Inference in a spatial autoregressive model with an extended coefficient range and a similarity-based weight matrix

Francesca Rossi, Offer Lieberman
Inference in a spatial autoregressive model with an extended coefficient range and a similarity-based weight matrix

Francesca Rossi∗ and Offer Lieberman†

December 22, 2021

Abstract

We provide in this paper asymptotic theory for a spatial autoregressive model (SAR, henceforth) in which the spatial coefficient, λ, is allowed to be less than or equal to unity, as well as consistent with a local to unit root (LUR) model and of the moderate deviation (MI) from unity type, and the spatial weights are allowed to be similarity-based and data driven. Other special cases of our setting include the random walk, a model in which all the weights are equal, the standard SAR model in which |λ| < 1 and the similarity based autoregression in which λ = 1 and data do not display a natural order. As the norming rates for the asymptotic theory are very different in the |λ| < 1 - compared with the λ = 1 and LUR cases, we resort to random norming that treats all cases in a uniform manner. It turns out that standard CLT results prevail in a large class of models in which the infinity norm of the inverse of the weighting structure that characterizes the reduced-form process is of \(O(n^γ)\), \(γ \in [0, 1)\), and is non-standard in the case \(γ = 1\). We use a penalized and shifted profile likelihood to obtain results which are valid for all cases. A small simulation experiment supports our findings and the usefulness of our model is illustrated with an empirical application of the Boston housing data set in which the estimate of λ appeared to be very close to unity.

Keywords: Spatial Autoregression; Similarity Function; Weight Matrix; Quasi-Maximum-Likelihood.

Paper’s JEL Classification: C21, C22; Francesca Rossi’s JEL Classification: C13, C21; Offer Lieberman’s JEL Classification: C13, C22.

1 Introduction

Given an \(n \times 1\) vector of observables \(y\), we consider the following data generating process (DGP)

\[
y_1 = \beta_0'z_1 + \varepsilon_1, \tag{1.1}
\]
\[
y_i = \beta_0'z_i + \lambda_0 \sum_{j=1, j \neq i}^{n} h_{i,j}y_j + \varepsilon_i, \quad i = 2, \ldots, n, \tag{1.2}
\]

where \(\varepsilon_i, i = 1, \ldots, n\), are i.i.d. random variables with zero mean and variance \(σ_0^2\), \(z_i\) is the transpose of the \(i\)-th row of an exogenous \(n \times m\) matrix \(Z\) of standard covariates (which might...
or might not include a column of ones),

\[ h_{i,j} = \frac{s(x_i, x_j; w_0)}{\sum_{j \neq i} s(x_i, x_j; w_0)}, \quad (1.3) \]

with \( s(x_i, x_j; w_0) \) being a similarity function which belongs to \( \mathbb{R}_+ \) and \( x_i, x_j \) being the transpose of \( i-th \) and \( j-th \) rows, respectively, of an \( n \times k \) matrix \( X \) of fixed explanatory variables.

Examples of well-defined similarity functions are given by the exponential and inverse similarity functions, viz.,

\[ s(x_i, x_j; w_0) = \exp \left( -\sum_{t=1}^{k} w_{0t} (x_{it} - x_{jt})^2 \right) \quad (1.4) \]

and

\[ s(x_i, x_j; w_0) = \frac{1}{1 + \sum_{t=1}^{k} w_{0t} (x_{it} - x_{jt})^2}, \quad (1.5) \]

respectively. The unknown parameters of the full model in (1.2) are the scalar \( \lambda_0 \in [-1, 1] \), the \( k \times 1 \) vector \( w_0 = (w_{10}, \ldots, w_{k0})' \), which is assumed to belong to a subset of \( \mathbb{R}_+^k \), the \( m \times 1 \) vector \( \beta_0 \) and \( \sigma_0^2 \), assumed to belong to suitable subsets of \( \mathbb{R}_+ \) and \( \mathbb{R}_m \), respectively.

We remark that in order to allow the possibility of \( \lambda_0 = 1 \) and at the same time to guarantee that model (1.2) retains an equilibrium, we need to introduce an “initial” condition on one observation, as displayed in (1.1). Since cross-sectional data do not have a natural order, we state such condition on \( y_1 \) without loss of generality. We stress that for sufficiently large \( n \) such a condition has no particular impact on the economic interpretation of results and it is not unreasonable to assume that one observation is not affected by the rest. When \( \beta_0 = 0 \), this starting condition mimics the common requirement that a process starts from the origin in the time series literature.

The model (1.2) contains both similarity based models and spatial autoregressions as special cases. Indeed, when \( \lambda_0 = 1 \) and \( \beta_0 = 0 \) a priori, model (1.2) represents an extension to the spatial setting of the similarity process originally axiomatized in Gilboa et al. (2006), whose asymptotic properties have been established in Lieberman (2010) in the case where the data is ordered, so that the sum in (1.2) extends over \( j < i \). By allowing \( \beta_0 \) to take non-zero values, model (1.2) allows each \( y_i \) to be distributed around a unit-specific regression function \( (\beta_0' z_i) \) plus a weighted average of all other values \( y_j \), with \( j \neq i \). On the other hand, for \( \lambda_0 \in (-1, 1) \), model (1.2) mimics the structure of a spatial autoregression model (e.g. Lee (2004), and references therein), with the additional feature that the so-called weight matrix is now parametrized in terms of a set of parameters \( w_0 \) that have to be estimated, rather than taken as exogenously chosen. To this extent, this paper contributes to the spatial econometric literature by relaxing the strong assumptions that \( \lambda_0 \in (-1, 1) \) and that the weights’ structure has to be fully conjectured ex-ante.

In principle, in the terminology of Gilboa et al. (2006), model (1.2) is a hybrid model, containing a rule based model, with a rule \( \beta_0' z_i \) associating the \( z_i \)‘s to the \( y_i \)‘s and with a weight of unity, and a case based component, \( \sum_{j \neq i} h_{i,j} y_j \), with a weight of \( \lambda_0 \). In standard spatial models the weight matrix is fixed, arbitrary and determined a priori. In the similarity model the weights are determined by the data. Specifically, the \( w \)‘s are data driven. More similar cases,
measured by how close $x_i$ is to $x_j$, will give a larger weight to $y_j$ in (1.2). The idea is somewhat similar to $k$-nearest neighbors ($k$NN), except that $k$NN is a fitting/estimation technique whereas here the similarity function is part of the DGP.

After Gilboa et al.’s (2006) introduction of similarity based models to economics, the relevance of this class of models to the definition of objective probabilities was discussed by Gilboa et al. (2010) and was considered for prediction by Gilboa et al. (2011). Gayer et al. (2007) applied the idea in the context of case-based modeling of real estate pricing. Asymptotic theory for model (1.2) has been established in Lieberman (2010), in the case where the data is ordered, so that the sum in (1.2) extends over $j < i$ and when $\beta_0 = 0$. This work has been extended by Lieberman (2012) and Lieberman and Phillips (2014) to the time-varying coefficient, non-stationary autoregression where the similarity function is possibly time varying. The latter model has been applied to Japanese dual stock data and to international Exchange Traded Funds. Gayer et al. (2019) established asymptotic theory for a similarity based model of categorical data by making extensive use of Markov chain theory. The concepts of similarity and contagion of views are central in Kapetanios et al. (2013), who constructed a nonlinear panel data model of cross-sectional dependence. Finally, Teitelbaum (2013) suggested similarity function for applications of empirical similarity theory in which the notion of similarity is asymmetric.

While the SAR and similarity based models are highly related, it appears that the literature on SAR has propagated completely independently of similarity models and as far as we know this paper is the first attempt to bring the two streams of literature together. Asymptotic theory for standard spatial models ranges from the well established (quasi-)maximum likelihood estimator (QMLE, henceforth) in, e.g., Cliff and Ord (1975) and Lee (2004) to two-stage least squares (2SLS) (e.g. Kelejian and Prucha (1998)) and generalized method of moments (GMM) (e.g. Kelejian and Prucha (1999)) with their respective numerous refinements, such as the extension to QMLE with heteroskedastic disturbances in Liu and Yang (2015), or the efficiency improvements in 2SLS/GMM established in Lee (2003, 2007). Even though the spatial literature has been extended to increasingly more complex models, such as, for instance, higher-order SAR (e.g. Gupta and Robinson (2015, 2018)) or panel data structures with heteroskedastic disturbances (e.g. Kelejian and Prucha (2010) and Lin and Lee (2010)), asymptotic theory has been developed under the condition that data are weakly dependent and spatial parameter(s) belong to the interior of a compact set, which in turn depends on the predetermined weight matrix choice (Kelejian and Prucha (2010)). In case of a row-normalized weight matrix such as that in (1.3) with known $w_0$, the parameter space typically results in any compact subset of $(-1, 1)$. In addition, even though several definitions of weak/strong spatial dependence are given in the literature (e.g. Robinson (2011), Chudik and Pesaran (2015) and Bailey et al. (2016)), standard SAR assumptions imply that the largest eigenvalue of the variance-covariance matrix of the dependent variable is bounded, such that every form of strong dependence is automatically ruled out.

In this paper we focus on developing the asymptotic theory for inference on

$$\theta_0 = (\sigma_{0}^{2}, \lambda_{0}, w_{10}, ..., w_{k0})^{'},$$
in model (1.2) when \( \beta_0 = 0 \) \textit{a priori}. As the norming rates for the asymptotic theory are very different in the \(|\lambda| < 1\) - compared with the \( \lambda = 1 \) case, we resort to random norming that treats all cases in a uniform manner without any requirement on weak dependence across spatial units. As expected, standard CLT results prevail in the \(|\lambda| < 1\) case but for \( \lambda = 1 \) the results are non-standard and, as expected, the asymptotic distribution of the profiled QMLE is not Gaussian. The inclusion of a regression component, as in model (1.2) is left for future investigation. We expect that once established the limit properties of estimators for \( \theta \), the inclusion of a regression function will only be a fairly minor modification since, for given \( \lambda_0 \) and \( w_0, \beta_0 \) can be concentrated-out and its estimator would enjoy a closed-form expression.

The plan for the paper is as follows. In Section 2 we introduce some notation and discuss some special cases which are of special interest. Limit Theory and the main results of the paper are given in Sections 3-5. Simulations and an empirical example follow in Sections 6 and 7, respectively. Section 8 concludes and proofs are given in Appendices A and B and in an Online Supplement.

2 Some special cases

For any generic \( p \times q \) matrix \( A \), we denote by \( a_{ij} \) its \((i, j)\)-th element and by \( a_i \) the transpose of its \(i\)-th row. Also \( b^{ij} \) denotes the \((i, j)\)-th elements of \( B^{-1} \) for any generic, square, invertible matrix \( B \). The symbol \( 1 = 1_n \) denotes an \( n \times 1 \) vector of ones, \( \| \cdot \|, \| \cdot \|_\infty \), and \( \| \cdot \|_F \) represent spectral, uniform absolute row sum and Frobenius norms, respectively. \( A' \) is the transpose of \( A \), and \( K > 0 \) is an arbitrary finite constant whose value may change in each location. For a generic square matrix, \( \eta_{\min}(B) \) and \( \eta_{\max}(B) \) denote minimum and maximum eigenvalues of \( B \), respectively. The symbol \( \sim \) indicates ‘asymptotic equivalence’. In the sequel, the subscript \((\cdot)_0\) indicates true values, or quantities evaluated at the true parameters’ values, while the absence of such subscript denotes parameters that are free to vary within the parameters’ space or quantities evaluated at generic values of the parameters\(^1\).

Model (1.2) with \( \beta_0 = 0 \) can be written as

\[
S_{n0}y_n = \varepsilon_n, \tag{2.1}
\]

where

\[
S_{n0} = S_n(\lambda_0, w_0) = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
-\lambda_0 h_{2,1} & 1 & \cdots & -\lambda_0 h_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
-\lambda_0 h_{n,1} & -\lambda_0 h_{n,2} & \cdots & 1
\end{pmatrix} = I - \lambda_0 C_n(\lambda_0, w_0) = I - \lambda_0 C_{n0}. \tag{2.2}
\]

In (2.1), as well as in (1.2), \( y = y_n, \varepsilon = \varepsilon_n, X = X_n, C_0 = C_{n0} \) and \( S_0 = S_{n0} \) are, in general, triangular arrays, but we omit the subscript \( n \) in the sequel for notational simplicity. The

\(^1\)Dependence on parameters is occasionally retained explicitly whenever the concise notation might create ambiguity or lack of clarity.
triangular array structure, other than a mere requisite of generality, is specifically induced by the fact that \( h_{i,j} = h_{i,j,n} \), for \( i,j = 1, \ldots, n \), contains a normalization that depends on sample size. This is similar to the various choices of normalization of the weight matrix in spatial autoregressions, which in turn imply dependence on sample size of the data generating process itself.

Provided that an equilibrium exists, the reduced form of (2.1) is conveniently defined as

\[
y = S_0^{-1} \varepsilon. \tag{2.3}
\]

For \( |\lambda_0| < 1 \) and for given \( w_0 \), model (2.1) formally corresponds to a SAR model with no exogenous regressors, and the theory for developing inference on \( \lambda_0 \) is well established under some suitable additional conditions. In this section, before focusing on the general case in (2.1) and (2.3) with \( |\lambda_0| \leq 1 \) and unknown \( w_0 \), we present some special cases of interest that help justifying our theory in the following sections.

### 2.1 The null model

We consider first the data generating process given by (2.1) with \( h_{i,j} = 1 / (n - 1) \), \( \forall i,j = 1, \ldots, n \) with \( i \neq j \), and \( \lambda_0 = 1 \). This is an important benchmark as it amounts to \( w_0 = 0 \) in the similarity functions (1.4) and (1.5), and more generally it corresponds to any similarity function for which \( s(x_i, x_j; 0) = 1 \), and thus meaning that the exogenous \( X \)'s do not play a role in the weighting of the \( y_i \)'s. Also, in view of model (1.2), \( \lambda_0 = 1 \) and \( \beta_0 = 0 \) represents a decision process that is entirely driven by case-based reasoning. Hence, we aim to estimate \( \lambda_0 \) in the model

\[
y = \lambda_0 C_0 y + \varepsilon \quad \text{with} \quad C_0 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\
\frac{1}{n-1} & 0 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{n-1} & \frac{1}{n-1} & \cdots & \cdots & 0 \end{pmatrix}, \tag{2.4}
\]

when \( \lambda_0 = 1 \). For sake of illustration only, we use ordinary least squares (OLS, henceforth), and obtain

\[
\hat{\lambda}_n = \hat{\lambda} = (y'C_0'y)^{-1} y'C_0'y \tag{2.5}
\]

and thus

\[
\hat{\lambda} - 1 = \frac{y'C_0'y - y'C_0'y}{y'C_0'y} = \frac{y'C_0'(y - C_0y)}{y'C_0'y} = \frac{y'C_0'\varepsilon}{y'C_0'y}. \tag{2.6}
\]

We deduce that, given \( \lambda_0 = 1 \) and \( w_0 = 0 \),

\[
S_0^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\
1 & 2(1 - \frac{1}{n}) & 1 - \frac{1}{n} & \cdots & 1 - \frac{1}{n} \\
\cdots & 1 - \frac{1}{n} & \cdots & \cdots & \cdots \\
1 & 1 - \frac{1}{n} & \cdots & 1 - \frac{1}{n} & 2(1 - \frac{1}{n}) \end{pmatrix}. \tag{2.7}
\]
and
\[
C_0 S_0^{-1} = \begin{pmatrix}
0 & 0 & \cdots & 0 \\
1 & 1 - \frac{2}{n} & 1 - \frac{1}{n} & \cdots & 1 - \frac{1}{n} \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
1 & 1 - \frac{1}{n} & \cdots & 1 - \frac{1}{n} & 1 - \frac{2}{n} \\
1 & 1 & \cdots & 1 & 1 - \frac{2}{n}
\end{pmatrix}
\] (2.8)

We show in the Online Supplement that
\[
n\left(\hat{\lambda}_n - 1\right) = n \frac{y'C_0'\varepsilon}{y'C_0'C_0y} \xrightarrow{d} 1.
\] (2.9)

From (2.9), we notice that when \(\lambda_0 = 1\) and for this specific choice of weights, the OLS estimator \(\hat{\lambda}\) converges in distribution to \(\lambda_0\) at rate \(n\) and the limiting distribution of the standardized estimator is degenerate.

2.2 The random walk model

A second special case that is worth investigating is the cross-sectional variant of a random walk model. Formally, the model is identical to the well-known random walk in the time series literature. In terms of cross-sectional data, such model can be interpreted as each unit depending on the neighbour “behind” or (equivalently) on the one “ahead”\(^2\).

We estimate \(\lambda_0\) in the process
\[
y = \lambda_0 C_0 y + \varepsilon \quad \text{with} \quad C_0 = \begin{pmatrix}
0 & 0 & 0 & \cdots & 0 \\
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & \cdots & 1 & 0
\end{pmatrix}
\] (2.10)

when \(\lambda_0 = 1\). In this case
\[
S_0^{-1} = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
1 & 1 & 0 & \cdots \\
1 & 1 & 1 & \cdots \\
\vdots & \ddots & \ddots & \ddots \\
1 & 1 & \cdots & 1 & 1
\end{pmatrix}
\] (2.11)

\(^2\)A popular spatial structure is given by a circulant Toeplitz matrix where each unit is related to “one ahead” and “one behind”.

6
Furthermore, as \( I = S_0 S_0^{-1} = (I - C_0) S_0^{-1} \),

\[
C_0 S_0^{-1} = S_0^{-1} - I = \begin{pmatrix}
0 & 0 & \cdots & 0 \\
1 & 0 & 0 & \cdots \\
1 & 1 & \cdots & \cdots \\
\vdots & \vdots & \ddots & 0 \\
1 & 1 & \cdots & 1 & 0
\end{pmatrix}.
\]

(2.12)

Thus,

\[
\frac{1}{n} y' C_0' \varepsilon = \frac{1}{n} \varepsilon' S_0^{-1} y' C_0' \varepsilon = \frac{1}{n} \begin{pmatrix}
0 \\
\varepsilon_1 \\
\vdots \\
\sum_{j=1}^{n-1} \varepsilon_j
\end{pmatrix}' \varepsilon \xrightarrow{d} \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon} \int_0^1 W(r) \, dW(r)
\]

(2.13)

and similarly,

\[
\frac{1}{n^2} y' C_0' C_0 y = \frac{1}{n^2} \varepsilon' S_0^{-1} y' C_0' S_0^{-1} \varepsilon \xrightarrow{d} \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon} \int_0^1 W^2(r) \, dr.
\]

(2.14)

As

\[
\int_0^1 W(r) \, dW(r) = \frac{\sigma_\varepsilon^2 (W^2(1) - 1)}{2},
\]

(2.15)

we obtain the well-known result (e.g. Phillips (1987))

\[
n \left( \hat{\lambda}_n - 1 \right) = n \left( \frac{y' C_0' \varepsilon}{y' C_0' C_0 y} - 1 \right) \xrightarrow{d} \frac{W^2(1) - 1}{2 \int_0^1 W^2(r) \, dr},
\]

(2.16)

which, similarly to what obtained for (2.4), shows that \( \hat{\lambda} \) is consistent at rate \( n \) for \( \lambda_0 \) and that its standardized version does not have a standard normal limiting distribution.

### 2.3 Local to unit root and moderate integration models

The time series model

\[
y_t = \lambda_n y_{t-1} + \varepsilon_t, \quad t = 2, \ldots, n, \quad \lambda_n = 1 - c/k_n,
\]

(2.17)

with \( c > 0 \), is called a local to unit root (LUR) model when \( k_n = n \) and a moderate integration (MI) model when \( k_n = n^\alpha \) and \( \alpha \in (0, 1) \). See, for instance, Phillips (1987) and Phillips and Magdalinos (2007), respectively. It appears that the distinction between these two models is critical to support the much more general asymptotic theory to be developed in this paper, in which a suitably normalized QMLE of \( \lambda \) will converge to a Gaussian variable in some cases and to a non-Gaussian limit in others. Specifically, in our notation the matrix \( C_0 \) is as in (2.10) and
we can write

\[
S_0^{-1} = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
\lambda_n & 1 & 0 & \cdots \\
\lambda_n^2 & \lambda_n & 1 & \cdots \\
\vdots & \vdots & \ddots & \ddots \\
\lambda_n^{n-1} & \lambda_n^{n-2} & \cdots & \lambda_n & 1
\end{pmatrix},
\]

(2.18)

from which follows that

\[
||S^{-1}(\theta)||_\infty = \frac{1}{1 - \lambda_n} = \frac{k_n}{c}.
\]

(2.19)

We know from Theorem 1(a) of Phillips (1987) that for the LUR model (2.17) the \(n\)-normalized OLS estimator of \(\lambda\) converges to a non-Gaussian limit whereas from Theorem 3.2(c) of Phillips and Magdalinos (1987) we learn that the \((nk_n)^{1/2}\)-normalized OLS estimator of \(\lambda\) in the MI model converges to a Gaussian limit. In view of (2.19), in the LUR case, \(||S^{-1}(\theta)||_\infty = O(n)|\), whereas in the MI case \(||S^{-1}(\theta)||_\infty = O(n^\alpha)\), \(\alpha \in (0, 1)\).

The conclusion from the special cases discussed above appears to be in sharp contrast with the usual \(\sqrt{n}\)-consistency and standard normal limiting distribution we generally obtain for SAR models with \(|\lambda_0| < 1\) (e.g. Lee (2004) and references therein). Limit theory for standard SAR models is generally obtained for \(|\lambda_0| < 1\) under the assumptions

\[
\sup_{\theta \in \Theta} (||C(\theta)||_\infty + ||C^\prime(\theta)||_\infty) < K
\]

(2.20)

and

\[
\sup_{\theta \in \Theta} (||S^{-1}(\theta)||_\infty + ||S^{-1}_{\ast}(\theta)||_\infty) < K,
\]

(2.21)

for a suitable choice of \(\Theta\), for \(n\) large enough. Condition (2.21) is typical in asymptotic theory for SAR models (e.g. Kelejian et al. (1998)), as it limits the degree of spatial correlation to a manageable amount\(^3\). Indeed, letting \(\Omega_y(\theta_0) = E(yy')\), (2.3) and (2.21) trivially implies \(||\Omega_y(\theta_0)||_\infty < K\) so that any form of strong dependence across components of \(y\) is therefore ruled out. To the best of our knowledge, condition (2.21) has never been relaxed in SAR literature, even though it has been discussed in, e.g., Robinson (2011) in the context of a nonparametric regression for spatial data, where spatial dependence is embedded in a general error structure that is not limited to weak dependence.

On the other hand, for the special cases discussed above we see from (2.7), (2.11) and (2.19) that (2.21) is violated. Since the value of \(\lambda_0\) appears to play a central role in determining whether condition (2.21) holds or otherwise, we can prove the following.

**Claim 1** Let \(C\) be a matrix whose elements are all nonnegative, \(\sum_{i=1}^n c_{ij} = 1\) \(\{i \geq 2\}\) and \(c_{i1} = O(n^{-\tau})\) for some \(\tau > 0\) and \(\forall i\). Let \(S = I - \lambda C\), \(|\lambda| < 1\). Then

\[
\lim_{n \to \infty} ||S^{-1}||_\infty = \frac{1}{(1 - \lambda)}.
\]

(2.22)

\(^3\)(2.21) trivially implies that the largest eigenvalue of \(E(yy') < K\) for all sufficiently large \(n\), and thus in the SAR literature condition (2.21) is often referred to as a “weak dependence” assumption.
The proof of Claim 1 is in the Online Supplement.

Furthermore, for any matrix $C(\cdot)$ such that $\|C(\theta)\|_\infty = 1$, such as any choice of $C(\theta)$ implied by the similarity structure in (2.2), we can write

$$\|S^{-1}(\theta)\|_\infty \leq \sum_{t=0}^\infty |\lambda|^t \|C(\theta)\|_\infty = \sum_{t=0}^\infty |\lambda|^t = \frac{1}{1-|\lambda|},$$

(2.23)

such that $\|S^{-1}(\theta)\|_\infty = O(1/(1-|\lambda|))$. In particular, this result is consistent with $\|S^{-1}(\theta)\|_\infty = O(n)$ and $\|S^{-1}(\theta)\|_\infty = O(n^\alpha)$, $\alpha \in (0, 1)$, for the LUR and MI models, as shown in (2.19). More generally, we shall separate between cases in which

$$\|S^{-1}(\theta)\|_\infty = O(n^\gamma),$$

(2.24)

for some $\gamma \in [0, 1)$ and the case $\gamma = 1$. The case $\gamma = 0$ corresponds to the standard SAR setup. We impose $\|S^{-1}(\theta)\|_\infty = O(\|S^{-1}(\theta)\|_\infty)$ such that, in case $\gamma = (\|S^{-1}(\theta)\|_\infty) = O(n^\gamma)$ with $\gamma > 0$, $\|S^{-1}(\theta')\|_\infty$ could be bounded or increasing without bound as well. In case $\gamma > 0$, the standard condition in (2.21) does not hold and standard limit theory for SAR models is not available. By allowing $\gamma > 0$ we relax the standard assumption of weak dependence across $y$ and we are also allowing $y_i$, for $i \in 1, ..., n$, to have a variance that increases with sample size, since it is straightforward to see that $\text{Var}(y_i) = O(\|S^{-1}(\theta)\|_\infty)$, unless we introduce the additional assumption that $s^{ij}$ are square-summable over $j$, a requirement that is not necessary in view of our development.

The special cases with $\lambda_0 = 1$, and the LUR model illustrated in Sections 2.1-2.3, correspond to a limit regime such that $\|S^{-1}_0\|_\infty \sim n$. For these cases, as expected, we do not obtain a standard limiting distribution with $\sqrt{n}$ rate, but rather a non-standard limit distribution and a rate of convergence equal to $n$. In Section 3 we will prove that, in line with the special cases reported in Sections 2.1-2.3 and with results for standard SAR models, under regularity conditions, consistency is not affected by the value of $\lambda_0$ and/or the value of $\gamma$, while the rate of convergence and the asymptotic distribution depend on the limit behaviour of $\|S^{-1}_0\|_\infty$. In particular, the limit distribution in case $\|S^{-1}_0\|_\infty \sim n$ cannot be obtained within a standard framework and needs to be considered on a case-by-case basis.

### 3 Identification and consistency of QML estimator

We focus on the model (2.1),

$$y = \lambda_0 C_0 y + \epsilon,$$

(3.1)

with $C_0 = C(\theta_0)$ defined as in (2.2). As introduced at the beginning of Section 2, the reduced form (2.3) is well defined as long as the model has an equilibrium. In this section we derive the statistical properties in the limit for the QML estimator of $\theta_0$ under the general asymptotic regime

$$\|S^{-1}(\theta_0)\|_\infty = \|S^{-1}_0\|_\infty = O(n^{\gamma}) \quad 0 \leq \gamma \leq 1$$

(3.2)

with $\|S^{-1}(\theta)\|_\infty = O(\|S^{-1}(\theta)\|_\infty)$. 

9
We impose the following Assumptions:

**Assumption 1** For all $n$ and for $i = 1, \ldots, n$, the $\{\epsilon_i\}$ are a set of independent random variables, with mean zero and unknown variances $\sigma^2 > 0$. In addition, for some $\delta > 0$,

$$E|\epsilon_i|^{4+\delta} \leq K \text{ for } i = 1, \ldots, n.$$  

**Assumption 2** There exists $\sigma^2_L > 0$, $\sigma^2_H < \infty$ and $w_H < \infty$ such that $\sigma^2_L < \sigma^2_0 < \sigma^2_H$ and, for all $i = 1, \ldots, k$, $0 \leq w_{i0} < w_H$. Also, $-1 \leq \lambda_0 \leq 1$.

**Assumption 3** The matrix $X$ is allowed to lie in the set of all $n \times k$ non-random, real matrices such that for all sufficiently large $n$

$$S'S \neq S_0'S_0 \text{ for } \theta \neq \theta_0.$$  \hspace{1cm} (3.3)

**Assumption 4** For all $n$, $\eta_{\min}(S) > 0$ for all $\theta$.

**Assumption 5** For all $n$, $S'S$ has bounded and continuous derivatives, uniformly in $\theta_2 \in \Theta_2$.

Let $C_r = C_r(w_1, \ldots, w_k) = \frac{\partial C(w_1, \ldots, w_k)}{\partial w_r}$ for $r = 1, \ldots, k$.

**Assumption 6**

a) $\sup_{\theta \in \Theta} (||C(\theta)||_\infty + ||C'(\theta)||_\infty) \leq K$.

b) $\sup_{\theta \in \Theta} (||C_r(\theta)||_\infty + ||C'_r(\theta)||_\infty) \leq K$.

The homoskedasticity requirement in Assumption 1 is a sensible starting point to derive the limit theory of QML estimate of $\theta_0$. In general, QML estimators in SAR models are not consistent in presence of heteroskedasticity of unknown form (e.g. Lin and Lee (2010)), but corrections to accommodate it are available in the literature (Liu and Yang (2015)). We leave the analysis of such corrections for our present model for future investigation. Existence of moments higher than four, on the other hand, is a standard assumption to establish a CLT for triangular arrays of martingale differences. Assumption 2 requires compactness of the parameter space $\Theta$, while Assumption 3 is an identification condition that needs to hold at least for large $n$. Assumption 4 ensures that the model in (2.1) has an equilibrium and thus, that the reduced form in (2.3) is well defined for all $n$. Assumption 5 is related to A3 in Lieberman (2010) and it is required to guarantee uniform convergence of key quantities in the proof of Theorem 1. below. By construction of $C(\cdot)$, $||C_0||_\infty = 1$, with $C(\cdot)$ that is generally not symmetric. Assumption 6a) requires that uniform boundedness in row and column sums is preserved, uniformly over $\theta$. Such requirement is standard in SAR literature and it is similar to the first part of A3 in Lieberman (2010). Assumption 6b) is new in the spatial econometrics literature, where the weights’ structure is not data-driven and does not contain any unknown parameters. Assumption 6b) is implied by the second part of A3 in Lieberman (2010).
Given \( y \), we define the shifted, normalized and negative pseudo-log-likelihood function as

\[
\mathcal{L}(\theta) = \log \left( \frac{\sigma^2}{y' y} \right) - \frac{2}{n} \log |S(\theta)| + \frac{yy'S'(\theta)S(\theta)y}{n\sigma^2} - \log \left( \frac{y'y}{n} \right)
\]

\[
= \log \left( \frac{\sigma^2}{y' y} \right) - \frac{2}{n} \log |S| + \frac{yy'S'y}{n\sigma^2} - \log \left( \frac{y'y}{n} \right)
\]  

(3.4)

and \( \hat{\theta} = \arg\min_{\theta \in \Theta} \mathcal{L}(\theta) \). The shifting term \( -\log (y'y/n) \) is introduced to allow us to accommodate both \( \gamma = 0 \) and \( \gamma > 0 \) cases, with \( \gamma \) defined in (3.2), without affecting \( \arg\min \mathcal{L}(\theta) \).

Let \( \theta = (\sigma^2, \lambda, w') = (\theta_1, \theta_2') \), with \( \theta_1 = \sigma^2 \) and \( \theta_2 = (\lambda, w') \). Given \( \theta_2 \), we obtain

\[
\hat{\sigma}^2 = \hat{\sigma}^2(\theta_2) = \frac{yy'S'y}{n}
\]

(3.5)

and we define

\[
\hat{\sigma}^*2 = \hat{\sigma}^*2(\theta_2) = \frac{yy'S'y}{y'y},
\]

(3.6)

such that the profile, shifted, quasi-log-likelihood is equal to

\[
\mathcal{L}'(\theta_2) = \log \left( \frac{\hat{\sigma}^2}{y'y} \right) - \frac{2}{n} \log |S| + \frac{yy'S'y}{n\hat{\sigma}^2} - \log \left( \frac{y'y}{n} \right)
\]

(3.7)

which, up to constant terms becomes

\[
\mathcal{L}'(\theta_2) = \log \left( \frac{yy'S'y}{y'y} \right) - \frac{2}{n} \log |S| = \log \left( \hat{\sigma}^*2 \right) - \frac{2}{n} \log |S|.
\]

(3.8)

Thus, by substituting (2.3) and by Lemma 1b), given in the Online Supplement, we observe that under Assumption 6a), \( y'y = O_p(n) \) and \( yy'S'y = O_p(n) \) for \( \gamma = 0 \), while \( y'y = O_p(n||S_0^{-1}||_\infty) \)

and \( yy'S'y = O_p(n||S_0^{-1}||_\infty) \) when \( \gamma > 0 \). Thus, the random norming \( y'y \), rather than \( n \), ensures that the first term on the RHS of (3.8) remains bounded under the general condition (3.2) and allows us to deal with \( \gamma = 0 \) and \( \gamma > 0 \) within a unified framework.

We further define

\[
\hat{\sigma}^*2(\theta_2) = \lim_{n \to \infty} \left( \hat{\sigma}^*2(\theta_2) \right),
\]

(3.9)

and

\[
\bar{\mathcal{L}}'(\theta_2) = \log \left( \hat{\sigma}^*2 \right) - \frac{2}{n} \log |S| = \log \left( \frac{yy'S'y}{y'y} \right) - \frac{2}{n} \log |S| + o_p(1),
\]

(3.10)

with \( \theta_{20} = \arg\min_{\theta_2 \in \Theta_2} \bar{\mathcal{L}}'(\theta_2) \). For \( 0 \leq \gamma < 1 \), from Lemma 3 we obtain

\[
\hat{\sigma}^*2(\theta_2) = \lim_{n \to \infty} \frac{tr(S_0^{-1'}S'SS_0^{-1})}{tr(S_0^{-1'}S_0^{-1})} = O(1),
\]

(3.11)

for each \( \theta_2 \in \Theta_2 \), while, from Lemma 1b), for \( \gamma = 1 \), \( \hat{\sigma}^*2(\theta_2) = O_p(1) \) for each \( \theta_2 \in \Theta_2 \), where the limit needs to be established on a case-by-case basis and, more generally, it is a random variable. We impose the additional assumption to ensure existence of probability limits of relevant quantities.
Assumption 7

\[ \hat{\sigma}^{*}(\theta_{2}) = \lim_{n \to \infty} \hat{\sigma}^{*2}(\theta_{2}) \] exists for all \( \theta_{2} \in \Theta_{2}, \]

\[ \lim_{n \to \infty} \frac{\partial}{\partial \lambda} \hat{\sigma}^{*2}(\theta_{2}) \] and \( \lim_{n \to \infty} \frac{\partial}{\partial w_{j}} \hat{\sigma}^{*2}(\theta_{2}), \) for \( j = 1, \ldots, k, \) exist for all \( \theta_{2} \in \Theta_{2}. \) (3.12)

We prove the following in Appendix A.

**Theorem 1.** Assume that model (2.1) and Assumptions 1-7 hold. Under (3.2) with \( 0 \leq \gamma \leq 1, \) \( \theta_{20} \) is identified and \( \hat{\theta}_{2} \xrightarrow{p} \theta_{20}. \)

Consistency of \( \hat{\sigma}^{2} \) to \( \sigma_{0}^{2} \) follows trivially from Theorem 1 and (3.5). Thus, parameters’ identification and consistency of \( \hat{\theta} \) to \( \theta_{0} \) are guaranteed under the general (3.2) under a unified and mild set of Assumptions. The contribution of Theorem 1 to the spatial econometric literature is two-fold: first, the well-established restriction to “weak dependence” given in (2.21) is not needed to achieve consistency of the spatial parameter. Secondly, the choice of the weight matrix is not exogenously taken, but rather it is data-driven and depends on a finite set of unknown parameters that can be consistently estimated alongside \( \lambda_{0}. \) In terms of the similarity-based literature, Theorem 1 extends the consistency result of Lieberman (2010) to a bilateral similarity structure that includes the standard time-series notion of ordered observations as a special case and with \( \lambda \leq 1. \) Also, we allow the similarity structure to enter into the data generating process with some weight \( \lambda \) that can be estimated itself.

4 Asymptotic distribution

In this section we derive the asymptotic distribution of \( \hat{\theta}_{2}, \) from which the distribution of \( \hat{\theta}_{1} \) can be deduced by routine arguments. As somewhat expected from the two illustrations in Sections 2.1 and 2.2, even though identification and consistency results are not affected by the value of \( \gamma \) in (3.2), the rate of convergence and the limit distribution differ substantially between the two cases. We define the two regimes for \( ||S_{0}^{-1}||_{\infty} \) as:

R1) \( 0 \leq \gamma < 1 \) in (3.2).

R2) \( \gamma = 1 \) in (3.2).

In Theorem 2 below we will show that under R1 a standard central limit theorem holds, with a rate of convergence that depends on \( ||S_{0}^{-1}||_{\infty} = O(n^{\gamma}) \) for \( 0 \leq \gamma < 1. \) On the other hand, under R2, a central limit theorem does not hold and the limit distribution has to be established on a case-by-case basis. A more articulated discussion on R2 will be reported in Section 5.

For any matrix \( A, \) let \( A = A + A'. \) Also, in line with our notation so far, we let \( C_{r,0} = \)
\[ \frac{\partial C(w_1, \ldots, w_k)}{\partial w_r} \bigg|_{\theta_0} \text{ for } r = 1, \ldots, k \text{ and } C_{rs,0} = \frac{\partial^2 C(w_1, \ldots, w_k)}{\partial w_r \partial w_s} \bigg|_{\theta_0}, \text{ for } r, s = 1, \ldots, k. \]

We define

\[
\Sigma_{10} = \lim_{n \to \infty} \frac{\sigma_0^4}{2n||S_0^{-1}||_\infty} \begin{pmatrix}
tr((C_0S_0^{-1})^2) & \frac{tr((C_0S_0^{-1})(C_0S_0^{-1}))}{tr((C_0S_0^{-1})^2)} & \ldots & \frac{tr((C_kS_k^{-1})(C_kS_k^{-1}))}{tr((C_kS_k^{-1})^2)} \\
\frac{tr((C_0S_0^{-1})(C_0S_0^{-1}))}{tr((C_0S_0^{-1})^2)} & \frac{tr((C_0S_0^{-1})(C_1,0S_1^{-1}))}{tr((C_0S_0^{-1})^2)} & \ldots & \frac{tr((C_kS_k^{-1})(C_0S_0^{-1}))}{tr((C_kS_k^{-1})^2)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{tr((C_0S_0^{-1})(C_0S_0^{-1}))}{tr((C_0S_0^{-1})^2)} & \frac{tr((C_0S_0^{-1})(C_1,0S_1^{-1}))}{tr((C_0S_0^{-1})^2)} & \ldots & \frac{tr((C_kS_k^{-1})(C_kS_k^{-1}))}{tr((C_kS_k^{-1})^2)} \\
\end{pmatrix}
\]

and

\[
\Sigma_{20} = \lim_{n \to \infty} \frac{(\mu_0^{(4)} - 3\sigma_0^4)}{4n||S_0^{-1}||_\infty} \begin{pmatrix}
\sum_i ((C_0S_0^{-1})^2)_{ii} & \sum_i (C_0S_0^{-1})_{ii}(C_1,0S_1^{-1})_{ii} & \ldots & \sum_i (C_0S_0^{-1})_{ii}(C_kS_k^{-1})_{ii} \\
\sum_i (C_0S_0^{-1})_{ii}(C_0S_0^{-1})_{ii} & \sum_i (C_1,0S_1^{-1})_{ii}(C_0S_0^{-1})_{ii} & \ldots & \sum_i (C_kS_k^{-1})_{ii}(C_0S_0^{-1})_{ii} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_i (C_0S_0^{-1})_{ii}(C_0S_0^{-1})_{ii} & \sum_i (C_0S_0^{-1})_{ii}(C_1,0S_1^{-1})_{ii} & \ldots & \sum_i (C_kS_k^{-1})_{ii}(C_kS_k^{-1})_{ii} \\
\end{pmatrix} \left( \frac{tr((C_0S_0^{-1}))}{tr((C_0S_0^{-1})^2)} \right) 
\]

where \( \mu_0^{(4)} = \mathbb{E}(\varepsilon_i^4) \). Also, we define

\[ V = \Sigma_{10} + \Sigma_{20}. \]

Elements of first and second term in \( \Sigma_{20} \) are \( O(1/||S_0^{-1}||_\infty) \) from (S.19) and (S.22) reported in the Online Supplement, respectively. By a very minor modification of the argument in (S.23) in the Online Supplement, each element of \( \Sigma_{10} \) is \( O(1) \). Also, let \( D_0 \) be the \((k + 1) \times (k + 1)\) matrix with elements

\[
d_{11,0} = \lim_{n \to \infty} \frac{1}{n||S_0^{-1}||_\infty} \left( \frac{tr(S_0^{-1}C_0' + C_0S_0^{-1} - tr(C_0S_0^{-1})2I/n)}{n} \right)^2,
\]

\[
d_{ij,0} = \lim_{n \to \infty} \frac{\lambda_0}{n||S_0^{-1}||_\infty} \left( 2tr(C_{ij,0}S_0^{-1}) - \lambda_0 tr(S_0^{-1}C_j,0C_i,0S_0^{-1}) + tr(S_0^{-1}C_i,0S_0^{-1}C_j,0) \right) - \frac{4\lambda_0^2}{n^2||S_0^{-1}||_\infty} tr(C_j,0S_0^{-1})tr(C_i,0S_0^{-1}), \quad i, j = 2, \ldots, k + 1
\]

and

\[
d_{i,i,0} = d_{i,1,0} = \lim_{n \to \infty} \frac{2}{n||S_0^{-1}||_\infty} \left( \lambda_0 tr(S_0^{-1}C_0' + C_0S_0^{-1}) + tr(S_0^{-1}C_i,0S_0^{-1}C_i,0) \right) - \frac{4\lambda_0}{n^2||S_0^{-1}||_\infty} tr(C_i,0S_0^{-1})tr(C_i,0S_0^{-1}),
\]

\[
i = 2, \ldots, k + 1.
\]
Each element of $D_0$ is $O(1)$ from (S.22), (S.23) and (S.26). Finally, let

$$F = \begin{pmatrix}
-\frac{2}{\sigma_0^2} & 0 & \ldots & 0 \\
0 & -\frac{2\lambda_0}{\sigma_0^2} & 0 & \ldots \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & -\frac{2\lambda_0}{\sigma_0^2}
\end{pmatrix}.$$  

(4.7)

We introduce some further conditions to establish the limit distribution of $\hat{\theta}_2$.

**Assumption 8** The limits in $\Sigma_{10}$, $\Sigma_{20}$ and $D_0$ exist. Furthermore $\eta_{\min}(\Sigma_{20}) > 0$ and $\eta_{\min}(D_0) > 0$.

**Assumption 9**

a) $\sup_{\theta \in \Theta} (||C_{rs}(\theta)||_{\infty} + ||C'_{rs}(\theta)||_{\infty}) \leq K$ for $r, s = 1, \ldots, k$.

b) $\sup_{\theta \in \Theta} (||C_{rst}(\theta)||_{\infty} + ||C'_{rst}(\theta)||_{\infty}) \leq K$ for $r, s, t = 1, \ldots, k$.

Assumption 8 is a standard existence and non-singularity condition to ensure that the limit distribution of Theorem 2 is well defined. Assumption 9 extends Assumption 6 to uniform boundedness in row and column sum of second- and third-order derivatives of $C(\cdot)$, similarly to Assumption A4 in Lieberman (2010). We establish the following.

**Theorem 2.** Assume that model (2.1) and Assumptions 1-9 hold. Under $R1$,

$$(n||S_0^{-1}||_{\infty})^{1/2} (\hat{\theta}_2 - \theta_{20}) \overset{d}{\rightarrow} \mathcal{N}(0, \mathcal{V}),$$

where $\mathcal{V} = D^{-1}FVFD^{-1}$.

The proof of Theorem 2 is in Appendix A. We remark that the result in the theorem is consistent with an $(n||S_0^{-1}||_{\infty})^{1/2} = (n^{1+\alpha})^{1/2}$-normalization for the MI model discussed in Section 2.3, $\alpha \in (0, 1)$, see Theorem 3.2(c) of Phillips and Magdalinos (1987). The variance-covariance matrix $\mathcal{V}$ exists and it is nonsingular under Assumption 8 for $\lambda_0 \neq 0^4$, and can be estimated by replacing unknown parameters by their consistent estimates $\hat{\lambda}$, $\hat{w}$ and $\hat{\sigma}^2$, and $\hat{\mu}^{(4)} = \sum_n^{\kappa} \hat{\epsilon}_i^4 / n$, with $\hat{\epsilon} = y - \hat{\lambda}C(\hat{w})$.

In a similar fashion to the discussion following Theorem 1, Theorem 2 offers a generalization of standard limit theory available for SAR models under (2.21) to more flexible regimes for $||S_0^{-1}||_{\infty}$, while allowing for $C(\cdot)$ to be data-driven rather than exogenously chosen. Theorem 2 is new in the similarity literature as Lieberman (2010) only considered limit theory for estimates of $w_0$ when $\lambda_0 = 1$ in unilateral models that do not display a standard normal limiting distribution as $n$ increases. In general, under some standard regularity conditions, Theorem 2 accommodates a similarity structure for $C(w_0)$ such that $||C(w_0)||_{\infty} = 1$ and $\lambda_n = 1 - \rho / n^\alpha$ with $\alpha < 1$ and/or any standard weight structure for $C_0$ in line with Assumptions 6 and 9, such that $||S_0^{-1}||_{\infty} = O(n^\gamma)$ with $\gamma < 1$, and fixed $|\lambda_0| < 1$. The case $\lambda_0 = 1$ does not imply an obvious limit regime for $||S_0^{-1}||_{\infty}$ and has to be considered on a case-by-case basis, according to the resulting $S_0^{-1}$. For instance, in the two examples discussed in Sections 2.1 and 2.2 we obtain

---

4When $\lambda_0 = 0$, the parameter vector $w_0$ is not identified.
$\|S_0^{-1}\|_\infty \sim n$, corresponding to R2. A discussion on some useful results under R2 will be developed in the following section.

## 5 A testing framework in non-standard cases

In general, even though from Theorem 1 we know that the QML estimate of $\theta_{20}$ is consistent under both R1 and R2, for distributional and testing purposes we cannot infer whether R1 or R2 hold from a finite dataset, after replacing $\theta_{20}$ with its consistent estimate in $\|S_0^{-1}\|_\infty$. The limit distribution under R2 is generally non-standard, and needs to be investigated on a case-by-case basis. The development of a suitable numerical procedure to approximate the limit distribution of $\hat{\theta}_{2}$ under R2 and derive, e.g., confidence intervals for $\theta_{20}$, is beyond our scope in the present paper and it is under investigation in separate work. In this section we focus instead on the derivation of a suitable testing framework that can deliver reliable results under R2. Given our similarity-based model in (2.1), from the power series representation in (2.23) and the discussion thereafter, we deduce that Theorem 2 provides the theoretical framework to construct tests of hypotheses such as $H_0 : \lambda_0 = \bar{\lambda}$ for any fixed $\bar{\lambda} < 1$ against one- or two-sided alternatives, but tests of $H_0 : \lambda_0 = 1$ need to be developed by allowing for R2. In this section we present a testing framework and a numerical procedure to perform a test of

$$H_0 : \lambda_0 = 1 \text{ against } H_0 : \lambda_0 < 1. \tag{5.1}$$

Our procedure, after minor modifications, can in principle be applied to test any hypothesis such as $H_0 : \lambda_0 = \bar{\lambda}$ against a general one- or two-sided alternative in standard SAR models with an exogenous weight matrix that is not necessarily row-normalized (unlike our similarity-based structure $C(\theta)$), but for which $\|S^{-1}(\lambda_0)\|_\infty \sim n$.

Before introducing our testing algorithm, we provide a result to ensure that for the similarity-based weight matrix defined in (2.2), $\lambda_0 = 1$ implies $\|S^{-1}(\theta_{20})\|_\infty \sim n$. To this extent, we introduce an additional condition.

**Assumption 10** Given $C_0 = C_{0n}$ as implied by (2.2) and a positive constant $\delta$, for at least one $k = 1, \ldots, n$, $\left[\sum_{j=0}^n C_{0,j,k}\right]_k > \delta > 0 \forall i$.

Although Assumption 10 seems high-level, we stress that for similarity-based structures such as (1.4) and (1.5), $C_{0,n}$ typically has nonzero entries almost everywhere for each $n$, so that Assumption 10 is satisfied. In the Online Supplement we prove the following.

**Proposition 1** Let $C_0 = C_{n0}$ as implied by (2.2). The following two claims hold

a) $\lambda_0 = 1$ implies $\lim n^{-1}\|S_0^{-1}\|_\infty \leq 1$.

b) $\lambda_0 = 1$ and Assumption 10 implies $\lim n^{-1}\|S_0^{-1}\|_\infty > 0$.

Proposition 1 implies that the correct rate to develop inference on parameters in model (2.1) when $\lambda_0 = 1$ is $\|S_0^{-1}\|_\infty \sim n$.

We aim to construct a test of $H_0$ in (5.1) and start by re-defining model (2.1) as

$$y - C_0 y = (\lambda_0 - 1) C_0 y + \epsilon, \tag{5.2}$$
or equivalently
\[ S_0y = \beta_0C_0y + \epsilon, \tag{5.3} \]
with \( S_0 \) corresponding to \( S_0 \) in (2.2) under \( H_0 \), i.e. \( S_0 = I - C_0 \) and \( \beta_0 = \lambda_0 - 1 \). Equivalently,
\[ y = \beta_0S_0^{-1}C_0y + S_0^{-1}\epsilon. \tag{5.4} \]

Our hypotheses in (5.1) are thus equivalent to
\[ H_0 : \beta_0 = 0 \text{ against } H_1 : \beta_0 < 0. \tag{5.5} \]

Given \( w_0 \), from (5.3) the OLS estimator of \( \beta_0 \) is
\[ \hat{\beta} = \frac{y'C_0S_0y}{y'C_0C_0y} \tag{5.6} \]
and thus, under \( H_0 \), from (5.4)
\[ \hat{\beta} = \frac{y'C_0\epsilon}{y'C_0C_0y} = \frac{\epsilon'S_0^{-1}C_0\epsilon}{\epsilon'S_0^{-1}C_0C_0S_0^{-1}\epsilon}. \tag{5.7} \]

From Lemma 1 a) and b) we obtain, respectively,
\[ \epsilon'S_0^{-1}C_0\epsilon = O_p(n) \tag{5.8} \]
and
\[ \epsilon'S_0^{-1}C_0C_0S_0^{-1}\epsilon = O_p(n||S_0^{-1}||_{\infty}) = O(n^2), \tag{5.9} \]
where the last equality follows from Proposition 1, under (5.5). From (5.8) and (5.9), \( \hat{\beta} \) converges in probability to zero under \( H_0 \) in (5.5).

The statistic in (5.6) is not normalized and infeasible, thus we define our test statistic as
\[ \hat{\beta}_F = n \frac{y'\hat{C}'\hat{S}y}{y'\hat{C}'\hat{C}y}, \tag{5.10} \]
where \( \hat{C} \) and \( \hat{S} \) are obtained by replacing the unknown \( w_0 \) with its estimate \( \hat{w} \) under \( H_0 \) in (5.1) and \( \lambda_0 \) as its value under \( H_0 \). Under \( H_0 \), the limit distribution of (5.10) is non-standard. However, the simple ratio of quadratic forms in \( \epsilon \), as displayed in (5.7), allows a simple bootstrap implementation by means of the following steps:

1) Under \( H_0 \) in (5.1), estimate \( w_0 \) by the consistent profile QMLE and obtain \( \hat{C} \) and \( \hat{S} \). Compute \( \hat{\beta}_F \).

2) Obtain the \( n \times 1 \) vector of restricted (i.e. under \( H_0 \)) residuals \( \hat{\epsilon} = y - \hat{C}y \) and its centred version \( \hat{\epsilon}_c = \hat{\epsilon} - \sum^n \hat{\epsilon}_i/n \).

3) Generate \( B \) vectors of bootstrap residuals \( \hat{\epsilon}_b^* \), for \( b = 1, \ldots, B \) by resampling with replacement from \( \hat{\epsilon}_c \).
4) From the RHS of (5.7) we generate $B$ bootstrap statistics as
\[
\hat{\beta}^*_b = n \frac{\hat{\epsilon}^*_b S^{-1/2} \hat{C}' \hat{\epsilon}^*_b}{\hat{\epsilon}^*_b S^{-1/2} \hat{C}' \hat{C} S^{-1/2} \hat{\epsilon}^*_b}
\] (5.11)
and we sort them in ascending order.

5) We compute the $\alpha$ bootstrap quantile $\omega^*_\alpha$ as the solution of
\[
\frac{1}{B} \sum_{b=1}^{B} 1(\hat{\beta}^*_b \leq \omega^*_\alpha) = \alpha.
\] (5.12)

6) We reject $H_0$ in (5.1) at level $\alpha$ if
\[
\hat{\beta}_F < \omega^*_\alpha.
\] (5.13)

Equivalently, rather than steps 5) and 6) we can directly calculate bootstrap p-values as
\[
\frac{1}{B} \sum_{b=1}^{B} 1(\hat{\beta}^*_b \leq \hat{\beta}_F).
\] (5.14)

Step 3) above does not require any prior knowledge of the distribution of $\epsilon_i$, $i = 1, ..., n$. If we know that $\epsilon_i \sim N(0, \sigma^2_0)$ we can implement a parametric bootstrap and generate $\hat{\epsilon}^*_{i,b}$ for $i = 1, ..., n$ and $b = 1, ..., B$ as an i.i.d sample from $N(0, \hat{\sigma}^2)$, with $\hat{\sigma}^2 = \hat{\epsilon} \hat{\epsilon}/n$, $\hat{\epsilon}$ being the restricted residuals generated in step 2) above.

Alternatively, under normality of $\epsilon_i$, $i = 1, ..., n$ the implementation of the numerical Imhof/Davies procedure (e.g. Lu and King (2002)) allows to establish p-values of the test in (5.1) based on $\hat{\beta}_F$ after observing that, from Theorem 1,
\[
\hat{\beta}_F = \frac{y' C' S y}{y' C' C y} + o_p(1). \tag{5.15}
\]
The (approximate) p-value can be written as
\[
\mathbb{P} \left( \frac{y' C' S y}{y' C' C y} \leq b_F \right) = \mathbb{P} \left( y' B y \leq 0 \right), \tag{5.16}
\]
where $B = nC_0 S_0 - b_F C_0 C_0$, $y \sim N(0, \sigma^2 S_0^{-1/2} S_0^{-1})$ with $S_0 = I - C_0$ under $H_0$ and $b_F$ being the observed value of $\hat{\beta}_F$. The details of the numerical procedure are reported in, e.g., Lu and King (2002). From (5.15) we can implement the numerical procedure by replacing the unknown $w_0$ in $B$ with their consistent estimates under $H_0$.

The last issue we address is to suggest a suitable numerical procedure to construct reliable confidence sets for $w_0$ in case we fail to reject the null hypothesis in (5.1). As already discussed, Theorem 2 offers the basis of reliable inference for R1, but it does not deliver the asymptotic distribution and standard errors under R2. Even though the theoretical discussion of a suitably general bootstrap algorithm to establish inference under R2 is beyond our scope in the present paper, we suggest a simple bootstrap routine based on a non-studentized statistic that offer
reliable confidence sets for \(w_0\) when we fail to reject \(H_0\) in (5.1). Improvements to such routine via studentized statistics is left for future investigation. We perform the following steps:

1) Under \(H_0\) in (5.1), estimate \(w_0\) by the consistent profile QMLE and obtain \(\hat{C}\) and \(\hat{S}\).

2) Obtain the \(n \times 1\) of restricted (i.e. under \(H_0\)) residuals \(\hat{\epsilon} = y - \hat{C}y\) and its centred version \(\hat{\epsilon}_c = \hat{\epsilon} - \sum^n_i \hat{\epsilon}_i / n\).

3) Generate \(B\) vectors of bootstrap residuals \(\hat{\epsilon}_b^*\), for \(b = 1, ..., B\) by resampling with replacement from \(\hat{\epsilon}_c\).

4) Generate the bootstrap sample of \(B \times 1\) vectors \(y^*_b = (I - \hat{C})^{-1}\hat{\epsilon}_b^*\).

5) For each bootstrap sample obtained in point 4), compute estimates \(\hat{w}_b^*\) by the profile QMLE under \(H_0\) and sort them, component by component, in ascending order.

6) For each component of \(\hat{w}_b^*, \hat{w}_{j,b}^*\), with \(j = 1, ..., k\), compute the \(\alpha\) and \(1 - \alpha\) quantiles \(\omega_{j,\alpha}^*\) and \(\omega_{j,1-\alpha}^*\) as the respective solutions of

\[
\frac{1}{B} \sum_{b=1}^B 1(\hat{w}_{j,b}^* \leq \omega_{j,\alpha}^*) = \alpha \quad \text{and} \quad \frac{1}{B} \sum_{b=1}^B 1(\hat{w}_{j,b}^* \leq \omega_{j,1-\alpha}^*) = 1 - \alpha. \tag{5.17}
\]

7) For each \(j = 1, ..., k\), construct the approximate confidence interval for \(w_{j0}\) as \((\hat{w}_j - \omega_{j,1-\alpha}^*, \hat{w}_j - \omega_{j,\alpha}^*)\).

An example of practical implementation of both test of \(H_0\) in (5.1) and construction of confidence sets for \(w_0\) will be discussed in Section 7.

6 Monte Carlo simulations

In this section we report results of a small Monte Carlo exercise to assess finite sample performance of our estimates of \(\lambda_0\) and \(w_0\) in model (2.1). We generate \(X\) as a \(n \times 2\) matrix of i.i.d random variables from a uniform on support \([0, 4]\), although results are similar for i.i.d. Gaussian \(Xs\). The matrix \(X\) is generated once and kept fixed across replications. The disturbances \(\epsilon_i\)s are i.i.d., generated from \(N(0, \sigma_0^2)\), with \(\sigma_0^2 = 0.8\). We consider two choices for the similarity function, as given in (1.4) and (1.5), i.e.

\[
s_0(i,j) = \frac{1}{1 + w_{10}(X_{i1} - X_{j1})^2 + w_{20}(X_{i2} - X_{j2})^2}, \quad w_{10} = 1, \quad w_{20} = 1, \quad s_0(i,i) = 0 \quad \tag{6.1}
\]

and

\[
s_0(i,j) = \exp\left(-w_{10}(X_{i1} - X_{j1})^2 - w_{20}(X_{i2} - X_{j2})^2\right), \quad w_{10} = 1, \quad w_{20} = 1, \quad s_0(i,i) = 0 \quad \tag{6.2}
\]

We consider three different values of \(\lambda_0 = 0.1, 0.5, 1\) and assess Monte Carlo bias and variance using 1000 Monte Carlo replications.

Across Tables 1-6, bias and variance of \(\hat{\lambda}\) decrease steadily as \(n\) increases, as expected. Moreover, results displayed in the first two columns of Tables 3 and 6 confirm quite clearly the faster
Table 1: Monte Carlo bias and variance of QMLE of parameters in model (2.1), with similarity function as in (6.1) and $\lambda_0 = 0.1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.0851</td>
<td>0.0469</td>
<td>0.0198</td>
<td>0.1281</td>
<td>-0.0780</td>
<td>0.1249</td>
</tr>
<tr>
<td>100</td>
<td>0.0807</td>
<td>0.0455</td>
<td>-0.0005</td>
<td>0.1242</td>
<td>-0.0084</td>
<td>0.1238</td>
</tr>
<tr>
<td>250</td>
<td>0.0674</td>
<td>0.0420</td>
<td>0.0247</td>
<td>0.1226</td>
<td>0.0228</td>
<td>0.1224</td>
</tr>
<tr>
<td>500</td>
<td>0.0668</td>
<td>0.0328</td>
<td>-0.0053</td>
<td>0.1165</td>
<td>-0.0027</td>
<td>0.1151</td>
</tr>
</tbody>
</table>

Table 2: Monte Carlo bias and variance of QMLE of parameters in model (2.1), with similarity function as in (6.1) and $\lambda_0 = 0.5$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>-0.1083</td>
<td>0.0582</td>
<td>-0.0365</td>
<td>0.2337</td>
<td>-0.0221</td>
<td>0.2360</td>
</tr>
<tr>
<td>100</td>
<td>-0.0979</td>
<td>0.0580</td>
<td>-0.0519</td>
<td>0.2308</td>
<td>-0.0201</td>
<td>0.2329</td>
</tr>
<tr>
<td>250</td>
<td>-0.0972</td>
<td>0.0578</td>
<td>-0.0161</td>
<td>0.2322</td>
<td>-0.0352</td>
<td>0.2285</td>
</tr>
<tr>
<td>500</td>
<td>-0.0924</td>
<td>0.0550</td>
<td>0.0029</td>
<td>0.2292</td>
<td>-0.0192</td>
<td>0.2196</td>
</tr>
</tbody>
</table>

Table 3: Monte Carlo bias and variance of QMLE of parameters in model (2.1), with similarity function as in (6.1) and $\lambda_0 = 1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.0759</td>
<td>0.0342</td>
<td>-0.0125</td>
<td>0.1374</td>
<td>0.0077</td>
<td>0.1311</td>
</tr>
<tr>
<td>100</td>
<td>0.0814</td>
<td>0.0344</td>
<td>-0.0388</td>
<td>0.1336</td>
<td>-0.0172</td>
<td>0.1379</td>
</tr>
<tr>
<td>250</td>
<td>0.0711</td>
<td>0.0317</td>
<td>0.0171</td>
<td>0.1353</td>
<td>-0.0139</td>
<td>0.1333</td>
</tr>
<tr>
<td>500</td>
<td>0.0668</td>
<td>0.0328</td>
<td>-0.0054</td>
<td>0.1325</td>
<td>0.0051</td>
<td>0.1263</td>
</tr>
</tbody>
</table>

Table 4: Monte Carlo bias and variance of QMLE of parameters in model (2.1), with similarity function as in (6.2) and $\lambda_0 = 0.1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
<th>$E(...)$</th>
<th>$Var(...)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>-0.0399</td>
<td>0.0455</td>
<td>-0.0015</td>
<td>0.1896</td>
<td>0.0037</td>
<td>0.1888</td>
</tr>
<tr>
<td>100</td>
<td>-0.0320</td>
<td>0.0478</td>
<td>0.0249</td>
<td>0.1868</td>
<td>0.0119</td>
<td>0.1855</td>
</tr>
<tr>
<td>250</td>
<td>-0.0328</td>
<td>0.0466</td>
<td>0.0280</td>
<td>0.1778</td>
<td>0.0125</td>
<td>0.1814</td>
</tr>
<tr>
<td>500</td>
<td>-0.0317</td>
<td>0.0407</td>
<td>0.0209</td>
<td>0.1740</td>
<td>-0.0058</td>
<td>0.1807</td>
</tr>
</tbody>
</table>

Table 5: Monte Carlo bias and variance of QMLE of parameters in model (2.1), with similarity function as in (6.2) and $\lambda_0 = 0.5$.

The rate of convergence when $\lambda_0 = 1$ compared to cases with $\lambda_0 < 1$. Results for Monte Carlo bias of $\hat{w}_1$ and $\hat{w}_2$, instead, display a less clearly decreasing pattern, although the bias tends to remain
Table 6: Monte Carlo bias and variance of QMLE of parameters in model (2.1), with similarity function as in (6.2) and $\lambda_0 = 1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$E(\hat{\lambda}) - \lambda_0$</th>
<th>$Var(\hat{\lambda})$</th>
<th>$E(\hat{w}<em>1) - w</em>{10}$</th>
<th>$Var(\hat{w}_1)$</th>
<th>$E(\hat{w}<em>2) - w</em>{20}$</th>
<th>$Var(\hat{w}_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>-0.0399</td>
<td>0.0059</td>
<td>0.0469</td>
<td>0.2396</td>
<td>0.0698</td>
<td>0.2472</td>
</tr>
<tr>
<td>100</td>
<td>-0.0231</td>
<td>0.0026</td>
<td>0.0338</td>
<td>0.2142</td>
<td>0.0552</td>
<td>0.2152</td>
</tr>
<tr>
<td>250</td>
<td>-0.0153</td>
<td>0.0013</td>
<td>-0.0427</td>
<td>0.1884</td>
<td>-0.0647</td>
<td>0.1774</td>
</tr>
<tr>
<td>500</td>
<td>-0.0062</td>
<td>0.0004</td>
<td>-0.0654</td>
<td>0.1244</td>
<td>-0.0691</td>
<td>0.1192</td>
</tr>
</tbody>
</table>

very small across all scenarios. Monte Carlo variances of $\hat{w}_1$ and $\hat{w}_2$ decrease as $n$ increase for all scenarios, with the decreasing rate appearing to be faster when $\lambda_0 = 1$ compared to $\lambda_0 < 1$, as expected, especially for the similarity structure in (6.2).

7 Empirical example

In this section we report a small empirical application of our methodology based on a data driven, similarity based weight matrix. This empirical analysis is intended as an illustration of the theoretical contribution of this paper and thus the investigation of model selection strategies as well as of practical implications of our analysis fall beyond our scope. We use the Boston house price data (Harrison and Rubinfeld (1978)) and its ‘corrected’ version (Gilley and Pace (1996)), which also includes information on LON (tract point longitudes in decimal degrees and LAT (tract point latitudes in decimal degrees) for the 506 census tracts in the Boston Standard Metropolitan Area during the early 1970s. The dependent variable of interest is $\log(MEDV)$, which is the logarithm of the median price (in thousands of dollars) for owner-occupied houses. The dataset contains additional information about various environmental and socio-economic variables.

Table 7: Dataset variables.

<table>
<thead>
<tr>
<th>variable</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>crime</td>
<td>per capita crime rate by town;</td>
</tr>
<tr>
<td>zn</td>
<td>proportion of residential land zoned for lots over 25,000 sq.ft;</td>
</tr>
<tr>
<td>indus</td>
<td>proportion of non-retail business acres per town;</td>
</tr>
<tr>
<td>chas</td>
<td>Charles River dummy variable (= 1 if tract bounds river; 0 otherwise);</td>
</tr>
<tr>
<td>nox</td>
<td>nitrogen oxides concentration (parts per 10 million);</td>
</tr>
<tr>
<td>rm</td>
<td>average number of rooms per dwelling;</td>
</tr>
<tr>
<td>age</td>
<td>proportion of owner-occupied units built prior to 1940;</td>
</tr>
<tr>
<td>dis</td>
<td>weighted mean of distances to five Boston employment centres;</td>
</tr>
<tr>
<td>rad</td>
<td>index of accessibility to radial highways;</td>
</tr>
<tr>
<td>tax</td>
<td>full-value property-tax rate per 10,000$;</td>
</tr>
<tr>
<td>ptdratio</td>
<td>pupil-teacher ratio by town;</td>
</tr>
<tr>
<td>black</td>
<td>1000 * ($Bk - 0.63)^2$, where $Bk$ is the proportion of blacks by town;</td>
</tr>
<tr>
<td>lstat</td>
<td>lower status of the population (percent).</td>
</tr>
</tbody>
</table>

For our empirical analysis we consider model (2.1) with $\log(MEDV)$ as the main variable of interest, i.e.

$$\log(MEDV) = \lambda_0 C_0 \log(MEDV) + \epsilon$$ (7.1)
Table 8: Left panel: Inference for parameters in model (7.1), with similarity function as in (6.2) and setup a). Middle panel: Inference for parameters in model (7.1), with similarity function as in (6.2) and setup b). Right panel: Inference for parameters in model (7.1), with “null” similarity function as in c). t-statistics are in brackets

<table>
<thead>
<tr>
<th>n</th>
<th>( \lambda )</th>
<th>( \psi_1 )</th>
<th>( \psi_2 )</th>
<th>( \beta_F )</th>
<th>( \lambda )</th>
<th>( \psi_1 )</th>
<th>( \psi_2 )</th>
<th>( \beta_F )</th>
<th>( \lambda )</th>
<th>( \beta_F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>506</td>
<td>0.9928</td>
<td>0.7540</td>
<td>0.0197</td>
<td>-0.7673</td>
<td>0.9955</td>
<td>0.0024</td>
<td>1.2997</td>
<td>1.6902</td>
<td>0.9922</td>
<td>-0.0175</td>
</tr>
<tr>
<td>( CV_{b,1} )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.2597</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0895</td>
<td>-</td>
<td>0.2082</td>
</tr>
<tr>
<td>( CV_{b,0.05} )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0412</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.3417</td>
<td>-</td>
<td>0.0903</td>
</tr>
<tr>
<td>( CV_{b,0.01} )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.8599</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1.7653</td>
<td>-</td>
<td>-2.3481</td>
</tr>
<tr>
<td>Imhof p-value</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0290</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.9051</td>
<td>-</td>
<td>0.0333</td>
</tr>
<tr>
<td>( LE_{b,0.05} )</td>
<td>-</td>
<td>0.2647</td>
<td>0.0090</td>
<td>-</td>
<td>-</td>
<td>0.0004</td>
<td>1.2962</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( UE_{b,0.05} )</td>
<td>-</td>
<td>0.9128</td>
<td>0.0203</td>
<td>-</td>
<td>-</td>
<td>0.0041</td>
<td>2.5998</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 8: Left panel: Inference for parameters in model (7.1), with similarity function as in (6.2) and setup a). Middle panel: Inference for parameters in model (7.1), with similarity function as in (6.2) and setup b). Right panel: Inference for parameters in model (7.1), with “null” similarity function as in c). t-statistics are in brackets.

with two different choices of data-driven weights based on the similarity function in (1.4):

a) \( X_1 \) and \( X_2 \) are \( n \times 1 \) vectors containing, respectively information on \( ptratio \) and \( crime \).

b) \( X_1 \) and \( X_2 \) are \( n \times 1 \) vectors containing, respectively information on \( age \) and \( rm \).

c) \( C_0 \) is generated as our “null model” in Section 2.1, i.e. each element of \( C_0 \) is equal to \( 1/(n-1) \).

The choice in a) captures socio-economic features, such that the median house price in census tract \( i \) is formed according to a similarity criterion based on tracts that have similar socio-economic features, such as similar schools’ quality and local crime rate. The choice in b), instead captures neighbour’s tracts with similar housing characteristics. Results for estimates and t-statistics (in brackets) are reported in the first row of Table 8.

From Theorem 1, the estimates reported in the first row of Table 8 are consistent. Although \( \sigma^2_0 \) is not the parameter of interest, we report that its estimate amounts to \( \hat{\sigma}^2 = 0.1133 \), \( \hat{\sigma}^2 = 0.1088 \) and \( \hat{\sigma}^2 = 0.1872 \) for scenarios a), b) and c), respectively. However, \( \hat{\lambda} \) appears to be very close to unity in all scenarios and this casts some doubts on the reliability of the t-statistics reported in brackets. Indeed, the t-statistics reported in Table 8 are derived based on Theorem 2 and might fail in case \( \lambda_0 = 1 \). We thus apply the testing procedure we outlined in Section 5 to test \( H_0 \) in (5.1). In the fifth, ninth and eleventh columns of the first row of Table 8 we report the value of \( \beta_F \) in (5.10) for scenarios a), b) and c), while the second, third and fourth row of Table 8 display bootstrap critical values at \( \alpha = 0.1, 0.05, 0.01 \), respectively.

From the first panel of Table 8, we fail to reject \( H_0 \) in (5.1) based on the rejection rule in (5.13) at \( \alpha = 0.01 \). Critical values reported in Table 8 have been obtained by resampling with replacement from centred residuals, although similar results are obtained by a parametric bootstrap under normality of error terms, i.e. when \( \hat{\epsilon}_{i,b} \) for \( i = 1, \ldots, n \) and \( b = 1, \ldots, B \) are generated as an i.i.d sample from \( \mathcal{N}(0, \hat{\sigma}^2) \), with \( \hat{\sigma}^2 = \hat{\sigma}^2 / n \), \( \hat{\sigma} \) being the restricted residuals. Specifically, in the latter case we obtain critical values 0.2642/0.0652/−0.7803 at levels 0.10/0.05/0.01 respectively. From the middle panel of Table 8 we see that we fail to reject \( H_0 \) according to (5.13) even at \( \alpha = 0.10 \). From a parametric bootstrap we obtain critical values 0.0898/−0.4024/−1.7868 at levels 0.10/0.05/0.01 respectively. Finally, from the last panel in Table 8, we fail to reject
\( \mathcal{H}_0 \) according to (5.13) at \( \alpha = 0.01 \), with the parametric bootstrap delivering again results that are similar to those displayed in Table 8 and equal to 0.2056/0.1046/−2.8988 at levels 0.10/0.05/0.01 respectively.

In Table 8 we also report the p-value of the test in (5.1) obtained by the implementation of the Imhof/Davies procedure outlined in the previous section. For all three cases, results are in line with those obtained by the bootstrap algorithm. In the numerical procedure (e.g. Lu and King (2002)) we set \( E_I = 10^{-5} \) and \( E_T = 0.9 \times 10^{-6} \), where \( E_I \) and \( E_T \) denote integration and truncation errors, respectively, and we adopt the Ansley et al. (1992) truncation bound.

The last two rows of Table 8 display lower end-points and upper end-points (LEP and UEP) of confidence intervals for \( w_{10} \) and \( w_{20} \) with \( 1 - \alpha = 95\% \), for both scenarios a) and b). The confidence intervals have been obtained by implementing the algorithm described in Section 5. We deduce from LEP and UEP that \( w_{10} \) and \( w_{20} \) are statistically significant at \( \alpha = 0.05 \) for both scenarios a) and b).

8 Conclusion

In this paper we develop a framework that merges standard literature on spatial autoregressions with the notion of empirical similarity. Although standard spatial autoregressions and empirical similarity models are formally similar, their respective theoretical results have been derived independently and, to the best of our knowledge, this is the first attempt to unify them. Formally, our model mimics a spatial autoregression with a data-driven weighting structure that depends on some unknown parameters, which have to be estimated alongside the spatial coefficient. In turn, the spatial coefficient can take values at the boundary of the parameters’ space, a feature that is ruled out in SAR literature, and the model’s assumptions allow strong forms of strong cross-sectional dependence which are not accommodated in the spatial econometrics literature.

In the context of the empirical similarity literature, our model extends the results in Lieberman (2010) to bilateral models, and hence to non-ordered data, and it allows the magnitude of the similarity structure (which corresponds to the spatial parameters in the SAR jargon) to be estimated explicitly and not being fixed to unity ex-ante. In the setup of a simple model with no exogenous regressors and homoskedastic disturbances, we provide the asymptotic theory to establish consistency and the limit distribution of parameters’ estimates in a unified regime that allows “weak” and “strong” forms of cross-sectional interactions. In the boundary cases of \( \lambda_0 = 1 \) or the local to unity regime with \( \lambda_n = 1 - c/n \), with \( c \) being an arbitrary positive constant, we are able to establish consistency of estimates, but the limiting distribution needs to be investigated on a case-by-case basis. It turns out that the order of magnitude of \( \| S^{-1}(\theta) \|_\infty \) is critical for the separation between cases. We provide a simple procedure to test \( H_0 : \lambda_0 = 1 \) and to derive confidence intervals for the remaining parameters when \( H_0 \) is not rejected. We apply our results to assess the house prices formation in the Boston area.
Appendix A: proofs of main Theorems

Proof or Theorem 1. To prove the identification condition we write, for $\theta_2 \neq \theta_{20}$,

$$\tilde{L}^p(\theta_2) - \tilde{L}^p(\theta_{20}) = \log \left( \frac{y' S y}{y' S_{0} S_{0} y} \right) - \frac{2}{n} \log |S| + \frac{2}{n} \log |S_0| + o_p(1) \quad (A.1)$$

$$= \log \left( \frac{y' S y}{y' S_{0} S_{0} y} \right) + \frac{1}{n} \log |S^{-1} S_{0} S_{0}^{-1}| + o_p(1). \quad (A.2)$$

Identification holds as long as either (A.1) or (A.2) is strictly positive. For $\gamma = 0$ (A.2) is strictly positive by Lemma 4a), while for $0 < \gamma \leq 1$ (A.1) is strictly positive by Lemma 4b).

In order to show consistency of $\hat{\theta}$ we proceed similarly to the proof of Theorem 1 in Delgado and Robinson (2015). Let $\mathcal{N}_\delta = \{ \theta : ||\theta - \theta_{20}|| < \delta \}$ for some $\delta > 0$, and $\bar{\mathcal{N}}_\delta$ its complement. In general, the following chain of inequalities holds:

$$\mathbb{P}(\hat{\theta} \in \bar{\mathcal{N}}_\delta) \leq \mathbb{P} \left( \inf_{\mathcal{N}_\delta} L^p(\theta_2) < L^p(\theta_{20}) \right) \leq \mathbb{P} \left( \sup_{\theta_2} |L^p(\theta_2) - \tilde{L}^p(\theta_2)| \geq \inf_{\mathcal{N}_\delta} |\tilde{L}^p(\theta_2) - \tilde{L}^p(\theta_{20})| \right). \quad (A.3)$$

Consistency of $\hat{\theta}$ would therefore follow once we show the following two statements:

$$\inf_{\mathcal{N}_\delta} (\tilde{L}^p(\theta_2) - \tilde{L}^p(\theta_{20})) > \epsilon, \text{ for all sufficiently large } n \text{ and for some } \epsilon > 0, \quad (A.4)$$

$$\sup_{\theta_2} |L^p(\theta_2) - \tilde{L}^p(\theta_2)| \xrightarrow{p} 0, \text{ as } n \to \infty. \quad (A.5)$$

The proofs of (A.4) and (A.5) are given in Lemmas 5 and 6, respectively. ■

Proof of Theorem 2. From (3.8), by the mean value theorem,

$$(\hat{\theta}_2 - \theta_{20}) = \left( \frac{1}{||S^{-1}||_{\infty}} \frac{\partial^2 L^p(\tilde{\theta}_2)}{\partial \theta_2 \partial \theta_2} \right)^{-1} \frac{1}{||S^{-1}||_{\infty}} \frac{\partial L^p(\theta_{20})}{\partial \theta_2}, \quad (A.6)$$

with $\tilde{\theta}_2$ satisfying $|\tilde{\theta}_2 - \theta_{20}| < |\hat{\theta}_2 - \theta_{20}|$ for $j = 1, ..., k + 1$. The proof of Theorem 2 follows from the proofs of Lemmas 7-9 and by Crámer’s theorem. ■

Appendix B: additional technical Lemmas

Lemmas 1, 2 and 3 and respective proofs are reported in the Online Supplement.

Lemma 4. Under Assumptions 1-7:

a) For $\gamma = 0$ in (3.2),

$$\log \left( \frac{y' S y}{y' S_{0} S_{0} y} \right) + \frac{1}{n} \log |S^{-1} S_{0} S_{0}^{-1}| + o_p(1) > 0. \quad (B.1)$$
b) For $0 < \gamma \leq 1$ in (3.2),

$$\log \left( \frac{y' S' S y}{y' S_0' S_0 y} \right) - \frac{2}{n} \log |S| + \frac{2}{n} \log |S_0| + o_p(1) > 0. \quad (B.2)$$

**Proof of Lemma 4.** Let $\Delta = S_0^{-1} S' S S_0^{-1}$, which is strictly positive definite for $\theta \neq \theta_0$. We consider the case a) first. By replacing $y = S_0^{-1} \epsilon$ into the first term at the LHS of (B.1), under Assumption 6a) and by Lemma 3 in the Online Supplement, we obtain

$$\log \left( \frac{y' S' S y}{y' S_0' S_0 y} \right) = \log \left( \frac{\epsilon' S_0^{-1} S' S S_0^{-1} \epsilon}{\epsilon' \epsilon} \right) = \log(\sigma^2 tr(\Delta)/\sigma^2 n) + o_p(1) \quad (B.3)$$

and hence the LHS of (B.1) becomes

$$\log \left( \frac{1}{n} tr(\Delta)|\Delta|^{-1/n} \right) + o_p(1) \quad (B.4)$$

By the arithmetic-geometric mean inequality, $\frac{1}{n} tr(\Delta) \geq |\Delta|^{1/n}$, with equality iff the eigenvalues of $\Delta$ are all the same. Denote by $\eta_i(\Delta)$ for $i = 1, ..., n$ the eigenvalues of $\Delta$. Now, $\Delta = S_0^{-1} S' S S_0^{-1}$ is symmetric and therefore it is diagonalizable as $\Delta = Q \Lambda Q'$ where $Q$ is the orthogonal matrix of eigenvectors of $\Delta$ satisfying $Q' Q = QQ' = I$, and $\Lambda$ is the diagonal matrix of eigenvalues. Suppose that $\eta_1(\Delta) = \cdots = \eta_n(\Delta) = \eta$. Then $\Delta = \eta I$. Without loss of generality, we can set $\eta = 1$. We need to show that Assumptions 3 and 4 are sufficient for

$$\Delta - I = S_0^{-1'} (S' S - S_0' S_0) S_0^{-1} \neq 0. \quad (B.5)$$

Let $A = S' S - S_0' S_0$ and assume that it is non-null. Then there is at least one non-zero column vector in $A$, say $(A')_j = (a')_j$ for $j = 1, ..., n$. The quantity $S_0^{-1'} (a')_j$ is a linear combination of the columns of $S_0^{-1'}$, which must be non-zero because $S_0^{-1}$ is full rank under Assumption 4. Thus, there must be at least one non-zero row vector in $Z = S_0^{-1'} A$, say $z'_j$. Now, $z'_j S_0^{-1}$ is a linear combinations of the rows of $S_0^{-1}$, which again cannot be zero, because $S_0^{-1}$ is full rank. Hence $\Delta - I = S_0^{-1'} (S' S - S_0' S_0) S_0^{-1} \neq 0$, proving the claim in part a).

To prove part b), we have that under Assumption 6a) and by Lemma 1b) in the Online Supplement,

$$\frac{y' S' S y}{y' S_0' S_0 y} = O_p(||S_0^{-1}||_\infty) = O_p(n^\gamma), \quad (B.6)$$

with the LHS of (B.6) being non-negative by construction. We need to ensure that the LHS of (B.2) remains strictly positive by establishing boundedness of the remaining terms in (B.2). By construction $C_{ii} = 0$ for each $i = 1, ..., n$, such that $tr(S) = n, \forall \theta_2$. The second term at the LHS of (B.2) is bounded since, by the arithmetic-geometric mean inequality

$$1 = \frac{1}{n} tr(S) \geq |S|^{1/n}, \quad (B.7)$$

so that $|S|^{2/n}$ is bounded by 1 for all $\theta_2$. The third term at the LHS of (B.2) is bounded under
Assumption 4 since
\[ 0 < \eta_{\text{min}}^2(S_0) \leq |S_0|^2, \] (B.8)
concluding the claim. ■

**Lemma 5.** Under Assumptions 1-7:
\[ \inf_{\mathcal{N}_\delta} (\hat{L}^p(\theta_2) - \hat{L}^p(\theta_{20})) > \epsilon, \text{ for all sufficiently large } n \text{ and for some } \epsilon > 0 \] (B.9)
with \( \hat{L}^p(\cdot) \) defined in (3.10).

**Proof of Lemma 5.** We prove (B.9) by using the inequality
\[
\inf_{\theta_2: \|\theta_2 - \theta_2^1\| < \eta, \theta_2 \in \Theta} \left( \hat{L}^p(\theta_2) - \hat{L}^p(\theta_{20}) \right) \geq \left( \hat{L}^p(\theta_2^1) - \hat{L}^p(\theta_{20}) \right) \nonumber \\
- \sup_{\theta_2: \|\theta_2 - \theta_2^1\| < \eta, \theta_2 \in \Theta} \left| \hat{L}^p(\theta_2) - \hat{L}^p(\theta_2^1) \right|, \tag{B.10}
\]
where \( \eta \) is any positive constant, \( \theta_2^1 \in \Theta \setminus \{\theta_{20}\} \), and \( \Theta \) is compact under Assumption 2 and hence it has a finite subcover. We need to show that the RHS of (B.10) is strictly positive as \( n \) increases. The first term on the RHS of (B.10) is strictly positive under Assumptions 3 and 4 for \( \gamma = 0 \) (from (A.1)/(A.2) and (B.4) - (B.5) and discussion below), or it diverges to \( +\infty \) for \( 0 < \gamma \leq 1 \) as \( n \) increases under the mild Assumption 4 (from (A.1)/(A.2) and (B.6) - (B.8)).

We start by showing that the RHS of (B.10) is strictly positive in the limit for \( 0 < \gamma \leq 1 \). We refer to (B.10) as
\[
\hat{L}^p(\theta_2) - \hat{L}^p(\theta_2^1) = \log \left( \frac{y'y' S' S y}{y'y' S' S^2 y} \right) - \frac{2}{n} \log |S(\theta_2)| + \frac{2}{n} \log |S^1| + o_p(1). \tag{B.11}
\]
The first term on the RHS of the last equation is bounded, since, under Assumption 6a) and by Lemma 1b) in the Online Supplement, both numerator and denominator in the argument of the logarithm are \( O_p(n||S_0^{-1}||_\infty) \) uniformly in \( \theta_2 \), so that the first term is \( O_p(1) \). Also, from the arithmetic-geometric mean inequality and under Assumption 4
\[ 0 < \eta_{\text{min}}(S) \leq |S|^{1/n} \leq 1, \] (B.12)
uniformly in \( \theta_2 \), such that second and third terms in the last displayed equation are bounded. We therefore conclude that the RHS of (B.10) increases without bound as \( n \to \infty \), when \( 0 < \gamma \leq 1 \). Next, we prove that the RHS of (B.10) is strictly positive for \( \gamma = 0 \). Under Assumptions 3 and 4, from Lemma 4a) the first term at the RHS of (B.10) is strictly positive. Also, from (3.11),
\[
\frac{y'y' S'y}{y'y} = \frac{tr(S_0^{-1} S'y S'y S_0^{-1})}{tr(S_0^{-1} S'y S_0^{-1})} + o_p(1). \tag{B.13}
\]
Thus, we write the second term at the RHS of (B.10) as

\[
\hat{L}^p(\theta_2) - \hat{L}^p(\theta_2^{\dagger}) = \log \left( \frac{\text{tr}(S_0^{-1} S S_0^{-1})}{\text{tr}(S_0^{-1} S_0^{-1})} \right) + \log |S'|^{-\frac{1}{n}} - \log \left( \frac{\text{tr}(S_0^{-1} S_0^{-1} S S_0^{-1})}{\text{tr}(S_0^{-1} S_0^{-1})} \right) - \log |S'|^{-\frac{1}{n}} + o_p(1)
\]

\[
= \log \left( \frac{\text{tr}(S_0^{-1} S_0^{-1} S S_0^{-1})}{\text{tr}(S_0^{-1} S_0^{-1})} \right) + \frac{1}{n} \log |S'|^{-\frac{1}{n}} - \log \left( \frac{\text{tr}(S_0^{-1} S_0^{-1} S S_0^{-1})}{\text{tr}(S_0^{-1} S_0^{-1})} \right)
\]

\[
= \log \left( \frac{\text{tr}(\Omega_0\Omega^{-1})}{\text{tr}(\Omega_0\Omega^{-1})} \right) + \frac{1}{n} \log |\Omega^{-1}| - \log \left( \frac{\text{tr}(\Omega_0\Omega^{-1})}{\text{tr}(\Omega_0\Omega^{-1})} \right)
\]

\[
= \log \left( 1 + \frac{\text{tr}(\Omega^{-1} - \Omega_0^{-1})}{\text{tr}(\Omega_0\Omega^{-1})} \right) + \frac{1}{n} \log |I_n + (\Omega^{-1} - \Omega^{-1})\Omega|
\]

(B.14)

where \( \Omega = (S'S)^{-1} \). Let \( \omega_j = \eta_j((\Omega^{-1} - \Omega^{-1})\Omega_0) \) and \( \nu_j = \eta_j((\Omega^{-1} - \Omega^{-1})\Omega) \), respectively, where \( \eta_j(A) \) as usual denotes the \( j \)-th eigenvalue of a generic matrix \( A \). Using the inequality \( x \geq \log (1 + x) \),

\[
\hat{L}^p(\theta_2) - \hat{L}^p(\theta_2^{\dagger}) \leq \frac{\sum |\omega_j|}{\text{tr}(\Omega_0\Omega^{-1})} + \frac{\sum |\nu_j|}{n}.
\]

We have the inequality

\[
\frac{\text{tr}(\Omega_0\Omega^{-1})}{n} \geq |\Omega_0\Omega^{-1}|^{1/n} = |\Omega_0|^{1/n} |\Omega^{-1}|^{1/n} \geq \eta_{\text{min}}(\Omega_0) \eta_{\text{min}}(\Omega_0^{-1}).
\]

(B.15)

Under Assumption 4 \( \eta_{\text{min}}(\Omega_0) > 0 \). Also,

\[
\eta_{\text{min}}(\Omega_0^{-1}) = \frac{1}{\eta_{\text{max}}(\Omega_0)} \geq \frac{1}{\|\Omega^{-1}\|^2},
\]

such that, for \( \gamma = 0 \) in (3.2),

\[
\text{tr}(\Omega_0\Omega^{-1}) \geq Kn, \quad 0 < K < \infty.
\]

(B.16)

Hence,

\[
\hat{L}^p(\theta_2) - \hat{L}^p(\theta_2^{\dagger}) \leq \frac{\sum |\omega_j|}{\text{tr}(\Omega_0\Omega^{-1})} + \frac{\sum |\nu_j|}{n} \leq K \frac{\sum |\omega_j|}{n} + \frac{\sum |\nu_j|}{n}
\]

\[
\leq Kn^{-1/2} \left( \left( \sum \omega_j^2 \right)^{1/2} + \left( \sum \nu_j^2 \right)^{1/2} \right)
\]

\[
= Kn^{-1/2} \left( \|\Omega^{-1} - \Omega_0^{-1}\|_{F} \right) \leq K \left( \|\Omega^{-1} - \Omega_0^{-1}\| \right),
\]

\[
(B.17)
\]

where we have used \( \|A\|_{F} \leq \sqrt{n} \|A\| \), with \( K \) being a generic constant that can change from step to step, and

\[
\|\Omega\| \leq \|S^{-1}\|^2 < K,
\]

uniformly in \( \theta_2 \). Under Assumption 5, \( \Omega^{-1} \) has uniformly bounded derivatives. Thus, Assump-
tion 5 guarantees that entries of $\Omega^{-1}$ are uniformly continuous functions of $\theta_2$. On the other hand, the eigenvalues of a matrix are continuous functions of the elements of the matrix and hence the eigenvalues are continuous functions of uniformly continuous functions of $\theta_2$. By the Heine Cantor theorem, continuity in addition to compactness of $\Theta_2$ are sufficient for uniform continuity of eigenvalues of $\Omega^{-1}$. We conclude the proof of (B.9) for $\gamma = 0$ in (3.2) as the first term on the RHS of (B.10) tends to a positive constant and the second term tends to zero uniformly in $\theta_2$ since, fixing $\delta > 0$, there exists $\zeta > 0$ such that for large enough $n$

$$\sup_{\theta_2: ||\theta_2 - \bar{\theta}_2|| < \zeta} \left\| \Omega^{-1} - \Omega_1^{-1} \right\| < \delta.$$ 

\[\Box\]

**Lemma 6.** Under Assumptions 1-7,

$$\sup_{\theta_2} |\mathcal{L}^p(\theta_2) - \bar{\mathcal{L}}^p(\bar{\theta}_2)| \xrightarrow{P} 0, \quad n \to \infty. \quad (B.21)$$

with $\mathcal{L}^p(\cdot)$ and $\bar{\mathcal{L}}^p(\cdot)$ defined respectively in (3.8) and (3.10).

**Proof of Lemma 6.** Let $\mathcal{N}(\theta_2, \delta)$ a $\delta$-neighborhood of $\theta_2$ such that

$$\mathcal{N}(\theta_2, \delta) = \{\theta_2^j : |\lambda^j - \lambda| < \delta/(k+1), |w_j^j - w_j| < \delta/(k+1) \text{ for each } j = 1, \ldots, k\}. \quad (B.22)$$

Let $\tilde{\theta}_2$ such that: $|\bar{\lambda} - \lambda| < |\lambda^j - \lambda|, |\bar{w}_j - w_j| < |w_j^j - w_j|$ for each $j$. Let $\bar{S} = S(\tilde{\theta}_2)$ and $\tilde{S} = S(\bar{\theta}_2)$, with analogous notation for $C(\cdot)$ and $C_r(\cdot)$ for $r = 1, \ldots, k$. Since $\Theta_2$ is compact under Assumption 2 and thus it has a finite sub-covering, we focus on

$$\sup_{\theta_2^j \in \mathcal{N}(\theta_2, \delta)} \left| \mathcal{L}^p(\theta_2^j) - \bar{\mathcal{L}}^p(\bar{\theta}_2) \right| \leq \sup_{\theta_2^j \in \mathcal{N}(\theta_2, \delta)} \left| \mathcal{L}^p(\theta_2^j) - \mathcal{L}^p(\theta_2) \right| + \left| \mathcal{L}^p(\theta_2) - \bar{\mathcal{L}}^p(\bar{\theta}_2) \right|$$

$$+ \sup_{\theta_2^j \in \mathcal{N}(\theta_2, \delta)} \left| \bar{\mathcal{L}}^p(\theta_2^j) - \bar{\mathcal{L}}^p(\bar{\theta}_2) \right| \quad (B.23)$$

Pointwise convergence in probability of $\mathcal{L}^p(\theta_2)$ to $\bar{\mathcal{L}}^p(\bar{\theta}_2)$ holds by definition of $\hat{\sigma}^2(\theta_2)$ so that the second term at the RHS of (B.23) is $o_p(1)$.

We start with the first term at the RHS of (B.23). From (3.8), by the mean value theorem we can write

$$y' S' S y = y' S' S y + \frac{\partial y' S' S y}{\partial \lambda} (\lambda^j - \lambda) + \sum_{j=1}^k \frac{\partial y' S' S y}{\partial w_j} (w_j^j - w_j)$$

$$= y' S' S y - 2y' S' \bar{C} y (\lambda^j - \lambda) - 2\lambda \sum_{j=1}^k y' S' \bar{C}_j y (w_j^j - w_j). \quad (B.24)$$
so that
\[
\Delta_n(\theta_2, \theta_2) : = \frac{|y' S'' y - y' S' S y|}{y'y} = \frac{2}{y'y} |y' S' C y(\lambda^2 - \lambda) + \lambda \sum_{j=1}^{k} |y' S' C_j y (w_j^2 - w_j)|
\]
\[
\leq \frac{K}{y'y} \left( |y' S' C y||\lambda^2 - \lambda| + \sum_{j=1}^{k} |y' S' C_j y||w_j^2 - w_j| \right) \tag{B.25}
\]

From Lemma 1b) in the Online Supplement, under Assumption 6, we deduce
\[
y' y = O_p \left( n \|S_0^{-1}\|_{\infty} \right) \\
y' S' C y = O_p \left( n \|S_0^{-1}\|_{\infty} \right) \\
y' S' C_j y = O_p \left( n \|S_0^{-1}\|_{\infty} \right). \tag{B.26}
\]

From the definition of stochastic equicontinuity \cite{Andrews (1994)}, (B.25) and (B.26) imply that for all \( \zeta_1 > 0 \) and \( \zeta_2 > 0 \) there exists a \( \delta > 0 \) such that
\[
\limsup_{n \to \infty} \Pr \left( \sup_{\theta_2^* \in N(\theta_2, \delta)} \Delta_n(\theta_2, \theta_2^*) > \zeta_1 \right) < \zeta_2, \tag{B.27}
\]
where \( \zeta_1, \zeta_2 \) and \( \delta \) do not depend on \( \theta_2 \). This proves that the first term of \( L^p(\theta_2) \) in (3.8) is stochastic equicontinuous.

Now we consider the second term of \( L^p(\theta_2) \) in (3.8). We have
\[
-2 \log |S| = -2 \log |S| + 2 tr(S^{-1} C)(\lambda^2 - \lambda) + 2 \lambda \sum_{j=1}^{k} tr(S^{-1} C_j)(w_j^2 - w_j) \tag{B.28}
\]
Under Assumption 6, by (S.22),
\[
|tr(S^{-1} C)| = O(n) \quad \text{and} \quad |tr(S^{-1} C_j)| = O(n) \quad \forall j = 1, \ldots, k. \tag{B.29}
\]
Hence, for every \( \nu > 0 \) there exists a neighborhood \( N(\theta_2, \delta) \) that does not depend on \( n \) such that for all \( n > N \),
\[
\sup_{\theta_2^* \in N(\theta_2, \delta)} \left| \frac{2 \log |S^2|}{n} - \frac{2 \log |S|}{n} \right| \leq K \left( (\lambda^2 - \lambda) + \sum_{j=1}^{k} (w_j^2 - w_j) \right) \leq K \delta \leq \nu. \tag{B.30}
\]
Thus, the second term of of \( L^p(\theta_2) \) in (3.8) is uniformly equicontinuous.

In order to conclude the proof we need to focus on the third term at the RHS of (B.23) and show stochastic equicontinuity of \( \tilde{\sigma}^{2*}(\theta_2) \) in (3.10), as equicontinuity of the second term in \( \tilde{L}^p(\theta_2) \) follows as in (B.28) - (B.30). By the MVT, under Assumption 5 and since the module is
a continuous function,

\[
\sup_{\theta_2 \in N(\theta_2, \delta)} \left| \lim_{n \to \infty} \left( \frac{-2y' \hat{S}' \hat{C} y (\lambda^2 - \lambda)}{y' y} - \frac{2\lambda \sum_{j=1}^{k} y' \hat{S}' C_j y (w_j^2 - w_j)}{y' y} \right) \right| \leq K \sup_{\theta_2 \in N(\theta_2, \delta)} \left| \lim_{n \to \infty} \left( \frac{|y' \hat{S}' \hat{C} y||\lambda^2 - \lambda|}{y' y} + \sum_{j=1}^{k} |y' \hat{S}' C_j y||w_j^2 - w_j|}{y' y} \right) \right| \leq K \delta \lim_{n \to \infty} \left( \frac{|y' \hat{S}' \hat{C} y|}{y' y} + \sum_{j=1}^{k} |y' \hat{S}' C_j y|}{y' y} \right)
\]

where \( K \), as usual, denotes a constant that can change value from step to step. Under Assumption 6, from Lemma 1b) in the Online Supplement, for each \( \zeta_2 \) there exists a \( \Delta \) such that

\[
Pr \left( \lim_{n \to \infty} \left( \frac{|y' \hat{S}' \hat{C} y|}{y' y} + \sum_{j=1}^{k} |y' \hat{S}' C_j y|}{y' y} > \Delta \right) < \zeta_2. \quad (B.32)
\]

Let \( \zeta_1 = \delta K \Delta \). We have

\[
Pr \left( \sup_{\theta_2 \in N(\theta_2, \delta)} \left| \lim_{n \to \infty} \left( \frac{-2y' \hat{S}' \hat{C} y (\lambda^2 - \lambda)}{y' y} - \frac{2\lambda \sum_{j=1}^{k} y' \hat{S}' C_j y (w_j^2 - w_j)}{y' y} \right) \right| > \zeta_1 \right) \leq Pr \left( \lim_{n \to \infty} \left( \frac{|y' \hat{S}' \hat{C} y|}{y' y} + \sum_{j=1}^{k} |y' \hat{S}' C_j y|}{y' y} > \Delta \right) < \zeta_2, \quad (B.33)
\]

concluding the proof.  ■

Lemma 7. Under Assumptions 1, 2 and 6:

\[
\frac{\partial \mathcal{L}^p(\theta_{20})}{\partial \lambda} = O_p \left( \left( \frac{||S_0^{-1}||}{n} \right)^{1/2} \right) \quad \text{and} \quad \frac{\partial \mathcal{L}^p(\theta_{20})}{\partial w_j} = O_p \left( \left( \frac{||S_0^{-1}||}{n} \right)^{1/2} \right), \quad j = 1, \ldots, k.
\]

Proof of Lemma 7. By standard algebra,

\[
\frac{\partial \mathcal{L}^p(\theta_{20})}{\partial \lambda} = -2 \frac{y' C_0' S_0 y}{y' S_0' S_0 y} + \frac{2}{n} \text{tr} \left( S_0^{-1} C_0 \right) = -2 \left( \frac{\epsilon' S_0^{-1} C_0' \epsilon}{\epsilon' \epsilon/n} - \text{tr} \left( S_0^{-1} C_0 \right) \right) = -2 \left( \frac{\epsilon' \epsilon}{n} \right)^{-1} \epsilon' \left( S_0^{-1} C_0 - \frac{I}{n} \text{tr}(S_0^{-1} C_0) \right) \epsilon = O_p \left( \left( \frac{||S_0^{-1}||}{n} \right)^{1/2} \right) \quad (B.34)
\]

29
and, for $j = 1, \ldots, k$,

$$
\frac{\partial^2 \mathcal{L}^p(\theta_{20})}{\partial \lambda^2} = -2 \lambda_n \frac{y' C'_{j,0} S_{0} y}{y' S_{0} y} + \frac{2 \lambda_n}{n} \operatorname{tr} \left( S_{0}^{-1} C_{j,0} \right) \\
= -2 \lambda_n \left( \frac{y' S_{0}^{-1} C'_{j,0} \epsilon}{\epsilon' \epsilon n} - \operatorname{tr} \left( S_{0}^{-1} C_{j,0} \right) \right) \\
= -2 \lambda_n \left( \frac{\epsilon' \epsilon}{n} \right)^{-1} \operatorname{tr} \left( S_{0}^{-1} C_{j,0} \right) \epsilon = O_p \left( \left( \frac{\|S_{0}^{-1}\|_\infty}{n} \right)^{1/2} \right),
$$

where the order of magnitude in both expressions follows from Lemma 2 reported in the Online Supplement.

**Lemma 8.** Under Assumptions 1-9,

$$
\frac{1}{\|S_{0}^{-1}\|_\infty} \bar{H} \overset{p}{\to} D_0 > 0,
$$

where the elements of $D_0$ are given in (4.4), (4.5) and (4.6).

**Proof of Lemma 8.** Let $\bar{S} = S(\bar{\theta}_2)$. Using standard algebra we derive the Hessian $\bar{H} = H(\bar{\theta})$ as

$$
\frac{\partial^2 \mathcal{L}^p(\theta_{20})}{\partial \lambda^2} = 2 \lambda_n \frac{y' C'_{j,0} S_{0} y}{y' S_{0} y} + \frac{2 \lambda_n}{n} \operatorname{tr} \left( S_{0}^{-1} C_{j,0} \right) \\
= 2 \lambda_n \left( \frac{\epsilon' S_{0}^{-1} C'_{j,0} \epsilon}{\epsilon' \epsilon / n} - \operatorname{tr} \left( S_{0}^{-1} C_{j,0} \right) \right) \\
= 2 \lambda_n \left( \frac{\epsilon' \epsilon}{n} \right)^{-1} \operatorname{tr} \left( S_{0}^{-1} C_{j,0} \right) \epsilon = O_p \left( \left( \frac{\|S_{0}^{-1}\|_\infty}{n} \right)^{1/2} \right),
$$

and

$$
\frac{\partial^2 \mathcal{L}^p(\theta_{20})}{\partial \lambda \partial \lambda} = -2 \lambda_n \frac{y' C'_{j,0} S_{0} y}{y' S_{0} y} + \frac{2 \lambda_n}{n} \operatorname{tr} \left( S_{0}^{-1} C_{j,0} \right) \\
= -2 \lambda_n \frac{\epsilon' S_{0}^{-1} C'_{j,0} \epsilon}{\epsilon' \epsilon / n} - \frac{2 \lambda_n}{n} \operatorname{tr} \left( S_{0}^{-1} C_{j,0} \right) \\
= -2 \lambda_n \left( \frac{\epsilon' \epsilon}{n} \right)^{-1} \operatorname{tr} \left( S_{0}^{-1} C_{j,0} \right) \epsilon = O_p \left( \left( \frac{\|S_{0}^{-1}\|_\infty}{n} \right)^{1/2} \right).
$$

We focus in detail on (B.37), but a similar argument holds for both (B.38) and (B.39). From Theorem 1,

$$
\hat{\theta}_2 - \theta_{20} = O_p(\alpha_n),
$$

with $\alpha_n$ being a deterministic sequence that converges to zero as $n$ increases. Since $\epsilon' \epsilon \sim n$ and from Lemma 1b), under Assumption 6 the first term in (B.37) is $O_p(\|S_{0}^{-1}\|_\infty)$ since
Since $\alpha$ as long as $\alpha$ trivially satisfied for $\gamma$

and $\bar{\theta}$ where the order of magnitude follow from (S.22), (S.23) and (S.26) in the Online Supplement.

$-4$ from Lemma 3 equivalent to $\Theta$ which is $\Theta$ to $\Theta$.

where $\lambda^*$ is $\lambda^*$ to $\lambda^*$.

\begin{align}
\frac{1}{n} \sum_{j=1}^k \epsilon_j' S_0^{-1} C_j' S_0^{-1} \epsilon_j (\bar{w}_j - w_{j0}) = O_p(1) + O_p(\|S_0^{-1}\|_\infty \alpha_n),
\end{align}

(B.41)

where the first equality follows by a mean value theorem, with $\theta^*_j$ being the intermediate point and $S^* = S(\theta_j^*)$ (with the same notation applying to similar quantities), such that, for each $j = 1, \ldots, k + 1$, $|\theta^*_j - \theta_20| < |\theta^*_j - \theta_2| < |\theta^*_j - \theta_20|$, while the second equality follows by Lemma 1. The second term in (B.37) is $O_p(1)$ provided that $\alpha_n = O(1/\|S_0^{-1}\|_\infty)$, and converges to

\begin{align}
\lim_{n \to \infty} \sigma^2 \frac{\lambda_n^2}{n} = O(1),
\end{align}

(B.42)

as long as $\alpha_n = o(1/\|S_0^{-1}\|_\infty)$. Note that both $\alpha_n = O(1/\|S_0^{-1}\|_\infty)$ and $\alpha_n = o(1/\|S_0^{-1}\|_\infty)$ are trivially satisfied for $T = 0$ in (3.2). A similar argument follows for the third term of (B.37), since

\begin{align}
\frac{1}{n} \epsilon (\bar{S}^{-1} \bar{C})^2 = O(\|S_0^{-1}\|_\infty) + O_p(\|S_0^{-1}\|_\infty^2 \alpha_n),
\end{align}

(B.43)

which is $O_p(\|S_0^{-1}\|_\infty)$ as long as $\alpha_n = O(1/\|S_0^{-1}\|_\infty)$ and we can write

\begin{align}
\frac{1}{n} \epsilon (\bar{S}^{-1} \bar{C})^2 \to \frac{1}{n} \epsilon (\bar{S}^{-1} \bar{C})^2 = O(1)
\end{align}

(B.44)

when $\alpha_n = o(1/\|S_0^{-1}\|_\infty)$. In order to avoid repetition we omit a similar argument for (B.38) and (B.39) and by standard algebra of quadratic forms in i.i.d. random variables, we obtain from Lemma 3

\begin{align}
\frac{1}{n} \frac{\partial^2 L(p) (\bar{\theta})}{\partial \lambda^2} \to \frac{1}{n} \frac{2}{n} \epsilon (\bar{S}^{-1} \bar{C})^2 = O(1),
\end{align}

(B.45)

\begin{align}
\frac{1}{n} \frac{2}{n} \epsilon (\bar{S}^{-1} \bar{C})^2 = O(1)
\end{align}

and

\begin{align}
\frac{1}{n} \frac{2}{n} \epsilon (\bar{S}^{-1} \bar{C})^2 = O(1),
\end{align}

(B.46)

where the order of magnitude follow from (S.22), (S.23) and (S.26) in the Online Supplement.
Lemma 9. Under Assumptions 1-9,
\[
(n||S_0^{-1}||_\infty)^{1/2} \begin{pmatrix}
\frac{1}{||S_0^{-1}||_\infty} \frac{\partial L_p(\theta_{20})}{\partial \lambda}
\frac{1}{||S_0^{-1}||_\infty} \frac{\partial L_p(\theta_{20})}{\partial \omega_1}
\vdots \\
\vdots \\
\frac{1}{||S_0^{-1}||_\infty} \frac{\partial L_p(\theta_{20})}{\partial \omega_k}
\end{pmatrix} \xrightarrow{d} N(0, V),
\] (B.47)

where V is a positive definite variance-covariance matrix given by (B.55), (B.56), (B.57), or, equivalently, by (4.1) and (4.2).

Proof of Lemma 9. In order to prove (B.47), we define
\[
U_n = U = \frac{1}{(n||S_0^{-1}||_\infty)^{1/2}} \begin{pmatrix}
\epsilon' \left( S_0^{-1}C_0' - \frac{1}{n} tr(C_0S_0^{-1}) \right) \\
\epsilon' \left( S_0^{-1}C_{1,0}' - \frac{1}{n} tr(C_{1,0}S_0^{-1}) \right) \\
\vdots \\
\epsilon' \left( S_0^{-1}C_{k,0}' - \frac{1}{n} tr(C_{k,0}S_0^{-1}) \right)
\end{pmatrix},
\]
so that (B.47) can be written as
\[
(n||S_0^{-1}||_\infty)^{1/2} \begin{pmatrix}
\frac{1}{||S_0^{-1}||_\infty} \frac{\partial L_p(\theta_{20})}{\partial \lambda}
\frac{1}{||S_0^{-1}||_\infty} \frac{\partial L_p(\theta_{20})}{\partial \omega_1}
\vdots \\
\vdots \\
\frac{1}{||S_0^{-1}||_\infty} \frac{\partial L_p(\theta_{20})}{\partial \omega_k}
\end{pmatrix} = FU + o_p(1),
\] (B.48)

with
\[
F = \begin{pmatrix}
\frac{-2}{\sigma_0^2} & 0 & \ldots & 0 \\
0 & \frac{-2\lambda}{\sigma_0^2} & 0 & \ldots \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \frac{-2\lambda}{\sigma_0^2}
\end{pmatrix}. \tag{B.49}
\]

We define \( \psi_{ij} = \psi_{ijn} \) the \((k+1) \times 1 \) vectors \( \psi_{ij}, \ldots, \psi_{(k+1)ij} \)' such that, for each \( i, j = 1, \ldots, n \),
\[
\psi_{ij} = \frac{1}{2} \begin{pmatrix}
(C_0S_0^{-1})_{ij} \\
(C_{1,0}S_0^{-1})_{ij} \\
\vdots \\
(C_{k,0}S_0^{-1})_{ij}
\end{pmatrix}, \tag{B.50}
\]

with \( A = A + A' \) for any generic matrix \( A \), and let \( \Psi_s \) be the \( n \times n \) matrix with \( \psi_{sij}, s = 1, \ldots, k+1 \).
as its \((i,j)\)-th component. We can write \(U = \sum_{i=1}^{n} u_i / (n||S_0^{-1}||_\infty)^{1/2}\), with

\[
u_i = u_{in} = (\epsilon_i^2 - \sigma_0^2)\psi_{ii} + 2\epsilon_i \sum_{j<i} \psi_{ij}\epsilon_j - \frac{1}{n} \sum_{j=1}^{n} \psi_{jj}(\epsilon_i^2 - \sigma_0^2).
\] (B.51)

So, \(\{u_i, 1 \leq i \leq n, n = 1, 2, \ldots, \} \) is a triangular array of martingale differences with respect to the filtration formed by the \(\sigma\)-field generated by \(\{\epsilon_j; j < i\}\).

In the sequel, all the summations will range from 1 to \(n\), unless otherwise specified, and we define \(\mu_0^{(4)} = \mathbb{E}(\epsilon_i^4), \mu_0^{(3)} = \mathbb{E}(\epsilon_i^3)\) for each \(i\). Let

\[
\Omega = \Omega_n = \text{Var}(U) = \frac{1}{n||S_0^{-1}||_\infty} \sum_{i=1}^{n} \text{Var}(u_i)
\]

\[
= (\mu_0^{(4)} - \sigma_0^4) \frac{1}{n||S_0^{-1}||_\infty} \left( \sum_{i} \psi_{ii}\psi_i' - \frac{1}{n} \sum_{i} \sum_{j} \psi_{ij}\psi_{jj}' \right) + \frac{4}{n||S_0^{-1}||_\infty}\sigma_0^4 \sum_{i} \sum_{j<i} \psi_{ij}\psi_{ij}'
\]

\[
= (\mu_0^{(4)} - 3\sigma_0^4) \frac{1}{n||S_0^{-1}||_\infty} \sum_{i} \psi_{ii}\psi_i' - (\mu_0^{(4)} - \sigma_0^4) \frac{2}{n^2||S_0^{-1}||_\infty} \sum_{i} \sum_{j} \psi_{ij}\psi_{jj}' + \frac{2}{n||S_0^{-1}||_\infty}\sigma_0^4 \sum_{i} \sum_{j} \psi_{ij}\psi_{ij}'.
\] (B.52)

and \(z_i = z_{in} = \zeta \Omega^{-1/2} u_i / (n||S_0^{-1}||_\infty)^{1/2}\), with \(\zeta\) being any deterministic \((k+1) \times 1\) vector that satisfies \(\zeta'\zeta = 1\). By Theorem 2 of Scott (1973), \(\sum_{i=1}^{n} z_i \to \mathcal{N}(0,1)\), as long as

\[
\sum_{i=1}^{n} \mathbb{E}(\epsilon_i^2 | \epsilon_j, j < i) \to 1
\] (B.53)

and

\[
\sum_{i=1}^{n} \mathbb{E}(z_i^2 1(|z_i| > \delta)) \to 0, \ \forall \delta > 0,
\] (B.54)

where \(1(\cdot)\) is the indicator function.

We define

\[
V = \lim_{n \to \infty} \Omega = \Sigma_{10} + \Sigma_{20},
\] (B.55)

with

\[
\Sigma_{10} = \lim_{n \to \infty} \frac{2\sigma_0^{(4)}}{n||S_0^{-1}||_\infty} \left( \begin{array}{cccc}
\sum_{i} \sum_{j} \psi_{ii}^1 \psi_{jj}^1 & \sum_{i} \sum_{j} \psi_{ii}^1 \psi_{jj}^2 & \cdots & \cdots \\
\sum_{i} \sum_{j} \psi_{ii}^2 \psi_{jj}^1 & \sum_{i} \sum_{j} \psi_{ii}^2 \psi_{jj}^2 & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \sum_{i} \sum_{j} \psi_{(k+1)ij} \psi_{(k+1)ji}
\end{array} \right)
\] (B.56)
and

\[ \Sigma_{20} = \lim_{n \to \infty} \left( \frac{\mu_0^{(4)} - 3\sigma_0^4}{n||S_0^{-1}||_\infty} \right) \left( \begin{array}{cccc}
\sum \psi_{1ii}^2 & \sum \psi_{1ii}\psi_{2ii} & \cdots \\
\sum \psi_{2ii}\psi_{1ii} & \sum \psi_{2ii}^2 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
\sum \psi_{(k+1)ii}\psi_{(k+1)ii} & \cdots & \sum \psi_{(k+1)ii}^2 & \sum \psi_{(k+1)ii}^2 \\
\end{array} \right) \]

\[ - \lim_{n \to \infty} \left( \frac{\mu_0^{(4)} - \sigma_0^4}{n^2||S_0^{-1}||_\infty} \right) \left( \begin{array}{cccc}
\sum \sum \psi_{1i1} & \sum \sum \psi_{1i2} & \cdots \\
\sum \sum \psi_{2i1} & \sum \sum \psi_{2i2} & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
\sum \sum \psi_{(k+1)i1} & \cdots & \sum \sum \psi_{(k+1)i2} & \sum \sum \psi_{(k+1)i2} \\
\end{array} \right) , \quad (B.57) \]

where the explicit forms of \( \Sigma_{10} \) and \( \Sigma_{20} \) are given in (4.1) and (4.2). Also, \( V \) is nonsingular under Assumption 8 and each element of \( \Sigma_{10} \) is \( O(1) \) by a very minor modification of the argument in (S.23) reported in the supplement, while elements of first and second term in \( \Sigma_{20} \) are \( O(1/||S_0^{-1}||_\infty) \) from (S.19) and (S.22), respectively.

We start by showing (B.53) by proving, equivalently,

\[ \sum_{i=1}^{n} E \left( \zeta^2_i \mid \epsilon_j, j < i \right) - \zeta' \Omega^{-1/2} \Omega^{-1/2} \zeta \to 0, \quad (B.58) \]

which is

\[ \zeta' \Omega^{-1/2} \left( \frac{1}{n||S_0^{-1}||_\infty} \sum_i E(u_iu_i' \mid \epsilon_j, j < i) - \Omega \right) \Omega^{-1/2} \zeta \]

\[ = \frac{4}{n||S_0^{-1}||_\infty} \zeta' \Omega^{-1/2} \left( \sigma_0^2 \sum_i \left( \sum_{j<i} \psi_{ij} \epsilon_j \right) \left( \sum_{j<i} \psi_{ij} \epsilon_j \right)' \right) - \sigma_0^4 \sum_i \sum_{j<i} \psi_{ij} \psi_{ij}' + \mu_0^{(3)} \sum_i \left( \psi_{ii} - \frac{1}{n} \sum_t \psi_{tt} \right) \sum_{j<i} \psi_{ij} \epsilon_j \right) \Omega^{-1/2} \zeta \to 0. \]

Since \( \Omega = O(1) \) as \( n \) increases, we need to show (for a typical element of the following matrices) that

\[ \frac{1}{n||S_0^{-1}||_\infty} \left( \sigma_0^2 \sum_i \left( \sum_{j<i} \psi_{ij} \epsilon_j \right) \left( \sum_{j<i} \psi_{ij} \epsilon_j \right)' - \sigma_0^4 \sum_i \sum_{j<i} \psi_{ij} \psi_{ij}' \right) \to 0 \quad (B.59) \]

and

\[ \frac{1}{n||S_0^{-1}||_\infty} \mu_0^{(3)} \sum_i \left( \psi_{ii} - \frac{1}{n} \sum_t \psi_{tt} \right) \sum_{j<i} \psi_{ij} \epsilon_j \to 0. \quad (B.60) \]
We begin by showing (B.59). We consider the typical elements of the LHS of (B.59)
\[
\frac{\sigma^2}{n||S_0^{-1}||_{\infty}} \left( \sum_{i} \sum_{j<i} \psi_{sij}^2 (\epsilon_j^2 - \sigma_0^2) + \sum_{i} \sum_{j<k, j \neq k} \psi_{sij} \psi_{sik} \epsilon_j \epsilon_k \right), \quad s = 1, \ldots, k+1, \tag{B.61}
\]
and
\[
\frac{\sigma^2}{n||S_0^{-1}||_{\infty}} \left( \sum_{i} \sum_{j<i} \psi_{sij} \psi_{tij} (\epsilon_j^2 - \sigma_0^2) + \sum_{i} \sum_{j<k, j \neq k} \psi_{sij} \psi_{tik} \epsilon_j \epsilon_k \right), \quad s, t = 1, \ldots, k+1, \quad s \neq t. \tag{B.62}
\]

The first term in (B.61) has mean zero and variance bounded by
\[
\frac{K}{n^2||S_0^{-1}||_{\infty}^2} \sum_{i} \sum_{k} \sum_{j<i,k} \psi_{sij}^2 \psi_{skj}^2 \leq \frac{K}{n^2||S_0^{-1}||_{\infty}^2} \sum_{i} \sum_{k} \sum_{j} \psi_{sij}^2 \psi_{skj}^2 \leq \frac{K}{n^2||S_0^{-1}||_{\infty}^2} \left( \max_j \sum_i \psi_{sij}^2 \right) \sum_k \sum_j \psi_{skj}^2 = O \left( \frac{||S_0^{-1}||_{\infty}}{n} \right), \tag{B.63}
\]
where the last equality follows from (S.23), since
\[
\sum_k \sum_j \psi_{skj}^2 = \frac{1}{4} tr((C_0 S_0^{-1} + S_0^{-1} C_0')^2), \quad \text{or} \quad \sum_k \sum_j \psi_{skj}^2 = \frac{1}{4} tr((C_0 S_0^{-1} + S_0^{-1} C_0')^2) \quad j = 1, \ldots, k, \tag{B.64}
\]
and letting $e_j$ to denote the $n \times 1$ vector with 1 in the $j$–th position and zero elsewhere,
\[
\sum_i \psi_{sij}^2 = e_j^t \Psi_s^2 e_j \leq ||\Psi_s||^2 = O(||S_0^{-1}||_{\infty}^2). \tag{B.65}
\]

Under R1 and by Markov’s inequality, the first term in (B.61) is $o_p(1)$.

The second term in (B.61) has again mean zero and variance bounded by
\[
\frac{K}{n^2||S_0^{-1}||_{\infty}^2} \left( \sum_{i} \sum_{p} \sum_{j<i,p} \sum_{k<i,p} |\psi_{sij} \psi_{sik} \psi_{spj} \psi_{skp}| \right) \\
\leq \frac{K}{n^2||S_0^{-1}||_{\infty}^2} \sum_{i} \sum_{p} \sum_{j} \sum_{k} |\psi_{sij} \psi_{sik}| (\psi_{spj}^2 + \psi_{skp}^2) \\
\leq \frac{K}{n^2||S_0^{-1}||_{\infty}^2} \left( \sup_i \sum_k |\psi_{sik}| \right) \left( \sup_j \sum_i |\psi_{sij}| \right) \sum_p \sum_j \psi_{spj}^2 \\
\quad + \frac{K}{n^2||S_0^{-1}||_{\infty}^2} \left( \sup_i \sum_j |\psi_{sij}| \right) \left( \sup_k \sum_i |\psi_{sik}| \right) \sum_p \sum_k \psi_{spk}^2 = O \left( \frac{||S_0^{-1}||_{\infty}}{n} \right), \tag{B.66}
\]
again from (S.23) and since, for $s = 1, \ldots, k+1$, $||\Psi_s||_{\infty} = O(||S_0^{-1}||_{\infty})$. Under R1 then, we conclude that the second term in (B.61) is $o_p(1)$. 

35
The proof that (B.62) is $o_p(1)$ is virtually identical and it is omitted to avoid repetitions. We prove (B.60) by observing that the typical element at the LHS of (B.60) is

$$
\frac{1}{n||S_0^{-1}||_{\infty}}\sum_{i}^{|\mathbb{H}(3)} \tilde{\psi}_{sii} \sum_{j<i} \psi_{tij} \epsilon_j, \quad s, t = 1, \ldots, k + 1,
$$

(B.67)

where

$$
\tilde{\psi}_{sii} = \psi_{sii} - \frac{1}{n} \sum_{t} \psi_{itt}.
$$

(B.68)

The term in (B.67) has mean zero and variance bounded by

$$
\frac{K}{n^2||S_0^{-1}||_{\infty}^2} \sum_{i} \sum_{k} \sum_{j<i,k} |\tilde{\psi}_{sii} \tilde{\psi}_{skk} \psi_{tij} \psi_{tjk}| \leq \frac{K}{n^2||S_0^{-1}||_{\infty}^2} \sum_{i} \sum_{j} \sum_{k} |\psi_{tij}||\psi_{tkj}|(\tilde{\psi}_{sii}^2 + \tilde{\psi}_{skk}^2)
$$

$$
\leq \frac{K}{n^2||S_0^{-1}||_{\infty}^2} \left( \left( \sup_{j} \sum_{k} |\psi_{tkj}| \right) \left( \sup_{i} \sum_{j} |\psi_{tij}| \right) \sum_{i} \tilde{\psi}_{sii}^2 + \left( \sup_{k} \sum_{j} |\psi_{tkj}| \right) \left( \sup_{i} \sum_{j} |\psi_{tij}| \right) \sum_{k} \tilde{\psi}_{skk}^2 \right)
$$

$$= O \left( \frac{||S_0^{-1}||_{\infty}}{n} \right),
$$

(B.69)

where the last equality follows since

$$
\sum_{i} \tilde{\psi}_{sii}^2 \leq \sum_{i} \sum_{j} \tilde{\psi}_{sij}^2 = tr(\tilde{\Psi}_s^2) = tr \left( \left( \Psi_s - tr(\Psi_s) I \right) n \right)^2 = O \left( \max(n, ||S_0^{-1}||_{\infty}) \right) = O(n||S_0^{-1}||_{\infty}),
$$

(B.70)

from (S.22) and (S.23) in the Online Supplement. Under R1 the term in (B.60) is thus $o_p(1)$, concluding the proof of (B.53).

We prove (B.54) by showing the sufficient Lyapunov condition

$$
\sum_{i} \mathbb{E}|z_i|^{2+\delta} \rightarrow 0, \quad \text{for some} \quad \delta > 0
$$

(B.71)

and showing, for a typical standardized element of $u_i$, $s = 1, \ldots, k + 1$,

$$
\left( \frac{1}{n||S_0^{-1}||_{\infty}^{1/2}} \right)^{2+\delta} \sum_{i} \mathbb{E}|u_{si}|^{2+\delta} = \left( \frac{1}{n||S_0^{-1}||_{\infty}^{1/2}} \right)^{2+\delta} \sum_{i} \mathbb{E}(\mathbb{E}|u_{si}|^{2+\delta} | \epsilon_j, j < i) \rightarrow 0.
$$

(B.72)

We have, by the $c_r$ inequality,

$$
\left( \frac{1}{n||S_0^{-1}||_{\infty}} \right)^{1+\delta/2} \sum_{i} \mathbb{E}(\mathbb{E}|u_{si}|^{2+\delta} | \epsilon_j, j < i)
$$

$$
\left( \frac{K}{n||S_0^{-1}||_{\infty}} \right)^{1+\delta/2} \sum_{i} \tilde{\psi}_{sii}^{2+\delta} + \left( \frac{K}{n||S_0^{-1}||_{\infty}} \right)^{1+\delta/2} \sum_{j<i} \mathbb{E}|\psi_{sij} \epsilon_j|^{2+\delta}.
$$

(B.73)

The first term of (B.73) is

$$
\left( \frac{K}{n||S_0^{-1}||_{\infty}} \right)^{1+\delta/2} \left( \sup_{i} \tilde{\psi}_{sii}^{\delta} \right) \sum_{i} \tilde{\psi}_{sii}^{2} = O \left( \frac{1}{n||S_0^{-1}||_{\infty}^{\delta/2}} \right) = o(1),
$$

(B.74)
since the second factor is $O(1)$, given (S.19) and (S.22) in the Online Supplement, and the third factor is $O(n\|S_0^{-1}\|_\infty)$ from (B.70). The second term in (B.73), by the Burkholder and von Bahr/Esseen inequality, is bounded by

$$\left(\frac{K}{n\|S_0^{-1}\|_\infty}\right)^{1+\delta/2} \sum_i \sum_{j<i} \psi_{sij}^{2}\delta^{1+\delta/2} \leq \left(\frac{K}{n\|S_0^{-1}\|_\infty}\right)^{1+\delta/2} \sum_i \psi_{sij}^{2+\delta} \sum_{j<i}\psi_{sij}$$

$$\left(\frac{K}{n\|S_0^{-1}\|_\infty}\right)^{1+\delta/2} \left(\sum_i \sum_{j<i} \psi_{sij}^{2}\right)^{\delta/2} \leq \left(\frac{K}{n\|S_0^{-1}\|_\infty}\right)^{1+\delta/2} \left(\sup_i \sum_j \psi_{sij}^{2}\right)^{\delta/2} \sum_i \sum_j \psi_{sij}$$

$$\leq \left(\frac{K}{n\|S_0^{-1}\|_\infty}\right)^{\delta/2} \left(\sup_i \sum_j \psi_{sij}^{2}\right)^{\delta/2} = O\left(\frac{1}{n\|S_0^{-1}\|_\infty}\right)^{\delta/2} = o(1)$$  \hspace{1cm} (B.75)

under R1 and using (B.65).

Therefore, we can conclude that $\sum_i z_i \rightarrow_d N(0, 1)$, and thus

$$\frac{1}{(n\|S_0^{-1}\|_\infty)^{1/2}} \sum_i u_i = U \rightarrow_d N(0, V),$$  \hspace{1cm} (B.76)

where $V = \Sigma_{10} + \Sigma_{20}$ defined in (4.1) and (4.2). Finally, we have

$$\left(\frac{1}{n\|S_0^{-1}\|_\infty}\right)^{1/2} \begin{pmatrix} \frac{\partial L^p(\theta_{20})}{\partial \lambda} \\ \frac{\partial L^p(\theta_{20})}{\partial w_1} \\ \vdots \\ \frac{\partial L^p(\theta_{20})}{\partial w_k} \\ \frac{\partial L^p(\theta_{20})}{\partial u} \end{pmatrix} = FU + o_p(1) \rightarrow_d N(0, FVF),$$  \hspace{1cm} (B.77)

as required. ■

References


[28] Lu, Z.H. and M.L. King (2002), Improving the numerical technique for computing the accumulated distribution of a quadratic form in normal variables, Econometric Reviews, 21, 149-165.


