



HAL
open science

Interaction matrix selection in spatial econometrics with an application to growth theory

Nicolas Debarsy, Cem Ertur

► **To cite this version:**

Nicolas Debarsy, Cem Ertur. Interaction matrix selection in spatial econometrics with an application to growth theory. 2016. halshs-01278545v1

HAL Id: halshs-01278545

<https://shs.hal.science/halshs-01278545v1>

Preprint submitted on 24 Feb 2016 (v1), last revised 30 Nov 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interaction matrix selection in spatial econometrics with an application to growth theory

Nicolas Debarsy*[†]

LEO UMR 7322, F45067, Orléans, France

and

Cem Ertur

LEO UMR 7322, F45067, Orléans, France

February 24, 2016

Abstract

The interaction matrix, or spatial weight matrix, is the fundamental tool to model cross-sectional interdependence between observations in spatial econometric models. However, it is most of the time not derived from theory, as it should be ideally, but chosen on an ad hoc basis. In this paper, we propose a modified version of the J test to formally select the interaction matrix. Our methodology is based on the application of the robust against unknown heteroskedasticity GMM estimation method, developed by Lin & Lee (2010). We then implement the testing procedure developed by Hagemann (2012) to overcome the decision problem inherent to non-nested models tests. An application is presented for the Schumpeterian growth model with worldwide interactions (Ertur & Koch 2011) using three different types of interaction matrix: genetic distance, linguistic distance and bilateral trade flows and we find that the interaction matrix based on trade flows is the most adequate. Furthermore, we propose a network based innovative representation of spatial econometric results.

Keywords: Bootstrap; GMM; Interaction matrix; J tests ; Non-nested models; Heteroscedasticity; Spatial autoregressive models

JEL classification: C21- C12 - C52 - O4

*Research associate of the CNRS.

[†]The authors acknowledge the financial support of the Region Centre Val de Loire (APR-IA 2014 00094514)

1 Introduction

The interaction matrix, or spatial weight matrix, is the fundamental tool needed to model cross-sectional interdependence between observations in spatial econometric models. It allows to specify the structure of interactions or connectivity within a system through a spatial econometric model and is presumed to be known and exogenous (Anselin & Bera 1998). However, it is most of the time not derived from theory, as it should be ideally, but chosen on an ad hoc basis. In practice, an applied researcher has to choose a priori the interaction matrix, most of the time without any theoretical guidance. As a result, in empirical studies, interaction matrices are generally based on geographical distance, whose exogeneity is widely admitted. Nevertheless, other concepts of similarity or proximity can be useful to explain interactions. Indeed, by analogy to Akerlof (1997), countries or regions can be seen as localized in some general socio-economic, institutional or political space defined by a set of factors. Implementation of spatial econometric models thus requires the proper definition of the similarity or proximity measure to be used for building the interaction matrix.

Moreover, the consequences of a misspecification of the interaction matrix can be severe. For example, Lee (2008) shows that the misspecification of the interaction matrix can cause a significant bias in the estimates of the spatial autoregressive parameter in a SAR model. If the interaction matrix is under-specified, then both the two stage least squares (2SLS) and maximum likelihood (ML) estimator of the spatial autoregressive parameter are biased downward. Inversely, if the interaction matrix is over-specified, the spatial autoregressive parameter is overestimated. Furthermore, the under-specification of the interaction matrix is more problematic than its over-specification in terms of both bias and RMSE.

A special attention should therefore be given to the choice of the interaction matrix since it is given its specific characteristics that spatial econometric models are estimated and cross-sectional interdependence as well as spillovers are evaluated. One of the key points in the specification of spatial econometric models is then the choice, given a criterion, of the most relevant interaction matrix in a subset of all possible matrices.

As different types of interaction structures can be specified to study the same phenomenon, it can be difficult to identify the most appropriate interaction matrix. The literature on the tests of non-nested models has nevertheless made possible the handling of such situations. The literature devoted to the tests of non-nested assumptions finds its origins in many papers including the articles of Cox (1961, 1962) and Atkinson (1970) and in particular those of Davidson & MacKinnon (1981), Pesaran (1974), Pesaran & Deaton (1978), Mackinnon et al. (1983), Godfrey (1983), Pesaran & Weeks (2001). The tests of non-nested models are classified in two broad categories. One finds in the first category the test of Cox (1961, 1962), which was adapted later to linear regression models by Pesaran (1974) as well as non-linear regression models by Pesaran & Deaton (1978). The second category of tests is due to Davidson & MacKinnon (1981). The principle of their test, called the J test, is to combine competing models in an artificial nesting model.

In the spatial econometrics literature, one can find a first contribution in Anselin (1988). More recently, Kelejian (2008) extends the J test of Davidson & MacKinnon (1981) to a spatial context. The principle of the J test is to estimate a model which contains the null model and the predictive power of alternative models. Then the significance of additional terms is tested. Since this test does not use available information in an efficient way, Kelejian & Piras (2011) and Burrridge (2012) suggest different improvements. Kelejian & Piras (2011) modify the way available information is used in determining the predictions and consequently improve the power of the J test. Burrridge (2012) advocates the use of parameter estimates obtained from maximum likelihood estimation. Finally in an alternative approach, Jin & Lee (2013) extend the Cox test (Cox 1961, 1962) to a spatial framework. Moreover, in the closely related econometric network literature, an extension of the Kelejian's (2008) J test is proposed by Liu et al. (2014) to differentiate between the local-aggregate and the local-average endogenous peer effects in an econometric network model with network fixed effects.

We propose in this paper a modified version of the J test based on the Robust to unknown heteroskedasticity Generalized Method of Moments (RGMM) estimation method elaborated by Lin & Lee (2010).

It is well known in the literature that all the non-nested hypotheses tests suffer from a decision problem: in some circumstances, they may not help discriminating between the null and alternative models. To overcome this decision problem, Hagemann (2012) recently suggests a simple test for non-nested models, namely the *MJ* test, that we implement here in the framework of spatial econometric models.

We apply the preceding non-nested hypothesis tests to the generalized Schumpeterian growth model proposed by Ertur & Koch (2011). These authors propose a theoretical and methodological framework taking into account technological interactions between countries to explain the growth process in a Schumpeterian perspective. Worldwide interdependence, implied by international R&D spillovers, are therefore structurally introduced in a growth model with technological interactions, which yields a spatial econometric reduced form. This model includes both factor accumulation and innovation caused by R&D investments as determinants of the growth process, with technological diffusion occurring between interacting pairs of countries, human capital, reflecting the capacity for absorption of technology and physical capital playing the usual part.

In this paper, we consider three different interaction schemes to model the links between countries. The first one is based on genealogic/genetic distance. This concept of distance, defined by Spolaore & Wacziarg (2009, 2013), based on Cavalli-Sforza et al. (1994), corresponds to the time elapsed since two populations shared a common ancestor. In this setting, genetic distance is considered as a proxy for cultural, institutional and linguistic similarities. The results of Spolaore & Wacziarg (2009) suggest that this general measure of genealogical relatedness between populations can explain cross-country income differences. Their interpretation is that genetic distance captures the obstacles to the diffusion of development. The second interaction matrix we consider is based on linguistic proximity following Melitz & Toubal (2014). Melitz & Toubal (2014) summarize the evidence about the linguistic influences in an index, named Common Language Index (CL), resting strictly on exogenous linguistic factors. This summary index (CL) is constructed from different proxies: Common Official Language (COL), Common Native Language (CNL) and Language Proximity (LP). The third interaction matrix used is based on trade flows. Indeed, trade flows may

proxy multi-country technological interactions.

The use of the MJ test provides considerable help in the selection of the interaction matrix. Indeed, in the context of this application, we show that the trade flows based interaction matrix is the most appropriate pattern to model interactions between countries, among the three types of interaction patterns we consider. Furthermore, we propose a network based innovative representation of spatial econometric results. The obtained graphs significantly contribute to the interpretation of the estimated model.

The remainder of the paper is organized as follows: in section 2, we present the RGMM estimation approach robust to unknown heteroskedasticity elaborated by Lin & Lee (2010). Our modified version of the J test based on RGMM estimation as well as the predictor selection are then presented in section 3. The MJ test framework recently developed by Hagemann (2012) is then applied to our specific context of interaction matrix selection in section 4. Section 5 presents the application to the generalized Schumpeterian growth model while section 6 concludes.

2 GMM robust to heteroskedasticity

The model considered in this paper is the heteroskedastic SAR specification

$$Y_n = X_n\beta_0 + \lambda_0 W_n Y_n + \varepsilon_n. \quad (1)$$

where X_n is a $n \times k$ matrix of exogenous and nonstochastic explanatory variables, W_n is a non-stochastic interaction matrix whose diagonal elements are zero, λ_0 is the true parameter measuring interactions' intensity while β_0 is the true $k \times 1$ vector of unknown parameters associated to the explanatory variables. Finally, ε_n is the vector of independent error terms with ε_{in} characterized by a zero mean and a variance σ_{in}^2 , $i = 1, \dots, n$. We also note $\theta_0 = [\lambda_0, \beta_0']'$ and $\theta = [\lambda, \beta']'$ for any value of λ and β .¹ Additionally, $S_n = (I_n - \lambda_0 W_n)$, $S_n(\lambda) = (I_n - \lambda W_n)$ and we de-

¹In this paper, we note λ_0 and β_0 as the true value of the parameters, $\hat{\lambda}$ and $\hat{\beta}$ as estimators of λ_0 and β_0 and λ and β as parameters on which we maximize the objective function. Hence, only $\hat{\lambda}$ and $\hat{\beta}$ are viewed as random variables.

fine $\varepsilon_n(\theta) = S_n(\lambda)Y_n - X_n\beta$ for any value of θ . Finally, for any square matrix A_n , we define $A_n^s = A_n + A_n'$. This last notation is used in the expressions for identification and asymptotic distribution.

Lin & Lee (2010) show that the (quasi-) maximum likelihood estimator of model (1) is generally inconsistent in presence of unknown heteroskedasticity. Alternative estimation methods are thus required. The Two-Stage Least Squares (2SLS) approach (Kelejian & Prucha 1998), the Best 2SLS (Lee 2003) and the series Best 2SLS (Kelejian et al. 2004) could be considered, but according to Kelejian & Prucha (2010, p.60), the asymptotic normality is not guaranteed under unknown heteroskedasticity. Kelejian & Prucha (2010) generalize the method introduced in Kelejian & Prucha (1999) to estimate a SAR model with SAR disturbances in the presence of unknown heteroskedasticity. They combine a generalized method of moments estimator (GMME) for the coefficient of spatial autocorrelation in the error term to a 2SLSE for the coefficient λ_0 associated to the spatial lag.

In this paper, we apply the approach developed by Lin & Lee (2010) to estimate (1).² The advantage of GMM procedure over 2SLS is that the former remains consistent even if none of the explanatory variable is relevant (Lee 2007). Besides, 2SLS, as will be shown later on, only uses the information contained in the deterministic part of the reduced form of (1), leaving the information in the stochastic part of the reduced form unexploited. Hence, the GMME will be more efficient than the 2SLSE. Lin & Lee (2010) generalize Lee (2007) and Liu et al. (2010) where the focus is on homoskedastic error terms. With that respect, the main differences in the estimation procedure due to relaxing the homoskedasticity assumption come from the hypothesis on quadratic moments and the absence of a Best Optimal GMME (BOGMME).

The objective of this section is to present the GMM estimation in the presence of unknown heteroskedasticity. The set of assumptions presented below corresponds to those developed in Lin & Lee (2010) and are reproduced for the sake of readability of the paper.

²The estimation procedure developed in Kelejian & Prucha (2010) could also have been used since they use the same moment conditions as Lin & Lee (2010).

Assumption 1 *The ε_{in} 's are independent $(0, \sigma_{in}^2)$ with finite moments larger than the fourth order such that $\mathbf{E}|\varepsilon_{in}|^{4+\eta}$ for some $\eta > 0$ are uniformly bounded for all n and i .*

As orthogonality conditions require the computation of mean and variances of quadratic forms, one needs to impose the uniform boundedness of moments of order higher than 4.

Assumption 2 *The elements of the $n \times k$ regressor matrix X_n are uniformly bounded constants, X_n has a full rank k and $\lim_{n \rightarrow \infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular.*

Assumption 3 *The interaction matrices W_n and the matrix S_n^{-1} are uniformly bounded in absolute value in both row and column sums.*

Assumption 3 means that the inverse of $S_n = I_n - \lambda_0 W_n$ should be uniformly bounded in absolute value (in addition to W_n). It is important to note that the uniform boundedness is only imposed at the true parameter value and not for any value of λ . Assuming the existence of S_n^{-1} , the reduced form of model (1) is as follows

$$Y_n = S_n^{-1} X_n \beta + S_n^{-1} \varepsilon_n, \quad (2)$$

where $S_n^{-1} = I_n + \lambda_0 W_n + \lambda_0^2 W_n^2 + \dots$.

From (2), we can express $W_n Y_n$ as follows:

$$W_n Y_n = G_n X_n \beta_0 + G_n \varepsilon_n, \quad (3)$$

with $G_n = W_n S_n^{-1}$.

2.1 Orthogonality conditions

In general, $W_n Y_n$ is correlated with ε_n since $\mathbf{E}((W_n Y_n)' \varepsilon_n) = \text{tr}(\Sigma_n G_n)$ where Σ_n is a diagonal matrix containing the residual variance $\sigma_{in}^2 = \mathbf{E}(\varepsilon_{in}^2)$ for $i = 1, \dots, n$. The GMM approach developed by Lin & Lee (2010) combines linear moments (identical to 2SLS method) and quadratic moments. Let Q_n be a $n \times k^*$ matrix of IV's, with $k^* \geq k + 1$, which contains the information of the

deterministic part $\mathbf{E}(W_n Y_n | X_n)$ of $W_n Y_n$ in (3). This matrix is constructed from X_n and functions of W_n and can be used to instrument $W_n Y_n$. For instance, $Q_n = [X_n, W_n X_n, W_n^2 X_n, \dots]_{LI}$, where LI stands for linearly independent columns.

Moreover, in many economic models, some explanatory variables are suspected to be endogenous. In that case, the SAR model to be estimated is as follows:

$$Y_n = \lambda_0 W_n Y_n + X_{1n} \beta + Y_{2n} \gamma + \varepsilon_n, \quad (4)$$

with X_{1n} , a matrix of exogenous covariates of dimension $n \times k_1$ and Y_{2n} , a matrix of dimension $n \times k_2$ of endogenous variables ($\mathbf{E}(Y_{2n}' \varepsilon_n) \neq 0$). Let X_{2n} be a matrix of excluded exogenous variables of dimension $n \times k_x$, with $k_x \geq k_2$. Then a matrix of linear moments that can be used to consistently estimate λ_0 , β_0 and γ_0 will be of the form: $Q_n = [X_{1n}, W_n X_{1n}, W_n^2 X_{1n}, \dots, X_{2n}]_{LI}$, with $\text{rank}(Q_n) \geq k_1 + k_2 + 1$.³ However, in this paper we will assume that all explanatory variables are exogenous and the model to be estimated is (1) with a reduced form presented in (2).

Lin & Lee (2010) consider additionally quadratic moments $(P_{jn} \varepsilon_n(\theta))' \varepsilon_n(\theta)$ that capture the information contained in the stochastic part of (3). These quadratic moments serve as IV's for $G_n \varepsilon_n$, the non-deterministic part of $W_n Y_n$. One needs to impose some conditions on P_{jn} matrices so that quadratic moments can act as orthogonality conditions ($\mathbf{E}[\varepsilon_n' P_{jn}' \varepsilon_n] = 0$), i.e. $P_{jn} \varepsilon_n$ is uncorrelated with ε_n . Let \mathcal{P}_{1n} be the class of constant $n \times n$ matrices with zero trace. Consider a subclass \mathcal{P}_{2n} of \mathcal{P}_{1n} that contains constant $n \times n$ matrices with zero diagonal. Depending on the assumptions made on the error term, matrices P_{jn} may belong to \mathcal{P}_{1n} or \mathcal{P}_{2n} .

Let us consider heteroskedastic error terms with, as variance-covariance matrix, the diagonal matrix Σ_n , made of σ_{in}^2 , the individual variances. The orthogonality condition associated to a given matrix P_{jn} is $\mathbf{E}(\varepsilon_n' P_{jn}' \varepsilon_n) = \text{tr}(\Sigma_n P_{jn})$. The only way for this trace to be 0 is that $P_{jn} \in \mathcal{P}_{2n}$. The intuition is that when P_{jn} has zero diagonal, the l^{th} element of $P_{jn} \varepsilon_n = \sum_{k=1}^n p_{lk,jn} \varepsilon_{kn}$ excludes the corresponding element of ε_n (ε_{ln}) from the sum since $p_{ll,jn} = 0$. So, each component of $P_{jn} \varepsilon_n$

³Liu & Saraiva (2015) propose a GMM approach for endogenous explanatory variables where they consider quadratic moments based on the correlation structure of the spatially lagged dependent variable with the model disturbance term and with the endogenous regressors in addition to linear moments to gain in efficiency.

is uncorrelated with the corresponding component of ε_n . If we impose homoskedastic error terms, $\mathbf{E}(\varepsilon_n' P_{jn}' \varepsilon_n) = \sigma_0^2 \text{tr}(P_{jn})$. Hence, P_{jn} is only required to have a zero trace for this expression to be 0 (Lee 2001, 2007).

However, no matter the hypothesis on the variance of the error term, we need $P_{jn}\varepsilon_n$ to be correlated with $W_n(I_n - \lambda_0 W_n)^{-1}$ for $P_{jn}\varepsilon_n$ to act as an instrument for $G_n\varepsilon_n$ and the closest P_{jn} is of $W_n(I_n - \lambda_0 W_n)^{-1}$, the better the instrument.

Quadratic moments are used for 2 reasons. As they capture the information contained in the stochastic part of (3), they make the obtained estimator more efficient compared to 2SLS or B2SLS estimators. Second, if all X_n are irrelevant, the set of linear moments will not be useful to instrument $W_n Y_n$. However, in that case, which corresponds to the pure SAR model, i.e. $Y_n = \lambda W_n Y_n + \varepsilon_n$, quadratic moments can still act as instruments for $W_n Y_n$.

Assumption 4 *The matrices P_{jn} 's whose diagonal elements are 0 ($P_{jn} \in \mathcal{P}_{2n}$) are uniformly bounded in both row and column sums, and elements of Q_n are uniformly bounded.*

Under heteroskedasticity, the set of moment functions, assuming m quadratic moments, is:

$$\begin{aligned} g_n(\theta) &= (P_{1n}\varepsilon_n(\theta), \dots, P_{mn}\varepsilon_n(\theta), Q_n)' \varepsilon_n(\theta) \\ &= (\varepsilon_n(\theta)' P_{1n}\varepsilon_n(\theta), \dots, \varepsilon_n(\theta)' P_{mn}\varepsilon_n(\theta), \varepsilon_n(\theta)' Q_n)' \end{aligned} \tag{5}$$

2.2 Asymptotic properties

The following assumption ensures the identification of θ_0 from the orthogonality conditions $E(g_n(\theta_0)) = 0$, for a sufficiently large n .

Assumption 5 *Either:*

- a) $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n' (G_n X_n \beta_0, X_n)$ has the full rank $k + 1$, or
- b) $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n' X_n$ has the full rank k , $\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(\Sigma_n G_n^s P_{jn}) \neq 0$ for some j and $\lim_{n \rightarrow \infty} \frac{1}{n} (\text{tr}(\Sigma_n G_n' P_{1n} G_n), \dots, \text{tr}(\Sigma_n G_n' P_{mn} G_n))'$ are linearly independent.

The next assumption focuses on the parameter space for θ .

Assumption 6 θ_0 is in the interior of the parameter space Θ , which is a bounded subset of \mathcal{R}^{k+1} .

By contrast to maximum likelihood estimation, λ is not required to belong to some compact parameter space.⁴ The constraints imposed on the parameter space ensure the positiveness of the Jacobian of the transformation $|I_n - \lambda W_n|$. However, in the GMM framework, such constraints are not relevant because the objective function is a polynomial function (of order 4) in θ . The condition on the boundedness of the parameter space is solely for the theoretical purpose of proving consistency of the GMME (Lin & Lee 2010). Hence, the GMM is a unconstrained estimation procedure.

Instead of minimizing the objective function $g_n(\theta)'g_n(\theta)$ with respect to θ , Lin & Lee (2010) propose to minimize a linear combination $a_n g_n(\theta)$ of the moment functions $g_n(\theta)$, so that the objective function enters the Hansen (1982) setting where each condition is weighted by its precision. The matrix a_n has a full rank greater than or equal to the dimension of θ . The objective function is thus $g_n(\theta)'a_n'a_n g_n(\theta)$ where the matrix $a_n'a_n$ is non-negative and contains the weight of each orthogonality condition and motivates the issue of optimal weighting.

To express the asymptotic distribution of the Robust (to the presence of unknown heteroskedasticity) GMME, we need the expression of the gradient of $g_n(\theta_0)$ as well as the variance matrix of the orthogonality conditions (also evaluated at the true parameters values):

$$D_n = \frac{\partial \mathbf{E}(g_n(\theta_0))}{\partial(\theta)} = - \begin{pmatrix} \text{tr}(\Sigma_n P_{1n}^s G_n) & 0 \\ \vdots & \vdots \\ \text{tr}(\Sigma_n P_{mn}^s G_n) & 0 \\ Q_n' G_n X_n \beta_0 & Q_n' X_n \end{pmatrix} \quad (6)$$

⁴For more information about the parameter space for λ , the interested reader may consult Kelejian & Prucha (2010).

$$\Omega_n = Var(g_n(\theta_0)) = \begin{pmatrix} \text{tr}(\Sigma_n P_{n,1}(\Sigma_n P_{1n})^s) & \text{tr}(\Sigma_n P_{1n}(\Sigma_n P_{2n})^s) & \cdots & 0 \\ \text{tr}(\Sigma_n P_{n,2}(\Sigma_n P_{1n})^s) & \text{tr}(\Sigma_n P_{2n}(\Sigma_n P_{2n})^s) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & Q_n' \Sigma_n Q_n \end{pmatrix} \quad (7)$$

Lin & Lee (2010) show that under Assumptions 1-6, and that $\lim_{n \rightarrow \infty} \frac{1}{n} a_n D_n$ exists and has the full rank $k + 1$, the Robust GMME $\hat{\theta}$ is a consistent estimator of θ_0 and $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, \Gamma)$, where

$$\Gamma = \lim_{n \rightarrow \infty} \frac{1}{n} (D_n' a_n' a_n D_n)^{-1} D_n' a_n' a_n \Omega_n a_n' a_n D_n (D_n' a_n' a_n D_n)^{-1} \quad (8)$$

2.3 Optimal RGMME

Even though the choice of the weighting matrix $a_n' a_n$ does not affect the consistency of RGMME, it plays a role in its efficiency. By setting $a_n' a_n$ equal to the inverse of the covariance matrix of orthogonality conditions (Ω_n), one obtains the optimal RGMME, i.e. the most efficient estimator given $g_n(\theta)$. In their proposition 3, Lin & Lee (2010) show that if one has a consistent estimator $\hat{\Omega}_n$, the asymptotic distribution of the feasible ‘‘optimal’’ RGMME obtained from minimizing $g_n(\theta)' \hat{\Omega}_n g_n(\theta)$ with respect to θ is as follows:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N \left(0, \left(\lim_{n \rightarrow \infty} \frac{1}{n} D_n' \hat{\Omega}_n^{-1} D_n \right)^{-1} \right) \quad (9)$$

The optimal estimator is the most efficient estimator that can be obtained given a set of orthogonality conditions. However, we could question the existence of the best choice of orthogonality conditions, i.e. the set of conditions which guarantees that the best optimal RGMME is the most efficient one. Unfortunately, in the unknown heteroskedasticity case, no such set of orthogonality conditions exists, since they would involve the unknown Σ_n . However, Lin & Lee (2010) suggest to use as orthogonality conditions those corresponding to the best conditions in the *iid* case when $P_{jn} \in \mathcal{P}_{n,2}$. These conditions, shown in Lee (2007), are $P_n = (G_n - \text{Diag}(G_n))$ and $Q_n = [G_n X_n \beta_0, X_n]$, where $\text{Diag}(A_n)$ refers to a diagonal matrix where diagonal elements corre-

respond to the diagonal of A_n .⁵ Hence, they suggest to use only one quadratic and $k + 1$ linear conditions. The vector of orthogonality conditions becomes:

$$g_n(\theta) = (\varepsilon_n'(\theta)P_n\varepsilon_n(\theta), \varepsilon_n'Q_n)' \quad (10)$$

Expressions for the gradient and the covariance matrix of $g_n(\theta)$ defined in (10), evaluated at the true parameter value are:

$$D_n = \frac{\partial \mathbf{E}(g_n(\theta_0))}{\partial(\theta)} = - \begin{pmatrix} \text{tr}(\Sigma_n P_{n,1}^s G_n) & 0 \\ Q_n' G_n X_n \beta_0 & Q_n' X_n \end{pmatrix} \quad (11)$$

$$\Omega_n = \text{Var}(g_n(\theta_0)) = \begin{pmatrix} \text{tr}(\Sigma_n P_{n,1}(\Sigma_n P_{n,1})^s) & 0 \\ 0 & Q_n' \Sigma_n Q_n \end{pmatrix} \quad (12)$$

As these orthogonality conditions depend on the unknown parameters λ_0 and β_0 , we need to proceed in 2 steps to get the optimal RGMME. The first stage consists in getting consistent estimators of λ_0 and β_0 . To do so, we compute RGMME based on orthogonality conditions not involving any unknown parameters but which satisfy the above assumptions. For instance, we set $P_n = W_n$ and $Q_n = [X_n, W_n X_n, W_n^2 X_n]$. Also we use as initial guess of the weight matrix $a_n' a_n$ the identity matrix. The obtained (first-step) estimators, $\hat{\lambda}^{fs}$ and $\hat{\beta}^{fs}$, are then used to compute the residuals which allow to have an estimator $\hat{\Sigma}_n$, whose i^{th} diagonal element is $\hat{\varepsilon}_{in}^2(\hat{\theta})$. We are now able to compute an estimator $\hat{\Omega}_n$ of Ω_n . Besides, we can construct the new orthogonality conditions $g_n^{st} = (\varepsilon_n'(\theta)P_n^{st}\varepsilon_n(\theta), \varepsilon_n'(\theta)Q_n^{st})'$ based on $P_n^{st} = (\hat{G}_n - \text{Diag}(\hat{G}_n))$ and $Q_n^{st} = [\hat{G}_n X_n \hat{\beta}^{fs}, X_n]$, where $\hat{G}_n = W_n(I_n - \hat{\lambda}^{fs} W_n)^{-1}$.

The second steps consists in minimizing the objective function shown in (13) with respect to θ :

$$g_n^{st}(\theta)' \hat{\Omega}_n^{-1} g_n^{st}(\theta) \quad (13)$$

⁵If quadratic moments P_n belong to \mathcal{P}_{n1} , Liu et al. (2010) derive the best moment conditions under homoskedasticity, no matter the distribution of the error terms.

Statistical inference will be based on this second step estimator $\hat{\theta}$, where its estimated covariance matrix will be $(D'_n \hat{\Omega}_n^{-1} D_n)^{-1}$.

3 The modified J test based on RGMM

For ease of exposition, let us consider that under the null hypothesis H_0 and the $M - 1$ alternative hypotheses $H_{1,m}, m = 2, \dots, M$, we have heteroskedastic SAR models which only differ by the specification of the interaction matrix.

$$H_0 : Y_n = X_n \beta_1 + \lambda_1 W_{1n} Y_n + \varepsilon_{1n}, \quad (14)$$

$$H_{1,m} : Y_n = X_n \beta_m + \lambda_m W_{mn} Y_n + \varepsilon_{mn}, \quad (15)$$

The general motivation of the J test is to estimate a model which contains the null model and the predictive power of alternative models in order to check if predictors from alternative models can significantly contribute to the explanatory power in the null model. For the homoskedastic case, different predictors have been derived by Kelejian & Prucha (2007) which are presented here for the sake of comparison with the heteroskedastic case. The first predictor $\hat{y}_{in|1}^m$ is obtained from the reduced form of the alternative model m , which is

$$\begin{aligned} \hat{y}_{in|1}^m &= \mathbf{E}(y_{in} | \Lambda_1) \\ &= (I_n - \lambda_m W_{mn})_i^{-1} X_n \beta_m \end{aligned} \quad (16)$$

where A_i means the i^{th} row of matrix A . This predictor is based on the smallest information set, namely $\Lambda_1 = \{X, W_{mn}\}$. The second predictor used is based on a larger information set, namely $\Lambda_2 = \{X, W_{mn}, W_{m,i} Y_n\}$. It is obtained from the structural form of the alternative model m and

is computed as follows:

$$\begin{aligned}\hat{y}_{in|2}^m &= \mathbf{E}(y_{in}|\Lambda_2) \\ &= \lambda_m W_{mn,i} Y_n + X_{i.,n} \beta_m + \frac{\text{cov}(\varepsilon_{in}, W_{mn,i} Y_n)}{\text{var}(W_{mn,i} Y_n)} [W_{mn,i} Y_n - \mathbf{E}(W_{mn,i} Y_n)]\end{aligned}\quad (17)$$

where $W_{mn,i}$ is the i^{th} row of W_{mn} . Also, the last term takes into account the correlation between the error term and the spatial lag of the dependent variable. The different elements involved in this correlation are computed as follows:

$$\begin{aligned}\mathbf{E}(W_{mn,i} Y_n) &= W_{mn,i} (I - \lambda_m W_{mn})^{-1} X_n \beta_m \\ \text{var}(W_{mn,i} Y_n) &= \sigma_n^2 W_{mn,i} \Sigma_m^y W'_{mn,i} \\ \Sigma_m^y &= (I - \lambda_m W_{mn})^{-1} (I - \lambda_m W'_{mn})^{-1} \\ \text{cov}(\varepsilon_{in}, W_{mn,i} Y_n) &= \sigma_n^2 (I - \lambda_m W'_{mn})^{-1} X_n \beta_m\end{aligned}\quad (18)$$

where σ_n^2 is the variance of the error term in the homoskedastic case. Since the second information set is larger than the first one, the second predictor is more efficient (Kelejian & Prucha 2007). A third predictor is also sometimes considered due to its intuitive appeal. It is also based on the information set Λ_2 but assumes a zero correlation between the error term and the spatial lag of the dependent variable:

$$\hat{y}_{in|3}^m = \lambda_m W_{mn,i} Y_n + X_{i.,n} \beta_m.$$

Kelejian & Prucha (2007) show that this predictor is biased since the spatial lag is assumed independent of the error term. We will thus do not consider this predictor.⁶

Let us now study the predictors in the heteroskedastic case. Theoretically, we could use the first two predictors above. The first one, $\hat{y}_{in|1}^m$, based on the information set Λ_1 is not affected by the presence of heteroskedasticity and takes the same form as for homoskedastic errors (equation 16). However, the second predictor, $\hat{y}_{in|2}^m$, based on the information set Λ_2 takes on the expression

⁶Finally, Kelejian & Prucha (2007) propose the full information predictor but is not more efficient than the second in the SAR specification.

in (19).

$$\begin{aligned}
\mathbf{E}(W_{m,i}.Y_n) &= W_{mn,i}.(I - \lambda_m W_{mn})^{-1} X_n \beta_m \\
\text{var}(W_{mn,i}.Y_n) &= \sigma_{i,mn}^2 W_{m,i}.\Sigma_m^y W'_{mn,i}. \\
\Sigma_m^y &= (I - \lambda_m W_{mn})^{-1} (I - \lambda_m W'_{mn})^{-1} \\
\text{cov}(\varepsilon_{in}, W_{mn,i}.Y_n) &= \sigma_{i,mn}^2 (I - \lambda_m W'_{mn})^{-1} X_n \beta_m
\end{aligned} \tag{19}$$

where $\sigma_{i,mn}^2$ is the variance of the error term of individual i in model m . The drawback of this predictor is that it involves the terms $\sigma_{i,mn}^2$ which are assessed by computing $\hat{\varepsilon}_{i,mn}^2$ but cannot be consistently estimated.

Due to this drawback, we propose to rely on the first predictor, based on the reduced form of the alternative model, to construct the J test. Theoretically, $\hat{y}_{in|1}^m$ is less efficient than $\hat{y}_{in|2}^m$ since Λ_1 is a smaller information set than Λ_2 . In Table 2, we report the results of Monte Carlo simulations done to assess the bias and Mean Square Error (MSE) of the first and second predictors under different scenario. We consider the following DGP:

$$Y_n = \lambda W_n Y_n + \iota_n + X_{1n} \beta_1 + X_{2n} \beta_2 + \varepsilon_n. \tag{20}$$

where ι_n is the unit vector, $X_{1n} \sim U(0, 10)$, $X_{2n} \sim N(0, 1)$ and the associated vector of parameters β_1 and β_2 are both set equal to 1. In these simulations, the regressors are kept constants across replications. The error term is assumed heteroskedastic and non-normally distributed. More precisely, it is designed in the following way:

$$\varepsilon_{in} = \bar{\sigma} \left(\frac{\nu_{in} - 2}{\sqrt{2}} \right) X_{1n}, \tag{21}$$

where ν_{in} follows a $\Gamma(2, 1)$ distribution. Besides, to make simulation results comparable across different values of parameters, we set the average dispersion of the error term ($\bar{\sigma}$) so that the signal to noise ratio is equal to 0.5. This signal to noise ratio is computed as the variance of $(X_{1n} \beta_1 + X_{2n} \beta_2)$ over the sum of the variances of $(X_{1n} \beta_1 + X_{2n} \beta_2)$ and ε_n (see Lee 2007, Liu et al. 2010). Ignoring spatial autocorrelation, this ratio would correspond to a goodness of fit of 0.5 in

a standard regression. The spatial autoregressive parameter λ varies from 0 to 0.8 by increment of 0.2 to capture low, moderate and high spatial autocorrelation. Two sample sizes, $n = 100$ and $n = 254$, are considered, corresponding respectively to the number of counties in North Carolina and in Texas. We also used different interaction matrices W_n to cover sparse and dense interaction schemes. The definitions of W_n considered are given in Table 1. Let us note that the W_n are based on real locations (the counties). Finally, for each case, we perform 1000 replications. Analyzing

Table 1: Definition of the interaction matrices

Abbreviation	Description	expression
Inv. Exp. Sqr. Dist.	Inverse exponential square distance	$w_{ij} = e^{-2d_{ij}}$
Inv. Dist.	Inverse distance	$w_{ij} = 1/d_{ij}$
Binary 10 nn	Binary 10 nn	$w_{ij} = \begin{cases} 1 & \text{if } j \text{ is in the } 10 \text{ } nn \text{ of } i \\ 0 & \text{otherwise} \end{cases}$
Inv. Dist. 10 nn	10 nn . with weights equal to Inverse distance	$w_{ij} = \begin{cases} 1/d_{ij} & \text{if } j \text{ is in the } 10 \text{ } nn \text{ of } i \\ 0 & \text{otherwise} \end{cases}$

Note: The distance between 2 observations is computed as the geodesic distance and nn stands for nearest neighbors.

Table 2, we indeed observe higher MSE for the first predictor but the magnitude is very similar to those associated to the second predictor, except in some situations where $\lambda = 0.8$. We also observe that the bias of both predictors are very similar and small. Finally, when sample size increases, we observe that MSE of both predictors decrease and that even for high spatial autocorrelation, no difference between MSE of the first and second predictor appears.

Let us denote $\hat{Y}_{n|1}^M = [\hat{y}_{n|1}^2, \dots, \hat{y}_{n|1}^M]$ the $n \times (M - 1)$ matrix containing the $M - 1$ predicted alternative models using first predictor. we then augment the model under the null (model 14) with the predictors $\hat{Y}_{n|1}^M$:

$$Y_n = X_{1n}\beta_1 + \lambda_1 W_{1n} Y_n + \hat{Y}_{n|1}^M \delta + \varepsilon_{1n}, \quad (22)$$

where δ is a $(M - 1) \times 1$ vector of parameters: $\delta' = [\delta_2, \dots, \delta_M]$. If the model under the null is the correct model, then: $\delta' = (\delta_2, \dots, \delta_M) = 0$. This test is the J statistic.

Our J -test procedure based on the RGMM estimation procedure is summarized in the following steps:

Table 2: Bias and MSE Comparison of the 2 predictors

λ	n=100				n=254			
	First predictor		Second predictor		First predictor		Second predictor	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
W_n : Inv. Dist.								
0	0.001	2.696	-0.000	2.677	0.000	1.302	0.000	1.302
0.2	0.000	2.784	-0.000	2.756	0.000	1.300	0.000	1.296
0.4	0.000	2.849	-0.000	2.760	0.000	1.336	0.000	1.321
0.6	0.000	2.906	0.000	2.669	-0.000	1.354	0.000	1.312
0.8	0.001	3.620	0.000	2.743	-0.000	1.615	-0.000	1.45
W_n : binary 10 nn.								
0	0.001	2.725	0.001	2.660	0.000	1.280	0.000	1.274
0.2	0.001	2.902	0.001	2.694	0.000	1.297	0.000	1.245
0.4	0.005	3.461	0.001	2.570	0.001	1.377	-0.000	1.192
0.6	0.001	3.302	0.001	2.365	0.002	1.555	-0.001	1.108
0.8	0.000	5.668	0.001	2.275	-0.003	3.163	-0.002	1.082
W_n : 10 nn. based on Inv. Dist.								
0	0.002	2.695	0.003	2.642	0.000	1.278	0.000	1.274
0.2	0.002	2.892	-0.001	2.771	0.000	1.281	0.001	1.241
0.4	0.003	3.006	-0.003	2.616	0.000	1.338	0.000	1.199
0.6	0.004	3.364	-0.003	2.420	0.000	1.420	0.000	1.116
0.8	0.007	4.854	-0.002	2.279	0.001	1.706	-0.001	1.061
W_n : Inv. Sqr. Dist.								
0	-0.000	2.834	0.003	2.778	0.000	1.280	0.000	1.278
0.2	0.003	2.834	-0.001	2.718	0.000	1.281	0.000	1.262
0.4	0.002	2.762	-0.001	2.596	0.000	1.300	0.000	1.232
0.6	0.001	2.897	-0.001	2.533	-0.000	1.339	0.000	1.193
0.8	.0010	3.330	-0.001	2.433	-0.001	1.567	0.000	1.223

Step 1: Estimate the parameters β_m and λ_m for $m = 2, \dots, M$ by the two-step optimal RGMM approach described in section 2.

Step 2: Estimate $\hat{y}_{n|1}^m$ for $m = 2, \dots, M$ and form $\hat{Y}_{n|1}^M = [\hat{y}_{n|1}^2, \dots, \hat{y}_{n|1}^M]$.

Step 3: Estimate the augmented model $Y_n = X_n\beta_1 + \lambda_1 W_{1n} Y_n + \hat{Y}_{n|1}^M \delta + \varepsilon_{1n}$ by the two-step optimal RGMM approach described in section 2.

Note that the predictors do not add any difficulties to the model and are simply included in the set of exogenous variables : $\tilde{X} = [X, \hat{Y}_{n|1}^M]$ (see Davidson & MacKinnon 1981, p.785).

Step 4: Construct the J-test

Let us define $\gamma_1 = (\beta_1', \lambda_1, \delta')'$. Let $R = (0_{(M-1) \times (k_1+1)}, I_{(M-1)})$ so that $R\gamma_1 = 0$. Then, the J statistic defined as a Wald statistic is:

$$J_{1,n} = (R\hat{\gamma}_1)' \left[R(\hat{D}'_n(\hat{\Omega}_n^{-1})\hat{D}_n)^{-1}R' \right]^{-1} (R\hat{\gamma}_1) \sim \chi^2(M-1). \quad (23)$$

where \hat{D}'_n and $\hat{\Omega}_n^{-1}$ are respectively the gradient and the inverse of the covariance matrix of the augmented two-step RGMM estimation procedure. In this paper, we follow Davidson & Mackinnon (1985) who suggest to use the variance-covariance of the augmented specification.⁷

3.1 Statistical inference based on bootstrap

Davidson & MacKinnon (2002) have shown that statistical inference based on asymptotic theory causes size distortion of the J test while a bootstrap approach performs better in small sample sizes. Burridge & Fingleton (2010) implement a semiparametric bootstrap approach to the J test applied to the choice of the interaction matrix in spatial econometrics.⁸ They show that statistical inference based on bootstrap outperforms statistical inference based on asymptotic distribution for the J statistic. Jin & Lee (2015) formally prove the consistency of the bootstrap approach in spatial econometrics for linear-quadratic forms of residuals test-statistics, such as Moran's I or J test. In this paper, to account for heteroskedasticity, we follow Hagemann (2012) and implement the wild bootstrap of Liu (1988) and Mammen (1992). It consists in perturbing the residuals with *iid* copies $\eta_1, \eta_2, \dots, \eta_n$ of a random variable η with $\mathbf{E}(\eta|y, \mathcal{F}) = 0$, $\mathbf{E}(\eta^2|y, \mathcal{F}) = 1$ and $\mathbf{E}(|\eta|^{2+\delta}|y, \mathcal{F}) < \infty$ for some $\delta > 0$. Hagemann (2012) indicates that the recommended choice for η in practice is a Rademacher variable that takes on the value 1 with probability 0.5 and -1 with probability 0.5. Finally, let H be a diagonal matrix having as diagonal elements the terms $\eta_1, \eta_2, \dots, \eta_n$.

⁷Alternatively, Davidson & Mackinnon (1985) also suggest to use the estimated covariance matrix of the first step of the J test procedure. However, the Monte Carlo simulations we performed indicate poor performance of this alternative method.

⁸Burridge & Fingleton (2010) consider a homoskedastic error term, a first step estimation procedure based on maximum likelihood and an augmented model estimated by IV. We differ from that paper by allowing unknown heteroskedasticity in the error term and estimating all models by Robust GMM.

Consider the null hypothesis to be tested is that model (14) is the true model.

- i*) Compute the J test associated to this model using the methodology presented above and label it $J_{1,n}$.
- ii*) Estimate model (14) to get $\hat{\beta}_1$, $\hat{\lambda}_1$ and the residuals $\hat{\varepsilon}_{1n}$
- iii*) Generate $\varepsilon_{1n}^* = H\hat{\varepsilon}_{1n}$
- iv*) Generate the bootstrapped sample $Y_n^* = (I_n - \hat{\lambda}_1 W_{1n})^{-1}[X_n \hat{\beta}_1 + \varepsilon_{1n}^*]$
- v*) Compute the bootstrapped J test for the model ?? using the bootstrapped dependent variable Y_n^* and label it $J_{1,n}^*$.

Repeat operations *ii*) to *v*) B times, each with a new realization of H . Reject the null hypothesis at the α threshold if $J_{1,n}$ is larger than $c_{1-\alpha}^*$, the $1 - \alpha$ empirical quantile of the distribution of $J_{1,n}^*$.

Finally, Figure 1 compares the value of the J test obtained using the 2 predictors defined earlier, namely the one based on the reduced form of the alternative specification (first predictor) and the one based on the structural form of the alternative (second predictor). In all panels of Figure 1, we report the value of the J test associated to the true model (i.e. the model under the null is the DGP). The interaction matrix used to generate the true model is a exponential inverse square distance matrix.

The first predictor is drawn in blue while the second one is represented in orange. We directly observe the much higher volatility and non reliability of the second predictor, compared to the first one. Besides, we note values of the J test using the second predictor that can peak up to 200, on this set of 100 replications.

This Figure comforts ourselves in the decision of using only the first predictor to construct the J test when errors are allowed to be heteroskedastic.

4 The MJ -test

The J test presented above is a procedure for testing a null model against non-nested alternatives. It is generally known that tests of non-nested models may cause to a problem of model selection when inverting the null and alternative hypotheses do not give clear-cut results. To overcome this drawback, Hagemann (2012) suggests a simple test, the MJ (minimum J) statistic which allows to always select a model between competing alternatives. The MJ statistic consists in selecting the model with the smallest associated J statistic value. In this paper, we extend the use of the MJ test procedure to a spatial framework and more precisely to the selection of the interaction matrix in spatial autoregressive models.

Assume that one wishes to find the most appropriate interaction matrix among a set of M spatial autoregressive models.

$$Y_n = X_n \beta_m + \lambda_m W_{mn} Y_n + \varepsilon_{mn}, \quad m \in \mathcal{M} = \{1, \dots, M\}, \quad (24)$$

with ε_{mn} , the $n \times 1$ vector of independent error terms with $\varepsilon_{i,mn}$ characterized by a zero mean and a variance $\sigma_{i,mn}^2$, $i = 1, \dots, n$ and $\theta_m = (\lambda_m, \beta_m')$ is the vector of regression parameters to be estimated for $m = 1, \dots, M$. The MJ test procedure is the following:

Step 1: For each $m \in \mathcal{M}$, compute the J test of the null hypothesis H_0 corresponding to model m against $M - 1$ alternatives non-nested hypotheses.

For this, it is necessary to estimate the augmented model:

$$Y_n = X_{mn} \beta_m + \lambda_m W_{mn} Y_n + \sum_{l \in \mathcal{M} \setminus \{m\}} \delta_l \hat{y}_n^l + \varepsilon_{mn}. \quad (25)$$

Construct then $J_{m,n}$, the J -test associated to model m , as a test of joint significance of $\delta_l, l \in \mathcal{M} \setminus \{m\}$. The augmented model can be estimated by the two-step optimal RGMM estimation method discussed above.

Step 2: Let $\mathcal{J}_n = \{J_{m,n} : m \in \mathcal{M}\}$ and define the least J statistic: $MJ_n = \min \mathcal{J}_n$. This step solves the decision problem inherent to the J test since the model to be preferred in the set of the M model considered will be the one with the smallest value of $J_{m,n}$, $m = 1, \dots, M$.⁹

5 Application to the generalized Schumpeterian growth model

The purpose of this section is to apply our testing procedure to the multicountry Schumpeterian growth model with worldwide interactions, proposed by Ertur & Koch (2011). These authors propose an integrated theoretical and methodological framework characterized by technological interactions to explain growth processes from a Schumpeterian perspective. Their generalized Schumpeterian multi-country growth model, extending the one elaborated by Howitt (2000), explicitly takes into account global technological interdependence implied by R&D spillovers. Moreover they show that a multi-country Solow growth model with technological interdependence similar to that proposed by Ertur & Koch (2007) is actually nested in the generalized model leading to a simple test. Therefore, their generalized model can be interpreted as a Schumpeterian extension of the Solow growth model since in addition to factor accumulation, they show that innovation caused by R&D investment plays a major role in explaining the growth process. The model includes both determinants, with technological diffusion occurring concretely between pairs of countries, human capital reflecting absorptive capacity along the lines of Nelson & Phelps (1966), and physical capital playing the usual role. More specifically, they show that when the R&D expenditures have no effect on the growth rate of technology, the multi-country Schumpeterian growth model reduces to the multi-country Solow growth model.¹⁰

Their empirical results show that the multi-country pure Solow growth model is rejected in favor of the multi-country Schumpeterian growth model, once both models are integrated in a unified

⁹Hagemann (2012) also proposes to test whether the correct model is part of the considered ones by performing statistical inference on the MJ_n . However, this theoretical consideration can only be assessed in Monte Carlo simulations where the DGP is fixed. The interested reader may consult their Monte Carlo results to obtain further details on the behavior of this test.

¹⁰The interested reader can refer to Ertur & Koch (2011) for the full development of the multi-country Schumpeterian growth model.

theoretical and methodological framework characterized by technological interdependence. They finally estimate the TFP equation implied by their multi-country Schumpeterian growth model as well as the international R&D spillovers implied by technological interdependence using two different interaction matrices based on geographical distance and trade flows. The main result they underline is the fact that their model allows to have multiple local technological leaders: besides USA, Germany and Japon are also playing that role in their respective area of influence.

5.1 The econometric specification

Due to space limitation, we will focus here more specifically on the TFP equation implied by the multi-country growth model developed by Ertur & Koch (2011, eq. 59, p.234):

$$\ln TFP_i = \beta_0 + \beta_1 \ln \frac{s_{K,i}}{n_i + 0.05} + \beta_2 \ln s_{A,i} + \beta_3 \ln n_i + \gamma H_i \sum_{j \neq i}^n v_{ij} \ln TFP_j + \varepsilon_i \quad (26)$$

where $\ln TFP_i$ is the logarithm of Total Factor Productivity of country i at steady state (taking one third as the capital's share in income); β_1 is the coefficients associated with the logarithm of the investment rate divided by the effective depreciation rate $\ln \frac{s_{K,i}}{n_i + 0.05}$, β_2 is the coefficient associated with the logarithm of the investment rate in the research sector $\ln s_{A,i}$ and β_3 is the coefficient associated to the logarithm of the working-age population growth rate n_i . H_i is the stock of human capital, which is designed to measure the capacity of absorption of new technology for country i . Finally γ is the autoregressive interaction parameter. In matrix form, the model is written as follows:

$$Y = \beta_0 \iota + \beta_1 S_K + \beta_2 S_A + \beta_3 N + \gamma W y + \varepsilon \quad (27)$$

where Y is the $(n \times 1)$ vector of the logarithms of TFP ; ι is the unit vector, S_K is the vector of the logarithms of the investment rates in physical capital divided by the effective depreciation rates, S_A is the vector of the logarithms of expenditures in the research sector and N is the vector of the logarithms of working-age population growth rates. W is the interaction matrix modeling technological interactions as a function of the new technology absorption capacity of the receiving

country and some ad hoc measure of similarity between countries with general term $w_{ij} = H_i v_{ij}$. We will consider in the application different definitions of similarity v_{ij} and select the most relevant one given the information contained in the sample.

5.2 Data and interaction matrices

For the sake of comparability we use the same database as Ertur & Koch (2011). The data are extracted from the Heston et al. (2006) Penn World Tables (PWT version 6.2), containing information on real income, investment and population for a large number of countries. Data on R&D expenditures come from the World Investment Report (2005) of the United Nations Conference on Trade and Development (UNCTD). The sample contains 58 countries and includes 21 OECD countries over the period 1990-2003.¹¹ The explanatory variables are constructed as follows. To measure n_i , for $i = 1, \dots, n$, as the average growth of the working-age population (ages 15 to 64), it is necessary to compute the number of workers as: $RGDPCH \times POP/RGDPW$, where $RGDPCH$ is real GDP per capita computed by the chain method, $RGDPW$ is real-chain GDP per worker, and POP is the total population. The saving rate $s_{K,i}$, for $i = 1, \dots, n$, is measured as the average share of gross investment in GDP over the period as in Mankiw et al. (1992). The variable $s_{A,i}$, is measured as the average share gross domestic expenditure on R&D (GERD) relative to GDP over the 1991-2001 period. As suggested by Mankiw et al. (1992) among others, we use $g_w + \delta = 0.05$. Finally, the human capital stock is measured using the Mincerian equation, as proposed by Hall & Jones (1999) or Caselli (2005), from the data of the database developed by Cohen & Soto (2007).¹² The dependent variable is defined in Ertur & Koch (p.234, 2011) and computed as $\ln TFP_i = \ln y_i - 0.5 \ln \frac{s_{K,i}}{n_i + 0.05}$ where y_i is the real income per worker in 2003.

The first interaction matrix $W1$ we consider is based on genealogical or "genetic" distance

¹¹Due to missing data for the computation of the linguistic interaction matrix, we exclude South Korea from the sample as Korean is generally admitted to be a language isolate.

¹²This database is available at:
<http://www.parisschoolofeconomics.eu/fr/cohen-daniel/base-de-donnees-internationale-education/>.

Table 3: Descriptive statistics

	Min	Max	Mean	Median	St. err.
$\ln TFP$	3.2444	4.5469	4.1123	4.1789	0.3223
H_i	1.0287	3.4632	2.3597	2.2724	0.6172
$\ln s_A$	-3.6815	-1.3694	-2.5357	-2.6789	0.6779
$\ln n$	-1.5016	-1.3463	-1.4333	-1.4397	0.0427
$\ln \frac{s_K}{n+0.05}$	-0.1363	0.3867	0.2169	0.2284	0.1229

following Spolaore & Wacziarg (2009, 2013).¹³ This notion of distance corresponds to the time elapsed since two populations shared a common ancestor. In other words, the genetic distance is a measure of relatedness between populations over time. It can be interpreted as a general metric for average differences in characteristics transmitted across generations. Moreover they argue that: “...by its very definition, genetic distance is an excellent *summary statistic* capturing divergence in the whole set of implicit beliefs, customs, habits, biases, conventions, etc. that are transmitted across generations - biologically and/or culturally - with high persistence.” (Spolaore & Wacziarg 2009, p.471). The general idea is that more closely related societies are more likely to learn from each other and adopt each other’s innovations. Spolaore & Wacziarg (2009, 2013) show that this general genealogical measure between populations may then explain income differences today. Their interpretation is that the genetic distance captures the barriers to the diffusion of development.

Several measures of genetic distance have been proposed in the literature related to population genetics. The most popular are FST distance or “coancestor coefficient” (Wright 1965) and Nei’s distance (Nei 1973). These two measures have slightly different theoretical properties but their correlation is very strong. Thus, the choice of one these should not affect the results. The basic unit of these measures is the allele, which is a form taken by a gene. Differences in allele frequencies hence form the basis of summary measures of calculating distance between populations. The genetic distance matrix $W1$ used here is based on the differentiation coefficient G_{ST} of Nei which corresponds to the probability that two randomly selected variants in a population are different.

¹³The term genetic distance comes from the literature related to population genetics and is also used in linguistics to refer to a measure of relatedness.

Nei (1973) propose indices based on allele frequencies that describe the level and organization of diversity within populations. These indices portray the total diversity (h_T), the average intra-population diversity (h_S) or differentiation between populations (G_{ST}). G_{ST} takes a value equal to zero if and only if the distributions of alleles are identical between the two populations, whereas it is positive when the distributions of alleles are different. A greater distance measure G_{ST}^W reflects a longer separation between populations. However, like many countries, such as United States and Australia, are composed of sub-populations that are genealogically distant, it is preferable to use a weighted measure of genealogical distance G_{ST} . Thus, the interaction matrix $W1$ is based on a decreasing function of the weighted genealogical distance G_{ST}^W between countries whose functional form is the negative exponential of the distance. The general term of this matrix $W1$ is defined as $w1_{ij} = H_i v1_{ij}$ with:

$$v1_{ij} = \begin{cases} 0 & \text{if } i = j \\ e^{-g_{ij}} / \sum_{j \neq i} e^{-g_{ij}} & \text{otherwise} \end{cases} \quad (28)$$

where g_{ij} represents the weighted genetic distance G_{ST}^W between countries i and j and H_i is the human capital stock of the receiving country i .

The second interaction matrix $W2$ we consider is based on a linguistic proximity concept following Melitz & Toubal (2014). They summarize the evidence about the linguistic influences in an aggregate index, named Common Language index (CL), resting strictly on exogenous linguistic factors. This summary index (CL) is obtained using three indices: Common Official Language (COL), Common Native Language (CNL) and Language Proximity (LP).¹⁴ The Common Language index (CL) ranges between 0 and 1. More precisely, CL takes a value equal to zero if the degree of linguistic proximity between two countries is null, whereas CL tends to 1 when the linguistic proximity between these two countries is high. The general term of this matrix $W2$ is defined as

¹⁴Data on Common Language are available at http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=19.

$w2_{ij} = H_i v2_{ij}$ with:

$$v2_{ij} = \begin{cases} 0 & \text{if } i = j \\ CL_{ij} / \sum_{j \neq i} CL_{ij} & \text{otherwise} \end{cases} \quad (29)$$

where CL_{ij} represents the Common Language Index (CL) between countries i and j and H_i is the human capital stock of the receiving country i .

The last considered interaction matrix $W3$ is based on trade flows. Indeed, trade flows may proxy multi-country technological interactions.¹⁵ The general term of this matrix is defined as $w3_{ij} = H_i v3_{ij}$ where:

$$v3_{ij} = \begin{cases} 0 & \text{if } i = j \\ m_{ij} / \sum_{j \neq i} m_{ij} & \text{otherwise} \end{cases} \quad (30)$$

where m_{ij} is defined as the average imports of country i coming from country j , using the bilateral trade flows database from CEPII, and H_i is the human capital stock of the receiving country i . Ertur & Koch (2011) use a different international trade database and define a block triangular interaction matrix assuming that the North club (OECD countries) diffuses its knowledge to the South club (the rest of the countries of their sample), but the opposite does not occur (see Appendix 2, p.250, Ertur & Koch (2011)). Here we relax this assumption and further allow the South club to diffuse its knowledge to the North club. Hence this interaction matrix is full.

5.3 Econometric results

The TFP equation (26) implied by the multi-country Schumpeterian growth model and the international R&D spillovers implied by technological interdependence, is estimated by the RGMM procedure described in section 2 using the three interaction matrices based on genealogical distance

¹⁵The choice of this matrix is justified by studies of Grossman & Helpman (1991), Coe & Helpman (1995) and Coe et al. (1997) which show that international trade may be seen as a major diffusion vector of technological progress.

($W1$), linguistic distance ($W2$), and trade flows ($W3$). Each matrix has been normalized by the minimum between the maximal values of the row sums and column sums, as suggested in Kelejian & Prucha (2010, Lemma 2, p.56). Estimation results are presented in Table 4. They are consistent with those obtained by Ertur & Koch (2011) using different interaction matrices. The coefficient associated with R&D expenditures is significant whatever the interaction matrix considered. Its estimated value ranges from 0.170 to 0.184. We also note that the spatial autocorrelation parameter is significant and has an estimated value ranging from 0.210 to 0.220. Thus, the TFP of one country cannot be considered as independent from that of other countries. The coefficients of the investment rate divided by the effective depreciation rate and of the working-age population growth rate are non-significant as in Ertur & Koch (2011, Table 2, p.242).

Table 4: Estimation Results

Estimation method Interaction matrix	SAR-RGMM		
	($W1$)	($W2$)	($W3$)
Constant	4.3673 (3.0078)	4.5860 (3.2079)	4.4174 (3.0574)
$\ln(s_K/(n + 0.05))$	-0.0371 (-0.0966)	-0.0187 (-0.0480)	-0.0648 (-0.1624)
$\ln s_A$	0.1702 (2.2574)	0.1837 (2.4286)	0.1785 (2.3586)
$\ln n$	0.3083 (0.3229)	0.4338 (0.4602)	0.3204 (0.3363)
γ	0.2199 (3.1267)	0.2147 (3.1373)	0.2101 (3.0787)
n	58	58	58
Pseudo- R^2	0.6346	0.6249	0.6257

t-stats between brackets. Pseudo- R^2 is the squared correlation between predicted and actual values.

5.3.1 Application of tests for non-nested spatial models

In order to provide considerable help in the selection of the interaction matrix, we apply the J tests to equation (26) and then use the MJ procedure to select the most adequate interaction matrix (amongst those considered). The results are presented in Table 5. We first observe that none of

the three J tests are significant, implying a decision problem with respect to the selection of the interaction matrix. Indeed, the (bootstrapped) p-values are all higher than 5%. To overcome this decision problem, we use the MJ test and select the model for which the associated J statistic has the minimum value. This minimum value is obtained for the model with $W3$, namely an interaction matrix based on trade flows. All interpretations in the rest of the paper will thus be based on this interaction matrix.

Table 5: MJ tests

H_0 :	$W1$	$W2$	$W3$
H_A :	$W2,W3$	$W1,W3$	$W1,W2$
J-test	3.536	3.225	2.835
	(0.378)	(0.380)	(0.508)

Bootstrapped p-values ($B = 399$) between brackets.

5.3.2 Interpretation of the results

In this section, we compute the elasticity of the TFP of countries with respect to their own and to foreign R&D expenditures using the reduced form of model (27):

$$Y = \beta_0(I - \gamma W)^{-1} + \beta_1(I - \gamma W)^{-1}S_K + \beta_2(I - \gamma W)^{-1}S_A + \beta_3(I - \gamma W)^{-1}N + (I - \gamma W)^{-1}\varepsilon \quad (31)$$

Taking the partial derivative of Y with respect to S_A , we then obtain, in matrix form :

$$\Xi_{TFP}^{s_A} = \frac{\partial Y}{\partial S_A} = \beta_2(I - \gamma W)^{-1} = \beta_2 \sum_{r=0}^{\infty} \gamma^r W^r \quad (32)$$

where $(I - \gamma W)^{-1}$ is interpreted as the technological multiplier in the theoretical framework developed by Ertur & Koch (2011). Using straightforward notations, we get:

$$\frac{\partial y_i}{\partial s_{A,i}} \equiv (\Xi_{TFP}^{s_A})_{ii} \quad \text{and} \quad \frac{\partial y_i}{\partial s_{A,j}} \equiv (\Xi_{TFP}^{s_A})_{ij}. \quad (33)$$

In our case, the interaction matrix is a function of the capacity to absorb new technologies, measured by the stock of human capital of the receiving country, and on trade flows ($W3$) since the results of the MJ test indicate that this interaction scheme is the most adequate amongst the three schemes considered. We then obtain the $n \times n$ estimated matrix of elasticities, using the estimated coefficients of the econometric model:

$$\widehat{\Xi}_{TFP}^{sA} = \hat{\beta}_2(I - \hat{\gamma}W3)^{-1} = \hat{\beta}_2(I + \hat{\gamma}W3 + \hat{\gamma}^2W3^2 + \dots) \quad (34)$$

The diagonal elements of the matrix, specified in equation (34), represents the direct elasticities which are heterogeneous in presence of spatial autocorrelation due to higher order feedback effects. More precisely, the own derivative for country i includes feedback effects where country i affects country j and country j also affects country i as well as longer paths which might go from country i to j to k and back to i . This is what Debarsy & Ertur (2010) call *interactive heterogeneity*. The magnitude of those direct effects depends on the degree of interactions between countries, which is governed by the $W3$ matrix, the estimated value of the autoregressive parameter $\hat{\gamma}$, which measures the strength of the spatial autocorrelation and the estimated value of the parameter $\hat{\beta}_2$. Note also that the magnitude of pure feedback effects are then given by $(\widehat{\Xi}_{TFP}^{sA})_{ii} - \hat{\beta}_2$, where $\hat{\beta}_2$ could be interpreted as representing the first order direct impact of a change of R&D expenditures in country i if there was no spatial autocorrelation, i.e. if $\hat{\gamma}$ was equal to zero. The off-diagonal elements represent the indirect effects. The sum of row i represents the total impact on the TFP of country i of a 1% change of R&D expenditures in each of the countries in the sample. The sum of column j gives the total impact on the TFP of all countries of a 1% change of R&D expenditures in country j , which is of particular interest in terms of interpretation here.

On the basis of econometric results in Table 4 and equation (34), we compute the bilateral impacts, i.e. the impacts of home and foreign R&D expenditures on the TFP of a given country, using $W3$, the interaction matrix based on trade flows. Results for OECD countries plus China and Brazil, displaying significant impacts are presented in Table 6. The flow of knowledge between

countries i and j goes from the country in column j to the country in row i . Figures 3 to 5 present results for significant direct impacts, indirect impacts and total impacts for the whole sample. A network representation of those significant impacts is also proposed in Figure 6, in terms of weighted out-degrees based on the estimated value of significant indirect elasticities, using the Fruchterman and Reingold force-directed algorithm (see Fruchterman & Reingold 1991).

Table 6 is divided in two parts: the upper part displays R&D spillovers within the club of OECD countries (North-North R&D spillovers) while the lower part shows R&D spillovers from OECD countries to countries not belonging to this club (North-South R&D spillovers). Results for China and Brasil are added in the last columns to give an insight into R&D spillovers from non OECD emerging countries to countries belonging to the North club (South-North R&D spillovers) and to the South club (South-South R&D spillovers). The highlighted array elements on the main diagonal correspond to the direct elasticities, that is the elasticity of a given country with respect to its own R&D expenditures.¹⁶

We first note that the direct elasticities are all significant and slightly higher than $\hat{\beta}_2 = 0.17855$ when we take into account spatial autocorrelation, because of higher order feedback effects in the expansion (34). It is interesting to note that those feedback effects are the highest for USA, Germany, Japan, Canada, France and United Kingdom (Figure 3). Indeed, international spillovers affect the level of TFP at steady state as in Ertur & Koch (2011). However, those direct impacts differ depending on the specific interaction between countries, i.e. feedback effects are therefore inherently heterogenous. For OECD countries, they range from 0.1786% for Greece or Portugal to 0.1810% for USA, which benefits of the highest feedback effects, while the average direct impact including feedback effects is 0.1791, i.e. on average, a 1% increase of their R&D expenditures yields a 0.1791% increase of TFP. For countries not belonging to this club, on average, a 1% increase of their R&D expenditures yields a 0.1786% increase in TFP, comparable to the direct impacts for Greece or Portugal.

Second, with respect to indirect elasticities, i.e. the off-diagonal terms of the matrix in Table 6,

¹⁶Complete numerical results for the whole sample are available from the authors upon request.

we note that the United States is the country, among OECD countries, which diffuses the most of its R&D to other countries, followed first by Germany and Japan and second by France, United-Kingdom and Italy (Figures 4 and 6). The indirect impact of a 1% increase of R&D expenditures in USA on the TFP of all other countries is 0.3117%, whereas this impact is 0.2265% for Germany, 0.1790% for Japan, 0.1180% for France, 0.1018% for United Kingdom and 0.1005% for Italy (see sums of the columns of Table 6 and Figure 4). Among OECD countries, countries which diffuse the least of their R&D are Greece, New-Zealand and Portugal. Note also that the share of total intra OECD impact in the total impact ranges from 58% for USA to 99% for Portugal. For non-OECD countries, it is interesting to note that for China for instance the indirect impact of a 1% increase of R&D expenditures on the TFP of all other countries is 0.054%, an impact that is higher than impacts reported for most of the OECD countries. This is also the case for Brazil (0.039%).

In addition, let us note that, for instance, a 1% increase in R&D expenditures in USA yields a 0.0248% increase in Canadian TFP, whereas a 1% increase in R&D expenditures in Canada only yields a 0.0075% increase in US TFP. This asymmetric effect is mainly due to the asymmetric nature of trade between USA and Canada. On average, for the period under study, only 19.82% of the total imports of USA come from Canada, whereas 67.27% of the total imports of Canada come from USA. Note also that the capacity of absorption of new technologies, as measured by human capital stocks, is slightly higher in USA (3.41) than in Canada (3.29).

Moreover, it is also possible to interpret the results in terms of clubs. Inside the North club (OECD countries), the average indirect impact is 0.0303, i.e. on average, a 1% increase of the R&D expenditures of a OECD country yields a 0.0342% increase of the TFP of all other OECD countries. Inside the South club (non-OECD countries), on average, a 1% increase of the R&D expenditures yields a much lower 0.00685% increase of the TFP of the other countries. Not surprisingly, the North-South average indirect impact is on average 0.106%, much higher than the South-North average indirect impact, which is 0.00165% (see Table 7).

Third, those results are essentially due to the weight of the United States in the international trade pattern. The United States R&D diffusion impacts are higher for countries belonging to

Central and South-America, such as Canada, Mexico, Trinidad and Tobago, Jamaica, Costa-Rica and Honduras for instance than for other countries belonging to the South Club.¹⁷ We also note the role played by the human capital stock in enhancing the absorption capacity of international R&D diffusion since the impact on Canada (0.0248%) is higher than the impact on Mexico (0.0186%) although the latter has an higher import share from the United States. These results for United States show the influence of this country in the American continent, as also underlined by Coe et al. (1997). The elasticities from Japan to South East Asian countries are also higher than the elasticities from Japan to other countries belonging to the South club. These results suggest that the United States is a natural technological leader for Central and Southern American countries and that Japan is the technological leader in South East Asia.

Finally, using the network representation in Figure 7, obtained by using in addition the coordinates of the country capitals for the localization of the nodes, we note that knowledge diffuses *locally* between European countries where indirect estimated elasticities are higher for larger emitting countries as Germany, France, United Kingdom or Italy than for smaller countries. Actually Germany appears as the European technological leader. We also note that the elasticities between European and African countries are relatively high, showing the importance of European countries (essentially France and to some extend United Kingdom) as technological leaders for African countries (see Table 6).

6 Conclusions

We propose a solution to the interaction matrix selection problem in spatial econometric models using the RGMM estimation method proposed by Lin & Lee (2010), which consists in a new, robust versus unknown heteroscedasticity, version of the J test first proposed by Kelejian (2008) and modified by Kelejian & Piras (2011). Moreover the common model selection problem that arises with non-nested hypotheses tests are overcome using the approach recently proposed by

¹⁷In our sample, Canada imports almost 67% from the United States, Mexico, 77%; Trinidad and Tobago, 51%; Jamaica, 54%; Honduras, 56% and Costa-Rica, 53%.

Hagemann (2012), called the MJ test.

We then apply this non-nested hypotheses testing procedure to the the multicountry Schumpeterian growth model proposed by Ertur & Koch (2011) in order to discriminate between three different types of interaction matrices: genetic distance, linguistic proximity and bilateral trade flows. The estimation results of the TFP equation implied by this model are robust with respect to the choice of the interaction matrix. However using the J test to select an interaction matrix leads to a model selection problem. To overcome this problem, we use the MJ test and its result shows that the interaction matrix based on trade flows is the most appropriate amongst the three interaction matrices we consider.

The obtained estimation results are consistent with those of Ertur & Koch (2011) and Coe et al. (1997) and highlight the heterogeneity of the international diffusion of knowledge. As already noted by Ertur & Koch (2011), this empirical evidence cannot be captured by the standard Howitt & Aghion (1998) and Howitt (2000) models, which assume that USA are the unique *global* technological leader, whereas the integrated multi-country Schumpeterian growth model allows the emergence of *local* technological leaders. Using the selected trade flows based interaction matrix, we show that indeed USA, but also Germany and Japan actually play the role of *local* technological leaders.

We believe that the suggested interaction matrix selection procedure will be helpful for applied researchers.

Acknowledgements

We would like to thank Marion Nicolucci, Matthias Pécot and Yvan Stroppa for valuable research assistance. We benefited from the support of the Centre de Calcul Scientifique en région Centre to perform all Monte Carlo simulations. The authors acknowledge the financial support of the Region Centre Val de Loire (APR-IA 2014 00094514). We are grateful to Lung-Fei Lee and Ingmar Prucha for helpful discussions and valuable comments as well as to the participants of the 14th international

workshop on Spatial econometrics and statistics in Paris, 7th Jean Paelinck Seminar in Saragossa, 26th EC² conference hosted in Edimburgh and 5th Shanghai Econometrics Workshop.

References

- Akerlof, G. A. (1997), 'Social distance and social decisions', *Econometrica* **65**, 1005–1027.
- Anselin, L. (1988), 'Model validation in spatial econometrics: a review and evaluation of alternative approaches', *International Regional Science Review* **11**, 279–316.
- Anselin, L. & Bera, A. (1998), Spatial dependence in linear regression models with an application to spatial econometrics, in A. Ullah & D. Giles, eds, 'Handbook of Applied Economics Statistics', Marcel Dekker, New York, pp. 237–289.
- Atkinson, A. (1970), 'A method for discriminating between models.', *Journal of the Royal Statistical Society, Serie B* **32**, 323–353.
- Burridge, P. (2012), 'Improving the J test in the SARAR model by likelihood-based estimation', *Spatial Economic Analysis* **7**, 75–107.
- Burridge, P. & Fingleton, B. (2010), 'Bootstrap inference in spatial econometrics: the J-test', *Spatial Economic Analysis* **5**, 93–119.
- Caselli, F. (2005), *Accounting for cross-country income differences*, Vol. 1, Elsevier, chapter 9, pp. 679–741.
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994), *The History and Geography of Human Genes*, Princeton University Press.
- Coe, D. T. & Helpman, E. (1995), 'International R & D spillovers', *European Economic Review* **39**, 859–887.
- Coe, D. T., Helpman, E. & Hoffmaister, A. (1997), 'North-south spillovers', *The Economic Journal* **107**, 134–149.
- Cohen, D. & Soto, M. (2007), 'Growth and human capital: good data, good results', *Journal of Economic Growth* **12**, 51–76.
- Cox, D. (1961), 'Tests of separate families of hypotheses', *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* **1**, 105–123.
- Cox, D. R. (1962), 'Further results on tests of separate families of hypothesis.', *Journal of the Royal Statistical Society, Serie B* **24**, 406–424.
- Davidson, R. & MacKinnon, J. (2002), 'Bootstrap J tests of nonnested linear regression models', *Journal of Econometrics* **109**, 167–193.

- Davidson, R. & Mackinnon, J. D. (1985), ‘Heteroskedasticity-robust tests in regression directions’, *Annales de L'insee* **59-60**, 183–218.
- Davidson, R. & MacKinnon, J. G. (1981), ‘Several tests for model specification in the presence of alternative hypotheses’, *Econometrica* **49**, 781–793.
- Debarys, N. & Ertur, C. (2010), ‘Testing for spatial autocorrelation in a fixed effects panel data model’, *Regional Science and Urban Economics* **40**, 453–470.
- Ertur, C. & Koch, W. (2007), ‘Growth, technological interdependence and spatial externalities: theory and evidence’, *Journal of Applied Econometrics* **22**, 1033–1062.
- Ertur, C. & Koch, W. (2011), ‘A contribution to the theory and empirics of schumpeterian growth with worldwide interactions’, *Journal of Economic Growth* **16**, 215–255.
- Fruchterman & Reingold (1991), ‘Graph drawing by force-directed placement’, *Software Practice and Experience* **21**(11), 1129–1164.
- Godfrey, L. G. (1983), ‘Testing non-nested models after estimation by instrumental variables or least squares.’, *Econometrica* **51**, 355–366.
- Grossman, G. M. & Helpman, E. (1991), *Innovation and growth in the global economy*, MIT Press, Cambridge, MA.
- Hagemann, A. (2012), ‘A simple test for regression specification with non-nested alternatives’, *Journal of Econometrics* **166**, 247–254.
- Hall, R. E. & Jones, C. I. (1999), ‘Why do some countries produce so much more output per worker than others?’, *The Quarterly Journal of Economics* **114**, 83–116.
- Hansen, L. P. (1982), ‘Large sample properties of generalized method of moments estimators’, *Econometrica* **50**, 1029–1054.
- Heston, A., Summers, R. & Aten, B. (2006), ‘Penn world table version 6.2’.
- Howitt, P. (2000), ‘Endogenous growth and cross-country income differences’, *The American Economic Review* **90**, 829–846.
- Howitt, P. & Aghion, P. (1998), ‘Capital accumulation and innovation as complementary factors in long-run growth’, *Journal of Economic Growth* **3**, 111–130.
- Jin, F. & Lee, L.-F. (2013), ‘Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances’, *Regional Science and Urban Economics* **43**, 590–616.
- Jin, F. & Lee, L.-f. (2015), ‘On the bootstrap for moran’s I test for spatial dependence’, *Journal of Econometrics* **184**, 295–314.
- Kelejian, H. H. (2008), ‘A spatial J-test for model specification against a single or a set of non-nested alternatives’, *Letters in Spatial and Resource Sciences* **1**, 3–11.

- Kelejian, H. H. & Piras, G. (2011), ‘An extension of Kelejian’s J-test for non-nested spatial models’, *Regional Science and Urban Economics* **41**, 281–292.
- Kelejian, H. H. & Prucha, I. R. (1998), ‘A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances’, *Journal of Real Estate Finance and Economics* **17**, 99–121.
- Kelejian, H. H. & Prucha, I. R. (1999), ‘A generalized moments estimator for the autoregressive parameter in a spatial model’, *International Economic Review* **40**, 509–533.
- Kelejian, H. H. & Prucha, I. R. (2007), ‘Relative efficiencies of various predictors in spatial econometric models containing spatial lags’, *Regional Science and Urban Economics* **37**, 363–374.
- Kelejian, H. H. & Prucha, I. R. (2010), ‘Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances’, *Journal of Econometrics* **157**, 53–67.
- Kelejian, H. H., Prucha, I. R. & Yuzevovich, Y. (2004), Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results, in J. LeSage & K. R. Pace, eds, ‘Spatial and Spatiotemporal Econometrics, Advance in Econometrics’, Vol. 18, Emerald Group Publishing, pp. 163–198.
- Lee, L.-F. (2001), Generalized method of moments estimation of spatial autoregressive processes, Technical report.
- Lee, L.-F. (2003), ‘Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances’, *Econometric Reviews* **22**, 307–335.
- Lee, L.-F. (2007), ‘GMM and 2SLS estimation of mixed regressive, spatial autoregressive models’, *Journal of Econometrics* **137**, 489–514.
- Lee, S.-Y. (2008), Three essays on spatial econometrics and empirical industrial organization, PhD thesis.
- Lin, X. & Lee, L.-F. (2010), ‘GMM estimation of spatial autoregressive models with unknown heteroskedasticity’, *Journal of Econometrics* **157**, 34–52.
- Liu, R. (1988), ‘Bootstrap procedures under some non-i.i.d. models’, *Annals of Statistics* **16**(4), 1696–1708.
- Liu, X., Lee, L.-F. & Bollinger, C. R. (2010), ‘An efficient gmm estimator of spatial autoregressive models’, *Journal of Econometrics* **159**, 302–319.
- Liu, X., Patacchini, E. & Zenou, Y. (2014), ‘Endogenous peer effects: local aggregate of global average?’, *Journal of Economic Behavior and Organization* **103**, 39–59.
- Liu, X. & Saraiva, P. (2015), ‘GMM estimation of sar models with endogenous regressors’, *Regional Science and Urban Economics* **55**, 68–79.

- Mackinnon, J. D., White, H. & Davidson, R. (1983), 'Tests for model specification in the presence of alternative hypotheses', *Journal of Econometrics* **21**, 53–70.
- Mammen, E. (1992), *When does bootstrap work? Asymptotic results and simulations*, Springer, New York.
- Mankiw, G. N., Romer, D. & Weil, D. N. (1992), 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics* **107**, 407–437.
- Melitz, J. & Toubal, f. (2014), 'Native language, spoken language, translation and trade', *Journal of International Economics* **93**, 351–363.
- Nei, M. (1973), 'Analysis of gene diversity in subdivided populations', *Proceedings of the National Academy of Sciences of the United States of America* **70**, 3321–3323.
- Nelson, R. R. & Phelps, E. S. (1966), 'Investment in humans, technological diffusion, and economic growth', *The American Economic Review* **56**, 69–75.
- Pesaran, H. M. (1974), 'On the general problem of model selection', *The Review of Economic Studies* **41**, 153–171.
- Pesaran, H. M. & Deaton, A. (1978), 'Testing non-nested nonlinear regression models', *Econometrica* **46**, 677–694.
- Pesaran, H. M. & Weeks, M. (2001), Non-nested hypothesis testing: an overview, in B. H. Baltagi, ed., 'A Companion to Theoretical Econometrics', Blackwell Publishers Ltd.
- Spolaore, E. & Wacziarg, R. (2009), 'The diffusion of development', *The Quarterly Journal of Economics* **124**, 469–529.
- Spolaore, E. & Wacziarg, R. (2013), 'How deep are the roots of economic development?', *Journal of Economic Literature* **51**, 325–369.
- Wright, S. (1965), 'The interpretation of population structure by f-statistics with special regard to systems of mating', *Evolution* .

Table 6: Matrix of significant elasticities at 5% for OECD countries using trade flows ($W3$)

	AUS	AUT	BEL	CAN	DNK	FIN	FRA	GER	GRC	IRL	ITA	JPN
AUS	0.1787	-	0.0000	-	-	0.0004	0.0017	0.0039	-	-	0.0018	0.0090
AUT	-	0.1788	0.0014	-	-	-	0.0025	0.0165	-	-	0.0036	0.0017
BEL	-	-	0.1790	-	-	0.0003	0.0057	0.0081	-	-	0.0021	0.0015
CAN	-	-	0.0000	0.1796	-	-	0.0012	0.0019	-	-	-	0.0042
DNK	-	-	0.0018	-	0.1787	0.0013	0.0029	0.0095	-	-	0.0022	0.0017
FIN	-	-	0.0015	-	0.0013	0.1787	0.0024	0.0076	-	-	0.0021	0.0030
FRA	-	-	0.0033	-	-	0.0004	0.1792	0.0079	-	-	0.0041	0.0019
GER	-	0.0021	0.0033	-	-	0.0006	0.0054	0.1801	-	-	0.0044	0.0032
GRC	-	-	0.0015	-	-	0.0004	0.0033	0.0063	0.1786	-	0.0056	0.0019
IRL	-	-	0.0000	-	-	-	0.0020	0.0032	-	0.1787	0.0012	0.0025
ITA	-	0.0010	0.0020	-	-	0.0003	0.0049	0.0079	-	-	0.1790	0.0013
JPN	0.0011	-	0.0005	0.0016	-	-	0.0015	0.0033	-	0.0005	0.0014	0.1798
NLD	-	-	0.0045	-	-	0.0005	0.0032	0.0092	-	-	0.0019	0.0024
NZL	0.0083	-	0.0000	0.0009	-	-	0.0013	0.0027	-	-	0.0013	0.0072
NOR	-	-	0.0013	0.0011	0.0029	0.0016	0.0022	0.0064	-	-	0.0019	0.0027
PRT	-	-	0.0011	-	-	0.0002	0.0032	0.0044	-	-	0.0025	0.0010
ESP	-	-	0.0015	-	-	0.0003	0.0059	0.0060	-	-	0.0035	0.0015
SWE	-	-	0.0018	-	0.0026	0.0025	0.0030	0.0086	-	-	0.0019	0.0021
CHE	-	0.0019	0.0019	-	-	-	0.0051	0.0146	-	-	0.0047	0.0021
GBR	-	-	0.0023	-	-	0.0006	0.0043	0.0071	-	0.0018	0.0026	0.0030
USA	-	-	0.0000	0.0075	-	-	0.0016	0.0034	-	-	0.0015	0.0089
ARG	-	-	0.0000	-	-	-	0.0020	0.0025	-	-	0.0021	0.0019
BOL	-	-	0.0000	0.0005	-	-	0.0005	0.0013	-	-	0.0007	0.0023
BRA	-	-	0.0000	0.0009	-	-	0.0012	0.0031	-	-	0.0017	0.0021
BFA	-	-	0.0005	-	0.0002	-	0.0042	0.0008	-	-	0.0005	0.0010
CHL	-	-	0.0000	0.0010	-	-	0.0016	0.0025	-	-	0.0015	0.0029
CHN	-	-	0.0003	0.0004	-	-	0.0011	0.0024	-	-	0.0010	0.0080
COL	-	-	0.0000	0.0009	-	-	0.0009	0.0017	-	-	0.0009	0.0020
CRI	-	-	0.0000	0.0008	-	-	0.0007	0.0011	-	-	-	0.0021
CUB	-	-	0.0000	0.0012	-	-	0.0014	0.0013	-	-	0.0017	0.0008
CYP	-	-	0.0000	-	-	-	0.0017	0.0040	0.0021	-	0.0028	0.0030
ECU	-	-	0.0000	0.0007	-	-	0.0010	0.0018	-	-	0.0012	0.0027
EGY	-	-	0.0006	-	-	0.0003	0.0022	0.0029	-	-	0.0025	0.0016
SLV	-	-	0.0000	0.0005	-	-	0.0008	0.0011	-	-	0.0005	0.0015
HND	-	-	0.0000	0.0006	-	-	-	0.0008	-	-	-	0.0016
HUN	-	0.0040	0.0014	-	-	-	0.0022	0.0116	-	-	0.0035	0.0018
IND	-	-	0.0004	0.0004	-	-	0.0011	0.0026	-	-	0.0010	0.0024
JAM	-	-	0.0000	0.0015	-	-	0.0008	0.0009	-	-	0.0005	0.0027
MDG	-	-	0.0006	-	-	-	0.0057	0.0013	-	-	0.0007	0.0013
MYS	0.0008	-	0.0000	0.0004	-	-	0.0009	0.0017	-	-	0.0006	0.0083
MUS	0.0008	-	0.0000	-	-	-	0.0043	0.0019	-	-	0.0012	0.0021
MEX	-	-	0.0000	0.0012	-	-	0.0006	0.0014	-	-	0.0005	0.0021
NIC	-	-	0.0000	0.0006	-	-	-	0.0008	-	-	-	0.0020
PAN	-	-	0.0000	-	-	-	0.0006	0.0008	-	-	-	0.0156
PRY	-	-	0.0000	-	-	-	0.0005	0.0010	-	-	0.0006	0.0020
PER	-	-	0.0000	0.0008	-	-	0.0008	0.0017	-	-	0.0010	0.0026
PHL	0.0009	-	0.0000	0.0005	-	-	0.0010	0.0017	-	-	0.0006	0.0090
ROM	-	0.0013	0.0010	-	-	-	0.0031	0.0082	-	-	0.0065	0.0008
SGP	0.0005	-	0.0000	-	-	-	0.0010	0.0016	-	-	0.0007	0.0072
ZAF	0.0005	-	0.0008	0.0004	-	-	0.0015	0.0046	-	0.0003	0.0014	0.0026
SYR	-	-	0.0010	-	-	-	0.0021	0.0031	-	-	0.0029	0.0019
THA	0.0005	-	0.0000	0.0004	-	-	0.0009	0.0018	-	-	0.0006	0.0090
TTO	-	-	0.0000	0.0021	-	-	0.0008	0.0018	-	-	0.0009	0.0026
TUN	-	-	0.0010	-	-	-	0.0049	0.0028	-	-	0.0035	0.0005
TUR	-	-	0.0009	-	-	-	0.0022	0.0050	-	-	0.0028	0.0014
UGA	-	-	0.0006	0.0003	0.0003	-	0.0008	0.0014	-	0.0002	0.0009	0.0018
URY	-	-	0.0000	-	-	-	0.0014	0.0016	-	-	0.0017	0.0013
VEN	-	-	0.0000	0.0010	-	-	0.0008	0.0016	-	-	0.0013	0.0017

- denotes non significant (at the 5% threshold) while all figures are significant impact at the 5% level.

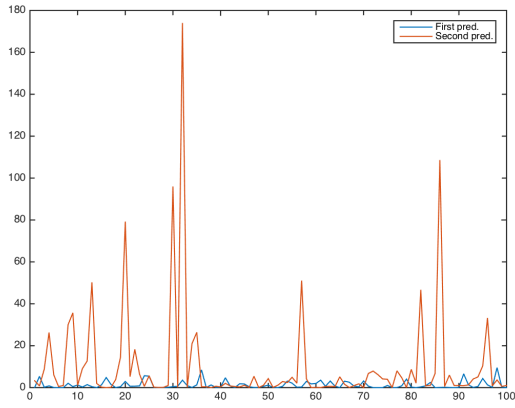
Table 6: Matrix of significant elasticities for OECD countries using trade flows ($W3$), Cont'd

	NLD	NZL	NOR	PRT	ESP	SWE	CHE	GBR	USA	BRA	CHN
AUS	-	0.0016	-	-	-	-	-	0.0033	0.0090	-	0.0027
AUT	0.0016	-	-	-	-	-	0.0016	0.0016	0.0023	-	0.0007
BEL	0.0059	-	-	-	-	0.0011	-	0.0029	0.0028	-	0.0006
CAN	-	-	-	-	-	-	-	0.0014	0.0248	-	0.0012
DNK	0.0030	-	0.0013	-	-	0.0048	-	0.0035	0.0024	-	0.0010
FIN	0.0018	-	0.0010	-	0.0007	0.0048	0.0009	0.0032	0.0035	-	0.0010
FRA	0.0021	-	-	-	0.0023	-	-	0.0031	0.0036	-	0.0009
GER	0.0038	-	-	-	0.0014	-	0.0020	0.0033	0.0040	-	0.0014
GRC	0.0024	-	-	-	0.0012	-	-	0.0024	0.0018	-	0.0007
IRL	0.0014	-	-	-	-	-	-	0.0120	0.0057	-	0.0006
ITA	0.0023	-	-	-	0.0015	-	0.0014	0.0025	0.0022	-	0.0008
JPN	-	-	-	-	-	0.0006	0.0007	0.0016	0.0120	0.0005	0.0062
NLD	0.1791	-	-	-	-	0.0012	-	0.0038	0.0038	-	0.0009
NZL	-	0.1786	-	-	-	-	-	0.0027	0.0067	-	0.0019
NOR	0.0021	-	0.1787	-	-	0.0063	-	0.0040	0.0035	-	0.0010
PRT	0.0015	-	-	0.1786	0.0051	-	-	0.0019	0.0010	-	0.0003
ESP	0.0015	-	-	0.0009	0.1788	-	-	0.0027	0.0023	-	0.0008
SWE	0.0028	-	0.0020	-	-	0.1789	-	0.0042	0.0035	-	0.0007
CHE	0.0024	-	-	-	-	-	0.1788	0.0024	0.0032	-	0.0008
GBR	0.0032	-	-	-	0.0012	0.0013	-	0.1792	0.0054	-	0.0009
USA	-	-	-	-	-	-	-	0.0020	0.1810	-	0.0024
ARG	-	-	-	-	0.0014	-	-	0.0009	0.0070	0.0068	0.0011
BOL	-	-	-	-	0.0006	0.0006	-	0.0005	0.0052	0.0057	-
BRA	-	-	-	-	-	-	-	0.0009	0.0071	0.1787	0.0006
BFA	0.0005	-	-	-	0.0004	-	-	0.0005	0.0009	-	0.0004
CHL	-	-	-	-	0.0013	-	-	0.0009	0.0084	0.0030	0.0014
CHN	0.0004	-	-	-	-	0.0005	0.0003	0.0007	0.0028	0.0003	0.1789
COL	-	-	-	-	0.0007	-	-	0.0007	0.0084	0.0010	0.0005
CRI	-	-	-	-	-	-	-	-	0.0118	0.0007	0.0004
CUB	-	-	-	-	0.0045	-	-	-	0.0008	0.0006	0.0045
CYP	0.0008	-	-	-	0.0010	-	-	0.0032	0.0033	-	0.0006
ECU	-	-	-	-	0.0009	-	-	0.0006	0.0083	0.0013	0.0005
EGY	0.0007	-	-	-	0.0006	0.0006	0.0006	0.0012	0.0028	-	0.0005
SLV	-	-	-	-	-	-	-	-	0.0082	-	0.0003
HND	-	-	-	-	0.0005	-	-	-	0.0112	0.0004	0.0004
HUN	0.0014	-	-	-	0.0007	0.0007	0.0010	0.0016	0.0019	-	0.0009
IND	0.0005	-	-	-	-	-	0.0007	0.0015	0.0029	-	0.0007
JAM	0.0004	-	-	-	-	-	-	0.0015	0.0147	0.0006	0.0006
MDG	0.0004	-	-	-	0.0003	-	-	0.0007	0.0010	-	0.0009
MYS	0.0004	-	-	-	-	0.0003	0.0004	0.0012	0.0042	-	-
MUS	-	-	-	-	-	-	-	0.0020	0.0011	-	0.0016
MEX	-	-	-	-	-	-	-	-	0.0186	-	0.0006
NIC	-	-	-	-	0.0006	-	-	-	0.0069	-	0.0003
PAN	-	-	-	-	-	-	-	-	0.0045	-	0.0012
PRY	-	-	-	-	-	-	-	0.0008	0.0044	0.0076	0.0004
PER	-	-	-	-	0.0009	-	-	0.0006	0.0080	0.0017	0.0010
PHL	0.0005	-	-	-	-	-	-	0.0009	0.0054	-	-
ROM	0.0014	-	-	-	-	-	0.0009	0.0018	0.0018	-	0.0006
SGP	0.0004	-	-	-	-	-	0.0003	0.0011	0.0044	-	-
ZAF	0.0009	-	-	-	0.0004	0.0004	0.0007	0.0030	0.0034	0.0003	0.0009
SYR	0.0008	-	-	-	0.0006	-	-	0.0011	0.0014	-	0.0008
THA	0.0004	-	-	-	-	-	0.0004	0.0008	0.0040	-	0.0014
TTO	0.0006	-	-	-	-	-	-	0.0027	0.0153	0.0007	0.0008
TUN	0.0006	-	-	-	0.0008	-	-	0.0006	0.0009	-	0.0002
TUR	0.0010	-	-	-	0.0008	0.0006	0.0007	0.0017	0.0023	-	0.0006
UGA	0.0005	-	-	-	0.0003	0.0002	0.0003	0.0027	0.0012	0.0002	0.0006
URY	-	-	-	-	0.0011	-	-	0.0008	0.0038	0.0067	0.0007
VEN	-	-	-	-	0.0006	-	-	0.0007	0.0098	0.0010	0.0003

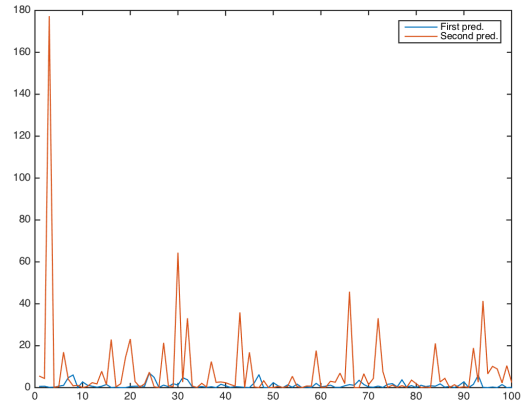
– denotes non significant (at the 5% threshold) while all figures are significant impacts at the 5% level.

Table 7: Average significant impacts for Clubs

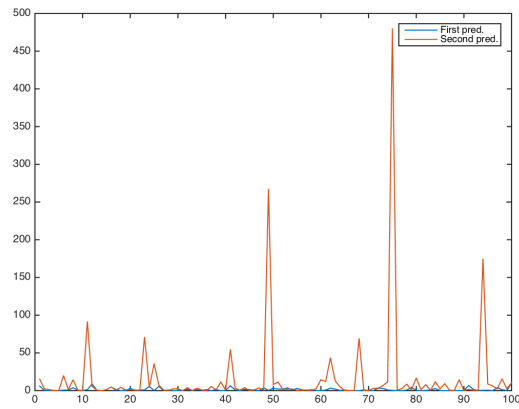
Club	Mean	Median
North - North		
Direct impact	0.1791	0.17889
Feedback effect	0.000526	0.000271
Indirect impact	0.03030	0.01021
Total impact	0.20938	0.18982
North - South		
Indirect impact	0.10602	0.10602
South - South		
Direct impact	0.1786	0.1786
Feedback effect	0.0000575	0.0000115
Indirect impact	0.00685	0.001306
Total impact	0.18543	0.17987
South - North		
Indirect impact	0.00165	0.0000



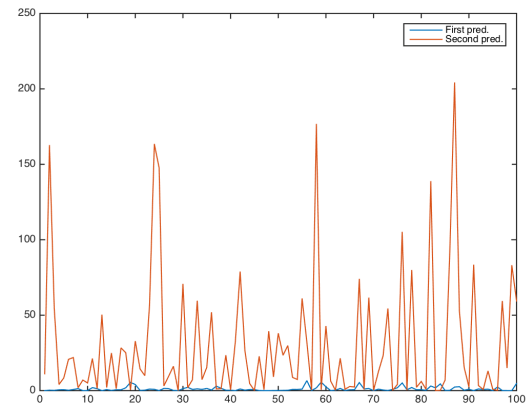
(a) $n = 100, \lambda = 0.3$



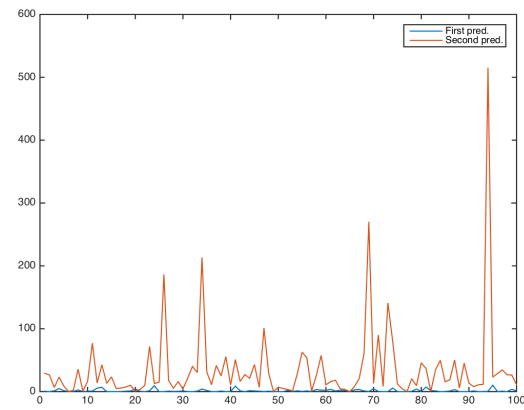
(b) $n = 254, \lambda = 0.3$



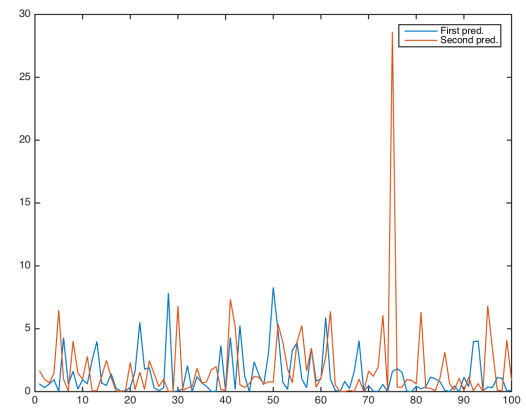
(c) $n = 100, \lambda = 0.5$



(d) $n = 254, \lambda = 0.5$



(e) $n = 100, \lambda = 0.7$



(f) $n = 254, \lambda = 0.7$

Figure 1: Value of the J test associated to the true model using the first and second predictors.

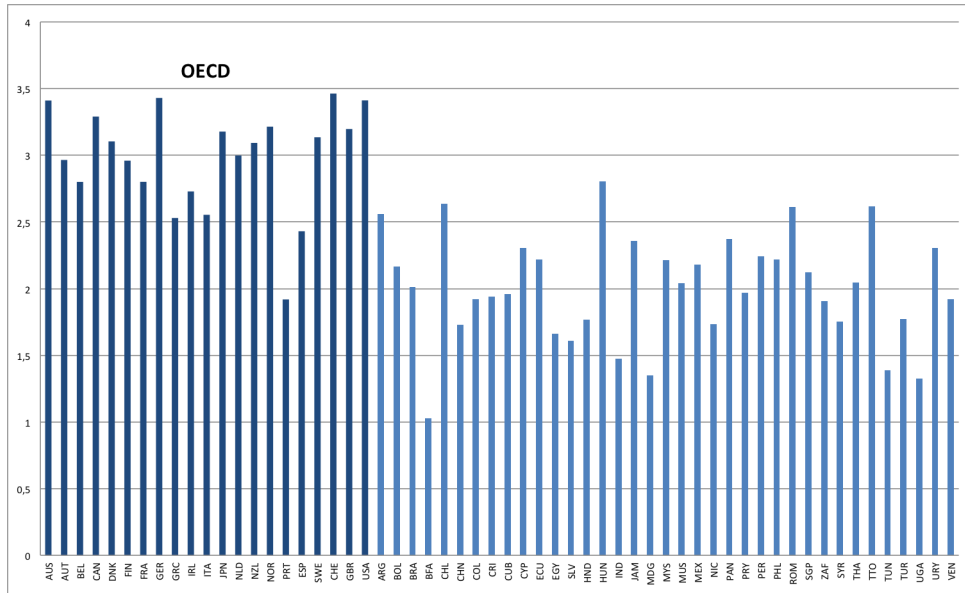


Figure 2: Human capital stock (h_i) measured using the Mincerian equation, as proposed by Hall & Jones (1999) or Caselli (2005), from the data of the database developed by Cohen & Soto (2007).

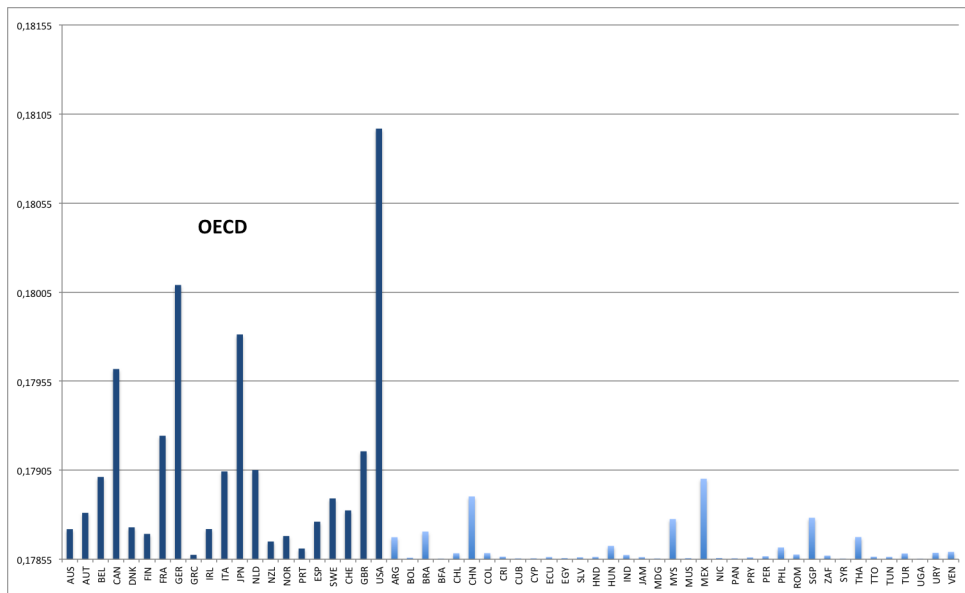


Figure 3: Direct impacts including heterogenous feedback effects of a change of R&D expenditures on the TFP of all countries using the trade based interaction matrix (W^3)

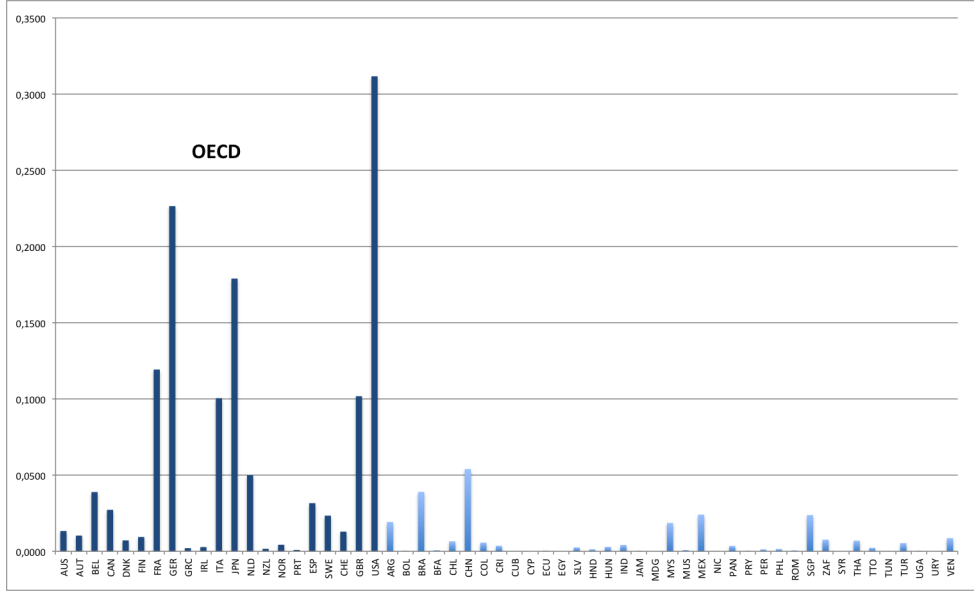


Figure 4: Indirect impacts of a change of R&D expenditures on the TFP of all countries using the trade based interaction matrix ($W3$)

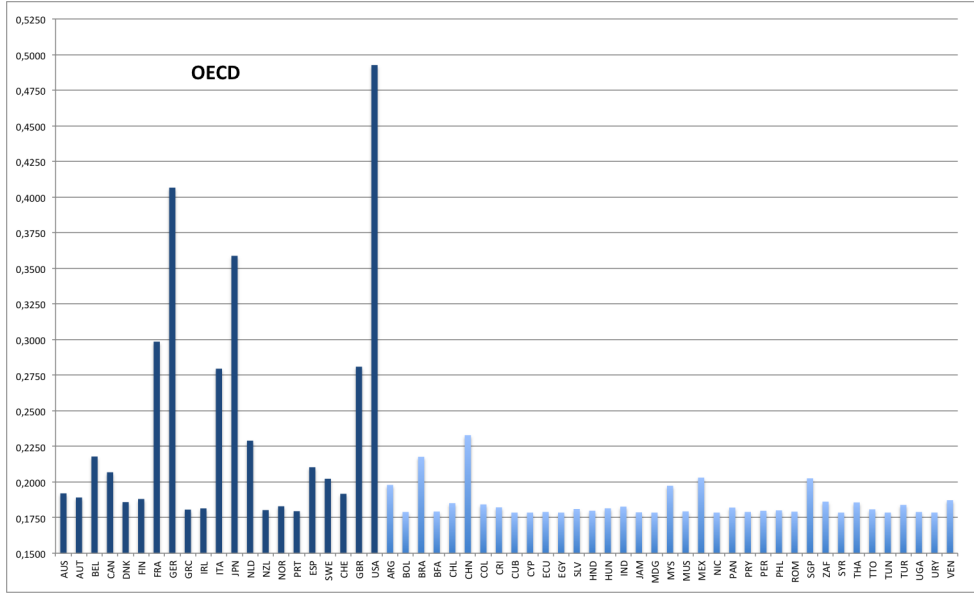


Figure 5: Total impacts of a change of R&D expenditures on the TFP of all countries using the trade based interaction matrix ($W3$)

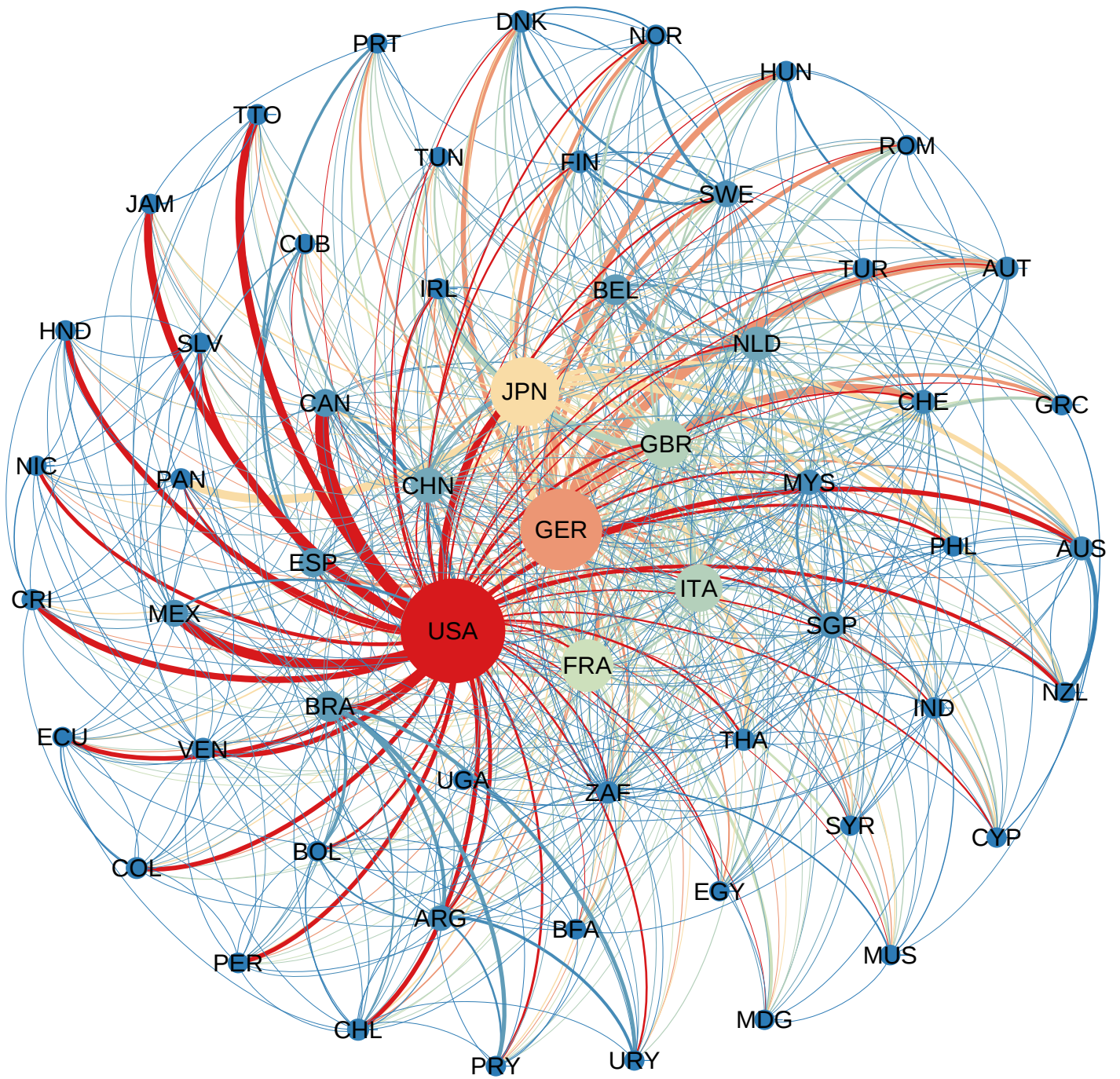


Figure 6: Network representation of significant indirect impacts (weighted out-degrees).

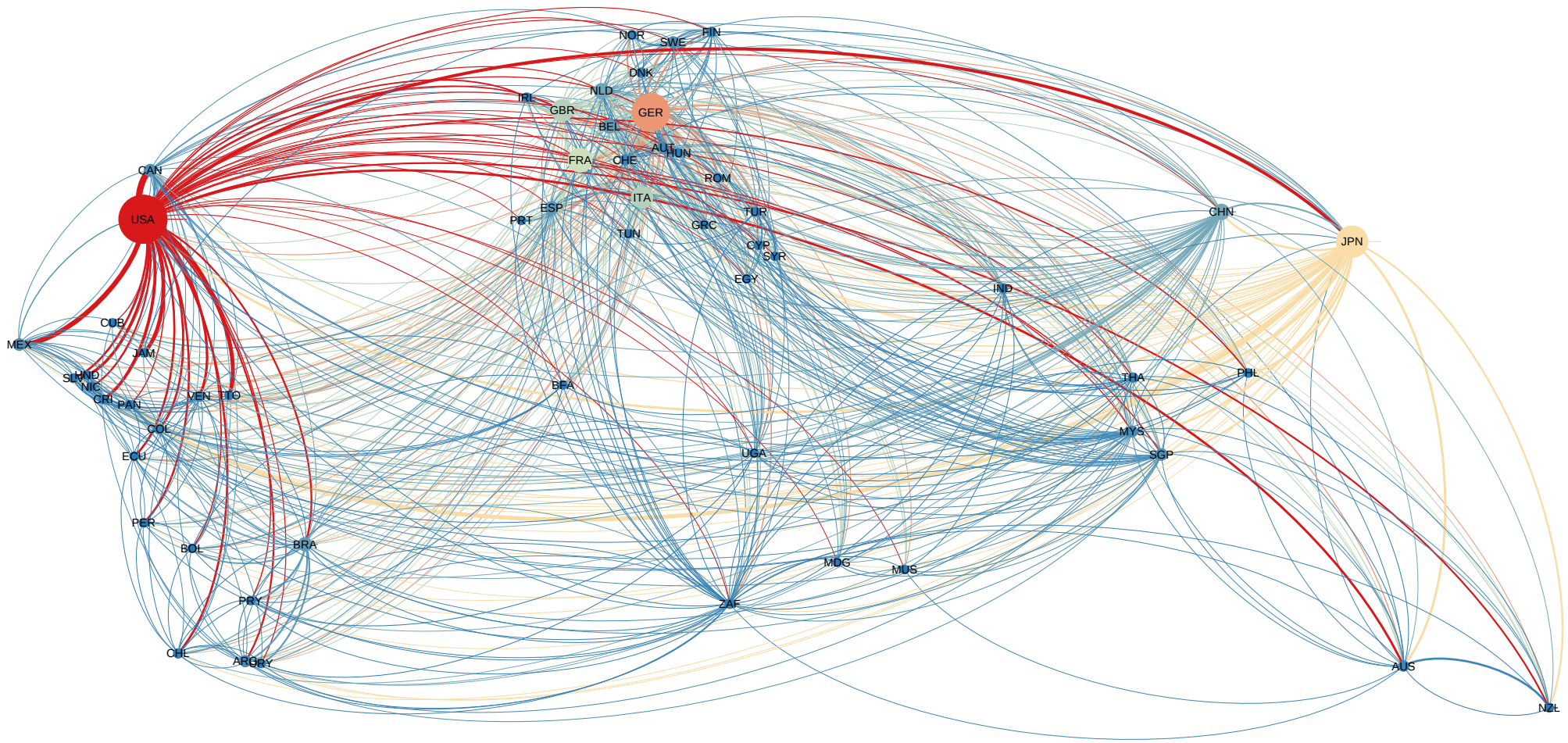


Figure 7: Network representation of significant indirect impacts (weighted out-degrees). (Nodes are localized using the coordinates of the capitals of the countries of the sample)