. Notice similarly that the model does not contain time effects but these can easily be incorporated into the model (by e.g. including time dummies among the regressors, as we illustrate in the empirical application of Section 4).

$W_n Y_{nt}$ in the right-hand side of 2.1 produces "global" spatial contagion (Anselin, 2003) in the individual effects, our interest here lies in the existence of "local" spatial contagion. In particular, the individual-specific effects and their spatially weighted counterparts need to be estimated in order to determine which units are "locally" affecting and which units are "locally" affected, respectively, by the spatial spillover of the individual effect, and how intense such a "local" spillover is with respect to the total effect (i.e., the partial derivative of the conditional expectation of the dependent variable with respect to the individual effect). We discuss this issue in detail below, but first it is important to notice that this is generally not possible because 2.1 is observationally equivalent to a model that only includes individual effects (Beer and Riedl, 2012).

In this paper we follow Miranda et al. (2017) in using a correlated random effects specification to identify the local spatial contagion in the individual effects. This means making use of the following correlation functions (Mundlak, 1978):

$$\begin{aligned} \mu_n &= l_n c_0 + \overline{X}_n \pi_{\mu_0} + \upsilon_{n\mu} \\ \alpha_n &= \overline{X}_n \pi_{\alpha_0} + \upsilon_{n\alpha}, \end{aligned} \tag{2.2}$$

where $\overline{X}_n = (\overline{X}'_{1\cdot}, \overline{X}'_{2\cdot}, \ldots, \overline{X}'_{n\cdot})'$ are composed of the period-means of the regressors, $\overline{X}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, π_{μ_0} and π_{α_0} are $K \times 1$ ("true") parameter vectors, l_n is the unit vector of dimension $n \times 1$, and c_0 is the constant term to be estimated. The error terms, $\upsilon_{n\mu}$ and $\upsilon_{n\alpha}$, are assumed to be random vectors of dimension n, with $\upsilon_{n\mu} \sim (0, \sigma^2_{\mu_0} I_n)$ and $\upsilon_{n\alpha} \sim (0, \sigma^2_{\alpha_0} I_n)$, uncorrelated with ε_{nt}. Notice, however, that $\upsilon_{n\mu}$ and $\upsilon_{n\alpha}$ are not assumed to be independent, the covariance, $\sigma_{\mu\alpha_0}$, being such that $E(\upsilon_{n\mu}\upsilon'_{n\alpha}) = \sigma_{\mu\alpha_0} I_n$ with E denoting the mathematical expectation. Notice also that although we assume that the correlation functions are linear and have the means of the regressors as their main component, this does not always need to be the case. Non-linear functions, different moments and/or other variables may be employed to construct the correlation functions (Chamberlain, 1984). For the sake of simplicity, however, in this paper we restrict the analysis to the linear-means case.

Plugging equations 2.2 into model 2.1 we obtain

$$Y_{nt} = l_n c_0 + \rho_0 Y_{n,t-1} + \lambda_0 W_n Y_{nt} + X_{nt}\beta_{10} + W_n X_{nt}\beta_{20} + \overline{X}_n \pi_{\mu 0} + W_n \overline{X}_n \pi_{\alpha 0} + \eta_{nt} \tag{2.3}$$

where $\eta_{nt} = \upsilon_{n\mu} + W_n \upsilon_{n\alpha} + \varepsilon_{nt} = V_n + \varepsilon_{nt}$ (see also Kapoor et al., 2007). Notice that the variance-covariance matrix of this error term is given by $E[\eta_{nt}\eta'_{nt}] = E[V_n V'_n] + \sigma^2_{\varepsilon_0} I_n$, where $E[V_n V'_n] = \sigma^2_{\mu_0} I_n + \sigma_{\mu\alpha_0}(W_n + W'_n) + \sigma^2_{\alpha_0} W_n W'_n$ is the variance-covariance matrix of the composed error term of the individual effects and their spatial spillovers, V_n. Thus, if we define $\Sigma_0 = \frac{1}{\sigma^2_{\varepsilon_0}}\left(\sigma^2_{\mu_0} I_n + \sigma_{\mu\alpha_0}(W_n + W'_n) + \sigma^2_{\alpha_0} W_n W'_n\right)$, then the variance-covariance matrix of the error term can be rewritten as $E[\eta_{nt}\eta'_{nt}] = \sigma^2_{\varepsilon_0}(\Sigma_0 + I_n)$.

It is also worth noting the alternative specifications that are nested in our error term structure. The most obvious, perhaps, is the standard "random effects" (without spatial contagion), which is derived from our model by imposing the constraints $\pi_{\mu_0} = \pi_{\alpha_0} = 0$, $\sigma^2_{\alpha_0} = 0$ and $\sigma^2_{\mu 0} \neq 0$ (see e.g. Mundlak 1978 and Chamberlain 1982). Notice, however, that we may alternatively consider a "random effects" specification with spatial contagion by imposing the constraints $\pi_{\mu_0} = \pi_{\alpha_0} = 0$, $\sigma^2_{\alpha_0} \neq 0$ and $\sigma^2_{\mu_0} \neq 0$ and $\sigma_{\mu\alpha_0} \neq 0$ and, as a particular case, a "random effects" specification with proportional spatial contagion by imposing the constraints $\pi_{\mu_0} = \pi_{\alpha_0} = 0$, $\sigma^2_{\mu_0} \neq 0$, $\sigma^2_{\alpha_0} = a^2\sigma^2_{\mu_0}$ and $\sigma_{\mu\alpha_0} = a\sigma^2_{\mu_0}$ (or simply $\pi_{\mu_0} = \pi_{\alpha_0} = 0$ and $\alpha_n = a\mu_n$), with $a \neq 0$ constant. These, in turn, can be seen as a simplified version of the error structure proposed by Kapoor et al. (2007). Interestingly, however, our model also covers "fixed effects" versions of the previously discussed structures ("fixed" in the sense of being correlated with – some of – the regressors). That is, by imposing alternative constraints we may derive: *i)* a "fixed effects" error term ($\pi_{\mu_0} \neq 0$, $\pi_{\alpha_0} = 0$, $\sigma^2_{\alpha_0} = 0$ and $\sigma^2_{\mu_0} \neq 0$) analogous to that discussed by Mundlak (1978) and Chamberlain (1982), $\overline{X}_n \pi_{\mu_0} + \upsilon_{n\mu}$; *ii)* a "fixed effects" error term with spatial contagion ($\pi_{\mu_0} \neq 0$, $\pi_{\alpha_0} \neq 0$, $\sigma^2_{\alpha_0} \neq 0$ and $\sigma^2_{\mu_0} \neq 0$) and, if we impose that $\sigma^2_{\alpha_0} = 0$, a fixed effect error term analogous to that discussed by Debarsy (2012), $\overline{X}_n \pi_{\mu_0} + W_n \overline{X}_n \pi_{\alpha 0} + \upsilon_{n\mu}$, in which we cannot guarantee the existence of spatial contagion in the individual effects[4]; and *iii)* a "fixed effects" error term with proportional spatial contagion ($\pi_{\alpha_0} = a\pi_{\mu_0} \neq 0$, $\sigma^2_{\mu 0} \neq 0$,

[4]Except if we impose, as we do, that the direct effect of the individual effects of a unit (see below) only depends on the characteristics of that unit and not on those of the other units.

$\sigma^2_{\alpha_0} = a^2\sigma^2_{\mu_0}$ and $\sigma_{\mu\alpha_0} = a\sigma^2_{\mu_0}$, with $a \neq 0$ constant; or, simply, $\pi_{\mu_0} \neq 0$ and $\alpha_n = a\mu_n$).

2.1 Marginal effects: spatial spillovers and diffusion effects

Thus, providing that an estimate of μ_n and α_n is available, our model specification allows us to consider the existence of both "local" and "global" (through λ_0) spatial contagion in the individual effects (Anselin, 2003). However, because of the presence of the dynamic term $Y_{n,t-1}$ in the model, we may also consider the existence of "diffusion effects" in the partial derivative of the (conditional expectation of the) dependent variable with respect to the individual effects (Debarsy et al., 2012). To see this, let us rewrite the model in 2.1 as (by repeated substitution):

$$Y_{nt} = \rho_0^t S_0^{-t} Y_{n,0} + \sum_{s=0}^{t-1} \rho_0^s S_0^{-(s+1)} \left[X_{n,t-s}\beta_{10} + W_n X_{n,t-s}\beta_{20} + \mu_n + W_n\alpha_n + \varepsilon_{n,t-s}\right]$$

where $S_0 = I_n - \lambda_0 W_n = S_n(\lambda_0)$.[5] In full matrix form:

$$\mathbf{Y} = \mathbf{G}_0 Y_{n,0} + \mathbf{C}_0\mathbf{X}\beta_{10} + \mathbf{C}_0\mathbf{W}\mathbf{X}\beta_{20} + \mathbf{C}_0(l_T \otimes I_n)\mu_n + \mathbf{C}_0\mathbf{W}(l_T \otimes I_n)\alpha_n + \mathbf{C}_0\boldsymbol{\varepsilon} \tag{2.4}$$

with $\mathbf{Y} = (Y'_{n1}, Y'_{n2}, \ldots, Y'_{nT})'$, $\mathbf{X} = (X'_{n1}, \ldots, X'_{nT})'$, $\boldsymbol{\varepsilon} = (\varepsilon'_{n1}, \ldots, \varepsilon'_{nT})'$, $\mathbf{W} = I_T \otimes W_n$, $\mathbf{G}_0 = \left(\rho_0(S_0^{-1})', \rho_0^2(S_0^{-2})', \ldots, \rho_0^T(S_0^{-T})'\right)'$ and

$$\mathbf{C}_0 = \begin{pmatrix} S_0^{-1} & 0 & 0 & \cdots & 0 \\ \rho_0 S_0^{-2} & S_0^{-1} & 0 & \cdots & 0 \\ \rho_0^2 S_0^{-3} & \rho_0 S_0^{-2} & S_0^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_0^{T-1} S_0^{-T} & \rho_0^{T-2} S_0^{-(T-1)} & \rho_0^{T-3} S_0^{-(T-2)} & \cdots & S_0^{-1} \end{pmatrix}$$

[5] We denote matrices and vectors depending on parameters of the model with the name of the matrix and vector, respectively, followed by the parameter(s) in brackets. For example, $S(\lambda) = S_n(\lambda)$. In particular, in the case of the "true" parameters we simply add the subindex zero to the name of the matrix. Thus, $S_0 = S_n(\lambda_0)$. Notice also that we use bold letters to denote $n \times T$ matrices (and similarly for $nT \times 1$ vectors), i.e., matrices resulting from stacking $n-$dimensional matrices. For example, $\mathbf{Y} = (Y'_{n1}, Y'_{n2}, \ldots, Y'_{nT})'$ and $\mathbf{X} = (X'_{n1}, \ldots, X'_{nT})'$, but also $\mathbf{S}(\lambda) = I_{nT} - \lambda(I_T \otimes W_n)$ and $\mathbf{S}_0 = I_T \otimes S_0$.

Lastly, let e_j be the j-th column of $l_T \otimes I_n$ with $j = 1, \ldots, n$. The marginal effects of the individual-specific effects are:

$$\frac{\partial}{\partial e_j} E(\mathbf{Y}|\mathbf{X}) = \mathbf{C}_0 \left[I_T \otimes (I_n \mu_j + W_n \alpha_j) \right] \tag{2.5}$$

where the diagonal elements of this matrix represent the *direct marginal effects* of unit j and the off-diagonal elements of this matrix represent the *spillovers or indirect marginal effects* of unit j (LeSage and Pace, 2009). Notice, however, that the dynamics of the model make direct and indirect effects stretch over time. That is, although the individual-specific effects are time-invariant, its marginal effects vary over time (to the extent that $\rho_0 \neq 0$). Yet we cannot interpret these variations as the result of "temporary" or "permanent" changes in the individual effects over time (which is the standard interpretation for regressors; see e.g. Debarsy et al. 2012). Bearing this in mind, the impact on the dependent variable in period $t = 1, ..., T$ is

$$\frac{\partial}{\partial e_j} E(Y_{nt}|\mathbf{X}) = \sum_{s=1}^{t} \rho_0^{s-1} S_0^{-s} (I_n \mu_j + W_n \alpha_j) \tag{2.6}$$

This expression can be interpreted as the "global" marginal effect (in period t), to the extent that it involves all the spatial units and not only at those considered to be neighbours by W_n (Anselin, 2003). However, if we rewrite 2.6 as

$$\frac{\partial}{\partial e_j} E(Y_{nt}|\mathbf{X}) = \sum_{s=1}^{t} \rho_0^{s-1} (I_n \mu_j + W_n \alpha_j) + \sum_{r=1}^{\infty} \lambda_0^r W_n^r \sum_{s=1}^{t} \rho_0^{s-1} \sum_{m=0}^{s-1} S_0^{-m} (I_n \mu_j + W_n \alpha_j),$$

we notice that the first term in this expression only involves the neighbouring units (as defined by W_n). Thus, we may interpret $\sum_{s=1}^{t} \rho_0^{s-1} (I_n \mu_j + W_n \alpha_j)$ as the "local" marginal effect (Anselin, 2003). In fact, this is the marginal effect when $\lambda_0 = 0$, since in that case $W_n Y_{nt}$ is missing from the model and there is no "global" spatial contagion.

In particular, the row i and column m elements of $\sum_{s=1}^{t} \rho_0^{s-1} S_0^{-s} (I_n \mu_j + W_n \alpha_j)$ and $\sum_{s=1}^{t} \rho_0^{s-1} (I_n \mu_j + W_n \alpha_j)$ can be interpreted as the global and local impact, respectively, on the outcome of unit i of unit m having the unobserved characteristics of unit j. Following Miranda et al. (2017), however, we find that is of greater interest to report the impact of unit m having its own unobserved characteristics (i.e., the unobserved characteristics of unit m) on the outcome of unit i. This means using the matrices

$$\sum_{s=1}^{t} \rho_0^{s-1} S_0^{-s} \left[diag(\mu_n) + W_n diag(\alpha_n)\right] \tag{2.7}$$

and

$$\sum_{s=1}^{t} \rho_0^{s-1} \left[diag(\mu_n) + W_n diag(\alpha_n)\right] \tag{2.8}$$

to compute the global and local marginal effects of interest, respectively. That is, the global and local marginal effects for each unit of all the other units having their own characteristics.

Thus, the main diagonal elements of these matrices provide, respectively, the direct global and local marginal effects (to reiterate, the impact on each unit of its own characteristics), whereas the off-diagonal elements of these matrices provide, respectively, the indirect global and local marginal effects (for a given time period t). We also obtain the spill-in and spill-out effects of the individual effects by respectively row- and column-summing the off-diagonal elements of these matrices (LeSage and Chih, 2016). In this vein the spill-in effect provides the global and local impact on the outcome of unit i of all the units neighbouring i having their unobserved characteristics, whereas the spill-out effect provides the global and local impact on the outcome of the units neighbouring i of the individual effect of unit i.

3 QML estimation: likelihood function and asymptotic properties

In this section we derive the quasi likelihood function of the spatial Durbin dynamic panel model with correlated random effects. We also study the consistency and asymptotic normality of the associated QML estimator. All results are obtained assuming that Y_{n0} is exogenous.

The endogenous case, which is more involved (see e.g. Su and Yang, 2015), is left for future research.[6]

3.1 The QML estimator

Following the notation introduced in 2.4, let us now define $\mathbf{Y}_{-1} = \left(Y'_{n0}, Y'_{n1}, \ldots, Y'_{n(T-1)}\right)'$, $\overline{\mathbf{X}} = l_T \otimes \overline{X}_n$, $\widetilde{\mathbf{X}} = \left(\boldsymbol{l}_{nT} \,\vdots\, \mathbf{Y}_{-1} \,\vdots\, \mathbf{X} \,\vdots\, \mathbf{WX} \,\vdots\, \overline{\mathbf{X}} \,\vdots\, \mathbf{W}\overline{\mathbf{X}} \right)$, and $\boldsymbol{\eta} = (\eta'_{n1}, \ldots, \eta'_{nT})'$. We can then rewrite the model in 2.3, evaluated at any parameter value and to include all nT observations, as

$$\mathbf{SY} = \tilde{\mathbf{X}}\theta + \boldsymbol{\eta} \tag{3.1}$$

with $\theta = \left(c, \rho, \beta'_1, \beta'_2, \pi'_\mu, \pi'_\alpha\right)'$. Further, let $\psi = \left(\theta', \sigma^2_\varepsilon, \delta'\right)'$, $\delta = (\sigma', \lambda)'$, $\sigma' = (\sigma_1, \sigma_2, \sigma_3)'$, $\boldsymbol{\eta}(\lambda, \theta) = \mathbf{S}(\lambda)\mathbf{Y} - \tilde{\mathbf{X}}\theta$ and $\sigma^2_\varepsilon \boldsymbol{\Omega}(\sigma) = \sigma^2_\varepsilon \left(J_T \otimes \Sigma(\sigma) + I_T \otimes I_n\right)$, with $\Sigma(\sigma) = \sigma_1 I_n + \sigma_2(W_n + W'_n) + \sigma_3 W_n W'_n$. Then, the quasi-loglikelihood function of the model in 3.1 can be written as

$$\mathcal{L}(\psi) = \ln|\mathbf{S}(\lambda)| - \frac{nT}{2}\ln(2\pi) - \frac{nT}{2}\ln(\sigma^2_\varepsilon) - \frac{1}{2}\ln|\boldsymbol{\Omega}(\sigma)| - \frac{1}{2\sigma^2_\varepsilon}\boldsymbol{\eta}'(\lambda,\theta)\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}(\lambda,\theta). \tag{3.2}$$

where $|\cdot|$ denotes the determinant of a matrix. Notice that, given δ, the values of θ and σ^2_ε that maximize 3.2 are given by:

$$\begin{aligned} \hat{\theta}(\delta) &= \left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{S}(\lambda)\mathbf{Y} \\ \hat{\sigma}^2_\varepsilon(\delta) &= \frac{1}{nT}\hat{\boldsymbol{\eta}}'(\delta)\boldsymbol{\Omega}^{-1}(\sigma)\hat{\boldsymbol{\eta}}(\delta), \end{aligned} \tag{3.3}$$

where $\hat{\boldsymbol{\eta}}(\delta) = \mathbf{S}(\lambda)\mathbf{Y} - \tilde{\mathbf{X}}\hat{\theta}(\delta)$. Thus, substituting 3.3 into 3.2 we obtain the concentrated quasi-loglikelihood function of δ:

$$\mathcal{L}_c(\delta) = \ln|\mathbf{S}(\lambda)| - \frac{nT}{2}\left(\ln(2\pi) + 1\right) - \frac{nT}{2}\ln\left(\hat{\sigma}^2_\varepsilon(\delta)\right) - \frac{1}{2}\ln|\boldsymbol{\Omega}(\sigma)| \tag{3.4}$$

[6] In any case, it is interesting to note that Monte Carlo evidence reported by Su and Yang (2015, p. 202-203) shows that, in the random effects case, estimating a model assuming that Y_{n0} is exogenous when it is actually not yields "estimates [that] are in general quite close to the true estimates except [when ρ is] large and positive" whereas, in the fixed effects model, "a wrong treatment on the initial values may lead to misleading results though to a much lesser degree as compared with the case of random effects model".

Maximising 3.4 yields the QML estimator of δ, $\hat{\delta} = (\hat{\sigma}', \hat{\lambda})'$, whereas the QMLE estimators of θ and σ_ε^2 are given by $\hat{\theta} \equiv \hat{\theta}(\hat{\delta})$ and $\hat{\sigma}_\varepsilon^2(\hat{\delta}) = \hat{\sigma}_\varepsilon^2$, respectively. Further, the QML estimator of $\left(\sigma_\mu^2, \sigma_{\mu\alpha}, \sigma_\alpha^2\right)$ is given by $\left(\hat{\sigma}_\mu^2, \hat{\sigma}_{\mu\alpha}, \hat{\sigma}_\alpha^2\right) = \hat{\sigma}_\varepsilon^2 \left(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3\right) = \hat{\sigma}_\varepsilon^2 \hat{\sigma}$. Therefore, $\hat{\psi} = \left(\hat{\theta}', \hat{\sigma}_\varepsilon^2, \hat{\delta}'\right)'$.

3.2 Asymptotic Properties

To derive the asymptotic properties of the QML estimator of the model, we must first ensure that $\psi = \left(\theta', \sigma_\varepsilon^2, \delta\right)'$ is identifiable. Notice, however, that given 3.3 it suffices to ensure that $\delta = (\sigma', \lambda)'$ is identifiable. To this end, let us define $\mathcal{L}_c^*(\delta) = \max_{\theta, \sigma_\varepsilon^2} E\left[\mathcal{L}(\psi)\right]$. It can be proved that the arguments that maximize $E\left[\mathcal{L}(\psi)\right]$ given δ are:

$$\tilde{\theta}(\delta) = \left[E(\widetilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\sigma)\widetilde{\mathbf{X}})\right]^{-1} E\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{S}(\lambda)\mathbf{Y}\right] \tag{3.5}$$

$$\tilde{\sigma}_\varepsilon^2(\delta) = \frac{1}{nT} E\left[\tilde{\boldsymbol{\eta}}'(\delta)\mathbf{\Omega}^{-1}(\sigma)\tilde{\boldsymbol{\eta}}(\delta)\right] \tag{3.6}$$

with $\tilde{\boldsymbol{\eta}}(\delta) \equiv \boldsymbol{\eta}(\tilde{\theta}(\delta), \lambda)$. Consequently:

$$\mathcal{L}_c^*(\delta) = \ln|\mathbf{S}(\lambda)| - \frac{nT}{2}\left(\ln(2\pi) + 1\right) - \frac{nT}{2}\ln\left(\tilde{\sigma}_\varepsilon^2(\delta)\right) - \frac{1}{2}\ln|\mathbf{\Omega}(\sigma)| \tag{3.7}$$

Notice also that, by using Lemma A.3, $\tilde{\theta}(\delta_0) = \theta_0$ and $\tilde{\sigma}_\varepsilon^2(\delta_0) = \sigma_{\varepsilon_0}^2$.

Let us now denote by $\Delta = \Delta_\sigma \times \Delta_\lambda$ the (compact) parameter space of δ, with Δ_σ and Δ_λ being the (compact) parameter spaces of σ and λ, respectively.[7] Further, let us redefine $\widehat{\delta} = \max_{\delta \in \Delta} \mathcal{L}_c(\delta)$. We then require the following assumptions to prove that the QML estimator of the model, $\widehat{\psi} = (\widehat{\theta}', \widehat{\sigma}_\varepsilon^2, \widehat{\delta}')'$, is consistent and asymptotically normally distributed:

Assumption 1. *The available observations are (y_{it}, x_{it}), $i = 1, ..., n$ and $t = 1, ..., T$, with $T \geq 2$ fixed and $n \to \infty$. Also, all the elements of x_{it} are independent across i, and have $4 + \epsilon_0$ moments for some $\epsilon_0 > 0$.*

[7] Notice that we do not require specific assumptions about the parametric space of ρ_0. In particular, since we concentrate on the case of T finite and Y_{n0} exogenous, we do not need to assume that $|\rho_0| < 1$ to derive the results obtained in the paper (see Su and Yang, 2015, p. 236).

Assumption 2. *The elements of the disturbance vector* ε_{it} *are i.i.d. for all* i *and* t*, with* $E(\varepsilon_{it}) = 0$*,* $Var(\varepsilon_{it}) = \sigma^2_{\varepsilon_0}$ *and* $E|\varepsilon_{it}|^{4+\epsilon_0} < \infty$ *for some* $\epsilon_0 > 0$*. Similarly,* $(\upsilon_{i\mu}, \upsilon_{i\alpha})$ *are i.i.d. with* $E(\upsilon_{i\mu}) = E(\upsilon_{i\alpha}) = 0$*,* $Var(\upsilon_{i\mu}) = \sigma^2_{\mu_0}$*,* $Var(\upsilon_{i\alpha}) = \sigma^2_{\alpha_0}$*,* $Cov(\upsilon_{i\mu}, \upsilon_{i\alpha}) = \sigma_{\mu\alpha_0}$ *and have* $4+\epsilon_0$ *finite moments for some* $\epsilon_0 > 0$*. Moreover,* ε_{it} *and* $(\upsilon_{j\mu}, \nu_{j\alpha})$ *are **i)** mutually independent, and **ii)** independent of* x_{sr} *for all* $i, j, s = 1 \ldots n$ *and* $r, t = 1 \ldots T$*. Lastly,* $\sigma_0 = (\sigma_{10}, \sigma_{20}, \sigma_{30})'$ *is in the interior of* Δ_σ*.*

Assumption 3. *The elements of* W_n*,* W_{nij}*, are at most of order* h_n^{-1}*, uniformly in all* i *and* j *with* $h_n/n \to 0$ *as* $n \to \infty$*.*

Assumption 4. *Matrix* $S(\lambda)$ *is nonsingular for all* $\lambda \in \Delta_\lambda$*, with* λ_0 *being in the interior of* Δ_λ*.*

Assumption 5. *The sequence of matrices* W_n *and* $S^{-1}(\lambda)$ *are uniformly bounded in both row and column sums and uniformly in* λ *in the compact parameter space* Δ_λ*.*[8]

Assumption 6. $\lim_{n\to\infty} \frac{1}{nT} \left\{ \ln\left|\sigma^2_{\varepsilon 0}\mathbf{S}_0^{-2}\mathbf{\Omega}_0\right| - \ln\left|\tilde{\sigma}^2_\varepsilon(\delta)\mathbf{S}(\lambda)^{-2}\mathbf{\Omega}(\sigma)\right| \right\} \neq 0$ *for any* $\delta \neq \delta_0$*. Also,* $\frac{1}{nT}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ *is positive definite almost surely for* n *sufficiently large.*

Assumption 7. *Let* $\boldsymbol{H}_n(\psi) = \frac{\partial^2}{\partial\psi\partial\psi'}\mathcal{L}(\psi)$ *be the hessian of the likelihood function and* $\mathcal{G}_n(\psi) = \frac{\partial}{\partial\psi}\mathcal{L}(\psi)\frac{\partial}{\partial\psi'}\mathcal{L}(\psi)$ *be the product of the score vector. Both* $\boldsymbol{H} = \lim_{N\to\infty}\frac{1}{nT}E[\boldsymbol{H}_n(\psi_0)]$ *and* $\mathcal{G} = \lim_{n\to\infty}\frac{1}{nT}E[\mathcal{G}_n(\psi_0)]$ *exist. Also,* $\mathcal{G}$ *and* $-\boldsymbol{H}$ *are positive definite matrices.*

Assumption 8. *Matrix* $\mathbf{\Omega}_0^{-1}$ *is uniformly bounded in both row and column sums.*

These assumptions are commonly used in the (spatial) panel data literature. In particular, Assumption 1 is standard for (dynamic) linear panel data models with large n and small T where Y_{n0} is exogenous. The first part of Assumption 2 is also rather standard in random-effects panel data models. What is not that common is the part that refers to the bivariate

[8] We say that a $k \times m$ matrix A (or a sequence of matrices A_n) is bounded in both row and column sums if there exists a constant $c < \infty$ such that $\max_j \sum_{i=1}^{k} A_{ij} < c$ and $\max_i \sum_{j=1}^{m} A_{ij} < c$.

random vector $(v_{i\mu}, v_{i\alpha})$, which is justified by the existence of spatial spillovers in the individual effects of our model.

As for the next three assumptions, they are widely used in spatial econometrics models. In particular, Assumption 3 is a necessary condition for Assumptions 6 and 7 that can be found in e.g. Lee (2004) and Su and Yang (2015). It is always satisfied if $\{h_n\}$ is a bounded sequence and essentially allows the weight matrices to be rather "general", "cover[ing] spatial weights matrices where elements are not restricted to be nonnegative and those that might not be row-normalized" (Lee, 2004, p. 1903). Assumptions 4 and 5 can be found in e.g. Lee (2004) and parallel Assumptions 3 and 5 of Yu et al. (2008). In particular, Assumption 5 was first employed by Kelejian and Prucha (1998, 2001). While Assumption 4 guarantees that $\mathbf{Y}$ can be expressed exclusively in terms of the exogenous variables, Assumption 5 essentially limits the spatial correlation. Notice also that Assumption 4 holds if $\lambda_0 \in \left(\frac{1}{\omega_{\min}}, \frac{1}{\omega_{\max}}\right)$, where $\omega_{\min}$ denotes the smallest and $\omega_{\max}$ denotes the largest characteristic root of the spatial weight matrix W_n $(\omega_{\min} < 0, \omega_{\max} > 0)$.

The last three assumptions have also been previously used to derive the asymptotic properties of a QML estimator in spatial econometrics models for cross-section and panel data (Lee, 2004; Su and Yang, 2015). Firstly, Assumption 6 basically provides conditions for the global identification of the estimator. More precisely, the first part is the identification uniqueness condition (White, 1994), while the second part guarantees that the regressors are not asymptotically multicollinear. In particular, in the second part of the assumption we can alternatively assume that $\frac{1}{nT}E(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})$ is positive definite for sufficiently large n. This is a softer condition that only requires some additional proof to be applied. Secondly, Assumption 7 guarantees the existence and positive definiteness of the Hessian and the variance covariance matrix of the score vector. It thus plays a basic role in the asymptotic normality results. Thirdly, Assumption 8 is necessary for the Central Limit Theorem we use to derive the asymptotic normality of the estimator (Kelejian and Prucha, 2001). In particular, it can be shown that this assumption also holds if $(I_n + T\Sigma(\sigma_0))^{-1}$ is uniformly bounded in both row and column sums.

Theorem 1. *Under assumptions 1 to 6, ψ_0 is globally identified and $\widehat{\psi}$ is a consistent estimator* *of ψ_0 with $\widehat{\psi} \xrightarrow{p} \psi_0$.*

Theorem 2. *Under assumptions 1 to 8, $\sqrt{nT}\left(\widehat{\psi} - \psi_0\right) \xrightarrow{d} N\left(0, \boldsymbol{H}^{-1}\mathcal{G}\boldsymbol{H}^{-1}\right)$.*

Remark 1. *Lee (2004), Yu et al. (2008) and Su and Yang (2015) use analog theorems to prove the consistency and asymptotic normality of their QML estimator in cross-section (Lee, 2004) and panel data (Yu et al., 2008; Su and Yang, 2015) models. In particular, Theorems 1 and 2 are similar to Theorems 3.1 and 3.2 of Lee (2004), Theorems 4 and 5 of Yu et al. (2008), and Theorems 4.1 and 4.2 of Su and Yang (2015), respectively. Because of the panel structure, our results are obviously closer to those of Yu et al. (2008), who analyse a spatial dynamic panel data model with fixed effects and no spatial contagion in the error term (and large T and n), and those of Su and Yang (2015), who analyse a dynamic panel data with spatially autocorrelated errors and both fixed and random effects (with small T and large n, as we do). This means that, on the one hand, our set of regressors is similar to that of Yu et al. (2008), except that we do not have the spatial lag of the lagged dependent variable and they do not have the spatially weighted exogenous variables (Su and Yang (2015) do not consider either spatially weighted regressors or the spatial lag of the – lagged – dependent variable). But, on the other hand, our error structure does have local spatial contagion, as Su and Yang's does (2015), although ours is in the individual-specific effects and theirs is in the idiosyncratic term (which in turn results in a variance-covariance matrix different from the ones assumed by these papers). Thus, our model specification is different, and so is the variance-covariance matrix, but the approach and the proof of our theorems largely follows their work (see Appendices A and B for details). In particular, the fact that our model specification includes the spatial lag of the endogenous variable makes the proof more involved than that of Su and Yang (2015). On the other hand, the scope of our proof is limited by the fact that we do not cover cases where Y_{n0} is endogenous, as they do.*

4 Empirical application

In this section we provide empirical evidence on a growth-initial level equation (see e.g. Islam 1995 and Elhorst et al. 2010) using the correlated random effects specification of the spatial

Durbin dynamic panel model presented in this paper. The principal aim of this empirical exercise is to show that *i)* we can (largely) replicate the results obtained by Lee and Yu (2016) using a standard spatial dynamic Durbin model (our benchmark); and *ii)* our model specification not only provides an estimate of the individual-specific effects but also of their spatial spillovers.

To this end, we use the data and (basic) model specification of Lee and Yu (2016). The dataset covers 28 OECD countries (see Ho et al. 2013 for details) over the period 1970 to 2005 (in time intervals of 5 years). The dependent variable, Y_{nt}, is the real GDP per capita (units of labour). As for the explanatory variables, $N_{nt} + 0.05$ is the sum of the annual average working-age population growth over the last 5 years (N_{nt}) and an approximation to the sum of the exogenous technical progress rate and the capital depreciation rate (see e.g. Ertur and Koch 2007 for details); S_{nt} is the average investment share in GDP; and $Y_{n,t-1}$ is the real GDP per capita lagged 5 years.

[Insert Table 1 about here]

The first column in Table 1 reports the results obtained by Lee and Yu (2016) using a weighting matrix W_n defined by the geographical distance between the capital of the countries. Notice that W_n is a row-normalized matrix with zeros in the diagonal. The second column provides the estimates of our model.[9] The parameter ρ measures the effect of the time-lagged real GDP ($Y_{n,t-1}$) on the dependent variable, whereas λ measures the intensity of its contemporaneous spatial interactions ($W_n Y_{nt}$). Also, the β-parameters measure the effect of the exogenous regressors (β_1 is the coefficient associated with $N_{nt} + 0.05$ and β_2 is the coefficient associated with S_{nt}), whereas the γ-parameters measure the intensity of the spatial contagion between the OECD countries arising from these exogenous regressors (γ_1 and γ_2 are the counterparts of β_1 and β_2). Lastly, the π-parameters are the coefficients associated with the variables included in the correlation functions. In particular, the π_μ-parameters correspond to those employed for the individual effects (π_{μ_1} is the coefficient of the mean of $N_{nt} + 0.05$

[9] Estimates were obtained using the optimizing routines of R and the log-likelihood function in 3.2.

and π_{μ_2} is that of the mean of S_{nt}) and the π_α-parameters to those employed for their spatial spillovers (π_{α_1} is the coefficient of the spatially weighted mean of $N_{nt} + 0.05$ and π_{α_2} is that of the spatially weighted mean of S_{nt}).

The first thing to notice is that our results largely concur with those of Lee and Yu (2016). This means that in both cases the coefficients of the working-age population growth rate (β_1) are negative and statistically significant at standard confidence levels, while the coefficients of the savings rate (β_2) are positive and statistically significant. Notice also that while the parameter associated with the time lagged real GDP is positive and statistically significant, the intensity of the contemporaneous spatial interactions of Y_{nt} is not statistically significant. This stands in contrast to the findings of Ertur and Koch (2007) and Elhorst et al. (2010).

It is also worth noting that only the coefficients associated with $N_{nt} + 0.05$ are – weakly – statistically significant in the correlations functions (the $p-$value of π_{μ_1} is 0.14, slightly above the standard 0.10).[10] This contrasts with the clear statistical significance of π_{α_1} (and the joint test for the π_α parameters), which supports the existence of spatial spillovers in the individual effects. However, the estimated variances indicate that the individual effects and their spatial counterparts do not have a significant random component. All in all, these results seem to be consistent with an error term specification analogous to the one proposed by Debarsy (2012).

Thus, if we interpret the estimated individual effects as a proxy for the unobserved productivity of the countries (see Islam 1995), our results suggest that the growth of some countries may be – weakly – related not only to their unobserved productivity, but also to the impact that the unobserved productivity of other countries have on their economies.[11] More generally, our results point to the importance of unobserved country-specific intrinsic features (economic, social, historical, etc.) in growth.

In order to further explore this idea and following the discussion in Section 2, we computed

[10] We also computed Wald tests for the joint significance of the coefficients in each correlation function. Results show that while the variables included in π_μ are not jointly significant (the p-value was 0.21), the variables included in π_α rejected the null hypothesis (the p-value was 0.01).

[11] Notice that, given the lack of statistical significance of σ_α, our results may also be consistent with the hypothesis (see Debarsy, 2012) that the growth of one country is linked to its unobserved productivity and this, in turn, is related to the (mean) characteristics of the other countries (but not to their unobserved productivities).

the direct and indirect global and local effects. However, since the λ coefficient is not statistically significant, the global and local effects coincide: the global effects are only of a local nature (Anselin, 2003). Thus, we interpret our results as local effects and, since the weight matrix is defined in terms of geographical distances, closer neighbours will have greater weight than distant neighbours in the indirect effects. In particular, we report the local direct effects for each period in Table 2 and the "spill-in" and "spill-out" effects of the estimated individual effects for each period in Tables 3 and 4, respectively.

The first column in Table 2 is the direct local effect in period one, which can be interpreted as the impact on the dependent variable (the log of real GDP per capita) of the estimated individual effects. In other words, these figures provide, for each country, an estimate of the difference in the log of real GDP per capita of having or not the unobserved heterogeneity term (i.e., having a zero value individual effect). As a caveat, notice that, given the weak statistical significance of the π_μ-parameters, these direct effects may not be statistically different than zero.

With this in mind, results indicate the existence of three groups of countries in our sample: those with a large individual effect, with values above the third quartile (Canada, Chile, Israel, Mexico, Netherlands, New Zealand, Turkey and the US); those with a small individual effect, with values below the first quartile (Austria, Belgium, Denmark, Finland, Greece, Italy, Japan, Korea, Norway, Portugal and Switzerland); and those with an intermediate individual effect (Australia, France, Iceland, Ireland, Spain, Sweden and the UK). It is also interesting to note that, for most countries, our ranking does not substantially differ from that of Islam (1995). However, in order to make meaningful comparisons, in the last two columns of Table 2 we report his estimated individual effects (obtained from a model without spatial interactions and for a sample of 192 countries over the period 1965 to 1985) and our equivalent estimate, $\hat{\mu} + W_n\hat{\alpha}$. We can see then that fifteen out of the 25 countries commonly analysed barely changed their ranking (Austria, Chile, Denmark, France, Greece, Israel, Italy, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland and the UK) and that, in fact, the most important differences arise from seven countries that dramatically changed their position in the rankings (Japan and Belgium, from the top of his ranking to the bottom of ours, and

Finland, Ireland, Korea, Mexico and Turkey, the other way round).

As for spill-in effects reported in Table 3, for each 5-year period the columns report the (local) impact on the log of real GDP per capita of each country associated with the unobserved characteristics of the other countries. The most affected countries (above the third quartile) are Austria, Finland, France, Ireland, Italy, Korea, Netherlands, Norway, Sweden and Switzerland, whereas the least affected countries (below the first quartile) are Australia, Canada, Chile, Iceland, Israel, Japan, Mexico, New Zealand, and the US. Notice that most of the countries with a small/large individual effect are among the most/least affected by their neighbours (in terms of geographical distance). Also, as expected figures in the other columns of the table show that, to a large extent, these groups remain stable over time.

Lastly, the columns in Table 4 contain, for each 5-year period, the estimated (local) impact on the log of real GDP per capita of the neighbouring countries associated with the unobserved characteristics of each country. However, rather than reporting the spill-out effect as described in Section 2, we simply report the estimated $\sum_{s=1}^{t} \rho^{s-1}\alpha_n$, which provides essentially the same picture.[12] Results show that the countries that impact least on their neighbours are Canada, Chile, Iceland, Israel, Korea, Mexico and New Zealand, whereas the countries that impact most on their neighbours are Austria, Belgium, Denmark, Finland, France, Italy, Japan, Sweden, Switzerland and the UK. Notice that countries that impact least/most on other countries tend to be those that are less/more affected by the spillovers from their neighbours (and generally have a larger/smaller individual effect). That is, there is a negative correlation between the estimated individual effects and the estimated spill-in (on average, -0.4) and spill-out (on average, -0.7) effects. Notice also that, as expected, these results largely hold for the seven periods considered.

[12] In particular, following Miranda et al. (2017, p. 4) we may interpret $\sum_{s=1}^{t} \rho^{s-1}\alpha_n$ "as the "potential" of the spatial spillovers of the individual effects" in each period (i.e., "a measure of the "potentiality of the spatial contagion" associated with the individual effect of [each] unit" in each period).

5 Conclusions

In this paper we consider a correlated random effects specification of the spatial Durbin dynamic panel model. We derive the likelihood function of the model and prove the consistency and asymptotic normality of the QML estimator under rather standard assumptions in the spatial econometrics literature. A major difference with respect to previous studies is that our model specification includes individual effects and their spatial spillovers.

Obtaining an estimate of the individual-specific effects (net of the spatially weighted effects) can be critical in certain applications, such as growth models in which a measure of the unobserved productivity of the geographical units under study can be obtained from the estimated individual effects and hence the existence of spatial spillovers in (unobserved) productivity can be analysed. We illustrate this point by estimating a growth-initial level equation using OECD data and providing evidence of spatial contagion in the individual effects.

Our results point to the importance of unobserved country-specific characteristics and their spatial spillovers in growth. In particular, we find that countries with a small/large estimated individual effect tend to be among the most/least affected by the impact of the estimated individual effects of their neighbours and among those whose individual effects impact most/least on the other countries (in terms of geographical distance). This means that, if we interpret the individual effect as a proxy for the unobserved productivity, more/less productive economies are less/more interrelated with the other economies. According to our estimates, examples of countries that fit into the first pattern include Chile, Israel, Mexico and New Zealand, whereas examples of countries that fit into the second pattern include Austria, Finland, Italy and Switzerland.

Table 1: **QML estimates**

Variable	Parameters	Lee and Yu (2016)	Our model
WY_t	λ	-0.040	-0.011
		(0.045)	(0.020)
Y_{t-1}	ρ	0.889^{***}	0.919^{***}
		(0.046)	(0.049)
$N_t + 0.05$	β_1	-0.198^{***}	-0.200^{***}
		(0.04)	(0.042)
S_t	β_2	0.143^{***}	0.141^{***}
		(0.047)	(0.048)
$W(N_t + 0.05)$	γ_1	0.102^{**}	0.108^{**}
		(0.047)	(0.048)
WS_t	γ_2	0.003	-0.001
		(0.057)	(0.057)
$\overline{N_t + 0.05}$	π_{μ_1}		0.115
			(0.079)
$(\overline{S_t})$	π_{μ_2}		-0.061
			(0.057)
$W(\overline{N_t + 0.05})$	π_{α_1}		-0.284^{***}
			(0.091)
$W\overline{S_t}$	π_{α_2}		-0.004
			(0.065)

Variance Components

σ^2_μ	σ^2_α	$\sigma_{\mu\alpha}$	σ^2_ε
0.0001	0.0000	0.0001	0.004^{***}
(0.0002)	(0.0002)	(0.0002)	(0.0005)

Note: *p-value<0.1; **p-value<0.05; ***p-value<0.01. We denote the time-mean of a variable with an upper bar.

Table 2: **Local Direct Effects**

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	Islam (1995)	$\hat{\mu}+W_n\hat{\alpha}$
Australia	-0.40	-0.77	-1.11	-1.42	-1.70	-1.97	-2.21	1.69	0.35
Austria	-0.42	-0.81	-1.16	-1.49	-1.79	-2.07	-2.32	1.72	0.38
Belgium	-0.42	-0.82	-1.17	-1.50	-1.81	-2.09	-2.34	1.75	0.35
Canada	-0.39	-0.74	-1.07	-1.37	-1.64	-1.90	-2.13	1.81	0.37
Chile	-0.38	-0.73	-1.05	-1.35	-1.62	-1.87	-2.10	1.49	0.34
Denmark	-0.42	-0.80	-1.15	-1.48	-1.77	-2.05	-2.30	1.74	0.38
Finland	-0.43	-0.82	-1.19	-1.52	-1.83	-2.11	-2.37	1.66	0.39
France	-0.41	-0.79	-1.14	-1.46	-1.75	-2.02	-2.27	1.75	0.39
Greece	-0.42	-0.80	-1.16	-1.48	-1.78	-2.06	-2.31	1.60	0.35
Iceland	-0.41	-0.78	-1.13	-1.45	-1.74	-2.01	-2.26	–	0.36
Ireland	-0.41	-0.78	-1.12	-1.44	-1.73	-1.99	-2.24	1.60	0.40
Israel	-0.39	-0.76	-1.09	-1.40	-1.68	-1.94	-2.18	1.70	0.37
Italy	-0.43	-0.82	-1.18	-1.52	-1.82	-2.10	-2.36	1.69	0.37
Japan	-0.44	-0.84	-1.22	-1.56	-1.87	-2.16	-2.43	1.75	0.29
Korea	-0.42	-0.80	-1.16	-1.48	-1.78	-2.05	-2.31	1.60	0.39
Mexico	-0.38	-0.73	-1.05	-1.35	-1.62	-1.87	-2.10	1.65	0.38
Netherlands	-0.39	-0.75	-1.08	-1.39	-1.67	-1.92	-2.16	1.73	0.41
New Zealand	-0.39	-0.75	-1.08	-1.38	-1.66	-1.91	-2.15	1.69	0.37
Norway	-0.42	-0.81	-1.16	-1.49	-1.79	-2.06	-2.32	1.77	0.39
Portugal	-0.42	-0.81	-1.17	-1.50	-1.81	-2.09	-2.34	1.58	0.35
Spain	-0.41	-0.79	-1.14	-1.46	-1.76	-2.03	-2.28	1.75	0.38
Sweden	-0.41	-0.79	-1.14	-1.46	-1.75	-2.02	-2.27	1.73	0.39
Switzerland	-0.43	-0.82	-1.18	-1.52	-1.82	-2.10	-2.36	1.70	0.37
Turkey	-0.38	-0.73	-1.05	-1.35	-1.62	-1.87	-2.10	1.53	0.39
United Kingdom	-0.40	-0.77	-1.11	-1.42	-1.70	-1.96	-2.21	1.73	0.39
United States	-0.39	-0.75	-1.09	-1.39	-1.67	-1.93	-2.17	1.80	0.36

The last two columns provide the estimated individual effects reported by Islam (1995) and our equivalent estimate, $\hat{\mu}+W_n\hat{\alpha}$.

Table 3: **Spill-in Effects**

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$
Australia	0.75	1.45	2.09	2.67	3.21	3.71	4.16
Austria	0.80	1.53	2.21	2.83	3.40	3.93	4.41
Belgium	0.78	1.50	2.15	2.76	3.32	3.83	4.30
Canada	0.76	1.46	2.10	2.69	3.24	3.74	4.19
Chile	0.72	1.38	1.99	2.55	3.06	3.54	3.97
Denmark	0.79	1.52	2.19	2.81	3.37	3.89	4.37
Finland	0.82	1.57	2.26	2.89	3.47	4.01	4.50
France	0.80	1.54	2.22	2.84	3.41	3.94	4.42
Greece	0.77	1.47	2.12	2.72	3.26	3.77	4.23
Iceland	0.76	1.47	2.11	2.71	3.26	3.76	4.22
Ireland	0.81	1.55	2.24	2.87	3.45	3.98	4.47
Israel	0.76	1.46	2.11	2.70	3.25	3.75	4.21
Italy	0.80	1.53	2.20	2.82	3.39	3.91	4.40
Japan	0.73	1.40	2.01	2.58	3.10	3.58	4.02
Korea	0.81	1.55	2.23	2.85	3.43	3.96	4.44
Mexico	0.76	1.46	2.10	2.69	3.24	3.74	4.19
Netherlands	0.80	1.54	2.22	2.85	3.43	3.95	4.44
New Zealand	0.76	1.45	2.09	2.68	3.22	3.72	4.17
Norway	0.81	1.56	2.25	2.89	3.47	4.00	4.50
Portugal	0.78	1.49	2.15	2.75	3.30	3.81	4.28
Spain	0.79	1.52	2.19	2.80	3.36	3.88	4.36
Sweden	0.80	1.54	2.22	2.84	3.41	3.94	4.42
Switzerland	0.80	1.53	2.21	2.83	3.40	3.92	4.40
Turkey	0.77	1.48	2.14	2.74	3.29	3.80	4.26
United Kingdom	0.79	1.51	2.18	2.79	3.36	3.87	4.35
United States	0.75	1.44	2.07	2.65	3.19	3.68	4.13

Table 4: **Spill-out Effects**

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$
Australia	0.76	1.45	2.09	2.68	3.22	3.72	4.17
Austria	0.80	1.53	2.21	2.83	3.40	3.92	4.41
Belgium	0.80	1.54	2.22	2.85	3.42	3.95	4.44
Canada	0.75	1.44	2.07	2.65	3.19	3.68	4.13
Chile	0.73	1.41	2.02	2.59	3.12	3.60	4.04
Denmark	0.81	1.56	2.24	2.88	3.46	3.99	4.48
Finland	0.82	1.57	2.25	2.89	3.47	4.01	4.50
France	0.80	1.53	2.21	2.83	3.40	3.92	4.41
Greece	0.79	1.52	2.19	2.80	3.37	3.89	4.37
Iceland	0.75	1.43	2.07	2.65	3.18	3.67	4.12
Ireland	0.76	1.46	2.11	2.70	3.24	3.75	4.21
Israel	0.72	1.38	1.98	2.54	3.05	3.52	3.95
Italy	0.82	1.57	2.26	2.90	3.49	4.02	4.52
Japan	0.81	1.55	2.23	2.85	3.43	3.96	4.44
Korea	0.73	1.40	2.01	2.58	3.10	3.58	4.02
Mexico	0.72	1.38	1.99	2.55	3.06	3.54	3.97
Netherlands	0.76	1.47	2.11	2.70	3.25	3.75	4.21
New Zealand	0.75	1.45	2.09	2.67	3.21	3.71	4.16
Norway	0.78	1.50	2.16	2.77	3.33	3.84	4.31
Portugal	0.79	1.52	2.18	2.80	3.36	3.88	4.36
Spain	0.78	1.49	2.15	2.75	3.30	3.81	4.28
Sweden	0.82	1.57	2.26	2.89	3.48	4.01	4.51
Switzerland	0.80	1.53	2.20	2.82	3.39	3.91	4.39
Turkey	0.76	1.46	2.11	2.70	3.24	3.74	4.20
United Kingdom	0.81	1.56	2.25	2.88	3.46	3.99	4.48
United States	0.76	1.46	2.10	2.69	3.24	3.74	4.19

A Lemmas

In this section we make extensive use of the following notation: $tr(A)$ denotes the trace of matrix A, $\tau_{\max}(A)$ the largest eigenvalue of matrix A, $\tau_{\min}(A)$ the smallest eigenvalue of matrix A, and $\|A\|_m$ the $m-$norm of matrix A with $m = 1, 2, \infty$ and F ($m = F$ being the Frobenius norm). Further, we use the term u.b.r.c.s. to refer to a matrix or sequence of matrices "uniformly bounded in both row and column sums".

We also make use of the following representation of the model in 2.3 and 2.4 (obtained by repeated substitution):

$$Y_{nt} = \rho_0^t S_0^{-t} Y_{n,0} + \sum_{j=0}^{t-1} \rho_0^j S_0^{-(j+1)} \mathbb{X}_{n,t-j}\phi_0 + \sum_{j=0}^{t-1} \rho_0^j S_0^{-(j+1)} \left(\upsilon_{n\mu} + W_n \upsilon_{n\alpha}\right) + \sum_{j=0}^{t-1} \rho_0^j S_0^{-(j+1)} \varepsilon_{n,t-j}$$

where $\phi_0 = (c_0, \beta_{10}', \beta_{20}', \pi_{\mu 0}', \pi_{\alpha 0}')'$, $\mathbb{X}_{nt} = \left(\begin{array}{c:c:c:c:c} l_n & X_{nt} & W_n X_{nt} & \overline{X}_n & W_n \overline{X}_n \end{array} \right)$ is an $n \times (4K+1)$ matrix, and the other elements are defined in Section 2. In full matrix notation:

$$\mathbf{Y} = \mathbf{G}_0 Y_{n,0} + \mathbf{C}_0 \mathbb{X}\phi_0 + \mathbf{L}_0 \left(\upsilon_{n\mu} + W_n \upsilon_{n\alpha}\right) + \mathbf{C}_0 \boldsymbol{\varepsilon} \tag{A.1}$$

with $\mathbf{L}_0 = \mathbf{C}_0(l_T \otimes I_n)$.

Lastly, some of the lemmas make use of the following property:

Property 1. *Let $D^{-1}(\sigma)$ be an $r \times r$ symmetric matrix, with $\sigma \in \Delta$ being a $p \times 1$ vector of parameters and Δ a compact parametric space. Then, there exists a matrix $A_k(\sigma, \overline{\sigma})$ such that*

i) $D^{-1}(\sigma) - D^{-1}(\overline{\sigma}) = \sum_{k=1}^{p} (\sigma_k - \overline{\sigma}_k) A_k(\sigma, \overline{\sigma})$ *for all* $\sigma, \overline{\sigma} \in \Delta$

ii) $\sup_{\sigma \in \Delta} \tau_{\max}\left(D^{-2}(\sigma)\right) \leq c_\tau < \infty$

iii) $\sup_{\sigma, \overline{\sigma} \in \Delta} \tau_{\max}\left(A_k(\sigma, \overline{\sigma}) A_k'(\sigma, \overline{\sigma})\right) \leq c_\tau < \infty$ *for* $k = 1, \ldots, p$

Lemma A.1. *Let A be a real symmetric $n \times n$ matrix and B a random $n \times m$ matrix. Then,*

$$\tau_{\min}\left(E\left(B'AB\right)\right) \geq \tau_{\min}\left(A\right)\tau_{\min}\left(E\left(B'B\right)\right)$$

Proof. By definition, $\tau_{\min}\left(E\left(B'AB\right)\right) = \min_{z \in R^m}\left\{z'E\left(B'AB\right)z \middle| z'z = 1\right\}$. Let $\overline{z}$ be such that $\tau_{\min}\left(E\left(B'AB\right)\right) = \overline{z}'E\left(B'AB\right)\overline{z}$. Let D_A be the diagonal matrix of eigenvalues of A. Since A is a real symmetric matrix, there exists Q such that $A = QDQ'$ and $QQ' = I_n$. Then,

$$\begin{aligned}\tau_{\min}\left(E\left(B'AB\right)\right) &= E\left(\overline{z}'B'QD_AQ'B\overline{z}\right)\\ &\geq \tau_{\min}(A)E\left(\overline{z}'B'QQ'B\overline{z}\right)\\ &\geq \tau_{\min}(A)\min_{z \in R^m}\left\{E\left(z'B'Bz\right) \middle| z'z = 1\right\}\\ &\geq \tau_{\min}(A)\tau_{\min}\left(E\left(B'B\right)\right)\end{aligned}$$

□

Lemma A.2. *Let A be a real positive semidefinite $n \times n$ matrix and B a real symmetric $n \times n$ matrix. Then,*

$$tr\left(AB\right) \leq \tau_{\max}\left(B\right)tr\left(A\right).$$

Proof. Since B is a real symmetric matrix, it can be diagonalized. Let P_B be the orthogonal matrix with the eigenvectors of B ($P_BP_B' = I_n$) and let D_B be the diagonal matrix of eigenvalues of B such that $B = P_BD_BP_B'$. Then,

$$tr\left(AB\right) = tr\left(AP_BD_BP_B'\right) = tr\left(P_B'AP_BD_B\right) = tr\left(CD_B\right)$$

where C is a symmetric positive semidefinite matrix (given that A is a positive semidefinite matrix and $y'P_B'AP_By = x'Ax \geq 0$). Using that $tr\left(C\right) = tr\left(A\right)$ and given that $c_{ii} \geq 0$ for $i = 1, ..., n$ (because of the positive definitiveness of C),

$$tr\left(AB\right) = tr\left(CD_B\right) = \sum_{i=1}^{n} c_{ii}\tau_i\left(B\right) \leq \tau_{\max}\left(B\right)\sum_{i=1}^{n}\left|c_{ii}\right| = \tau_{\max}\left(B\right)tr\left(C\right) = \tau_{\max}\left(B\right)tr\left(A\right)$$

$\square$

Lemma A.3. *Under assumptions 1 to 6,* $E\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right]=0.$

Proof. We start by noting that, given that $\widetilde{\mathbf{X}}=\left[\begin{array}{c:c:c:c:c:c} \boldsymbol{l}_{nT} & \mathbf{Y}_{-1} & \mathbf{X} & \mathbf{WX} & \overline{\mathbf{X}} & \mathbf{W}\overline{\mathbf{X}}\end{array}\right]$, we only need to prove that $E\left[\mathbf{Y}_{-1}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right]=0$, since $E\left[\mathbf{Z}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right]=0$ for $\mathbf{Z}=\boldsymbol{l}_{nT},\mathbf{X},\mathbf{WX},\overline{\mathbf{X}}$ and $\mathbf{W}\overline{\mathbf{X}}$ by the strict exogeneity of $\mathbf{X}$. Notice also that, by using equation A.1, we have that

$$\mathbf{Y}_{-1}=\mathbf{G}_0^{-}Y_{n,0}+\mathbf{C}_0^{-}\mathbb{X}_{-1}\phi_0+\mathbf{L}_0^{-}\left(\upsilon_{n\mu}+W_n\upsilon_{n\alpha}\right)+\mathbf{C}_0^{-}\boldsymbol{\varepsilon}, \tag{A.2}$$

with $\mathbb{X}_{-1}=\left(0,\mathbb{X}_{n1}',\cdots,\mathbb{X}_{n,T-1}'\right)'$, $\mathbf{G}_0^{-}=\left(I_n,\rho_0S_0^{-1\prime},\ldots,\rho_0^{T-2}S_0^{-(T-1)\prime}\right)'$, $\mathbf{L}_0^{-}=\mathbf{C}_0^{-}\left(l_T\otimes I_n\right)$ and

$$\mathbf{C}_0^{-}=\begin{pmatrix} 0 & 0 & 0 & \cdots & 0\\ S_0^{-1} & 0 & 0 & \cdots & 0\\ \rho_0S_0^{-2} & S_0^{-1} & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \rho_0^{T-2}S_0^{-(T-1)} & \rho_0^{T-3}S_0^{-(T-2)} & \rho_0^{T-4}S_0^{-(T-3)} & \cdots & 0\end{pmatrix}$$

Thus,

$$\mathbf{Y}_{-1}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}=Y_{n0}'\mathbf{G}_0^{-}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+\phi'\mathbb{X}_{-1}'\mathbf{C}_0^{-\prime}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+\varepsilon'\mathbf{C}_0^{-\prime}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+\left(\upsilon_{n\mu}'+\upsilon_{n\alpha}'W_n'\right)\mathbf{L}_0^{-\prime}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}$$

First, it is easy to show that $E\left(Y_{n0}'\mathbf{G}_0^{-}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right)=0=E\left(\phi'\mathbb{X}_{-1}'\mathbf{C}_0^{-\prime}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right)$. Second, notice that we can write $E\left(\boldsymbol{\varepsilon}'\mathbf{C}_0^{-\prime}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right)=\sigma_\varepsilon^2tr\left(\mathbf{\Omega}_0^{-1}\mathbf{C}_0^{-\prime}\right)$ and, given that $\mathbf{L}_0^{-\prime}=\left(l_T'\otimes I_n\right)\boldsymbol{C}_0^{-\prime}$, $J_T=l_Tl_T'$ and $E\left[\left(\upsilon_\mu+W_n\upsilon_{n\alpha}\right)\left(\upsilon_\mu+W_n\upsilon_{n\alpha}\right)'\right]=\sigma_{\varepsilon_0}^2\Sigma_0$,

$$E\left(\left(\upsilon_{n\mu}'+\upsilon_{n\alpha}'W_n'\right)\mathbf{L}_0^{-\prime}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right)=\sigma_{\varepsilon0}^2tr\left[\mathbf{\Omega}_0^{-1}\left(J_T\otimes\Sigma_0\right)\mathbf{C}_0^{-\prime}\right]$$

Also, following Magnus (1982), we can rewrite $\mathbf{\Omega}_0^{-1}$ as $\mathbf{\Omega}_0^{-1}=\left(I_T\otimes I_n\right)-\frac{1}{T}J_T\otimes\left[I_n-\left(I_n+T\Sigma_0\right)^{-1}\right]$, which means that

$$\sigma_{\varepsilon_0}^2tr\left(\mathbf{\Omega}_0^{-1}\mathbf{C}_0^{-\prime}\right)+\sigma_{\varepsilon_0}^2tr\left[\mathbf{\Omega}_0^{-1}\left(J_T\otimes\Sigma_0\right)\mathbf{C}_0^{-\prime}\right]=\sigma_{\varepsilon_0}^2tr\left[\mathbf{C}_0^{-\prime}\right]+\sigma_{\varepsilon_0}^2tr\left[\mathbf{A}\mathbf{C}_0^{-\prime}\right]=0$$

since

$$\begin{aligned}\mathbf{A} &= -\frac{1}{T}\left(J_T \otimes \left[I_n - (I_n + T\Sigma_0)^{-1}\right]\right) + \frac{1}{T}\left(J_T \otimes T\Sigma_0\right) \\ &- \left(\frac{1}{T}J_T \otimes \left[I_n - (I_n + T\Sigma_0)^{-1}\right]\right)\left(J_T \otimes \Sigma_0\right) \\ &= \frac{1}{T}J_T \otimes \left[-I_n + (I_n + T\Sigma_0)^{-1}(I_n + T\Sigma_0)\right] = 0\end{aligned}$$

and $tr\left[\mathbf{C}_0'\right] = 0$ because of the structure of $\mathbf{C}_0'$. □

Lemma A.4. *Let A, B and C be real constant matrices of order $(n \times r)$, $(r \times r)$ and $(r \times n)$ respectively, with A and C u.b.r.c.s. and B being a symmetric matrix with $\tau_{\max}(B^2) < \infty$. Then, for $Q = ABC$:*

i) $tr\left(QQ'\right) = O(\min(r, n))$

ii) $l_n' QQ' l_n = O\left(n\right)$*, where l_n is a unit vector of dimension $n \times 1$*

iii) $\sum_{i=1}^{n} Q_{ii}^2 = O(min(r, n))$ *and* $tr\left(QQ\right) = O(n)$*.*

Proof. Firstly, by the Cauchy-Schwarz inequality and Lemma A.2,

$$\begin{aligned}tr\left(QQ'\right) = tr\left(ABCC'BA'\right) &= tr\left(BCC'BA'A\right) \\ &\le \left[tr\left(BCC'CC'B\right)\right]^{1/2}\left[tr\left(BAA'AA'B\right)\right]^{1/2} \\ &\le \tau_{\max}(B^2)\left[tr\left(C'CC'C\right)\right]^{1/2}\left[tr\left(A'AA'A\right)\right]^{1/2}\end{aligned}$$

Then, by using the second part of Lemma B.1 in Su and Yang (2015), we can show that $\tau_{\max}(B^2)\left[tr\left(C'CC'C\right)\right]^{1/2}\left[tr\left(A'AA'A\right)\right]^{1/2} = O(\min(r, n))$.

Secondly,

$$\begin{aligned}l_n' QQ' l_n &= tr\left(l_n' ABCC'BA' l_n\right) = tr\left(BCC'BA' l_n l_n' A\right) \\ &\le \tau_{\max}\left(BCC'B\right) tr\left(A' l_n l_n' A\right) \le \tau_{\max}\left(BCC'B\right) tr\left(l_n' AA' l_n\right) = O\left(n\right),\end{aligned}$$

where the last equality holds because

- given that C is u.b.r.c.s., $\|C'\|_2^2 \leq \|C'\|_1^2 \|C'\|_\infty^2 \leq c^2$, with $\max_i \sum_{j=1}^n |c_{ij}| \leq c$, $\max_j \sum_{i=1}^n |c_{ij}| \leq c$ and $c < \infty$, and $\|B\|_2^2 = \tau_{\max}\left(B^2\right)$; then, since $\|.\|_2$ is a sub-multiplicative norm[13], $\tau_{\max}\left(BCC'B\right) = \|C'B\|_2^2 \leq \|C'\|_2^2 \|B\|_2^2 \leq c^2 \tau_{\max}\left(B^2\right)$,

- given that A is u.b.r.c.s., $\max_i \sum_{j=1}^n |a_{ij}| \leq a$, $\max_j \sum_{i=1}^n |a_{ij}| \leq a$ and $a < \infty$; then, $l_n' AA' l_n \leq a^2 l_n' l_n \leq a^2 n$.

Thirdly, $\sum_{i=1}^n Q_{ii}^2 \leq \sum_{i=1}^n \sum_{j=1}^n |Q_{ij}|^2 = \|Q\|_F^2 \leq tr\left(Q'Q\right)$, which, because of result i), is $O(\min(r,n))$. Also using result i), $tr(QQ) \leq tr(QQ')^{1/2} tr(QQ')^{1/2} = tr(QQ') = O(\min(r,n))$. □

Lemma A.5. *Let $a = (a_1, ..., a_n)'$ and $b = (b_1, ..., b_n)'$. Also, let $\{(a_i, b_i)\}_{i=1}^n$ be an i.i.d. sequence of random vector variables with $E(a_i) = E(b_i) = 0$ and finite second moments. Lastly, let P be an $n \times n$ constant matrix and let $\Omega = E(ab') = \mu_{ab} I_n$ such that $(a'Pb - tr(P\Omega)) = tr\left(Pba' - P\Omega\right) = \sum_{i=1}^n \sum_{j=1}^n P_{ij}(a_i b_j - \Omega_{ij})$. Then,*

$$E\left[\left(a'Pb - tr(P\Omega)\right)^2\right] = (\sigma_{ab}^2 - \sigma_a^2\sigma_b^2 - \mu_{ab}^2) \sum_{i=1}^n P_{ii}^2 + \sigma_a^2\sigma_b^2 tr\left(PP'\right) + \sigma_{ab}^2 tr\left(PP\right)$$

where $\sigma_a^2 = E(a_i^2)$, $\sigma_b^2 = E(b_i^2)$, $E(a_i b_i) = \mu_{ab}$ and $E[(a_i b_i - \mu_{ab})^2] = \sigma_{ab}^2$. Notice that, if a and b are independent, $E\left[\left(a'Pb - tr(P\Omega)\right)^2\right] = \sigma_a^2\sigma_b^2 tr(PP')$. Notice also that if $a = b$, then $E\left[\left(a'Pb - tr(P\Omega)\right)^2\right] = (\sigma_a^{(4)} - 2\sigma_a^4) \sum_{i=1}^n P_{ii}^2 + \sigma_a^4 tr\left(PP'\right) + \sigma_a^{(4)} tr\left(PP\right)$, with $\sigma_a^{(4)} = E[(a_i^2 - \sigma_a^2)^2]$.

Proof. Notice that $E\left[\left(a'Pb - tr(P\Omega)\right)^2\right] = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n P_{ij} P_{kl} E\left[(a_i b_j - \Omega_{ij})(a_k b_l - \Omega_{kl})\right]$. Also, given the independence of (a_i, b_i) and (a_j, b_j) for $i \neq j$, $E\left[(a_i b_j - \Omega_{ij})(a_k b_l - \Omega_{kl})\right] \neq 0$

[13] This means that, for any two matrices A and B, $\|AB\|_2 \leq \|A\|_2 \|B\|_2$.

only for $i=j=k=l$, $i=k\neq j=l$ and $i=l\neq j=k$. Thus,

$$
\begin{aligned}
E\left[(a'Pb-tr(P\Omega))^2\right] &= \sum_{i=1}^{n} P_{ii}^2 E\left[(a_ib_i-\mu_{ab})^2\right] + \sum_{i=1}^{n}\sum_{j\neq i}^{n} P_{ij}^2 E\left[(a_ib_j)^2\right] \\
&+ \sum_{i=1}^{n}\sum_{j\neq i}^{n} P_{ij}P_{ji} E\left[(a_ib_i)(a_jb_j)\right] \\
&= (\sigma_{ab}^2-\sigma_a^2\sigma_b^2-\mu_{ab}^2)\sum_{i=1}^{n} P_{ii}^2 + \left(\sigma_a^2\sigma_b^2 tr(PP') + \mu_{ab}^2 tr(PP)\right)
\end{aligned}
$$

□

Lemma A.6. *Let $a=(a_1,...,a_n)'$ and $b=(b_1,...,b_n)'$, with $\{(a_i,b_i)\}_{i=1}^n$ i.i.d. sequences of random vector variables with finite second moments. Let P_n and Q_n be $n\times r$ constant matrices u.b.r.c.s.. Lastly, let $D(\sigma)$ be an $r\times r$ constant symmetric matrix that satisfies Property 1, with $\sigma\in\Delta$ being a $p\times 1$ vector of parameters. Then,*

$$
\sup_{\sigma\in\Delta}\left|E\left(a'P_nD^{-1}(\sigma)Q_nb\right)\right| = O(n)
$$

Note that the Lemma still holds if $a=b$ and $P_n=Q_n$.

Proof. By the Cauchy-Schwarz inequality and Lemma A.2,

$$
\begin{aligned}
\sup_{\sigma\in\Delta}\left|E\left(a'P_nD^{-1}(\sigma)Q_nb\right)\right| &\leq \sup_{\sigma\in\Delta} E\left(tr\left(a'P_nD^{-2}(\sigma)P_n'a\right)^{1/2} tr\left(b'Q_n'(\sigma)Q_nb\right)^{1/2}\right) \\
&\leq \left[\sup_{\sigma\in\Delta}\tau_{\max}\left(D^{-2}(\sigma)\right)\right]^{1/2}\left[E\left(tr\left(P_n'aa'P_n\right)^{1/2} tr\left(Q_nbb'Q_n'\right)^{1/2}\right)\right] \\
&\leq \left[\sup_{\sigma\in\Delta}\tau_{\max}\left(D^{-2}(\sigma)\right)\right]^{1/2}\tau_{\max}^{1/2}\left(P_nP_n'\right)\tau_{\max}^{1/2}\left(Q_n'Q_n\right)\left[E\left(tr\left(aa'\right)\right)E\left(tr\left(bb'\right)\right)\right]^{1/2} \\
&\leq C\left[E\left(tr\left(aa'\right)\right)E\left(tr\left(bb'\right)\right)\right]^{1/2}
\end{aligned}
$$

with $C<\infty$ given that $D^{-1}(\sigma)$ satisfies Property 1, $\tau_{\max}^{1/2}(A'A)=\|A\|_2\leq(\|A\|_1\|A\|_\infty)^{1/2}$, and P_n and Q_n are u.b.r.c.s.. Also, $E\left(tr\left(aa'\right)\right)=E\left[\sum_{i=1}^{n} a_i^2\right]\leq nE(a_i^2)$. Thus, given that a and b have finite second moments, the lemma is proved. □

Lemma A.7. *Let* $a = (a_1, ..., a_n)'$ *and* $b = (b_1, ..., b_n)'$, *with* $\{(a_i, b_i)\}_{i=1}^n$ *i.i.d. sequences of random vector variables with finite second moments. Let* P_n *and* Q_n *be* $n \times r$ *constant matrices u.b.r.c.s.. Lastly, let* $D(\sigma)$ *be an* $r \times r$ *constant symmetric matrix that satisfies Property 1, with* $\sigma \in \Delta$ *being a* $p \times 1$ *vector of parameters. Then,*

$$\frac{1}{\max(n,r)} \left\{ a' P_n D^{-1}(\sigma) Q_n b - E\left[a' P_n D^{-1}(\sigma) Q_n b\right] \right\} \xrightarrow{p} 0 \text{ uniformly in } \sigma \in \Delta$$

Note that the Lemma still holds if $a = b$.

Proof. Let us denote $E(a_i) = \mu_a, E(b_j) = \mu_b$, $E\left[(a_i - \mu_a)^2\right] = \sigma_a^2$, $E\left[(b_j - \mu_b)^2\right] = \sigma_b^2$ for all i and j. We start by proving that

$$\frac{1}{\max(n,r)} \left[\mu_a l_n' P_n D^{-1}(\sigma) Q_n (b - \mu_b)\right] \xrightarrow{p} 0 \text{ uniformly in } \sigma \in \Delta$$

To prove the uniform convergence (see e.g. Theorem 21.9 of Davidson 1994), we prove that $l_n' P_n D^{-1}(\sigma) Q_n (b - \mu_b)$ is stochastically equicontinuous and, for a given σ, satisfies a Law of Large Numbers (LLN hereafter). First we prove the convergence for a given σ. Given that $E\left[l_n' P_n D^{-1}(\sigma) Q_n (b - \mu_b)\right] = 0$, to derive a LLN it is enough to prove that

$$\frac{1}{\max(n,r)^2} Var\left[l_n' P_n D^{-1}(\sigma) Q_n (b - \mu_b)\right] \longrightarrow 0$$

It is straightforward to prove that $Var\left[l_n' P_n D^{-1}(\sigma) Q_n (b - \mu_b)\right] = \sigma_b^2 l_n' P_n D^{-1}(\sigma) Q_n Q_n' D^{-1}(\sigma) P_n' l_n$ and, by Lemma A.4, $l_n' P_n D^{-1}(\sigma) Q_n Q_n' D^{-1}(\sigma) P_n l_n = O(n)$, so that

$$\begin{aligned}
\frac{1}{\max(n,r)^2} Var\left[l_n' P_n D^{-1}(\sigma) Q_n (b - \mu_b)\right] &\leq \frac{1}{\max(n,r)^2} l_n' P_n D^{-1}(\sigma) Q_n Q_n' D^{-1}(\sigma) P_n l_n \\
&\leq \frac{O(n)}{\max(n,r)^2} = o(1)
\end{aligned}$$

which proves the LLN. To prove the stochastic equicontinuity, note that, by Property 1, the

Cauchy-Schwarz inequality and Lemma $A.2$,

$$
\begin{aligned}
&\left|l_n' P_n D^{-1}(\sigma) Q_n (b-\mu_b) - l_n' P_n D^{-1}(\overline{\sigma}) Q_n (b-\mu_b)\right| \leq \left|l_n' P_n \left(D^{-1}(\sigma) - D^{-1}(\overline{\sigma})\right) Q_n (b-\mu_b)\right| \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| tr^{1/2} \left(l_n' P_n A_k(\sigma,\overline{\sigma}) A_k'(\sigma,\overline{\sigma}) P_n' l_n\right) tr^{1/2} \left((b-\mu_b)' Q_n' Q_n (b-\mu_b)\right) \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| \tau_{\max} \left(A_k(\sigma,\overline{\sigma}) A_k'(\sigma,\overline{\sigma})\right) tr^{1/2} \left(l_n' P_n P_n' l_n\right) tr^{1/2} \left((b-\mu_b)' Q_n' Q_n (b-\mu_b)\right) \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| c_\tau tr^{1/2} \left(l_n' P_n P_n' l_n\right) | tr^{1/2} \left((b-\mu_b)' Q_n' Q_n (b-\mu_b)\right)
\end{aligned}
$$

with $c_\tau < \infty$. Also, by Lemma $A.4$, $tr\left(l_n' P_n P_n' l_n\right) = O(n)$ and, by Lemma $A.6$, $tr\left((b-\mu_b)' Q_n' Q_n (b-\mu_b)\right) = O_p(n)$, so we can apply Theorem 21.10 of Davidson (1994) to prove the stochastic equicontinuity and Theorem 21.9 of Davidson (1994) to prove the uniform convergence.

Next we prove the case $E(a_i) = E(b_i) = 0$. We first prove the convergence in probability given σ. To this end, notice that $E\left\{a' P_n D^{-1}(\sigma) Q_n b - E\left[a' P_n D^{-1}(\sigma) Q_n b\right]\right\} = 0$ and, from Lemmas $A.4$ and $A.5$, $E\left\{\left(a' P_n D^{-1}(\sigma) Q_n b - E\left[a' P_n D^{-1}(\sigma) Q_n b\right]\right)^2\right\} = O(n)$, so that

$$
\lim_{n\to\infty} \frac{1}{\max(n,r)^2} E\left\{\left(a' P_n D^{-1}(\sigma) Q_n b - E\left[a' P_n D^{-1}(\sigma) Q_n b\right]\right)^2\right\} = 0
$$

which proves the convergence given σ. To prove the stochastic equicontinuity, note that, by Property 1, the Cauchy-Schwarz inequality and Lemma $A.2$,

$$
\begin{aligned}
\left|a' P_n D^{-1}(\sigma) Q_n b - a' P_n D^{-1}(\overline{\sigma}) Q_n b\right| &\leq \left|a' P_n \left(D^{-1}(\sigma) - D^{-1}(\overline{\sigma})\right) Q_n b\right| \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| \left|a' P_n A_k(\sigma,\overline{\sigma}) Q_n b\right| \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| \, tr^{1/2} \left(a' P_n A_k(\sigma,\overline{\sigma}) A_k'(\sigma,\overline{\sigma}) P_n' a\right) tr^{1/2} \left(b' Q_n' Q_n b\right) \\
&\leq c_\tau tr^{1/2} \left(a' P_n P_n' a\right) tr^{1/2} \left(b' Q_n' Q_n b\right) \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k|
\end{aligned}
$$

Also, by Lemma A.6, $E\left[tr\left(a'P_nP_n'a\right)\right]=O(n)$ and $E\left[tr\left(b'Q_n'Q_nb\right)\right]=O(n)$. Then,

$$\frac{1}{\max(n,r)}\left|a'P_nD^{-1}(\sigma)Q_nb-a'P_nD^{-1}(\overline{\sigma})Q_nb\right|=O_p(1)$$
$$\frac{1}{\max(n,r)}\left|E\left(a'P_nD^{-1}(\sigma)Q_nb\right)-E\left(a'P_nD^{-1}(\overline{\sigma})Q_nb\right)\right|=O(1)$$

and we can apply Theorems 21.9 and 21.10 of Davidson (1994) to prove uniform convergence. Further, the most general case $E(a_i)\neq 0$ and $E(b_i)\neq 0$ follows straightforward by noting that

$$\left\{a'P_nD^{-1}(\sigma)Q_nb-E\left[a'P_nD^{-1}(\sigma)Q_nb\right]\right\}=a^{*\prime}P_nD^{-1}(\sigma)Q_nb^*-E\left[a^{*\prime}P_nD^{-1}(\sigma)Q_nb^*\right]+$$
$$a^{*\prime}P_nD^{-1}(\sigma)Q_nE(b)+E(a)'P_nD^{-1}(\sigma)Q_nb^*$$

with $a^*=a-\mu_a$ and $b^*=b-\mu_b$.

□

Lemma A.8. *Let* $G_{nt}=\rho_0^tS_0^{-t}$, $C_{nt}=G_{nt}S_0^{-1}$ *and* $L_{nt}=\sum_{j=0}^{t-1}\rho_0^jS_0^{-(j+1)}$. *Under Assumption 5,* W_nL_{nt}, W_nG_{nt} *and* W_nC_{nt} *are all u.b.r.c.s. for* $t=1,2,...,T$ *and* $\mathbf{WL}_0,\mathbf{WG}_0$ *and* $\mathbf{WC}_0$ *are all u.b.r.c.s..*

Proof. First note that if A and B are two matrices u.b.r.c.s., $A+B$ and AB are also u.b.r.c.s. (see Remark A2 in Kapoor et al. 2007). With this result, under Assumption 5 it is easy to prove that G_{nt}, C_{nt} and L_{nt} are u.b.r.c.s.. Further, given that $T<\infty$, it is easy to prove that $\mathbf{WL}_0$, $\mathbf{WG}_0$ and $\mathbf{WC}_0$ are all u.b.r.c.s.. □

Lemma A.9. *Let* $\mathbf{\Omega}(\sigma)=(I_T\otimes I_n)+(J_T\otimes\Sigma(\sigma))$ *and* $\Sigma(\sigma)=\sum_{k=1}^{3}\sigma_k\Sigma_k=\sigma_1I_n+\sigma_2(W_n+W_n')+\sigma_3W_nW_n'$, *with* W_n *u.b.r.c.s. and* $(\sigma_1,\sigma_2,\sigma_3)\in\Delta$, *being* Δ *a compact space such that* $\Sigma(\sigma)$ *is positive semidefinite for any* $\sigma\in\Delta$. *Then,* $\Omega^{-1}(\sigma)$ *satisfies Property 1 for* $A_k(\sigma,\overline{\sigma})=\Omega^{-1}(\sigma)(J_T\otimes\Sigma_k)\Omega^{-1}(\overline{\sigma})$ *and any* $\sigma,\overline{\sigma}\in\Delta$. *Moreover,* $\exists\, c_\tau<\infty$ *such that* $\sup_{\sigma\in\Delta}\tau_{\max}(\mathbf{\Omega}(\sigma))<c_\tau$.

Proof. We start by proving that $\exists\, c_\tau<\infty$ such that $\sup_{\sigma\in\Delta}\tau_{\max}(\mathbf{\Omega}(\sigma))<c_\tau$. To this end, note that the eigenvalues of the matrix (I_n+B) are $1+\tau_i(B)$, with $\tau_i(B)$ being the $i=1,...,n$ eigenvalues

of B. Then, by definition, $\sup_{\sigma\in\Delta}\tau_{\max}(\mathbf{\Omega}(\sigma)) = 1+\sup_{\sigma\in\Delta}\tau_{\max}((J_T\otimes\Sigma(\sigma))) = 1+T\sup_{\sigma\in\Delta}\tau_{\max}(\Sigma(\sigma))$. Further, using that $\Sigma(\sigma)$ is a symmetric positive semidefinite matrix, $\sup_{\sigma\in\Delta}\tau_{\max}(\Sigma(\sigma)) = \sup_{\sigma\in\Delta}\|\Sigma(\sigma)\|_2$. Then,

$$\sup_{\sigma\in\Delta}\|\Sigma(\sigma)\|_2 \leq \sum_{k=1}^{3}\sup_{\sigma\in\Delta}|\sigma_k|\|\Sigma_k\|_2 \leq \sup_{\sigma\in\Delta}|\sigma_1|+\sup_{\sigma\in\Delta}|\sigma_2|\|W_n+W_n'\|_2+\sup_{\sigma\in\Delta}|\sigma_3|\|W_nW_n'\|_2$$

Given that W_n is u.b.r.c.s., W_n+W_n' and W_nW_n' are u.b.r.c.s., too (see Remark A2 in Kapoor et al. 2007). Further, $(\|W_n+W_n'\|_1\,\|W_n+W_n'\|_\infty)<\infty$ and $(\|W_nW_n'\|_1\,\|W_nW_n'\|_\infty)<\infty$. Then, $\|W_nW_n'\|_2\leq(\|W_nW_n'\|_1\,\|W_nW_n'\|_\infty)^{1/2}<\infty$ and $\|W_n+W_n'\|_2\leq\|W_n\|_2+\|W_n'\|_2\leq 2\left(\|W_n\|_1\,\|W_n\|_\infty\right)^{1/2}<\infty$. Finally, given that $\sigma\in\Delta$ and Δ is compact, $\sup_k\sup_{\sigma\in\Delta}\sigma_k<\infty$. Then, $\exists\, c<\infty$ such that $\sup_{\sigma\in\Delta}\tau_{\max}(\Sigma(\sigma))<c$ and $\sup_{\sigma\in\Delta}\tau_{\max}(\mathbf{\Omega}(\sigma))<1+Tc<\infty$.

Next we prove that $\Omega^{-1}(\sigma)$ satisfies Property 1 for $A_k(\sigma,\overline{\sigma})=\Omega^{-1}(\sigma)(J_T\otimes\Sigma_k)\Omega(\overline{\sigma})$ and any $\sigma,\overline{\sigma}\in\Delta$. To this end, we need to prove that: *i)* $\mathbf{\Omega}^{-1}(\sigma)-\mathbf{\Omega}^{-1}(\overline{\sigma})=\sum_{k=1}^{3}(\sigma_k-\overline{\sigma}_k)A_k(\sigma,\overline{\sigma})$; *ii)* $\sup_{\sigma\in\Delta}\tau_{\max}(\mathbf{\Omega}^{-2}(\sigma))<c_\tau<\infty$ and *iii)* $\sup_{\sigma,\overline{\sigma}\in\Delta}\tau_{\max}(A_k(\sigma,\overline{\sigma})A_k'(\sigma,\overline{\sigma}))<c_\tau<\infty$ for $k=1,2,3$.

To prove *i)*, note that

$$\begin{aligned}\mathbf{\Omega}^{-1}(\sigma)-\mathbf{\Omega}^{-1}(\overline{\sigma}) &= \mathbf{\Omega}^{-1}(\sigma)\left[\mathbf{\Omega}(\overline{\sigma})-\mathbf{\Omega}(\sigma)\right]\mathbf{\Omega}^{-1}(\overline{\sigma})\\ &= \mathbf{\Omega}^{-1}(\sigma)\left[J_T\otimes(\Sigma(\overline{\sigma})-\Sigma(\sigma))\right]\mathbf{\Omega}^{-1}(\overline{\sigma})\\ &= \sum_{k=1}^{3}(\overline{\sigma}_k-\sigma_k)\mathbf{\Omega}^{-1}(\sigma)\left[J_T\otimes\Sigma_k\right]\mathbf{\Omega}^{-1}(\overline{\sigma}) = \sum_{k=1}^{3}(\sigma_k-\overline{\sigma}_k)A_k(\sigma,\overline{\sigma}).\end{aligned}$$

To prove *ii)*, note that, given that $\Sigma(\sigma)$ is a positive semidefinite matrix for all $\sigma\in\Delta$, $\inf_{\sigma\in\Delta}\tau_{\min}(\Sigma(\sigma))\geq 0$. Then, using that $\mathbf{\Omega}^{-1}(\sigma)$ is positive semidefinite for all $\sigma\in\Delta$ (since all the eigenvalues of J_T and $\Sigma(\sigma)$ are equal to or bigger than zero for all $\sigma\in\Delta$, all the eigenvalues of $\mathbf{\Omega}(\sigma)$ are bigger or equal than 1), $\sup_{\sigma\in\Delta}\tau_{\max}\left(\mathbf{\Omega}^{-2}(\sigma)\right)=\sup_{\sigma\in\Delta}\left[\tau_{\max}\left(\mathbf{\Omega}^{-1}(\sigma)\right)\right]^2 = \left[\inf_{\sigma\in\Delta}\tau_{\min}\left(\mathbf{\Omega}(\sigma)\right)\right]^{-2}\leq\left[1+T\inf_{\sigma\in\Delta}\tau_{\min}\left(\Sigma(\sigma)\right)\right]^{-2}\leq 1.$

To prove *iii)*, note that $\|.\|_2$ is a sub-multiplicative norm (see footnote 13). Thus,

$$\begin{aligned}
\sup_{\sigma,\overline{\sigma}\in\Delta} \tau_{\max}(A_k(\sigma,\overline{\sigma})A_k'(\sigma,\overline{\sigma})) &= \sup_{\sigma,\overline{\sigma}\in\Delta} \left\|\mathbf{\Omega}^{-1}(\sigma)\left(J_T\otimes\Sigma_k\right)\mathbf{\Omega}^{-1}(\overline{\sigma})\right\|_2^2 \\
&\leq \sup_{\sigma\in\Delta} \left\|\mathbf{\Omega}^{-1}(\sigma)\right\|_2^4 \left\|(J_T\otimes\Sigma_k)\right\|_2^2 \\
&\leq \sup_{\sigma\in\Delta} \left[\tau_{\max}\left(\mathbf{\Omega}^{-1}(\sigma)\right)\right]^4 \tau_{\max}\left(J_T\otimes\Sigma_k\right) \\
&\leq \sup_{\sigma\in\Delta} \left[\tau_{\max}\left(\mathbf{\Omega}^{-1}(\sigma)\right)\right]^4 T\tau_{\max}\left(\Sigma_k\right) < c_\tau < \infty
\end{aligned}$$

using ii) and $\sup_k \tau_{\max}\left(\Sigma_k\right) < c_\tau < \infty$ (from the first part of the proof).

□

Lemma A.10. *Let $B_n^{-1}(\sigma) = I_n - (I_n + T\Sigma(\sigma))^{-1}$ and $\Sigma(\sigma) = \sum_{k=1}^{3}\sigma_k\Sigma_k = \sigma_1 I_n + \sigma_2(W_n + W_n') + \sigma_2 W_n W_n'$, with W_n u.b.r.c.s. and $(\sigma_1,\sigma_2,\sigma_3) \in \Delta$, being Δ a compact space such that $\Sigma(\sigma)$ is positive semidefinite for any $\sigma \in \Delta$. Then, $B_n^{-1}(\sigma)$ satisfies Property 1 for $A_k(\sigma,\overline{\sigma}) = TB_n^*(\overline{\sigma})\left(J_T\otimes\Sigma_k\right)B_n^*(\sigma)$ and any $\sigma,\overline{\sigma}\in\Delta$ with $B_n^*(\sigma) = (I_n + T\Sigma(\sigma))^{-1}$.*

Proof. To prove that $B_n^{-1}(\sigma)$ satisfies Property 1 for $A_k(\sigma,\overline{\sigma}) = TB_n^*(\overline{\sigma})\left(J_T\otimes\Sigma_k\right)B_n^*(\sigma)$, we need to prove that: *i)* $B_n(\sigma) - B_n(\overline{\sigma}) = \sum_{k=1}^{3}(\sigma_k - \overline{\sigma}_k)A_k(\sigma,\overline{\sigma})$; *ii)* $\sup_{\sigma\in\Delta}\tau_{\max}(B_n^{-2}(\sigma)) < c_\tau < \infty$ and *iii)* $\sup_{\sigma,\overline{\sigma}\in\Delta}\tau_{\max}(A_k(\sigma,\overline{\sigma})A_k'(\sigma,\overline{\sigma})) < c_\tau < \infty$ for $k = 1,2,3$.

To prove *i)*, note that

$$\begin{aligned}
B_n^{-1}(\sigma) - B_n^{-1}(\overline{\sigma}) &= \left(I_n + T\Sigma(\overline{\sigma})\right)^{-1} - \left(I_n + T\Sigma(\sigma)\right)^{-1} \\
&= T\left(I_n + T\Sigma(\overline{\sigma})\right)^{-1}\left(\Sigma(\sigma) - \Sigma(\overline{\sigma})\right)\left(I_n + T\Sigma(\sigma)\right)^{-1} \\
&= T\sum_{k=1}^{3}(\sigma_k - \overline{\sigma})\left(I_n + T\Sigma(\overline{\sigma})\right)^{-1}\Sigma_k\left(I_n + T\Sigma(\sigma)\right)^{-1} \\
&= T\sum_{k=1}^{3}(\sigma_k - \overline{\sigma})B_n^*(\overline{\sigma})^{-1}\Sigma_k B_n^*(\sigma)
\end{aligned}$$

To prove *ii)*, note first that $B_n^{-1}(\sigma)$ is a positive semidefinite matrix for all $\sigma \in \Delta$, since $\inf_{\sigma\in\Delta}\tau_{\min}\left(I_n + T\Sigma(\sigma)\right) \geq 1$, $\sup_{\sigma\in\Delta}\tau_{\max}\left(I_n + T\Sigma(\sigma)\right)^{-1} \leq 1$, and $\inf_{\sigma\in\Delta}\tau_{\min}\left(B_n^{-1}(\sigma)\right) \geq 0$.

Note also that $\sup_{\sigma\in\Delta}\tau_{\max}(B_n^{-2}(\sigma)) = \left[\sup_{\sigma\in\Delta}\tau_{\max}(B_n^{-1}(\sigma))\right]^2$, and, since $(I_n+T\Sigma(\sigma))^{-1}$ is a positive semidefinite matrix and $\inf_{\sigma\in\Delta}\tau_{\min}\left[(I_n+T\Sigma(\sigma))^{-1}\right] \geq 0$, then $\sup_{\sigma\in\Delta}\tau_{\max}(B_n^{-1}(\sigma)) \leq 1-\inf_{\sigma\in\Delta}\tau_{\min}\left[(I_n+T\Sigma(\sigma))^{-1}\right] \leq 1$.

To prove *iii)*, note that, given that $\tau_{\max}(\Sigma_k) \leq c_\tau < \infty$ (proved in Lemma A.9) and $\sup_{\sigma\in\Delta}\|B_n^*(\sigma)\|_2 = \sup_{\sigma\in\Delta}\left\|(I_n+T\Sigma(\overline{\sigma}))^{-1}\right\|_2 \leq \left[\inf_{\sigma\in\Delta}\tau_{\min}\left(I_n+T\Sigma(\overline{\sigma})\right)\right]^{-1} \leq 1$, then

$$\sup_{\sigma,\overline{\sigma}\in\Delta}\tau_{\max}(A_k(\sigma,\overline{\sigma})A_k'(\sigma,\overline{\sigma})) = \sup_{\sigma,\overline{\sigma}\in\Delta}\|A_k'(\sigma,\overline{\sigma})\|_2^2 \leq \sup_{\sigma\in\Delta}\|B_n^*(\sigma)\|_2^4\sup_{\sigma,\overline{\sigma}\in\Delta}\|\Sigma_k\|_2^2 \leq c_\tau < \infty$$

□

Lemma A.11. *Let* $\mathbf{\Pi}_{\mathbf{a},\mathbf{b}}(\sigma) = \frac{1}{nT}\left\{\mathbf{a}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{b} - E\left[\mathbf{a}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{b}\right]\right\}$, *with* $\boldsymbol{a},\boldsymbol{b} = \boldsymbol{\eta},\mathbf{WY}$. *Under Assumptions 1 to 6,*

$$\mathbf{\Pi}_{\mathbf{a},\mathbf{b}}(\sigma)\xrightarrow{p}0 \text{ uniformly in } \sigma$$

Proof. We provide the proof for the most involved case, $\boldsymbol{a},\boldsymbol{b} = \mathbf{WY}$. The proof of the other cases is similar. From expression A.1 we have that

$$\begin{aligned}
\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{WY} = {} & Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{G}_0Y_{n,0} + 2Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\mathbb{X}\phi_0 \\
& + 2Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0\left(\upsilon_{n\mu}+W_n\upsilon_{n\alpha}\right) + 2Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} \\
& + \phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\mathbb{X}\phi_0 + 2\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0\left(\upsilon_{n\mu}+W_n\upsilon_{n\alpha}\right) \\
& + 2\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} + \left(\upsilon_{n\mu}'+\upsilon_{n\alpha}'W_n'\right)\mathbf{L}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0\left(\upsilon_{\mu}+W_n\upsilon_{n\alpha}\right) \\
& 2\left(\upsilon_{n\mu}'+\upsilon_{n\alpha}'W_n'\right)\mathbf{L}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}'\mathbf{C}_0\boldsymbol{\varepsilon}
\end{aligned}$$

The proof of the Lemma follows from proving that each of the previous summands, minus its expected value, converge in probability to 0 uniformly in σ. Following Magnus (1982), we have that $\mathbf{\Omega}^{-1}(\sigma) = I_T\otimes I_n - \frac{1}{T}J_T\otimes B_n^{-1}(\sigma)$ with $B_n^{-1}(\sigma) = I_n-(I_n+T\Sigma(\sigma))^{-1}$. Also, let $G_{nt} = \rho_0^tS_0^{-t}$, $C_{nt} = G_{nt}S_0^{-1}$ and $L_{nt} = \sum_{j=0}^{t-1}\rho_0^jS_0^{-(j+1)}$ (see also Lemma A.8). Thus, we may rewrite the summands in the previous expression as follows:

$$\begin{aligned}
Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\mathbb{X}\phi_0 &= \sum_{t=1}^{T}\sum_{j=1}^{t} Y_{n,0}'G_{nt}'W_n'W_nC_{nt-j}\mathbb{X}_{n,j}\phi_0 \\
&- \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{t} Y_{n,0}'G_{ns}'W_n'B_n^{-1}(\sigma)W_nC_{n,t-j}\mathbb{X}_{n,j}\phi_0
\end{aligned}$$

$$\begin{aligned}
Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} &= \sum_{t=1}^{T}\sum_{j=1}^{t} Y_{n,0}'G_{nt}'W_n'W_nC_{n,t-j}\varepsilon_{n,j} \\
&- \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{t} Y_{n,0}'G_{ns}'W_n'B_n^{-1}(\sigma)W_nC_{n,t-j}\varepsilon_{n,j}
\end{aligned}$$

$$\begin{aligned}
\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\mathbb{X}\phi_0 &= \sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{l=1}^{t} \phi_0'\mathbb{X}_{nj}'C_{n,t-j}'W_n'W_nC_{n,t-l}\mathbb{X}_{n,l}\phi_0 \\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\sum_{l=1}^{s} \phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'B_n^{-1}\left(\sigma\right)W_nC_{n,s-l}\mathbb{X}_{n,l}\phi_0
\end{aligned}$$

$$\begin{aligned}
\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} &= \sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{l=1}^{t} \phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'W_nC_{n,t-l}\varepsilon_{n,l} \\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\sum_{l=1}^{s} \phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'B_n^{-1}\left(\sigma\right)W_nC_{n,s-l}\varepsilon_{n,s-l}
\end{aligned}$$

$$
\begin{aligned}
\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0\left(\upsilon_{n\mu}+W_n\upsilon_{n\alpha}\right) &= \sum_{t=1}^{T}\sum_{j=1}^{t}\phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'W_nL_{nt}\upsilon_{n\mu} \\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nL_{ns}\upsilon_{n\mu} \\
&+ \sum_{t=1}^{T}\sum_{j=1}^{t}\phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'W_nL_{nt}W_n\upsilon_{n\alpha} \\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nL_{ns}W_n\upsilon_{n\alpha}
\end{aligned}
$$

And, finally,

$$
\begin{aligned}
\boldsymbol{\varepsilon}'\mathbf{C}_0'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0\left(\upsilon_{n\mu}+W_n\upsilon_{n\alpha}\right) &= \sum_{t=1}^{T}\sum_{j=1}^{t}\varepsilon_{n,j}'C_{n,t-j}'W_n'W_nL_{nt}\upsilon_{n\mu} \\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\varepsilon_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nL_{ns}\upsilon_{n\mu} \\
&+ \sum_{t=1}^{T}\sum_{j=1}^{t}\varepsilon_{n,j}'C_{n,t-j}'W_n'W_nL_{nt}W_n\upsilon_{n\alpha} \\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\varepsilon_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nL_{ns}W_n\upsilon_{n\alpha}
\end{aligned}
$$

Notice that each of the summands in the previous expressions, minus its expected value, can be written as $\frac{1}{nT}\left[a'P_nD^{-1}(\sigma)Q_nb - E\left(a'P_nD^{-1}(\sigma)Q_nb\right)\right]$ with $a,b = Y_{n,0}, X_{n,j}\beta_{10}, X_{n,j}\beta_{20}, \overline{X}_n\pi_{\mu 0}, \overline{X}_n\pi_{\alpha 0}, \varepsilon_{nt}, \upsilon_{n\mu}, \upsilon_{n\alpha}$; $P_n, Q_n = W_nL_{nt}, W_nG_{nt}, W_nC_{nt}, W_nC_{nt}W_n$, $\mathbf{WL}_0, \mathbf{WG}_0, \mathbf{WC}_0$; and $D^{-1}(\sigma) = \mathbf{I}_{nT}, \boldsymbol{\Omega}^{-1}(\sigma)$, $B_n^{-1}(\sigma)$. This means that, if we can apply Lemma *A.7*, the Lemma is proved. To apply these lemmas, P_n and Q_n must be u.b.r.c.s. which is proved for all the cases in Lemma *A.8*. Also, $D(\sigma)^{-1}$ must satisfy Property 1, which is proved in Lemma *A.9* for $\boldsymbol{\Omega}^{-1}(\sigma)$ and in Lemma *A.10* for $B_n^{-1}(\sigma)$. Lastly, a and b must be an i.i.d. sequence with finite second moments, which is guaranteed by Assumptions 1 and 2.

□

Lemma A.12. *Let*

$$\boldsymbol{\Upsilon}_{\mathbf{a},\mathbf{b}}(\sigma) = \mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{a}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{b}}(\sigma) - E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{a}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{b}}(\sigma)\right]$$

with $\mathbf{a},\mathbf{b} = \boldsymbol{\eta}, \mathbf{WY}$ *and* $\mathbf{Q}_{\mathbf{A},\mathbf{B}}(\delta) = \frac{1}{nT}\mathbf{A}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{B}$. *Under Assumptions 1 to 6,*

$$\boldsymbol{\Upsilon}_{\mathbf{a},\mathbf{b}}(\sigma) \xrightarrow{p} 0 \text{ uniformly in } \sigma$$

Proof. The proof of this Lemma is similar to the proof of Lemma *A*.11. Thus, we only provide the proof for the case $\mathbf{a} = \mathbf{b} = \mathbf{WY}$ (the others are similar). We start by decomposing $\boldsymbol{\Upsilon}_{\mathbf{WY},\mathbf{WY}}(\sigma)$:

$$\begin{aligned}
\boldsymbol{\Upsilon}_{\mathbf{WY},\mathbf{WY}}(\sigma) &= \mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right] \\
&= \left\{\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\}\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) + \\
&E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\left\{\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right] - \mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right\}\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) + \\
&E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\left\{\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\}
\end{aligned}$$

First we need to prove that $\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]$ and $\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]$ converges elementwise to 0 uniformly in σ. The proof follows the same steps as that of Lemma *A*.11 (note that all the elements of $\tilde{\mathbf{X}}$ are in $\mathbf{WY}$, including, as shown in 3.1 and *A*.1, $\mathbf{Y}_{-1}$), so it is not reproduced here. The elementwise convergence implies, by the Slutsky theorem, that $\left\|\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]\right\|_F = o_p(1)$ uniformly in σ and $\left\|\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right)\right]\right\|_F = o_p(1)$ uniformly in σ. Then, by using the properties of the matrix norm, $\left\|\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]\right\|_2 \leq \left\|\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]\right\|_F = o_p(1)$ uniformly in σ and $\left\|\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right)\right]\right\|_2 \leq \left\|\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right)\right]\right\|_F = o_p(1)$ uniformly in σ.

Next we prove that $\sup_\sigma \left\|\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\right\|_2 = O(1)$ and $\sup_\sigma \left\|E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\|_2 = O(1)$.

Let us first consider

$$\sup_{\sigma} \left\| \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right) \right]^{-1} \right\|_2 = \sup_{\sigma} \tau_{\max} \left\{ E\left(\frac{1}{nT} \tilde{\mathbf{X}}' \mathbf{\Omega}^{-1}(\sigma) \tilde{\mathbf{X}} \right)^{-1} \right\}$$
$$= \left(\inf_{\sigma} \tau_{\min} \left\{ \frac{1}{nT} E\left(\tilde{\mathbf{X}}' \mathbf{\Omega}^{-1}(\sigma) \tilde{\mathbf{X}} \right) \right\} \right)^{-1}$$

Note that, since $\mathbf{\Omega}^{-1}(\sigma)$ is a symmetric definite positive matrix, we can apply Lemma $A.1$ to obtain

$$\inf_{\sigma} \tau_{\min} \left\{ \frac{1}{nT} E\left(\tilde{\mathbf{X}}' \mathbf{\Omega}^{-1}(\sigma) \tilde{\mathbf{X}} \right) \right\} \geq \inf_{\sigma} \tau_{\min} \left\{ \mathbf{\Omega}^{-1}(\sigma) \right\} \tau_{\min} \left\{ \frac{1}{nT} E\left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right) \right\}$$
$$\geq \left[\sup_{\sigma} \tau_{\max} \left(\mathbf{\Omega}(\sigma) \right) \right]^{-1} \tau_{\min} \left\{ \frac{1}{nT} E\left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right) \right\}$$

From Lemma $A.9$, $\sup_{\sigma} \tau_{\max} \left(\mathbf{\Omega}(\sigma) \right) < c_{\tau} < \infty$ and, from Assumption 6, $\tau_{\min} \left\{ \frac{1}{nT} E\left(\tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right) \right\} > 0$ for sufficiently large n. Then, $\sup_{\sigma} \left\| \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right) \right]^{-1} \right\|_2 < C < \infty \Rightarrow \sup_{\sigma} \left\| \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right) \right]^{-1} \right\|_2 = O(1)$.

As for $\sup_{\sigma} \left\| E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right] \right\|_2$, notice that $\left\| E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right] \right\|_2 \leq \left\| E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right] \right\|_F$ and

$$\left\| E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right] \right\|_F = tr \left[E\left(\frac{1}{nT} \mathbf{Y}'\mathbf{W}' \mathbf{\Omega}^{-1}(\sigma) \tilde{\mathbf{X}} \right) E\left(\frac{1}{nT} \tilde{\mathbf{X}} \mathbf{\Omega}^{-1}(\sigma) \mathbf{WY} \right) \right]^{1/2}$$
$$= \left(\sum_{k=1}^{K} \left[E\left(\frac{1}{nT} \tilde{\mathbf{X}}'_k \mathbf{\Omega}^{-1}(\sigma) \mathbf{WY} \right) \right]^2 \right)^{1/2},$$

where the last expression is $O(1)$ uniformly in σ if $\sup_{\sigma} \left| E\left(\tilde{\mathbf{X}}'_k \mathbf{\Omega}^{-1}(\sigma) \mathbf{WY} \right) \right| = O(n)$. To prove that $\sup_{\sigma} \left| E\left(\tilde{\mathbf{X}}'_k \mathbf{\Omega}^{-1}(\sigma) \mathbf{WY} \right) \right| = O(n)$, we follow the same steps as in Lemma $A.11$. Thus, we decompose the term $\tilde{\mathbf{X}}'_k \mathbf{\Omega}^{-1}(\sigma) \mathbf{WY}$ in a finite sum of terms that can be written as $a' P_n D^{-1}(\sigma) Q_n b$, with a, b, P_n, Q_n and $D^{-1}(\sigma)$ satisfying the conditions of Lemma $A.6$. This provides the proof that $\sup_{\sigma} \left| E\left(a' P_n D^{-1}(\sigma) Q_n b \right) \right| = O(n)$ and so that of $\sup_{\sigma} \left| E\left(\tilde{\mathbf{X}}'_k \mathbf{\Omega}(\sigma)^{-1} \mathbf{WY} \right) \right| = O(n)$.

Moreover, given that $\left\| \mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right) \right] \right\|_2 = o_p(1)$ uniformly in σ and

$\sup_{\sigma} \left\| E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right) \right\|_2 = O(1)$, then $\sup_{\sigma} \left\| \mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right\|_2 = O_p(1)$.

Finally, we need to prove that $\sup_{\sigma} \left\| \mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right\|_2 = O_p(1)$. To this end, notice that

$$
\begin{aligned}
\sup_{\sigma} \left\| \mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right\|_2 &= \sup_{\sigma} \tau_{\max}\left(\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right) = \sup_{\sigma} \tau_{\max}\left(\left[\frac{1}{nT} \tilde{\mathbf{X}}' \mathbf{\Omega}^{-1}(\sigma) \tilde{\mathbf{X}} \right]^{-1} \right) \\
&= \left[\inf_{\sigma} \tau_{\min}\left(\frac{1}{nT} \tilde{\mathbf{X}}' \mathbf{\Omega}^{-1}(\sigma) \tilde{\mathbf{X}} \right) \right]^{-1} \\
&\leq \left[\inf_{\sigma} \tau_{\min}(\mathbf{\Omega}^{-1}(\sigma)) \tau_{\min}\left(\frac{1}{nT} \tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right) \right]^{-1} \\
&\leq \sup_{\sigma} \tau_{\max}(\mathbf{\Omega}(\sigma)) \left[\tau_{\min}\left(\frac{1}{nT} \tilde{\mathbf{X}}' \tilde{\mathbf{X}} \right) \right]^{-1},
\end{aligned}
$$

which is $O_p(1)$ given that, by Lemma A.9, $\sup_{\sigma} \tau_{\max}(\mathbf{\Omega}(\sigma)) < c_\tau < \infty$, and, by Assumption 6, $\tau_{min}\left(\frac{1}{nT} \tilde{X}' \tilde{X} \right) > 0$ almost surely for sufficiently large n.

Then,

$$
\begin{aligned}
&\left\{ \mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right] \right\} \mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) = o_p(1) O_p(1) O_p(1) \\
&E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right] \mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \left\{ \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right) \right] - \mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right\} \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right) \right]^{-1} \mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) = O_p(1) o_p(1) O_p(1)
\end{aligned}
$$

and

$$
E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right] \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) \right) \right]^{-1} \left\{ \mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \right] \right\} = O(1) O(1) o_p(1),
$$

all the cases uniformly in σ. This proves that $\Upsilon_{\mathbf{WY},\mathbf{WY}}(\sigma) = o_p(1)$ uniformly in σ (and the proof is analogous for the rest of cases). □

Lemma A.13. *Under Assumptions 1 to 7,*

$$
\frac{1}{nT} \left[\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial \psi \partial \psi'} - E\left(\frac{1}{nT} \frac{\partial^2 \mathcal{L}(\psi_0)}{\partial \psi \partial \psi'} \right) \right] = o_p(1)
$$

Proof. It can be proved, following the proof of Lemmas A.11 and A.12, that each element of the matrix $\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial \psi \partial \psi'} - E\left(\frac{1}{nT} \frac{\partial^2 \mathcal{L}(\psi_0)}{\partial \psi \partial \psi'} \right)$ can be written as the finite sum of $\frac{1}{nT} \left[a' P_n D^{-1}(\sigma_0) Q_n b - E\left(a' P_n D^{-1}(\sigma_0) Q_n b \right) \right]$, with $a, b = Y_{n,0}$, $\mathbb{X}_{n,j}$, ε_{nt}, $\upsilon_{n\mu}$ and $\upsilon_{n\alpha}$;

$P_n, Q_n = W_n L_{nt},\ W_n G_{nt},\ W_n C_{nt}$; and $D^{-1}(\sigma_0) = \boldsymbol{A}_i \left[\boldsymbol{B}_\kappa \boldsymbol{A}_j \boldsymbol{B}_\varrho + \boldsymbol{B}_\varrho \boldsymbol{A}_j \boldsymbol{B}_\kappa\right] \boldsymbol{A}_i,\ \boldsymbol{A}_1 \boldsymbol{B}_\kappa \boldsymbol{A}_j \boldsymbol{B}_\varrho + \boldsymbol{B}_\varrho \boldsymbol{A}_j \boldsymbol{B}_\kappa \boldsymbol{A}_1$ for $i, j = 0, 1$ and $\kappa, \varrho = 0, 1, 2, 3$ with $\boldsymbol{A}_0 = \boldsymbol{B}_0 = I_n$, $\boldsymbol{A}_1 = B_n^{-1}(\sigma_0)$, $\boldsymbol{B}_\kappa = \Sigma_\kappa$ for $\kappa = 1, 2, 3$ and $B_n^{-1}(\sigma_0)$ and Σ_κ defined in Lemmas A.9 and A.10 (see Appendix C for details on the elements of $\dfrac{\partial^2 \mathcal{L}(\psi_0)}{\partial \psi \partial \psi'}$). This means that if a, b, P_n, Q_n and $D^{-1}(\sigma_0)$ satisfy the conditions of Lemma A.7 in all the cases, then $\dfrac{1}{nT}\left[a' P_n D(\sigma_0) Q_m b - E(a' P_n D(\sigma_0) Q_m b)\right] = o_p(1)$ in all the cases, which proves the Lemma. Notice also that we do not need to prove the uniform convergence because these second derivatives and their expectations are evaluated at the true parameters of the model. It is therefore enough to prove that $D^{-1}(\sigma_0)$ is a symmetric matrix with $\tau_{\max}\left(D^{-2}(\sigma_0)\right) < \infty$.

Firstly, Lemmas A.11 and A.12 show that all the possible cases of a, b, P_n and Q_n satisfy the conditions of Lemma A.7. Secondly, given that $D^{-1}(\sigma_0)$ is by definition symmetric, $\|I_n\|_2 = 1$, $\max_\kappa \|\Sigma_\kappa\|_2 < c_\tau < \infty$ (the bound is provided by Lemma A.9) and $\left\|B_n^{-1}(\sigma_0)\right\|_2 = \tau_{\max}\left(B_n^{-1}(\sigma_0)\right) < c_\tau < \infty$ (the bound is provided by Lemma A.10),

$$\tau_{\max}\left(D^{-2}(\sigma_0)\right) = \left\|D^{-1}(\sigma_0)\right\|_2^2 \leq 2 \max_i \|\boldsymbol{A}_i\|_2^6 \max_\kappa \|\boldsymbol{B}_\kappa\|_2^4 \leq c_\tau < \infty.$$

□

Lemma A.14. *Let $a_t = \{a_{i,t}\}_{i=1}^n$, $b_t = \{b_{i,t}\}_{i=1}^n$ be $n \times 1$ zero-mean random vectors independent in i. Let us also define $Q_n = \sum_{t=1}^T a_t' P_{t,n} b_t$ with $P_{t,n}$ $n \times n$ real matrices and $T < \infty$. Lastly, let us denote $\mu_{Q_n} = E(Q_n)$ and $s_{Q_n}^2 = E[(Q_n - \mu_{Q_n})^2]$. If $P_{t,n}$ for $t = 1, ..., T$ are u.b.r.c.s. and $\{(a_t, b_t)\}_{t=1}^T$ has $4 + \epsilon_1$ finite moments for some $\epsilon_1 > 0$ and $n^{-1} s_{Q_n}^2 \geq c > 0$, then*

$$\frac{Q_n - \mu_{Q_n}}{s_{Q_n}} \xrightarrow{d} N(0, 1)$$

Proof. The proof of this Lemma follows the proof of Theorem 1 in Kelejian and Prucha (2001, p. 243). First note that, given the independence in i of a_t and b_t, $\mu_{Q_n} = \sum_{t=1}^T \sum_{i=1}^n P_{t,n}[i,i] E(a_{i,t} b_{i,t})$, where we use the somewhat abusive notation $P_{t,n}[i,j]$ to refer to the row i and column j element of the matrix $P_{t,n}$. Notice also that $Q_n - \mu_{Q_n} = \sum_{t=1}^{T}\sum_{i=1}^{n} Y_{i,t}$ with

$$Y_{i,t} = P_{t,n}[i,i]\left(a_{i,t}b_{i,t} - E(a_{i,t}b_{i,t})\right) + a_{i,t}\sum_{j=1}^{i-1} P_{t,n}[j,i]b_{j,t} + b_{i,t}\sum_{j=1}^{i-1} P_{t,n}[i,j]a_{j,t}$$

for $i = 1, 2, ..., n$.

Let us now consider the $\sigma-$fields $\mathcal{F}_{0,n} = \{\emptyset, \Omega\}$ and $\mathcal{F}_{i,n} = \sigma\left(a_i, b_i, a_{i-1}, b_{i-1}, ..., a_1, b_1,\right)$, with $a_i = \{a_{i,t}\}_{t=1}^{T}$, $b_i = \{b_{i,t}\}_{t=1}^{T}$ and $1 \leq i \leq n$. By construction, $\mathcal{F}_{i-1,n} \subset \mathcal{F}_{i,n}$ and $Y_{i,t}$ is $\mathcal{F}_{i,n}-$measurable. It can also be shown that $E\left(Y_{i,t}|\mathcal{F}_{i-1,n}\right) = 0$. Therefore, $\{Y_{i,t}, \mathcal{F}_{i,n}, 1 \leq i \leq n, n \geq 1\}$ forms a martingale difference array and so $s_{Q_n}^2 = \sum_{i=1}^{n}\left(\sum_{t=1}^{T} E\left(Y_{i,t}^2\right) + 2\sum_{t=2}^{T}\sum_{s=1}^{t-1} E\left(Y_{i,t}Y_{i,s}\right)\right)$. Thus, the expression for the variance of Q_n follows from

$$\begin{aligned} E\left(Y_{i,t}Y_{i,s}\right) &= P_{t,n}[i,i]P_{s,n}[i,i]\sigma_{c,t,s}^{(2)} + \sigma_{a,t,s}^2\sigma_{b,t,s}^2\sum_{j=1}^{i-1}(P_{t,n}[i,j]P_{s,n}[i,j] + P_{t,n}[j,i]P_{s,n}[j,i]) \\ &\quad + \sigma_{c,t,s}\sigma_{c,s,t}\sum_{j=1}^{i-1}\left(P_{t,n}[i,j]P_{s,n}[j,i] + P_{s,n}[i,j]P_{t,n}[j,i]\right) \end{aligned} \tag{A.3}$$

with $\sigma_{c,t,s} = E\left(a_{i,t}b_{i,s}\right)$, $\sigma_{c,t,s}^{(2)} = E\left[(a_{i,t}b_{i,t} - \sigma_{c,t,t})(a_{i,s}b_{i,s} - \sigma_{c,s,s})\right]$, $\sigma_{a,t,s}^2 = E\left[a_{i,t}a_{i,s}\right]$ and $\sigma_{b,t,s}^2 = E\left[b_{i,t}b_{i,s}\right]$. Also, if we define $X_{i,t} = Y_{i,t}/s_{Q_n}$, then $\{X_{i,t}, \mathcal{F}_{i,n}, 1 \leq i \leq n, n \geq 1\}$ forms a martingale difference array.

In what follows we prove that

$$\frac{Q_n - \mu_{Q_n}}{s_{Q_n}} = \sum_{i=1}^{n}\sum_{t=1}^{T} X_{i,t} \xrightarrow{d} N(0,1)$$

by showing that $X_{i,n} = \sum_{t=1}^{T} X_{i,t}$ satisfies the remaining conditions of the Central Limit Theorem of Gänsler and Stute (1977, p. 365). In particular, we demonstrate that $X_{i,n}$ satisfies the condition:

$$\sum_{i=1}^{k_n} E\left\{E\left[|X_{i,n}|^{2+\delta}\middle| \digamma_{i-1,n}\right]\right\} \longrightarrow 0 \tag{A.4}$$

for some $\delta > 0$, which in turn is sufficient for

$$\sum_{i=1}^{k_n} E\left[|X_{i,n}|^2 \mathbf{1}\left(|X_{i,n} > \varepsilon|\right)\middle| \digamma_{i-1,n}\right] \xrightarrow{p} 0$$

for all $\varepsilon > 0$ and with $\mathbf{1}(\cdot)$ being an indicator function. Then we prove that $X_{i,n}$ satisfies

$$\sum_{i=1}^{k_n} E\left[X_{i,n}^2\middle| \digamma_{i-1,n}\right] \xrightarrow{p} 1 \tag{A.5}$$

Let us take $0 < \delta \leq \epsilon_1/2$. We note that, under the maintained moment assumptions on $\{(a_t, b_t)\}_{t=1}^{T}$, there exists a finite constant, $C_e \geq 1$, such that $E\left(\left|a_{i,t}^{r_1} b_{i,t}^{r_2} a_{i,s}^{r_3} b_{i,s}^{r_4}\right|\right) \leq C_e$ for $\sum_{l=1}^{4} r_l \leq 4 + 2\delta$, $r_l \geq 0$, $t = 1, 2, ..., T$ and $i = 1, 2, ..., n$. We further note that, under the maintained assumptions on the matrices $P_{t,n}$, there exists a finite constant, $C_m \geq 1$, such that $\sum_{j=1}^{n} (|P_{t,n}[i,j]| + |P_{t,n}[j,i]|) < C_m$ for $t = 1, 2, ..., T$ and $i = 1, ..., n$. Lastly, note that $\sum_{j=1}^{n} (|P_{t,n}[i,j]| + |P_{t,n}[j,i]|)^r \leq C_m^r$ for $r \geq 1$ and

$$\sum_{j=1}^{n} (|P_{t,n}[i,j]| + |P_{t,n}[j,i]|)(|P_{s,n}[k,j]| + |P_{s,n}[j,k]|) \leq \\ \sum_{j=1}^{n} (|P_{t,n}[i,j]| + |P_{t,n}[j,i]|) \sum_{j=1}^{n} (|P_{s,n}[k,j]| + |P_{s,n}[j,k]|) \leq C_m^2$$

for $t, s = 1, 2, ..., T$.

Let us now take $q = 2 + \delta$ and let $1/q + 1/p = 1$. We note that $\left|\sum_{t=1}^{T} Y_{i,t}\right|^q \leq T^q \sum_{t=1}^{T} |Y_{i,t}|^q$.

Also, using the triangle and Hölder's inequalities, we have that

$$
\begin{aligned}
|Y_{i,t}|^q &= \left| P_{t,n}[i,i]\left(a_{i,t}b_{i,t} - \sigma_{c,t,t}\right)) + a_{i,t}\sum_{j=1}^{i-1} P_{t,n}[j,i]b_{j,t} + b_{i,t}\sum_{j=1}^{i-1} P_{t,n}[i,j]a_{j,t} \right|^q \\
&\leq 2^q \left| 1/2 P_{t,n}[i,i]^{1/p} P_{t,n}[i,i]^{1/q}\left(a_{i,t}b_{i,t} - \sigma_{c,t,t}\right) + a_{i,t}\sum_{j=1}^{i-1} P_{t,n}[j,i]^{1/p} P_{t,n}[j,i]^{1/q} b_{j,t} \right|^q + \\
&2^q \left| 1/2 P_{t,n}[i,i]^{1/p} P_{t,n}[i,i]^{1/q}\left(a_{i,t}b_{i,t} - \sigma_{c,t,t}\right) + b_{i,t}\sum_{j=1}^{i-1} P_{t,n}[i,j]^{1/p} P_{t,n}[i,j]^{1/q} a_{j,t} \right|^q \\
&\leq 2^q \left[\sum_{j=1}^{i} |P_{t,n}[j,i]|\right]^{q/p} \left| 2^{-q}\left|P_{t,n}[i,i]\right|\left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q + |a_{i,t}|^q \sum_{j=1}^{i-1}\left|P_{t,n}[j,i]\right|\left|b_{j,t}\right|^q \right|^{q/q} + \\
&2^q \left[\sum_{j=1}^{i} |P_{t,n}[j,i]|\right]^{q/p} \left| 2^{-q}\left|P_{t,n}[i,i]\right|\left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q + |b_{i,t}|^q \sum_{j=1}^{i-1}\left|P_{t,n}[i,j]\right|\left|a_{j,t}\right|^q \right|^{q/q} \\
&\leq 2^q C_m^{q/p} \left(2^{-q}\left|P_{t,n}[i,i]\right|\left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q + |a_{i,t}|^q \sum_{j=1}^{i-1}\left|P_{t,n}[j,i]\right|\left|b_{j,t}\right|^q \right) + \\
&2^q C_m^{q/p} \left(2^{-q}\left|P_{t,n}[i,i]\right|\left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q + |b_{i,t}|^q \sum_{j=1}^{i-1}\left|P_{t,n}[i,j]\right|\left|a_{j,t}\right|^q \right).
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
&\sum_{i=1}^{n} E\left\{E\left[|Y1_{i,t}|^q \middle| \mathcal{F}_{i-1,n}\right]\right\} \leq \\
&\quad \sum_{i=1}^{n} 2^q C_m^{q/p} \left(\left|P_{t,n}[i,i]\right| E\left[\left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q\right] + E\left[|a_{i,t}|^q\right] \sum_{j=1}^{i-1}\left|P_{t,n}[j,i]\right| E\left[|b_{j,t}|^q\right] \right) \\
&\quad + \sum_{i=1}^{n} 2^q C_m^{q/p} \left(\left|P_{t,n}[i,i]\right| E\left[\left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q\right] + E\left[|b_{i,t}|^q\right] \sum_{j=1}^{i-1}\left|P_{t,n}[i,j]\right| E\left[|a_{j,t}|^q\right] \right) \\
&\quad \leq \sum_{i=1}^{n} 2^q C_m^{q/p} C_e \left(2\left|P_{t,n}[i,i]\right| + \sum_{j=1}^{i-1}\left|P_{t,n}[j,i]\right| + \sum_{j=1}^{i-1}\left|P_{t,n}[i,j]\right| \right) \leq n 2^{q+1} C_m^{q/p+1} C_e
\end{aligned}
$$

Thus,

$$\sum_{i=1}^{n} E\left\{E\left[\left|X_{i,n}\right|^{q}\right| F_{i-1,n}]\right\} = \frac{1}{s_{Q_n}^{q}} \sum_{i=1}^{n}\sum_{t=1}^{T} E\left\{E\left[\left|Y_{i,t}\right|^{q}\right| F_{i-1,n}]\right\}$$
$$= \frac{1}{\left[n^{-1}s_{Q_n}^{2}\right]^{1+\delta/2}} \frac{1}{n^{1+\delta/2}} \sum_{i=1}^{n}\sum_{t=1}^{T} E\left\{E\left[\left|Y_{i,t}\right|^{q}\right| F_{i-1,n}]\right\}$$
$$\leq \frac{1}{\left[n^{-1}s_{Q_n}^{2}\right]^{1+\delta/2}} \left\{\frac{1}{n^{\delta/2}} 2^{q+1} C_m^{q/p+1} C_e\right\}$$

Since $n^{-1}s_{Q_n}^{2} \geq c > 0$, the right-hand side of the last inequality goes to zero as $n \to \infty$, which proves that condition A.4 holds.

Now, using $s_{Q_n}^{2} = \sum_{i=1}^{n}\left(\sum_{t=1}^{T} E\left(Y_{i,t}^{2}\right) + 2\sum_{t=2}^{T}\sum_{s=1}^{t-1} E\left(Y_{i,t}Y_{i,s}\right)\right)$ and the definition of $X_{i,n}$ we obtain that

$$\sum_{i=1}^{n} E\left[X_{i,n}^{2}\right| F_{i-1,n}] - 1 = \frac{1}{n^{-1}s_{Q_n}^{2}} \frac{1}{n} \sum_{i=1}^{n}\sum_{t=1}^{T}\left[E\left(Y_{i,t}^{2}\right| F_{i-1,n}) - E\left(Y_{i,t}^{2}\right)\right] +$$
$$\frac{2}{n^{-1}s_{Q_n}^{2}} \frac{1}{n} \sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left[E\left(Y_{i,t}Y_{i,s}\right| F_{i-1,n}) - E\left(Y_{i,t}Y_{i,s}\right)\right]$$

This means that, since $n^{-1}s_{Q_n}^{2} \geq c > 0$, we can prove condition A.5 by proving that

$$\frac{1}{n}\sum_{i=1}^{n}\left[E\left(Y_{i,t}Y_{i,s}\right| F_{i-1,n}) - E\left(Y_{i,t}Y_{i,s}\right)\right] \xrightarrow{p} 0$$

for $t, s = 1, 2, ..., T$. We start the proof by noting that, since $(a_{i,t}, b_{i,t})$ are independent with zero mean, it follows that

$$\begin{aligned}
\left[E\left(Y_{i,t}Y_{i,s}\right|\mathcal{F}_{i-1,n}) - E\left(Y_{i,t}Y_{i,s}\right)\right] =& \sigma_{c,a,t,s}P_{t,n}[i,i]\sum_{j=1}^{i-1}P_{s,n}[j,i]b_{j,s} + \sigma_{c,b,t,s}P_{t,n}[i,i]\sum_{j=1}^{i-1}P_{s,n}[i,j]a_{j,s}+ \\
& \sigma_{c,a,s,t}P_{s,n}[i,i]\sum_{j=1}^{i-1}P_{t,n}[j,i]b_{j,t} + \sigma_{c,b,s,t}P_{s,n}[i,i]\sum_{j=1}^{i-1}P_{t,n}[i,j]a_{j,t}+ \\
& \sigma_{a,t,s}\sum_{j=1}^{i-1}\sum_{l=1}^{i-1}P_{t,n}[j,i]P_{s,n}[l,i]\left[b_{j,t}b_{l,s} - 1(j=l)\sigma_{b,t,s}\right] + \\
& \sigma_{c,t,s}\sum_{j=1}^{i-1}\sum_{l=1}^{i-1}P_{t,n}[j,i]P_{s,n}[i,l]\left[b_{j,t}a_{l,s} - 1(j=l)\sigma_{c,s,t}\right] + \\
& \sigma_{c,s,t}\sum_{j=1}^{i-1}\sum_{l=1}^{i-1}P_{s,n}[j,i]P_{t,n}[i,l]\left[b_{j,s}a_{l,t} - 1(j=l)\sigma_{c,t,s}\right] + \\
& \sigma_{b,t,s}\sum_{j=1}^{i-1}\sum_{l=1}^{i-1}P_{t,n}[i,j]P_{s,n}[i,l]\left[a_{j,t}a_{l,s} - 1(j=l)\sigma_{a,t,s}\right]
\end{aligned}$$

with $\sigma_{c,a,t,s} = E(a_{i,t}b_{i,t}a_{i,s})$ and $\sigma_{c,b,t,s} = E(a_{i,t}b_{i,t}b_{i,s})$, and so

$$\frac{1}{n}\sum_{i=1}^{n}\left[E\left(Y_{i,t}Y_{i,s}\right|\mathcal{F}_{i-1,n}) - E\left(Y_{i,t}Y_{i,s}\right)\right] = \sum_{k=1}^{8}H_{k,n}$$

where the subindex $1,\ldots,8$ indicates, in order of appearance, a summand in the expression above. Thus, to prove that $\frac{1}{s_{Q_n}^2}\sum_{i=1}^{n}\left[E\left(Y_{i,t}Y_{i,s}\right|\mathcal{F}_{i-1,n}) - E\left(Y_{i,t}Y_{i,s}\right)\right] \xrightarrow{p} 0$, next we prove that $H_{k,n} \xrightarrow{p} 0$ for $k = 1, ..., 8$.

To prove that $H_{1,n} = \sum_{i=1}^{n-1}\varphi_{i,n}b_{i,s}$ with $\varphi_{i,n} = n^{-1}\sigma_{c,a,t,s}P_{t,n}[i,i]\sum_{j=i+1}^{n}P_{s,n}[j,i]$, notice that, given that the $b_{i,s}$ are independent with mean zero, $E\left|b_{i,s}\right|^{1+\delta} \leq C_e$ for $\delta > 0$, $\lim\sup_{n\to\infty}\sum_{i=1}^{n-1}\varphi_{i,n} =$

$$\limsup_{n\to\infty} n^{-1}\sigma_{c,a,t,s}P_{t,n}[i,i]\sum_{j=i+1}^{n}P_{s,n}[j,i] \leq C_e C_m^2 < \infty, \text{ and}$$

$$\begin{aligned}\limsup_{n\to\infty}\sum_{i=1}^{n-1}\varphi_{i,n}^2 &= \limsup_{n\to\infty} n^{-2}\sigma_{c,a,t,s}^2\sum_{i=1}^{n-1}P_{t,n}[i,i]^2\left[\sum_{j=i+1}^{n}P_{s,n}[j,i]\right]^2 \\ &\leq n^{-1}C_e^2C_m^2 n^{-1}\sum_{i=1}^{n-1}C_m^2 \leq n^{-1}C_e^2C_m^4 \to 0\end{aligned}$$

Then, $H_{1,n}\xrightarrow{p}0$ by Davidson (1994, p. 299). Further, the cases $H_{k,n}$ for $k=2,3,4$ can be proved in the same way.

For $H_{5,n}$, notice that

$$\begin{aligned}H_{5,n} =&\sigma_{a,t,s}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{l=1}^{i-1}P_{t,n}[j,i]P_{s,n}[l,i]\left[b_{j,t}b_{l,s}-1(j=l)\sigma_{b,t,s}\right] \\ =&\sigma_{a,t,s}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{i-1}P_{t,n}[j,i]P_{s,n}[j,i]\left[b_{j,t}b_{j,s}-\sigma_{b,t,s}\right]+\sigma_{a,t,s}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{l=1}^{j-1}P_{t,n}[j,i]P_{s,n}[l,i]b_{j,t}b_{l,s}+ \\ &\sigma_{a,t,s}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{l=1}^{j-1}P_{t,n}[l,i]P_{s,n}[j,i]b_{l,t}b_{j,s} \\ =&H1_{5,n}+H2_{5,n}+H3_{5,n}.\end{aligned}$$

To prove that $H1_{5,n}\xrightarrow{p}0$, we follow the same steps as in $H_{1,n}$. Notice that $H1_{5,n} = \sum_{i=1}^{n-1}\phi_{i,n}\left(b_{i,t}b_{i,s}-\sigma_{b,t,s}\right)$ with $\phi_{i,n} = n^{-1}\sigma_{a,t,s}\sum_{j=i+1}^{n}P_{t,n}[j,i]P_{s,n}[j,i]$. Then, given that $(b_{i,t}b_{i,s}-\sigma_{b,t,s})$ are independent with mean zero, $E\left|b_{i,t}b_{i,s}-\sigma_{b,t,s}\right|^{1+\delta} \leq C_e$ for $\delta>0$, $\limsup_{n\to\infty}\sum_{i=1}^{n-1}\phi_{i,n} = \limsup_{n\to\infty} n^{-1}\sigma_{a,t,s}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}P_{t,n}[j,i]P_{s,n}[j,i] \leq \sigma_{a,t,s}C_m^2$ and $\limsup_{n\to\infty}\sum_{i=1}^{n-1}\phi_{i,n}^2 = \limsup_{n\to\infty} n^{-2}\sigma_{a,t,s}^2\sum_{i=1}^{n-1}\left[\sum_{j=i+1}^{n}P_{t,n}[j,i]P_{s,n}[j,i]\right]^2 \leq \limsup_{n\to\infty} n^{-1}\sigma_{a,t,s}^2C_m^4 = 0$. Thus, $H1_{5,n}\xrightarrow{p}0$ by Davidson (1994, p. 299).

Similarly, for $H2_{5,n}$, given that the $b_{i,t}b_{j,s}$ are independent with zero mean, it is not difficult to see that

$$
\begin{aligned}
E\left(H2_{5,n}^{2}\right) \leq & n^{-2}C_e\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{l=1}^{j-1}P_{t,n}[j,i]^2P_{s,n}[l,i]^2+4n^{-2}C_e\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{r=1}^{j-1}P_{t,n}[j,i]P_{t,n}[r,i]P_{s,n}[j,i]P_{s,n}[r,i]\\
&+2n^{-2}C_e\sum_{i=1}^{n}\sum_{k=1}^{i-1}\sum_{j=1}^{k-1}\sum_{r=1}^{k-1}P_{t,n}[j,i]P_{s,n}[r,i]P_{t,n}[j,k]P_{s,n}[r,k]\\
\leq & n^{-2}C_e\sum_{i=1}^{n}C_m^4+4n^{-2}C_e\sum_{i=1}^{n}C_m^4+2n^{-2}C_e\sum_{i=1}^{n}\sum_{j=1}^{k-1}P_{t,n}[j,i]\sum_{k=1}^{i-1}P_{t,n}[j,k]\sum_{r=1}^{k-1}P_{s,n}[r,i]P_{s,n}[r,k]\\
\leq & 7n^{-1}C_eC_m^4\longrightarrow 0
\end{aligned}
$$

Then, given that $E\left(H2_{5,n}\right)=0$, $H2_{5,n}\xrightarrow{p}0$. Also, the proof of $H3_{5,n}\xrightarrow{p}0$ follows the same steps. This proves that $H_{5,n}\xrightarrow{p}0$. Lastly, the cases $H_{k,n}$ for $k=6,7,8$ can be proved in the same way. This concludes our proof of A.5, and hence that of the Lemma. □

Lemma A.15. *Under Assumptions 1 to 8,*

$$
\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}\left(\psi_0\right)}{\partial\psi}\xrightarrow{d}N\left(0,E\left(\frac{1}{nT}\frac{\partial\mathcal{L}\left(\psi_0\right)}{\partial\psi}\frac{\partial\mathcal{L}\left(\psi_0\right)'}{\partial\psi}\right)\right)
$$

Proof. The key to the proof is to show that $\frac{1}{\sqrt{nT}}\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\xrightarrow{d}N\left(0,\mathcal{G}_{11}\right)$, with $\mathcal{G}_{11}=\lim_{n\to\infty}\frac{1}{nT}E\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\widetilde{\mathbf{X}}\right]$. In particular, by the Cramér-Wold device, it suffices to show that for any $c=(c_1',c_2',c_3)'\in\mathbb{R}^{4K+2}\times\mathbb{R}$ with $\|c\|=1$, $\frac{1}{\sqrt{nT}}c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\xrightarrow{d}N\left(0,c'\mathcal{G}_{11}c\right)$.

Let us define $\mathbb{X}1=\left[\begin{array}{c:c:c}\boldsymbol{l}_{nT} & \mathbf{X} & \overline{\mathbf{X}}\end{array}\right]$, $\mathbb{X}2=\left[\begin{array}{c:c}\mathbf{X} & \overline{\mathbf{X}}\end{array}\right]$, $\phi_{10}=(c_0,\beta_{10}',\pi_{\mu0}')'$ and $\phi_{20}=(\beta_{20}',\pi_{\alpha0}')'$. From *A.2* we have that:

$$
\begin{aligned}
c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}=&c_1'\mathbb{X}1'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+c_2'\mathbb{X}2'\mathbf{W}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+c_3\mathbf{Y}_{-1}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\\
=&c_1'\mathbb{X}1'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+c_2'\mathbb{X}2'\mathbf{W}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+c_3\mathbf{Y}_{n0}'\boldsymbol{G}_{\mathbf{0}}^{-'}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+\\
&c_3\phi_{10}'\mathbb{X}1'\boldsymbol{C}_{\mathbf{0}}^{-'}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+c_3\phi_{20}'\mathbb{X}2'\boldsymbol{C}_{\mathbf{0}}^{-'}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}+c_3\boldsymbol{\eta}'\boldsymbol{C}_{\mathbf{0}}^{-'}\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}.
\end{aligned}
$$

Following the steps of Lemma A.11, we can write the summands of the previous expression as sums of quadratic forms:

$$c_1'\mathbb{X}1'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} = \sum_{t=1}^{T} c_1'\mathbb{X}1_{nt}'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} c_1'\mathbb{X}1_{nt}'B_{n0}^{-1}\xi_{ns} + \sum_{t=1}^{T} c_1'\mathbb{X}1_{nt}'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T} c_1'\mathbb{X}1_{nt}'B_{n0}^{-1}W_n\upsilon_{n\alpha}$$

$$c_2'\mathbb{X}2'\mathbf{W}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} = \sum_{t=1}^{T} c_2'\mathbb{X}2_{nt}'W_n'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} c_2'\mathbb{X}2_{nt}'W_n'B_{n0}^{-1}\xi_{ns} + \sum_{t=1}^{T} c_2'\mathbb{X}2_{nt}'W_n'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T} c_2'\mathbb{X}2_{nt}'W_n'B_{n0}^{-1}W_n\upsilon_{n\alpha}$$

$$\mathbf{Y}_{n0}'\boldsymbol{G}_{\mathbf{0}}^{-}{}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} = \sum_{t=1}^{T} \mathbf{Y}_{n0}'G_{nt-1}'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathbf{Y}_{n0}'G_{nt-1}'B_{n0}^{-1}\xi_{ns} + \sum_{t=1}^{T} \mathbf{Y}_{n0}'G_{nt-1}'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T} \mathbf{Y}_{n0}'G_{nt-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha}$$

$$\phi_{10}'\mathbb{X}1'\boldsymbol{C}_{\mathbf{0}}^{-}{}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} = \sum_{t=1}^{T}\sum_{j=1}^{t} \phi_{10}'\mathbb{X}1_{nj}'C_{n,t-j-1}'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{s=1}^{T} \phi_{10}'\mathbb{X}1_{nj}'C_{n,t-j-1}'B_{n0}^{-1}\xi_{ns} + \sum_{t=1}^{T}\sum_{j=1}^{t} \phi_{10}'\mathbb{X}1_{nj}'C_{n,t-j-1}'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T}\sum_{j=1}^{t} \phi_{10}'\mathbb{X}1_{nj}'C_{n,t-j-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha}$$

$$\phi_{20}'\mathbb{X}2'W_n'\boldsymbol{C}_{\mathbf{0}}^{-}{}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} = \sum_{t=1}^{T}\sum_{j=1}^{t} \phi_{20}\mathbb{X}2_{nj}'W_n'C_{n,t-j-1}'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{s=1}^{T} \phi_{20}'\mathbb{X}2_{nj}'W_n'C_{n,t-j-1}'B_{n0}^{-1}\xi_{ns} + \sum_{t=1}^{T}\sum_{j=1}^{t} \phi_{20}'\mathbb{X}2_{nj}'W_n'C_{n,t-j-1}'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T}\sum_{j=1}^{t} \phi_{20}'\mathbb{X}2_{nj}'W_n'C_{n,t-j-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha}$$

$$\boldsymbol{\eta}'\boldsymbol{C}_{\mathbf{0}}^{-}{}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} = \sum_{t=1}^{T}\sum_{j=1}^{t} \xi_{nj}'C_{n,t-j-1}'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{s=1}^{T} \xi_{nj}'C_{n,t-j-1}'B_{n0}^{-1}\xi_{ns} + \sum_{t=1}^{T}\sum_{j=1}^{t} \upsilon_{n\alpha}'W_n'C_{n,t-j-1}'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{s=1}^{T} \upsilon_{n\alpha}'W_n'C_{n,t-j-1}'B_{n0}^{-1}\xi_{ns} + \sum_{t=1}^{T}\sum_{j=1}^{t} \xi_{nj}'C_{n,t-j-1}'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T}\sum_{j=1}^{t} \xi_{nj}'C_{n,t-j-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha} + \sum_{t=1}^{T}\sum_{j=1}^{t} \upsilon_{n\alpha}'W_n'C_{n,t-j-1}'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T}\sum_{j=1}^{t} \upsilon_{n\alpha}'W_n'C_{n,t-j-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha}$$

with $\xi_{nt} = (\varepsilon_{nt} + \upsilon_{n\mu})$, $C_{n,-1} = 0_{n\times n}$ and $B_{n0}^{-1} = B_n(\sigma_0)^{-1}$ (see Lemma A.10 for the definition of $B_n(\sigma)^{-1}$).

We can thus write $c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} = \sum_{l=1}^{L} a_l' P_{n,l} b_l$ with $L < \infty$. Then, it is easy to verify that a_l, b_l and $P_{n,l}$ for $l = 1, 2, ..., L$ satisfy the conditions of Lemma A.14 and, by Assumption 7, that $n^{-1}Var(c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}) \geq \bar{c} > 0$, so that $\frac{c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}}{\left[Var(c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta})\right]^{1/2}} \xrightarrow{d} N(0, 1)$, which in turn implies that $\frac{1}{\sqrt{nT}} c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \xrightarrow{d} N(0, c'\mathcal{G}_{11}c)$ for any $c \in \mathbb{R}^{4K+2} \times \mathbb{R}$ with $\|c\| = 1$. This proves the convergence for the first term of the gradient.

To conclude the proof, we note that each component of $\frac{\partial \mathcal{L}(\psi_0)}{\partial \psi}$ (see Appendix C for details) can be written as a finite sum of quadratic forms, so that the proof for these cases proceeds by closely following the previous steps. We consequently omit the details of these proofs. □

B Proof of Theorems.

We start by proving the consistence of the QML estimator (Theorem 1). The proof of normality comes next (Theorem 2).

B.1 Consistency

Proof of Theorem 1. The consistency proof closely follows the proof of Theorem 4.1 of Su and Yang (2015). In particular, by Theorem 3.4 of White (1994), it suffices to show that:

(1.) $\frac{1}{nT}\left[\mathcal{L}_c^*(\delta) - \mathcal{L}_c(\delta)\right] \xrightarrow{p} 0$ uniformly in $\delta \in \Delta = \Delta_\sigma \times \Delta_\lambda$

and

(2.) $\limsup_{n\to\infty} \max_{\delta \in N_\epsilon^c(\delta_0)} \frac{1}{nT}\left[\mathcal{L}_c^*(\delta) - \mathcal{L}_c^*(\delta_0)\right] < 0$ for any $\epsilon > 0$, where $N_\epsilon^c(\delta_0)$ is the complement of an open neighbourhood of δ_0 on Δ of radius ϵ.

To show that **(1.)** holds, it is sufficient to show that the following conditions hold: **(1.a)** $\widehat{\sigma}_\varepsilon^2(\delta) - \widetilde{\sigma}_\varepsilon^2(\delta) \xrightarrow{p} 0$ uniformly in $\delta \in \Delta$ and **(1.b)** $\widetilde{\sigma}_\varepsilon^2(\delta)$ is uniformly bounded away from zero on

Δ. Since **(1b)** will be checked in the proof of **(2.)**, next we concentrate on the proof of **(1.a)**.

By definition of our model, $\hat{\boldsymbol{\eta}}(\delta) = \mathbf{S}(\lambda)\mathbf{Y} - \tilde{\mathbf{X}}\hat{\theta}(\delta) = \boldsymbol{\Omega}^{1/2}(\sigma)\mathbf{M}(\sigma)\boldsymbol{\Omega}^{-1/2}(\sigma)\mathbf{S}(\lambda)\mathbf{Y}$, where $\mathbf{M}(\sigma) = \mathbf{I}_{nT} - \boldsymbol{\Omega}^{-1/2}(\sigma)\widetilde{\mathbf{X}}\left(\widetilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1/2}(\sigma)$. This means that

$$\begin{aligned}
\widehat{\sigma}^2_\varepsilon(\delta) &= \frac{1}{nT}\left(\mathbf{Y}'\mathbf{S}'(\lambda)\boldsymbol{\Omega}^{-1/2}(\sigma)\mathbf{M}(\sigma)\boldsymbol{\Omega}^{1/2}(\sigma)\right)\boldsymbol{\Omega}^{-1}(\sigma)\left(\boldsymbol{\Omega}^{1/2}(\sigma)\mathbf{M}(\sigma)\boldsymbol{\Omega}^{-1/2}(\sigma)\mathbf{S}(\lambda)\mathbf{Y}\right)\\
&= \frac{1}{nT}\boldsymbol{\eta}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta} - \frac{1}{nT}\mathbf{Q}'_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\\
&- (\lambda-\lambda_0)\frac{1}{nT}\boldsymbol{\eta}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{Y} + (\lambda-\lambda_0)\frac{1}{nT}\mathbf{Q}'_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\\
&- (\lambda-\lambda_0)\frac{1}{nT}\mathbf{Y}'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta} + (\lambda-\lambda_0)\frac{1}{nT}\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\\
&+ (\lambda-\lambda_0)^2\frac{1}{nT}\mathbf{Y}'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{Y} - (\lambda-\lambda_0)^2\frac{1}{nT}\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)
\end{aligned}$$

with $\mathbf{Q}_{\mathbf{A},\mathbf{B}}(\delta) = \mathbf{A}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{B}$.

From $\max\limits_{\theta,\sigma^2_\varepsilon} E\left[\mathcal{L}(\psi)\right]$,

$$\begin{aligned}
\tilde{\theta}(\delta) &= \left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{S}(\lambda)\mathbf{Y}\right]\\
&= \theta_0 + \left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}\right]\\
&- (\lambda-\lambda_0)\left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{Y}\right]
\end{aligned}$$

Then,

$$\begin{aligned}
\eta(\tilde{\theta}(\delta)) \equiv \tilde{\eta}(\delta) &= \mathbf{S}(\lambda)\mathbf{Y} - \tilde{\mathbf{X}}\tilde{\theta}(\delta)\\
&= \eta - (\lambda-\lambda_0)\mathbf{W}\mathbf{Y} - \tilde{\mathbf{X}}\left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}\right]\\
&+ (\lambda-\lambda_0)\tilde{\mathbf{X}}\left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{Y}\right]
\end{aligned}$$

and, consequently,

$$
\begin{aligned}
\tilde{\sigma}_{\varepsilon}(\delta) = & \frac{1}{nT} E\left[\boldsymbol{\eta}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}\right] - \frac{1}{nT}(\lambda-\lambda_0) E\left[\boldsymbol{\eta}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{WY}\right] \\
& - \frac{1}{nT} E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1} E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\right] \\
& + \frac{1}{nT}(\lambda-\lambda_0) E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1} E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right] \\
& - \frac{1}{nT}(\lambda-\lambda_0) E\left[\mathbf{Y}'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}\right] + \frac{1}{nT}(\lambda-\lambda_0)^2 E\left[\mathbf{Y}'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{WY}\right] \\
& + \frac{1}{nT}(\lambda-\lambda_0) E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1} E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\right] \\
& - \frac{1}{nT}(\lambda-\lambda_0)^2 E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1} E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]
\end{aligned}
$$

Let us also define

$$
\begin{aligned}
\boldsymbol{\Pi}_{\mathbf{a},\mathbf{b}}(\sigma) &= \frac{1}{nT}\left\{\mathbf{a}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{b} - E\left[\mathbf{a}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{b}'\right]\right\} \\
\boldsymbol{\Upsilon}_{\mathbf{a},\mathbf{b}}(\sigma) &= \frac{1}{nT}\left\{\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{a}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{b}}(\sigma) - E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{a}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1} E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{b}}(\sigma)\right]\right\},
\end{aligned}
$$

where $\mathbf{a}, \mathbf{b} = \boldsymbol{\eta}, \mathbf{WY}$. By using these when calculating the difference between $\tilde{\sigma}_{\varepsilon}(\delta)$ and $\hat{\sigma}_{\varepsilon}(\delta)$ we obtain:

$$
\begin{aligned}
\hat{\sigma}_{\varepsilon}(\delta) - \tilde{\sigma}_{\varepsilon}(\delta) = & \boldsymbol{\Pi}_{\boldsymbol{\eta},\boldsymbol{\eta}}(\sigma) - (\lambda-\lambda_0)\boldsymbol{\Pi}_{\boldsymbol{\eta},\mathbf{WY}}(\sigma) - (\lambda-\lambda_0)\boldsymbol{\Pi}'_{\boldsymbol{\eta},\mathbf{WY}}(\sigma) + (\lambda-\lambda_0)^2\boldsymbol{\Pi}_{\mathbf{WY},\mathbf{WY}}(\sigma) \\
& - \boldsymbol{\Upsilon}_{\boldsymbol{\eta},\boldsymbol{\eta}}(\sigma) + (\lambda-\lambda_0)\boldsymbol{\Upsilon}_{\boldsymbol{\eta},\mathbf{WY}}(\sigma) + (\lambda-\lambda_0)\boldsymbol{\Upsilon}'_{\boldsymbol{\eta},\mathbf{WY}}(\sigma) - (\lambda-\lambda_0)^2\boldsymbol{\Upsilon}_{\mathbf{WY},\mathbf{WY}}(\sigma)
\end{aligned}
$$

and, therefore, condition **(1.a)** follows by using Lemmas *A*.11 and *A*.12.

To show that condition **(2.)** holds, we closely follow the literature (Lee, 2004; Yu et al., 2008; Su and Yang, 2015) and use an auxiliary process to show, using Jensen inequality and $\tilde{\sigma}^2_{\varepsilon}(\delta_0) = \frac{\sigma^2_{\varepsilon 0}}{nT} tr(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Omega}_0) = \sigma^2_{\varepsilon 0}$ (which follows from the definition of $\tilde{\sigma}^2_{\varepsilon}(\delta)$ and Lemma *A*.3), that

$$
\mathcal{L}^*_c(\delta) \leq \mathcal{L}^*_c(\delta_0) \tag{B.1}
$$

Next we prove that $\frac{1}{nT}\mathcal{L}^*_c(\delta)$ is uniformly equicontinuous on $\delta \in \Delta$ by showing the uniform equicontinuity of $\frac{1}{nT}\ln|\mathbf{S}(\lambda)|$, then that of $\frac{1}{nT}\ln|\mathbf{\Omega}(\sigma)|$ and finally that of $\ln(\tilde{\sigma}_{\varepsilon}^{2}(\delta))$ on $\delta\in\Delta$.

Firstly, by the mean value theorem, $\ln|\mathbf{S}(\lambda^{*})|-\ln|\mathbf{S}(\lambda^{**})|=\left(\frac{\partial}{\partial\lambda}\ln|\mathbf{S}(\overline{\lambda})|\right)(\lambda^{*}-\lambda^{**})$ with $\overline{\lambda}\in(\lambda^{*},\lambda^{**})$. Also,

$$\frac{1}{nT}\frac{\partial}{\partial\lambda}\ln|\mathbf{S}(\overline{\lambda})|=\frac{1}{nT}tr\left[\mathbf{S}^{-1}(\overline{\lambda})\mathbf{W}\right]=O(1),$$

since $\mathbf{S}^{-1}(\lambda)\mathbf{W}$ is u.b.r.c.s. uniformly in λ and hence $tr\left[\mathbf{S}^{-1}(\overline{\lambda})\mathbf{W}\right]=O(nT)$. Thus, $\ln|\mathbf{S}(\lambda)|$ is uniformly equicontinuous in λ on Δ_{λ}.

Secondly, by the mean value theorem, $\ln|\mathbf{\Omega}(\sigma^{*})|-\ln|\mathbf{\Omega}(\sigma^{**})|=\sum_{k=1}^{3}\left(\frac{\partial}{\partial\sigma_k}\ln|\mathbf{\Omega}(\overline{\sigma})|\right)(\sigma_k^{*}-\sigma_k^{**})$, with $\overline{\sigma}$ lying elementwise between σ^{*} and σ^{**}. Also,

$$\begin{aligned}\frac{1}{nT}\frac{\partial}{\partial\sigma_1}\ln|\mathbf{\Omega}(\overline{\sigma})|&=\frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}(\overline{\sigma})(J_T\otimes I_n)\right]\\ \frac{1}{nT}\frac{\partial}{\partial\sigma_2}\ln|\mathbf{\Omega}(\overline{\sigma})|&=\frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}(\overline{\sigma})\left(J_T\otimes(W_n+W_n')\right)\right]\\ \frac{1}{nT}\frac{\partial}{\partial\sigma_3}\ln|\mathbf{\Omega}(\overline{\sigma})|&=\frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}(\overline{\sigma})(J_T\otimes W_nW_n')\right]\end{aligned}$$

Notice that, by Lemma A.2, and, given that $tr(W_n+W_n')=O(n)$ (see Remark A2 in Kapoor et al. 2007) and $\sup_{\sigma}\tau_{\max}(\Omega^{-1}(\sigma))<c_{\tau}<\infty$ (by Lemma A.9), we can show that $\frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}(\overline{\sigma})\left(J_T\otimes(W_n+W_n')\right)\right]\leq\frac{1}{nT}\left[\sup_{\sigma}\tau_{\max}(\Omega^{-1}(\sigma))\right]^{-1}tr\left(J_T\otimes(W_n+W_n')\right)\leq\frac{1}{nT}c_{\tau}tr(J_T)tr(W_n+W_n')=O(1)$ uniformly on Δ, and similarly for the other two cases (since, by Remark A2 in Kapoor et al. 2007, $tr(W_nW_n')=O(n)$). Thus, $\ln|\mathbf{\Omega}(\sigma)|$ s uniformly equicontinuous in σ on Δ_{λ}.

Thirdly, to show that $\ln[\tilde{\sigma}_{\varepsilon}^{2}(\delta)]$ is uniformly equicontinuous on Δ it suffices to show that $\tilde{\sigma}_{\varepsilon}^{2}(\delta)$ is uniformly equicontinuous and uniformly bounded away from zero on Δ. Thus, we start by noting from the definition of $\tilde{\sigma}_{\varepsilon}^{2}(\delta)$ used in the proof of **(1.a)** that all its elements appear in $\mathbf{\Pi}_{\mathbf{a},\mathbf{b}}(\sigma)$ and $\mathbf{\Upsilon}_{\mathbf{a},\mathbf{b}}(\sigma)$, which, using the same arguments as in Lemmas A.11 and A.12, and the results in Lemma A.7, proves the uniform equicontinuity of $\tilde{\sigma}_{\varepsilon}(\delta)$. Next, to show that $\tilde{\sigma}_{\varepsilon}^{2}(\delta)$ is uniformly bounded away from zero, we follow Su and Yang (2015) and establish the claim by a counter argument based on making its dependence on n explicit. To this end, we include the subindex n in $\tilde{\sigma}^2_\varepsilon(\delta)$, so that it then becomes $\tilde{\sigma}^2_{\varepsilon,n}(\delta)$.

If $\tilde{\sigma}^2_{\varepsilon,n}(\delta)$ is not uniformly bounded away from zero on Δ, then there exists a sequence $\{\delta_n\}$ in Δ such that $\lim_{n\to\infty} \tilde{\sigma}^2_{\varepsilon,n}(\delta) = 0$. Now, by $B.1$ we have that, for all δ,

$$-\ln\left[\tilde{\sigma}^2_\varepsilon(\delta)\right] \leq -\ln\left[\tilde{\sigma}^2_\varepsilon(\delta_0)\right] + \frac{1}{nT}\left[\ln|\mathbf{S}_0| - \ln|\mathbf{S}(\lambda)|\right] + \frac{2}{nT}\left[\ln|\mathbf{\Omega}(\sigma)| - \ln|\mathbf{\Omega}_0|\right]$$

Using the mean value theorem, as we previously did, it can be proved that $\frac{1}{nT}\left[\ln|\mathbf{S}_0| - \ln|\mathbf{S}(\lambda)|\right] = O(1)$ and $\frac{2}{nT}\left[\ln|\mathbf{\Omega}(\sigma)| - \ln|\mathbf{\Omega}_0|\right] = O(1)$ uniformly in Δ. This implies that $-\ln\left[\tilde{\sigma}^2_\varepsilon(\delta)\right]$ is bounded above, which is a contradiction, so we conclude that $\tilde{\sigma}^2_{\varepsilon,n}(\delta)$ is uniformly bounded away from zero on Δ.

Finally, the identification uniqueness also follows by contradiction. Using $\tilde{\sigma}^2_\varepsilon(\delta_0) = \sigma^2_{\varepsilon 0}$ (see Lemma $A.3$) we have that

$$\begin{aligned}\frac{1}{nT}\left[\mathcal{L}^*_c(\delta) - \mathcal{L}^*_c(\delta_0)\right] &= \frac{1}{2nT}\left\{\ln|\mathbf{\Omega}_0| - \ln|\mathbf{\Omega}(\sigma)|\right\} + \frac{1}{2}\left\{\ln\left[\sigma^2_{\varepsilon 0}\right] - \ln\left[\tilde{\sigma}^2_\varepsilon(\delta)\right]\right\} \\ &+ \frac{1}{nT}\left\{\ln|\mathbf{S}(\lambda)| - \ln|\mathbf{S}_0|\right\} \\ &= \frac{1}{2nT}\left\{\ln\left|\sigma^2_{\varepsilon 0}\mathbf{S}_0^{-2}\mathbf{\Omega}_0\right] - \ln\left|\tilde{\sigma}^2_\varepsilon(\delta)\mathbf{S}^{-2}(\lambda)\mathbf{\Omega}(\sigma)\right|\right\}\end{aligned}$$

If the identification uniqueness condition does not hold, then there exists an $\epsilon > 0$ and a sequence $\{\delta_n\}$ in $N^c_\epsilon(\delta_0)$ such that

$$\lim_{n\to\infty} \frac{1}{nT}\left[\mathcal{L}^*_{c,n}(\delta) - \mathcal{L}^*_{c,n}(\delta_0)\right] = 0,$$

where we have written $\mathcal{L}^*_{c,n}(.)$ for $\mathcal{L}^*_c(.)$ to stress its dependence on n. However, by the compactness of $N^c_\epsilon(\delta_0)$, there exists a convergent subsequence $\{\delta_{n_k}\}$ of $\{\delta_n\}$ with the limit δ_+ of δ_{n_k} being in $N^c_\epsilon(\delta_0)$. This implies that $\delta_+ \neq \delta_0$. Furthermore, by the uniform equicontinuity of $\frac{1}{nT}\mathcal{L}^*_{c,n}(\delta)$, $\lim_{n\to\infty} \frac{1}{n_k T}\left[\mathcal{L}^*_{c,n}(\delta_+) - \mathcal{L}^*_{c,n}(\delta_0)\right] = 0$. Yet this contradicts Assumption 6, since it amounts to $\lim_{n\to\infty} \frac{1}{nT}\left[\mathcal{L}^*_{c,n}(\delta) - \mathcal{L}^*_{c,n}(\delta_0)\right] \neq 0$ for any $\delta \neq \delta_0$. This completes the proof of the theorem. □

B.2 Asymptotic normality

Proof of Theorem 2. By Taylor series expansion,

$$0 = \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}(\psi)}{\partial \psi}\bigg|_{\widehat{\psi}} = \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}(\psi)}{\partial \psi}\bigg|_{\psi_0} + \frac{1}{nT} \frac{\partial^2 \mathcal{L}(\psi)}{\partial \psi \partial \psi'}\bigg|_{\overline{\psi}} \sqrt{nT}\left(\widehat{\psi} - \psi_0\right)$$

where the elements of $\overline{\psi} = \left(\overline{\theta}', \overline{\sigma}_\varepsilon^2, \overline{\lambda}, \overline{\sigma}'\right)'$ lie in the segment joining the corresponding elements of $\widehat{\psi}$ and ψ_0. Thus,

$$\sqrt{nT}\left(\widehat{\psi} - \psi_0\right) = \left[- \frac{1}{nT} \frac{\partial^2 \mathcal{L}(\psi)}{\partial \psi \partial \psi'}\bigg|_{\overline{\psi}}\right]^{-1} \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}(\psi)}{\partial \psi}\bigg|_{\psi_0} = \left[- \frac{1}{nT} \frac{\partial^2 \mathcal{L}\left(\overline{\psi}\right)}{\partial \psi \partial \psi'}\right]^{-1} \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}(\psi_0)}{\partial \psi}$$

By Theorem 1, $\widehat{\psi} \xrightarrow{p} \psi_0$, and so $\overline{\psi} \xrightarrow{p} \psi_0$. Therefore, it suffices to show that:

(i) $\dfrac{1}{nT} \dfrac{\partial^2 \mathcal{L}\left(\overline{\psi}\right)}{\partial \psi \partial \psi'} - \dfrac{1}{nT} \dfrac{\partial^2 \mathcal{L}(\psi_0)}{\partial \psi \partial \psi'} = o_p(1)$,

(ii) $\dfrac{1}{nT} \dfrac{\partial^2 \mathcal{L}(\psi_0)}{\partial \psi \partial \psi'} \xrightarrow{p} E\left(\dfrac{1}{nT} \dfrac{\partial^2 \mathcal{L}(\psi_0)}{\partial \psi \partial \psi'}\right)$, and

(iii) $\dfrac{1}{\sqrt{nT}} \dfrac{\partial \mathcal{L}(\psi_0)}{\partial \psi} \xrightarrow{d} N\left(0, E\left(\dfrac{1}{nT} \dfrac{\partial \mathcal{L}(\psi_0)}{\partial \psi} \dfrac{\partial \mathcal{L}(\psi_0)'}{\partial \psi}\right)\right)$.

Since ii) and (iii) follow from Lemmas *A*.13 and *A*.15, respectively, only (i) is left to be shown. In particular, given the expression of $\dfrac{\partial^2 \mathcal{L}(\psi)}{\partial \psi \partial \psi'}$ provided in Appendix *C*, it suffices to show that $\dfrac{1}{nT} \dfrac{\partial^2 \mathcal{L}\left(\overline{\psi}\right)}{\partial \omega \partial \varpi'} - \dfrac{1}{nT} \dfrac{\partial^2 \mathcal{L}(\psi_0)}{\partial \omega \partial \varpi'} = o_p(1)$ for $\omega, \varpi = \theta, \sigma_\varepsilon^2, \lambda$ and σ. However, we only show this for $(\omega, \varpi) = (\theta, \theta), \left(\theta, \sigma_\varepsilon^2\right), (\lambda, \lambda)$ and $(\sigma_\kappa, \sigma_\varrho)$, with $\kappa, \varrho = 1, 2, 3$, for the other cases can be shown in an analogous way.

For the (θ, θ) case, notice that

$$\frac{1}{nT}\left[\frac{\partial^2 \mathcal{L}\left(\overline{\psi}\right)}{\partial \theta \partial \theta'} - \frac{\partial^2 \mathcal{L}(\psi_0)}{\partial \theta \partial \theta'}\right] = -\frac{1}{nT} \frac{1}{\overline{\sigma}_\varepsilon^2} \widetilde{\mathbf{X}}' \mathbf{\Omega}^{-1}(\overline{\sigma}) \widetilde{\mathbf{X}} + \frac{1}{nT} \frac{1}{\sigma_{\varepsilon 0}^2} \widetilde{\mathbf{X}}' \mathbf{\Omega}_0^{-1} \widetilde{\mathbf{X}}$$

$$= \left(\frac{\overline{\sigma}_\varepsilon^2 - \sigma_{\varepsilon 0}^2}{\sigma_{\varepsilon 0}^2 \overline{\sigma}_\varepsilon^2}\right) \frac{1}{nT} \widetilde{\mathbf{X}}' \mathbf{\Omega}_0^{-1} \widetilde{\mathbf{X}} + \frac{1}{\overline{\sigma}_\varepsilon^2} \frac{1}{nT}\left[\widetilde{\mathbf{X}}' \mathbf{\Omega}_0^{-1}\left(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0\right) \mathbf{\Omega}^{-1}(\overline{\sigma}) \widetilde{\mathbf{X}}\right] \quad \text{(B.2)}$$

Given that $\sigma_{\varepsilon 0}^2 > 0$ and $\overline{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_{\varepsilon 0}^2$, $\left(\frac{\overline{\sigma}_\varepsilon^2 - \sigma_{\varepsilon 0}^2}{\sigma_{\varepsilon 0}^2 \overline{\sigma}_\varepsilon^2}\right) = o_p(1)$, from the proof of Lemma $A.12$ we can show that $\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\widetilde{\mathbf{X}} = O_p(1)$. As for the second term in the r.h.s. of $B.2$, note that $\tau_{\max}^{1/2}\left((\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)^2\right) = O_p(\|\overline{\sigma} - \sigma\|) = o_p(1)$. To prove this, notice that, since $\tau_{\max}(A \otimes B) \leq \tau_{\max}(A)\,\tau_{\max}(B)$, then $\tau_{\max}^{1/2}\left((\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)^2\right) = T\tau_{\max}^{1/2}\left((\mathbf{\Sigma}(\overline{\sigma}) - \mathbf{\Sigma}_0)^2\right)$. Further,

$$
\begin{aligned}
\tau_{\max}^{1/2}\left((\mathbf{\Sigma}(\overline{\sigma}) - \mathbf{\Sigma}_0)^2\right) &= \|(\mathbf{\Sigma}(\overline{\sigma}) - \mathbf{\Sigma}_0)\|_2 \\
&= \|(\overline{\sigma}_1 - \sigma_{10})\, I_n + (\overline{\sigma}_2 - \sigma_{20})\,(W_n + W_n') + (\overline{\sigma}_3 - \sigma_{30})\, W_n W_n'\|_2 \\
&\leq |\overline{\sigma}_1 - \sigma_{10}|\, \|I_n\|_2 + |\overline{\sigma}_2 - \sigma_{20})\, \||(W_n + W_n')\|_2 + |\overline{\sigma}_3 - \sigma_{30}|\, \|W_n W_n'\|_2
\end{aligned}
$$

Then, given that W_n is u.b.r.c.s., $\|W_n + W_n'\|_2 \leq \left(\|W_n + W_n'\|_1 \|W_n + W_n'\|_\infty\right)^{1/2} < \infty$ and $\|W_n W_n'\|_2 \leq \left(\|W_n W_n'\|_1 \|W_n W_n'\|_\infty\right)^{1/2} < \infty$. Thus, $\tau_{\max}^{1/2}\left((\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)^2\right) \leq [|\overline{\sigma}_1 - \sigma_{10}| + |\overline{\sigma}_2 - \sigma_{20}| + |\overline{\sigma}_2 - \sigma_{20}|]\, T c_\tau$ with $c_\tau < \infty$.

Let c be an arbitrary column vector in $\mathbb{R}^{4K+2}$. Then, by the Cauchy-Schwarz inequality, Lemmas $A.9$ and $A.2$, and $\frac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right| = O_p(1)$ (which can be proved following the same steps as in Lemma $A.12$), we have that

$$
\begin{aligned}
&\frac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)\,\mathbf{\Omega}^{-1}(\overline{\sigma})\,\widetilde{\mathbf{X}}c\right| \\
&\leq \frac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{\Omega}_0^{-1}\widetilde{\mathbf{X}}c\right|^{1/2}\left|c'\widetilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\overline{\sigma})\,(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)\,(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)\,\mathbf{\Omega}^{-1}(\overline{\sigma})\,\widetilde{\mathbf{X}}c\right|^{1/2} \\
&\leq \tau_{\max}\left(\mathbf{\Omega}_0^{-1}\right)\tau_{\max}\left(\mathbf{\Omega}^{-1}(\overline{\sigma})\right)\tau_{\max}^{1/2}\left((\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)^2\right)\frac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right| \\
&\leq O_p(\|\overline{\sigma} - \sigma\|)\, O_p(1) = o_p(1)
\end{aligned}
$$

Since $\frac{1}{\overline{\sigma}_\varepsilon^2} = O_p(1)$, it follows that the second term in the r.h.s. of $B.2$ is $o_p(1)$.

For the $(\theta, \sigma_{\varepsilon}^{2})$ case, notice that

$$\frac{1}{nT}\left[\frac{\partial^{2}\mathcal{L}(\overline{\psi})}{\partial\theta\partial\sigma_{\varepsilon}^{2}} - \frac{\partial^{2}\mathcal{L}(\psi_{0})}{\partial\theta\partial\sigma_{\varepsilon}^{2}}\right] = \frac{1}{nT}\left[\frac{1}{\overline{\sigma}_{\varepsilon}^{4}}\widetilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\overline{\sigma})\,\boldsymbol{\eta}(\overline{\theta},\overline{\lambda}) - \frac{1}{\sigma_{\varepsilon}^{4}}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\boldsymbol{\eta}\right]$$
$$= \left(\frac{1}{\overline{\sigma}_{\varepsilon}^{4}} - \frac{1}{\sigma_{\varepsilon}^{4}}\right)\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\boldsymbol{\eta} + \frac{1}{\overline{\sigma}_{\varepsilon}^{4}}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\left(\mathbf{\Omega}^{-1}(\overline{\sigma}) - \Omega_{0}^{-1}\right)\boldsymbol{\eta}(\overline{\theta},\overline{\lambda})\right]$$
$$+ \frac{1}{\overline{\sigma}_{\varepsilon}^{4}}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\left(\boldsymbol{\eta}(\overline{\theta},\overline{\lambda}) - \boldsymbol{\eta}\right)\right]$$

Following the same steps as in Lemmas $A.12$ and $A.13$, it can be proved that $\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\boldsymbol{\eta} = o_{p}(1)$, $\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\mathbf{Y} = O_{p}(1)$ and $\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\widetilde{\mathbf{X}} = O_{p}(1)$. Thus, by using $\boldsymbol{\eta}(\overline{\theta},\overline{\lambda}) = (\lambda_{0} - \overline{\lambda})\,\mathbf{Y} + \boldsymbol{\eta} + \widetilde{\mathbf{X}}(\overline{\theta} - \theta_{0})$, it can be proved that the three summands in the previous expression are $o_{p}(1)$.

For the (λ, λ) case, notice that

$$\frac{1}{nT}\left[\frac{\partial^{2}\mathcal{L}(\overline{\psi})}{\partial\lambda\partial\lambda} - \frac{\partial^{2}\mathcal{L}(\psi_{0})}{\partial\lambda\partial\lambda}\right]$$
$$= \frac{1}{nT}\left[tr\left(\left(\mathbf{S}_{0}^{-1}\mathbf{W}\right)^{2}\right) - tr\left(\left(\mathbf{S}^{-1}(\overline{\lambda})\,\mathbf{W}\right)^{2}\right) + \mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_{0}^{-1}\mathbf{W}\mathbf{Y} - \mathbf{Y}'\mathbf{W}'\mathbf{\Omega}^{-1}(\overline{\sigma})\,\mathbf{W}\mathbf{Y}\right]$$
$$= \frac{1}{nT}\left[tr\left(\left(\mathbf{S}_{0}^{-1}\mathbf{W}\right)^{2} - \left(\mathbf{S}^{-1}(\overline{\lambda})\,\mathbf{W}\right)^{2}\right)\right] + \frac{1}{nT}\left[\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_{0}\right)\mathbf{W}\mathbf{Y}\right]$$

Let us now consider the first term of the previous expression. Given that $\mathbf{S}^{-1}(\lambda)$ and $\mathbf{W}$ are u.b.r.c.s. uniformly in λ, then

$$\frac{1}{nT}tr\left(\left(\mathbf{S}_{0}^{-1}\mathbf{W}\right)^{2} - \left(\mathbf{S}^{-1}(\overline{\lambda})\,\mathbf{W}\right)^{2}\right) \leq \left|\overline{\lambda} - \lambda_{0}\right|\frac{1}{nT}tr\left(\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}^{-1}(\overline{\lambda})\,\mathbf{W}\mathbf{S}^{-1}(\overline{\lambda})\,\mathbf{W}\right)$$
$$+ \left|\overline{\lambda} - \lambda_{0}\right|\frac{1}{nT}tr\left(\mathbf{S}_{0}^{-1}W\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}^{-1}(\overline{\lambda})\,\mathbf{W}\right)$$
$$\leq o_{p}(1)\,O(1)\,,$$

where the second inequality holds because $tr\left(\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}^{-1}(\overline{\lambda})\,\mathbf{W}\mathbf{S}^{-1}(\overline{\lambda})\,\mathbf{W}\right) = O(nT)$ and $tr\left(\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}^{-1}(\overline{\lambda})\,\mathbf{W}\right) = O(nT)$. As for the second term, $\frac{1}{nT}\left[\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_{0}\right)\mathbf{W}\mathbf{Y}\right] = \frac{1}{nT}\left[\theta_{0}'\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_{0}\right)\widetilde{\mathbf{X}}\theta_{0}\right] + \frac{1}{nT}\left[\boldsymbol{\eta}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_{0}\right)\boldsymbol{\eta}\right] + 2\frac{1}{nT}\left[\theta_{0}'\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_{0}\right)\boldsymbol{\eta}\right]$, so that, using arguments analogous to the ones used in previous cases, it can be proved that

$\frac{1}{nT}\left[\mathbf{Y}'\mathbf{W}'\boldsymbol{\Omega}_0^{-1}\left(\boldsymbol{\Omega}\left(\overline{\sigma}\right)-\boldsymbol{\Omega}_0\right)\mathbf{W}\mathbf{Y}\right]=o_p\left(1\right).$

For the $(\sigma_\kappa,\sigma_\kappa)$ case, notice that

$$\begin{aligned}
\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}\left(\overline{\psi}\right)}{\partial\sigma_\kappa\partial\sigma_\kappa}-\frac{\partial^2\mathcal{L}\left(\psi_0\right)}{\partial\sigma_\kappa\partial\sigma_\kappa}\right] &= \frac{1}{nT}\left[\frac{1}{\overline{\sigma}_\varepsilon^2}\widetilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)\boldsymbol{\eta}\left(\overline{\theta}\right)-\frac{1}{\sigma_\varepsilon^2}\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}\right]\\
&=\left(\frac{1}{\overline{\sigma}_\varepsilon^2}-\frac{1}{\sigma_\varepsilon^2}\right)\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}\right]\\
&+\frac{1}{\overline{\sigma}_\varepsilon^2}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)-\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}\right]\\
&=\left(\frac{1}{\overline{\sigma}_\varepsilon^2}-\frac{1}{\sigma_\varepsilon^2}\right)\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}\right]\\
&+\frac{1}{\overline{\sigma}_\varepsilon^2}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\left(\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right]\\
&+\frac{1}{\overline{\sigma}_\varepsilon^2}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\widetilde{\mathbf{X}}\left(\overline{\theta}-\theta_0\right)\right]
\end{aligned}$$

The first and third summands can be proved to be $o_p\left(1\right)$ using arguments analogous to the ones used in previous cases. For the second one, note that

$$\tau_{\max}^{1/2}\left(\left(\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right)^2\right)=T^{3/2}\tau_{\max}^{1/2}\left(\left(\Sigma^{-1}\left(\overline{\sigma}\right)\Sigma_\kappa\Sigma^{-1}\left(\overline{\sigma}\right)-\Sigma_0^{-1}\Sigma_\kappa\Sigma_0^{-1}\right)^2\right)$$

Also, since $\left\|\Sigma^{-1}\left(\overline{\sigma}\right)-\Sigma_0^{-1}\right\|_2=o_p\left(1\right)$ and previous results show that $\exists c_\tau<\infty$ such that $\tau_{\max}\left(\Sigma_0^{-1}\right)\le c_\tau$, $\tau_{\max}\left(\Sigma^{-1}\left(\overline{\sigma}\right)\right)\le c_\tau$ and $\tau_{\max}\left(\Sigma_\kappa\right)\le c_\tau$ for $\kappa=1,2,3$,

$$\begin{aligned}
\tau_{\max}^{1/2}\left(\left(\Sigma^{-1}\left(\overline{\sigma}\right)\Sigma_\kappa\Sigma^{-1}\left(\overline{\sigma}\right)-\Sigma_0^{-1}\Sigma_\kappa\Sigma_0^{-1}\right)^2\right)&=\left\|\left(\Sigma^{-1}\left(\overline{\sigma}\right)\Sigma_\kappa\Sigma^{-1}\left(\overline{\sigma}\right)-\Sigma_0^{-1}\Sigma_\kappa\Sigma_0^{-1}\right)\right\|_2\\
&\le\left\|\left(\Sigma^{-1}\left(\overline{\sigma}\right)-\Sigma_0^{-1}\right)\Sigma_\kappa\Sigma^{-1}\left(\overline{\sigma}\right)\right\|_2+\\
&\quad\left\|\Sigma_0^{-1}\Sigma_\kappa\left(\Sigma^{-1}\left(\overline{\sigma}\right)-\Sigma_0^{-1}\right)\right\|_2\\
&\le\left\|\Sigma^{-1}\left(\overline{\sigma}\right)-\Sigma_0^{-1}\right\|_2\left\|\Sigma_\kappa\right\|_2\left(\left\|\Sigma^{-1}\left(\overline{\sigma}\right)\right\|_2+\left\|\Sigma_0^{-1}\right\|_2\right)\\
&\le\left\|\Sigma^{-1}\left(\overline{\sigma}\right)-\Sigma_0^{-1}\right\|_2c_\tau=o_p(1)
\end{aligned}$$

Thus, $\tau_{\max}^{1/2}\left(\left(\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}\left(\overline{\sigma}\right)-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right)^2\right)=o_p\left(1\right)$.

Further, let c be an arbitrary column vector in $\mathbb{R}^{4K+2}$. Then, by the Cauchy-Schwarz inequality, Lemma A.2, the fact that $\frac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right| = O_p(1)$ and $\frac{1}{nT}\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)'\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right) = O_p(1)$ (which can be proved following the same steps as in Lemmas A.12 and A.13),

$$\begin{aligned}
&\frac{1}{nT}\left[c'\widetilde{\mathbf{X}}'\left(\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right]\\
&\leq\left[\frac{1}{nT}c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right]^{1/2}\left[\frac{1}{nT}\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)'\left(\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right)\right.\\
&\qquad\qquad\left.\left(\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})-\Omega_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right]^{1/2}\\
&\leq\tau_{\max}^{1/2}\left(\left(\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right)^2\right)\left[\frac{1}{nT}c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right]^{1/2}\left[\frac{1}{nT}\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)'\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right]^{1/2},
\end{aligned}$$

so that, given the previous result showing that $\tau_{\max}^{1/2}\left(\left(\boldsymbol{\Omega}^{-1}(\overline{\sigma})\Sigma_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right)^2\right) = o_p(1)$, $\frac{1}{\overline{\sigma}_\varepsilon^2}\frac{1}{nT}\left[\frac{1}{\overline{\sigma}_\varepsilon^2}\widetilde{\mathbf{X}}'\left(\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right] = o_p(1)$.

Finally, to prove the $(\sigma_\kappa,\sigma_\varrho)$ case notice that $\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}\left(\overline{\psi}\right)}{\partial\sigma_\kappa\partial\sigma_\varrho}-\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\sigma_\kappa\partial\sigma_\varrho}\right]$ can be expressed as

$$\begin{aligned}
&\frac{1}{2}\frac{1}{nT}tr\left[\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\varrho-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\varrho\right]+\\
&\frac{1}{nT}\left[\frac{1}{\sigma_{\varepsilon 0}^2}\boldsymbol{\eta}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\varrho\Omega_0^{-1}\boldsymbol{\eta}-\frac{1}{\overline{\sigma}_\varepsilon^2}\boldsymbol{\eta}\left(\overline{\lambda},\overline{\theta}\right)'\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\varrho\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\eta}\left(\overline{\lambda},\overline{\theta}\right)\right]
\end{aligned}$$

Note also that $\tau_{\max}^{1/2}\left(\left(\left[\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right]\right)^2\right) = o_p(1)$ and, given that $\boldsymbol{\Sigma}_\kappa$ is u.b.r.c.s., $tr\left(\boldsymbol{\Sigma}_\kappa^2\right) = O(nT)$. Then, $\frac{1}{nT}tr\left[\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\varrho-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\varrho\right] \leq \frac{1}{nT}\tau_{\max}^{1/2}\left(\left(\left[\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})-\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\right]\right)^2\right)(nT)^{1/2}tr^{1/2}\left(\boldsymbol{\Sigma}_\varrho^2\right) = o_p(1)O(1) = o_p(1)$ and

$$\begin{aligned}
&\tau_{\max}^{1/2}\left(\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\varrho\boldsymbol{\Omega}_0^{-1}-\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\varrho\boldsymbol{\Omega}^{-1}(\overline{\sigma})\right)^2\right)\\
&\quad\leq\left\|\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}_0^{-1}-\Omega^{-1}(\overline{\sigma})\boldsymbol{\Sigma}_\kappa\boldsymbol{\Omega}^{-1}(\overline{\sigma})\right\|_2\tau_{\max}\left(\boldsymbol{\Sigma}_\varrho\right)\tau_{\max}\left(\boldsymbol{\Omega}^{-1}(\overline{\sigma})\right)+\\
&\qquad\left\|\boldsymbol{\Omega}_0^{-1}-\boldsymbol{\Omega}^{-1}(\overline{\sigma})\right\|_2\tau_{\max}\left(\boldsymbol{\Omega}_0^{-1}\right)^2\tau_{\max}\left(\boldsymbol{\Sigma}_\kappa\right)\tau_{\max}\left(\boldsymbol{\Sigma}_\varrho\right) = o_p(1)O(1).
\end{aligned}$$

Therefore, using arguments analogous to the ones used in previous cases, $\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}\left(\overline{\psi}\right)}{\partial\sigma_\kappa\partial\sigma_\varrho}-\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\sigma_\kappa\partial\sigma_\varrho}\right] = o_p(1)$.

We conclude the proof by noting that the $o_p(1)$ of the other components of $\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}(\psi)}{\partial\psi\partial\psi'} - \frac{\partial^2\mathcal{L}(\psi)}{\partial\psi\partial\psi'}\right]$ can be proved using previous results and arguments analogous to the ones used in the cases considered here. We consequently omit the details of these proofs.

□

C Gradient and Hessian of the QML function

C.1 Gradient

The gradient function $\bigtriangledown\mathcal{L}(\psi_0) = \frac{\partial\mathcal{L}(\psi_0)}{\partial\psi}$ has the following elements:

$$
\begin{aligned}
\frac{\partial\mathcal{L}(\psi_0)}{\partial\theta} &= \frac{1}{\sigma^2_{\varepsilon_0}}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial\mathcal{L}(\psi_0)}{\partial\sigma^2_{\varepsilon_0}} &= -\frac{nT}{2\sigma^2_{\varepsilon_0}} + \frac{1}{2\sigma^4_{\varepsilon_0}}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial\mathcal{L}(\psi_0)}{\partial\sigma_\kappa} &= -\frac{1}{2}tr\left(\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\right) + \frac{1}{2\sigma^2_{\varepsilon_0}}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial\mathcal{L}(\psi_0)}{\partial\lambda} &= -tr\left(\mathbf{S}^{-1}\mathbf{W}\right) + \frac{1}{\sigma^2_\varepsilon}\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}
\end{aligned}
$$

where $\kappa = 1, 2, 3$ and $\mathbf{\Sigma}_\kappa = \frac{\partial\mathbf{\Omega}(\sigma)}{\partial\sigma_\kappa}$. Also, $\mathbf{\Sigma}_1 = J_T \otimes \Sigma_1 = J_T \otimes I_n$, $\mathbf{\Sigma}_2 = J_T \otimes \Sigma_2 = J_T \otimes (W_n + W_n')$ and $\mathbf{\Sigma}_3 = J_T \otimes \Sigma_3 = J_T \otimes W_n W_n'$.

C.2 Hessian matrix

The Hessian of the likelihood function in 3.2 is:

$$
\boldsymbol{H}_n(\psi_0) = \begin{pmatrix}
\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\theta\partial\theta'} & \frac{\partial^2\mathcal{L}(\psi_0)}{\partial\theta\partial\sigma^2_\varepsilon} & \frac{\partial^2\mathcal{L}(\psi_0)}{\partial\theta\partial\delta'} \\
 & \frac{\partial^2\mathcal{L}(\psi_0)}{\partial\sigma^2_\varepsilon\partial\sigma^2_\varepsilon} & \frac{\partial^2\mathcal{L}(\psi_0)}{\partial\sigma^2_\varepsilon\partial\delta'} \\
 & & \frac{\partial^2\mathcal{L}(\psi_0)}{\partial\delta\partial\delta'}
\end{pmatrix}
$$

Next we provide detailed results for each row of the Hessian matrix. Thus, the first row of the Hessian matrix is

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\theta\partial\theta'} &= -\frac{1}{\sigma_\varepsilon^2}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\tilde{\mathbf{X}} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\theta\partial\sigma_\varepsilon^2} &= -\frac{1}{\sigma_\varepsilon^4}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\theta\partial\sigma_\kappa} &= -\frac{1}{\sigma_\varepsilon^2}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\theta\partial\lambda} &= -\frac{1}{\sigma_\varepsilon^2}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{W}\mathbf{Y},
\end{aligned}$$

while the second row of the Hessian matrix is

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\varepsilon^2\partial\sigma_\varepsilon^2} &= \frac{nT}{2\sigma_\varepsilon^4} - \frac{1}{\sigma_\varepsilon^6}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\varepsilon^2\partial\lambda} &= -\frac{1}{\sigma_\varepsilon^4}\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\varepsilon^2\partial\sigma_\kappa} &= -\frac{1}{2\sigma_\varepsilon^4}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}
\end{aligned}$$

and the third row of the Hessian matrix is

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\lambda\partial\lambda} &= -tr\left(\left(\mathbf{S}_0^{-1}\mathbf{W}\right)^2\right) - \mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_0^{-1}\mathbf{W}\mathbf{Y} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\lambda\partial\sigma_\kappa} &= -\frac{1}{\sigma_\varepsilon^2}\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\kappa\partial\sigma_\kappa} &= \frac{1}{2}tr\left[\left(\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\right)^2\right] - \frac{1}{\sigma_\varepsilon^2}\boldsymbol{\eta}'\left(\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\right)^2\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\kappa\partial\sigma_\varrho} &= \frac{1}{2}tr\left(\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\varrho\right) \\
&\quad - \frac{1}{2\sigma_\varepsilon^2}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\left[\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\varrho + \mathbf{\Sigma}_\varrho\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\right]\mathbf{\Omega}_0^{-1}\boldsymbol{\eta},
\end{aligned}$$

with $\kappa \neq \varrho$ and $\varrho = 1, 2, 3$.

References

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2):153–166.

Beer, C. and Riedl, A. (2012). Modelling spatial externalities in panel data: The spatial Durbin model revisited. *Papers in Regional Science*, 91(2):299–318.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5–46.

Chamberlain, G. (1984). Panel data. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2, chapter 22, pages 1247–1318. Elsevier, 1 edition.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

Debarsy, N. (2012). The mundlak approach in the spatial Durbin panel data model. *Spatial Economic Analysis*, 7(1):109–131.

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

Elhorst, J. P. (2012). Dynamic spatial panels: models, methods, and inferences. *Journal of Geographical Systems*, 14(1):5–28.

Elhorst, P., Piras, G., and Arbia, G. (2010). Growth and convergence in a multiregional model with spacetime dynamics. *Geographical Analysis*, 42(3):338–355.

Ertur, C. and Koch, W. (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics*, 22(6):1033–1062.

Gänsler, P. and Stute, W. (1977). *Wahrscheinlichkeitstheorie*. Springer Verlag.

Ho, C.-Y., Wang, W., and Yu, J. (2013). Growth spillover through trade: A spatial dynamic panel data approach. *Economics Letters*, 120(3):450–453.

Islam, N. (1995). Growth empirics: A panel data approach. *The Quarterly Journal of Economics*, 110(4):1127–1170.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1):97 – 130.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121.

Kelejian, H. H. and Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104(2):219 – 257.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925.

Lee, L.-F. and Yu, J. (2016). Identification of spatial Durbin panel models. *Journal of Applied Econometrics*, 31(1):133–162.

LeSage, J. (2014). What regional scientists need to know about spatial econometrics. *The Review of Regional Studies*, 44(1):13–323.

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*. CRC Press.

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

Magnus, J. R. (1982). Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. *Journal of Econometrics*, 19(2):239 – 285.

Miranda, K., Martínez-Ibañez, O., and Manjón-Antolín, M. (2017). Estimating individual effects and their spatial spillovers in linear panel data models: Public capital spillovers after all? *Spatial Statistics*, 22(Part 1):1–17.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.

Su, L. and Yang, Z. (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics*, 185(1):230 – 258.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.

Yu, J., de Jong, R., and Lee, L.-F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146(1):118 – 134.

Yu, J. and Lee, L.-F. (2012). Convergence: A spatial dynamic panel data approach. *Global Journal of Economics*, 1(1):1–36.