UNIVERSITY OF
CAMBRIDGE

Cambridge Working Papers in Economics

A Two Stage Approach to Spatiotemporal Analysis with Strong and Weak Cross-Sectional Dependence

Natalia Bailey, Sean Holly and M. Hashem Pesaran

CWPE 1362

A Two Stage Approach to Spatiotemporal Analysis with Strong and Weak Cross-Sectional Dependence

Natalia Bailey, Sean Holly and M. Hashem Pesaran

December 2013

CWPE 1362

A Two Stage Approach to Spatiotemporal Analysis with Strong and Weak Cross-Sectional Dependence*

Natalia Bailey
Queen Mary, University of London

Sean Holly
University of Cambridge

M. Hashem Pesaran
University of Southern California and Trinity College, Cambridge

December 2013

Abstract

An understanding of the spatial dimension of economic and social activity requires methods that can separate out the relationship between spatial units that is due to the effect of common factors from that which is purely spatial even in an abstract sense. The same applies to the empirical analysis of networks in general. We are able to distinguish between cross-sectional strong dependence and weak dependence. Strong dependence in turn suggests that there are common factors. We use cross unit averages to extract common factors and contrast this to a principal components approach widely used in the literature. We then use a multiple testing procedure to determine significant bilateral correlations (signifying connections) between spatial units and compare this to an approach that just uses distance to determine units that are neighbours. We apply these methods to real house price changes at the level of Metropolitan Statistical Areas in the USA, and estimate a heterogeneous spatiotemporal model for the de-factored real house price changes and obtain significant evidence of spatial connections, both positive and negative.

Keywords: Spatial and factor dependence, spatiotemporal models, house price changes.

JEL Classification: C21, C23

*The authors would like to thank Alex Chudik, George Kapetanios, Ron Smith, Vanessa Smith and participants at the 2012 SEA conference, Salvador, Brazil, the 2013 conference on Cross-sectional Dependence in Panel Data Models, Cambridge, UK, the 2013 Econometrics of Social Interaction Symposium, York, UK, and the 2013 SEA conference, Washington DC, US, for valuable comments and suggestions. The authors also wish to acknowledge financial support under ESRC Grant ES/I031626/1.

1 Introduction

The nature and degree of spatial dependence in economic, geographical, epidemiological and ecological systems has long been the focus of intensive study. Geographers regard the fundamental question in economic geography to be what explains the uneven pattern of economic activity in space. Indeed the New Economic Geography starting with Krugman (1991) addresses exactly this question. But where we have a data rich environment with observations on many spatial units over many time periods there may be obstacles to understanding these uneven patterns in spatial data because of complex dependencies between spatial units that reflect both local (clustering) and common factors. Recent developments in spatial econometrics have generated a growing literature on methods for modelling and measuring spatial or cross section dependence in data sets with a panel structure where there are observations over time (T) and over space (N). This in turn has identified a number of central research questions. What is the source of dependencies in space? To what extent are the observed dependencies between different spatial units due to common factors - for example, aggregate shocks - that affect different units rather than being the result of local interactions that generate spatial spill-over effects? Is the implementation of estimation procedures of panels that implicitly assume spatially correlated units justifiable when the degree of their cross dependence has not been established? Do existing methods of identifying neighbouring relationships fully reflect the actual spatial structure of the underlying data studied?

Currently, there are two main approaches to modelling cross sectional dependence in large panels: spatial processes and factor structures. Spatial processes were pioneered by Whittle (1954) and developed further in econometrics by Anselin (1988), Kelejian and Prucha (1999), and Lee (2002), amongst others. Factor models were introduced by Hotelling (1933) and first applied in economics by Stone (1947). They have been applied extensively in finance - Chamberlain and Rothschild (1983), Connor and Korajczyk (1993), Stock and Watson (1998), and Kapetanios and Pesaran (2007) -, and in macroeconomics as in Forni and Reichlin (1998) and Stock and Watson (2002a,b).

Factors can be represented by cross-sectional averages at regional and/or national levels (Pesaran, 2006), or can be estimated by Principle Components (PCs). The number of principal components can be determined using the information criteria proposed by Bai and Ng (2002), amongst others. Estimation of panels with spatially correlated errors include the use of parametric methods based on maximum likelihood - Lee (2004), Yu, de Jong and Lee (2008), Lee and Yu (2010), or the GMM approach proposed by Kelejian and Prucha (1999, 2010), Kapoor, Kelejian and Prucha (2007), and Lin and Lee (2010). Furthermore, non-parametric methods using spatial HAC estimators have been applied by Conley (1999), Kelejian and Prucha (2007), and Bester, Conley and Hansen (2011). Chudik and Pesaran (2013) provide a comprehensive review of recent literature on estimation and inference in large panel data models with cross-sectional dependence.

The factor and spatial econometric approaches tend to complement one another, with the factor approach more suited to modelling strong cross-sectional dependence, whilst the spatial approach generally requires the spatial dependence to be weak. See, for example, Chudik, Pesaran and Tosetti (2011). This presents a challenge as most panel data sets are subject to a combination of strong and weak cross dependencies, and a methodology that is capable of identifying and dealing with both forms of cross dependence is needed. This paper proposes a two-stage estimation and inference strategy, whereby

in the first step tests of cross sectional dependence are applied to ascertain if the cross sectional dependence is weak. If the null of weak cross-sectional dependence is rejected, the implied strong cross-sectional dependence is modelled by means of a factor model. Residuals from such factor models, referred to as de-factored observations, are then used to estimate possible connections amongst pairs of cross section units, and ultimately to model the remaining weak cross dependencies, making use of extant techniques from spatial econometrics.

In addition to using spatial weights matrices based on contiguity and geodesic distance, we also consider the use of pair-wise correlation of the de-factored observations to identify if a given pair of cross section units are connected. To avoid the multi-testing problem that such an approach entails we employ the Holm (1979) procedure discussed and justified in Bailey, Pesaran and Smith (2013) for consistent estimation of large correlation matrices.[1]

Finally, we provide a detailed application of the proposed two-step methodology to real house price changes across different Metropolitan Statistical Areas (MSAs) in the US. To estimate the correlation-based measures of connections, we begin with the estimation of a de-factored set of observations from a hierarchical spatiotemporal model or by means of the PC analysis. We then apply the Holm procedure to the pair-wise correlations of the de-factored observations to estimate the $N \times N$ connections matrix, $\hat{\mathbf{W}}$, which we then decompose into $\hat{\mathbf{W}}^{+}$ (representing positive connections) and $\hat{\mathbf{W}}^{-}$ (representing negative connections).[2] These positive and negative connection matrices are then compared to geodesic based spatial matrices, $\mathbf{W}_d$, (d being the selected distance measure between different MSAs) and their closeness examined by means of contingency tables. A spatiotemporal model of house prices is then estimated by the quasi maximum likelihood (QML) approach developed in Aquaro, Bailey and Pesaran (2013) for the analysis of heterogeneous spatiotemporal panel data models. The QML estimates confirm important dynamics in the de-factored house price changes as well as statistically significant positive and negative spill-over effects across the MSAs, with the positive effects being more prevalent.

The rest of the paper is organised as follows: Section 2 motivates and describes the first stage of the proposed two-step spatiotemporal modelling strategy. Section 3 focuses on the second stage of the proposed approach and suggests a correlation based method for approximating network connections using de-factored observations from the first stage. Section 4 presents the empirical application to the US real house price changes. Finally, Section 5 concludes. Data specifications and sources are relegated to the Appendix.

Notation: The largest and the smallest eigenvalues of the $N \times N$ matrix $\boldsymbol{A} = (a_{ij})$, are denoted by $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ respectively, $\|\boldsymbol{A}\| = \lambda^{1/2}_{\max}(\boldsymbol{A}'\boldsymbol{A})$ is the spectral (or operator norm) of $\mathbf{A}$, $\|\boldsymbol{A}\|_1 = \max_{1\leq j\leq N}\left\{\sum_{i=1}^{N}|a_{ij}|\right\}$ is its maximum absolute column sum norm, and $\|\boldsymbol{A}\|_\infty = \max_{1\leq i\leq N}\left\{\sum_{j=1}^{N}|a_{ij}|\right\}$ is its maximum absolute row sum norm.

[1] For a related field see Barigozzi and Brownlees (2013). Their empirical application infers just one common factor in the form of the market return.

[2] Note that $\hat{\mathbf{W}} = \hat{\mathbf{W}}^{+} + \hat{\mathbf{W}}^{-}$.

2 Spatial econometric models

The standard spatial econometric model can be written as

$$\boldsymbol{x}_{\circ t} = \psi \mathbf{W} \boldsymbol{x}_{\circ t} + \mathbf{u}_{\circ t}, \tag{1}$$

where $\boldsymbol{x}_{\circ t} = (x_{1t}, ..., x_{Nt})'$, $\mathbf{u}_{\circ t} = (u_{1t}, ..., u_{Nt})'$, and $\mathbf{W}$ is a given spatial weights matrix. The error terms u_{it} are assumed to be independently distributed over both i and t with zero mean and variances $var(u_{it}) = \sigma_{u_i}^2$, so that $var(\mathbf{u}_{\circ t}) = \mathbf{\Sigma}_u = Diag\left(\boldsymbol{\sigma}_{u_\circ}^2\right)$, where $\boldsymbol{\sigma}_{u_\circ}^2 = (\sigma_{u_1}^2, \ldots, \sigma_{u_N}^2)'$, $0 < \sigma_{u_i}^2 < K < \infty$ on its i^{th} diagonal and K is a finite generic constant independent of N. Hence, (1) can be re-written as

$$\boldsymbol{x}_{\circ t} = \psi \mathbf{W} \boldsymbol{x}_{\circ t} + \mathbf{\Sigma}_u^{1/2} \tilde{\mathbf{u}}_{\circ t}, \tag{2}$$

where $\tilde{\mathbf{u}}_{\circ t} = (\tilde{u}_{1t}, ..., \tilde{u}_{Nt})'$ and $\tilde{u}_{it} \sim IIDN(0, 1)$, for $i = 1, ..., N$. This is commonly referred to in the spatial econometrics literature as a *spatial autoregressive* model, or a *spatially lagged dependent variable* model. The usual approach is to specify the $\mathbf{W}$ matrix *a priori* and then estimate equation (1) directly.[3] One possible problem with this is that apparent cross sectional dependence that is meant to be captured by $\mathbf{W}$ may actually be due, in part, to common effects from the exogenous variables. We propose a two-stage modelling strategy below where we first purge the data of potential common effects, using a factor model, and then focus attention on the resulting residuals to identify possible spatial patterns through a multiple testing analysis of the correlation matrix.

In the first stage of our procedure we use the cross-sectional average approach of Pesaran (2006) to approximate the factors creating strong cross dependence at a national and regional level. As explained in the introduction, different methods can be used to eradicate the data from common effects, such as maximum likelihood - Robertson and Symons (2007) - or principal components - Bai (2003). Our preference towards the cross-sectional averages alternative mainly arises from the fact that in this case the factors have clear economic interpretations. On the other hand, under the principal components approach for instance, an infinite number of factor rotations are possible rendering representation of well-defined, economically meaningful factors difficult.

For the second stage of our strategy, there is a related literature that addresses the issue of identification of neighbours by estimating the corresponding sparse covariance matrix of the data set. The first approach uses Markov networks as its basis. This is defined as a graphical model that represents variables as nodes and 'conditional' dependencies (partial correlations) between variables as (undirected) edges. Estimation then amounts to setting elements of the inverse covariance matrix to zero - Dempster (1972). A number of estimation approaches have followed involving lasso regressions or graphical lasso (penalised ML with a lasso penalty) performed on the inverse covariance matrix. The main problem with this approach is that once de-factoring has taken place the interpretation of the resulting inverse covariance matrix is ambiguous. Furthermore, finding a good estimate of the inverse covariance matrix especially when $N > T$ can be challenging. The second approach (and less well-known) uses the so-called covariance graph at its centre which is the corresponding graphical model for 'marginal' dependencies (marginal correlations). Methods of estimating the covariance matrix involve pre-specifying a

[3] The regional science literature has long been aware of the potential problems with the prior specification of the $\boldsymbol{W}$ matrix. For recent contributions see Corrado and Fingleton (2012).

zero-pattern - Chaudhuri, Drton and Richardson (2007), Bayesian priors - Khare and Rajaratnam (2011), thresholding - Butte, Tamayo, Slonim, Golub and Kohane (2000) and shrinkage - Rothman, Bickel, Levina and Zhu (2008, 2010). In general neither approach shows relative superiority over the other.

The multiple testing method developed in Bailey, Pesaran and Smith (2013) belongs to the second line of thought of approximating 'marginal' dependencies. As shown in Section 3, it is practical and simple to implement, and it is invariant to the ordering of the underlying units. Furthermore, it circumvents the challenge of evaluating the theoretical constant, C, arising in the rate of convergence of other thresholding estimators. By using the inverse of the normal distribution at a pre-specified significance level, it avoids the computationally intensive cross validation procedure typically employed for the estimation of C.

2.1 Cross-sectional dependence (CSD) in panels

2.1.1 Spatial dependence - a form of weak CSD

Conventionally, spatial dependence is characterised by use of a predetermined metric such as space or 'economic distance' - Lee and Pesaran (1993), Conley and Dupor (2003), Conley and Topa (2003), Pesaran, Schuermann and Weiner (2004) and the review of spatial econometrics by Anselin (2001). However, often in economic applications these may not be appropriate measures. In some instances trade flows might be relevant, whilst in the case of inter-industry dependencies input-output matrices might provide the appropriate 'spatial' metric - Holly and Petrella (2012). Alternatively, there may be dependencies between geographical areas that reflect cultural similarity, and migration or commuting relationships.[4]

Irrespective of the measure used, spatial dependence relates to spill-over effects that are not pervasive in nature. In other words it conforms to the notion of cross-sectional weak dependence (CWD) as defined in Chudik, Pesaran and Tosetti (2011). To see why, consider as an example the first-order spatial autoregressive, SAR(1), model defined in (2). Assuming that $(\mathbf{I}_N - \psi\mathbf{W})$ is invertible, we have

$$\boldsymbol{x}_{\circ t} = \mathbf{G}\tilde{\mathbf{u}}_{\circ t}, \tag{3}$$

where

$$\mathbf{G} = (\mathbf{I}_N - \psi\mathbf{W})^{-1}\,\boldsymbol{\Sigma}_u^{1/2}.$$

In the spatial literature, $\mathbf{W}$ is assumed to have non-negative elements and is typically row-standardized so that $\|\mathbf{W}\|_\infty = 1$. Under these assumptions, $|\psi| < 1$ ensures that $|\psi|\,\|\mathbf{W}\|_\infty < 1$, and we have

$$\begin{aligned}\|\mathbf{G}\|_\infty &= \left\|\boldsymbol{\Sigma}_u^{1/2}\right\|_\infty \left\|\mathbf{I}_N + \psi\mathbf{W} + \psi^2\mathbf{W}^2 + ...\right\|_\infty \\ &\leq \left\|\boldsymbol{\Sigma}_u^{1/2}\right\|_\infty \left[1 + |\psi|\,\|\mathbf{W}\|_\infty + |\psi|^2\,\|\mathbf{W}\|_\infty^2 + ...\right] = \frac{\max_i(\sigma_{u_i})}{1 - |\psi|\,\|\mathbf{W}\|_\infty} < K < \infty.\end{aligned}$$

[4]Interactions in social networks can also be 'spatial' in an entirely abstract sense. For example Bhattacharjee and Holly (2013) explore interactions among members of a committee using a spatial analogy.

Similarly, $\|\mathbf{G}\|_1 < K < \infty$, if it is further assumed that $|\psi| \|\mathbf{W}\|_1 < 1$. In general, $(\mathbf{I}_N - \psi\mathbf{W})^{-1} \boldsymbol{\Sigma}_u^{1/2}$ has bounded row and column sum matrix norms if $|\psi| < \min(1/\|\mathbf{W}\|_1, 1/\|\mathbf{W}\|_\infty)$. See Chudik and Pesaran (2013) for further details.

Therefore, if the above condition on $|\psi|$ is met the covariance matrix of (3), $\boldsymbol{\Sigma} = \mathbf{G}\mathbf{G}'$, will also be row (column) bounded:

$$\|\boldsymbol{\Sigma}\|_1 = \|\mathbf{G}\mathbf{G}'\|_1 \leq \|\mathbf{G}\|_1 \|\mathbf{G}'\|_1 = \|\mathbf{G}\|_1 \|\mathbf{G}\|_\infty < K < \infty.$$

Similarly, assuming that $var(x_{it}) = \sigma_i^2 > 0$ is bounded away from zero, for the correlation matrix of (3), $\mathbf{R} = \mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}$, where $\mathbf{D} = diag(\sigma_i^2, i = 1, 2, ..., N)$ we have

$$\|\mathbf{R}\|_1 = \left\|\mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}\right\|_1 \leq \frac{1}{\min_i(\sigma_i^2)} \|\boldsymbol{\Sigma}\|_1 < K < \infty. \tag{4}$$

Also, $\lambda_{\max}(\mathbf{R}) \leq \|\mathbf{R}\|_1 < K$, where $\lambda_{\max}(\mathbf{R})$ is the largest eigenvalue of $\mathbf{R}$.

The degree of cross-sectional dependence among the N units can be summarised conveniently by their average cross-correlation (excluding the diagonal elements),

$$\bar{\rho}_N = \frac{\boldsymbol{\tau}'\mathbf{R}\boldsymbol{\tau} - N}{N(N-1)} = \frac{\boldsymbol{\tau}'\mathbf{R}\boldsymbol{\tau}}{N(N-1)} - \frac{1}{N-1}, \tag{5}$$

where $\boldsymbol{\tau}$ is an $N \times 1$ vector of ones. In general, noting that $(\boldsymbol{\tau}'\boldsymbol{\tau})\, \lambda_{\min}(\mathbf{R}) \leq \boldsymbol{\tau}'\mathbf{R}\boldsymbol{\tau} \leq (\boldsymbol{\tau}'\boldsymbol{\tau})\, \lambda_{\max}(\mathbf{R})$, where $\lambda_{\min}(\mathbf{R})$ is the smallest eigenvalue of $\mathbf{R}$, we have

$$\frac{\lambda_{\min}(\mathbf{R})}{(N-1)} \leq \frac{\boldsymbol{\tau}'\mathbf{R}\boldsymbol{\tau}}{N(N-1)} \leq \frac{\lambda_{\max}(\mathbf{R})}{(N-1)},$$

and

$$\frac{\lambda_{\min}(\mathbf{R}) - 1}{(N-1)} \leq \bar{\rho}_N \leq \frac{\lambda_{\max}(\mathbf{R}) - 1}{(N-1)}.$$

Therefore, in the case of weakly cross correlated processes, such as the spatial autoregressive models, where $\lambda_{\max}(\mathbf{R})$ is bounded in N, $\bar{\rho}_N \to 0$, as $N \to \infty$, and standard spatial econometric models cannot deal with cases where $\bar{\rho}_N$ differs from zero even for sufficiently large N.[5] In cases where $\bar{\rho}_N$ tends to a non-zero value, other modelling strategies such as the common factor models with pervasive effects across all units are needed.

2.1.2 The factor model - a form of strong CSD

To this end we draw from the analysis in Pesaran (2013) to test for weak cross-sectional dependence. Suppose that $\boldsymbol{x}_{\circ t}$ are now generated according to the following factor model

$$\boldsymbol{x}_{\circ t} = \boldsymbol{\Gamma}\boldsymbol{f}_t + \boldsymbol{\Omega}^{1/2}\tilde{\boldsymbol{\varepsilon}}_{\circ t}, \tag{6}$$

where $\boldsymbol{f}_t = (f_{1t}, f_{2t}, ..., f_{\ell t})'$ is the $\ell \times 1$ vector of unobserved common factors (ℓ being fixed) with $E(\boldsymbol{f}_t) = \mathbf{0}$, $\boldsymbol{\Sigma}_{ff} = Cov(\boldsymbol{f}_t) = \mathbf{I}_\ell$, and $\boldsymbol{\Gamma}$ is the $N \times \ell$ matrix of the factor loadings

[5] In cases where the degree of cross-sectional dependence is relatively high, one would expect $\lambda_{\max}(\mathbf{R})$ associated with the correlation matrix of the spatial model to be very large when $\mathbf{W}$ is row-standardized and ψ is close to unity. Using simulations we can confirm that in such cases $\lambda_{\max}(\mathbf{R})$ rises with N but at a slower rate, such that $\alpha_N = \ln(\lambda_{\max}(\mathbf{R}))/\ln(N)$ tends to a value which is below 1/2. Also, as to be expected, for each N, α_N rises with $|\psi|$ - see Bailey, Kapetanios and Pesaran (2013) for details regarding the specification of α_N.

$\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{i\ell})'$, for $i = 1, ..., N$, $\tilde{\boldsymbol{\varepsilon}}_{\circ t} = (\tilde{\varepsilon}_{1t}, ..., \tilde{\varepsilon}_{Nt})'$ are idiosyncratic errors that are cross-sectionally and serially independent, namely $\tilde{\varepsilon}_{it} \sim IID(0, 1)$, $i = 1, ..., N$. The error variance components are collected in $\boldsymbol{\Omega} = Diag\left(\omega_i^2,\ i = 1, ..., N\right)$ so that $\varepsilon_{it} = \omega_i \tilde{\varepsilon}_{it}$ is then distributed as $\varepsilon_{it} \sim IID(0, \omega_i^2)$, $i = 1, ..., N$. As before, the degree of cross-sectional dependence of $\boldsymbol{x}_{\circ t}$ is governed by the largest eigenvalue of the correlation matrix, $\mathbf{R}$, which bounds the rate at which the average pair-wise error correlation coefficient, $\bar{\rho}_N$, defined by (5), tends to zero in N.

In the case of the above factor model, $Var(x_{it}) = \sigma_i^2 = \omega_i^2 + \boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i$, $\rho_{ij} = Corr(x_{it}, x_{jt}) = \boldsymbol{\delta}_i'\boldsymbol{\delta}_j$, for $i \neq j$, where

$$\boldsymbol{\delta}_i = \frac{\boldsymbol{\gamma}_i}{\sqrt{1 + \boldsymbol{\gamma}_i\boldsymbol{\gamma}_i'}}, \tag{7}$$

and $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2}, ..., \delta_{i\ell})'$. Then,

$$\bar{\rho}_N = \left(\frac{N}{N-1}\right)\left(\bar{\boldsymbol{\delta}}_N'\bar{\boldsymbol{\delta}}_N - \frac{\sum_{i=1}^N \boldsymbol{\delta}_i'\boldsymbol{\delta}_i}{N^2}\right), \tag{8}$$

where $\bar{\boldsymbol{\delta}}_N = N^{-1}\sum_{i=1}^N \boldsymbol{\delta}_i$.

Consider now the effects of the j^{th} factor, f_{jt}, on the i^{th} unit, x_{it}, as measured by γ_{ij}, and suppose that these factor loadings take non-zero values for M_j out of the N cross-section units. Then, following Bailey, Kapetanios and Pesaran (2013 - BKP), the degree of cross-sectional dependence due to the j^{th} factor can be measured by $\alpha_j = \ln(M_j)/\ln(N)$, and the overall degree of cross-sectional dependence by $\alpha = \max_j(\alpha_j)$. They define α as the exponent of N that gives the maximum number of x_{it} units, $M = \max_j(M_j)$, that are pair-wise correlated. The remaining $N - M$ units will only be partially correlated. BKP refer to α as the exponent of cross-sectional dependence and it can take any value in the range 0 to 1, with 1 indicating the highest degree of cross-sectional dependence. Considering that $\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i = O(\ell)$ where ℓ is fixed as $N \to \infty$, the exponent of cross-sectional dependence of the units can be equivalently defined in terms of the scaled factor loadings, $\boldsymbol{\delta}_i$. Without loss of generality, suppose that only the first M_j elements of δ_{ij} over i are non-zero, and note that[6]

$$\bar{\delta}_{j,N} = \frac{1}{N}\left(\sum_{i=1}^{M_j}\delta_{ij} + \sum_{i=M_j+1}^{N}\delta_{ij}\right) = \frac{M_j}{N}\left(M_j^{-1}\sum_{i=1}^{M_j}\delta_{ij}\right) = N^{\alpha_j - 1}\mu_j = O(N^{\alpha_j - 1}),$$

where $\mu_j = \left(M_j^{-1}\sum_{i=1}^{M_j}\delta_{ij}\right) \neq 0$, for a finite M_j and as $M_j \to \infty$. Similarly, $N^{-2}\sum_{i=1}^N \delta_{ij}^2 = O(N^{\alpha_j - 2})$, and using (8) we have

$$\bar{\rho}_N = O(N^{2\alpha - 2}).$$

The values of α in the range $[0, 1/2)$ correspond to different degrees of cross-sectional weak dependence, as compared to values of α in the range $(1/2, 1]$ that relate to distinct degrees of cross-sectional strong dependence. Under the SAR(1) model specification of (3), $\bar{\rho}_N \to 0$ and $\|\mathbf{R}\|_1 = O\left(1\right)$, indicating that α must fall in the range $[0, 1/2)$, for N sufficiently large.

[6]The main results in Pesaran (2013) and Bailey, Kapetanios and Pesaran (2013) remain valid even if $\sum_{i=M_j+1}^N \delta_{ij} = O(1)$. But for expositional simplicity we maintain the assumption that $\sum_{i=M_j+1}^N \delta_{ij} = 0$.

2.1.3 Identifying the degree of cross-sectional dependence

In many applications, cross-sectional dependence could be due to common factors as well as spatial or network dependence, and it is important that both sources of cross-sectional dependence are taken into account. Mistaking factor dependence, as in (6), for spatial dependence can lead to spurious inference as to the pervasiveness and the degree of the cross-sectional dependence. In consequence, identifying the strength of such dependence is of special significance.

Suppose observations $\boldsymbol{x}_{\circ t} = (x_{1t}, ..., x_{Nt})'$, $t = 1, 2, ..., T$, are available and the aim is to model the cross-dependence between x_{it} and x_{jt} across $i, j = 1, ..., N$, with N and T relatively large. A first step requires one to evaluate the strength of cross-sectional correlation in $\boldsymbol{x}_{\circ t}$. The application of spatial methods should only be considered if the cross-sectional exponent of the observations, α, is sufficiently small, and particularly not close to unity. Regarding temporal dependence, this can be modelled through common factors or unit-specific dynamics using autoregressive distributed lag models or GVAR specifications (Pesaran, Schuermann and Weiner (2004), and Dees, di Mauro, Pesaran and Smith (2007)).

A two-step procedure suggests itself:

1. Apply the cross section dependence (CD) test developed in Pesaran (2013) to $\boldsymbol{x}_{\circ t}$, $t = 1, ..., T$, as shown in (10).

   (a) Only proceed to spatial modelling if the null of weak cross dependence is not rejected.

   (b) If the null of weak dependence is rejected, model the (semi-) strong dependence by use of factor models or cross section averages, and check that the residuals from (6), denoted by $\hat{\boldsymbol{\varepsilon}}_{\circ t} = (\hat{\varepsilon}_{1t}, ..., \hat{\varepsilon}_{Nt})'$, are weakly cross-correlated (by applying the CD test to $\hat{\boldsymbol{\varepsilon}}_{\circ t}$, $t = 1, ..., T$).

2. Apply spatial or network modelling techniques to $\hat{\boldsymbol{\varepsilon}}_{\circ t}$ and/or identify local connections for the spatial weights matrix $\mathbf{W}$.

In order to test for weak or spatial dependence, denote the sample estimates of the pair-wise correlations of (i, j) units of $\boldsymbol{x}_{\circ t}$, $t = 1, ..., T$, by

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)}{\left(\sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2\right)^{1/2} \left(\sum_{t=1}^{T} (x_{jt} - \bar{x}_j)^2\right)^{1/2}}, \tag{9}$$

where $\bar{x}_i = N^{-1} \sum_{i=1}^{N} x_{it}$. The CD statistic is then defined by

$$CD = \left[\frac{TN(N-1)}{2}\right]^{1/2} \widehat{\bar{\rho}}_N, \tag{10}$$

where

$$\widehat{\bar{\rho}}_N = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \hat{\rho}_{ij}. \tag{11}$$

Pesaran (2013) shows that $CD \rightarrow N(0,1)$, under the null hypothesis that the cross-sectional exponent of $\boldsymbol{x}_{\circ t}$, $t = 1, ..., T$, is $\alpha < (2-\epsilon)/4$ as $N \rightarrow \infty$, such that $T = \kappa N^{\epsilon}$, for some $0 \leq \epsilon \leq 1$, and a finite $\kappa > 0$.

If H_0 of weak dependence is rejected for $\boldsymbol{x}_{\circ t}$ in step 1 of the above procedure, then according to BKP the exponent of cross-sectional dependence, α, can be estimated. There are different ways of estimating this exponent if $1/2 < \alpha < 1$. We refer to section 3.1 of Bailey, Kapetanios and Pesaran (2013) for details of their estimation of α. Once step 1 is complete then we can be confident that our data have been stripped sufficiently of common effects and step 2 of the analysis can then begin. In principle, it is also possible to combine the two steps in one meta approach that simultaneously deals with factor and spatial dependence. Such an approach is beyond the scope of the present paper.

3 Correlation-based specification of spatial weights matrices

We now revert back to (3) and focus our attention on the choice of the spatial weights matrix, $\mathbf{W}$. Typically, this is constructed using geodesic, demographic or economic information brought in exogenously, and not contained in the data set under consideration, here $\boldsymbol{x}_{\circ t}$, $t = 1, ..., T$. In economic applications, economic measures, such as commuting times, trade and migratory flows across geographical areas have been used. For example, in GVAR modelling trade weights are used in the construction of link matrices that relate individual economies to their trading partners in the global economy - Pesaran, Schuermann and Weiner (2004). Alternatively geographical contiguity can be used as in Holly, Pesaran and Yamagata (2011a,b). Such measures are often preferable over geodesic measures - since they are closer to the decisions that underlie the observations, x_{it}, and they allow also for possible time variations in the weighting matrix which of course is not possible if we use only physical distance measures in the construction of $\mathbf{W}$.

The use of economic distance, however, might not be possible in practice, and it is desirable to see if $\mathbf{W}$ can be constructed without recourse to such exogenous information. In applications where the time dimension is reasonably large (around 50-80), it is possible to identify the non-zero elements of $\mathbf{W}$ with those elements of $\hat{\rho}_{ij}$, as expressed in (9), that are different from zero at a suitable significance level - Barigozzi and Brownlees (2013).[7] But since there are a large number of such statistical tests, multiple testing procedures that control the overall size of the tests have to be used.

The multiple testing problem arises when we are faced with a number of (possibly) dependent tests and our aim is to control the size of the overall test. Suppose we are interested in a family of null hypotheses, $H_{01}, H_{02}, ..., H_{0m}$ and we are provided with corresponding test statistics, $Z_{1T}, Z_{2T}, ..., Z_{mT}$, with separate rejection rules given by (using a two sided alternative)

$$\Pr\left(|Z_{iT}| > CV_{iT} \,|H_{0i}\right) \leq p_{iT},$$

[7] A related literature addresses the issue of approximating 'marginal' dependencies via thresholding or shrinking the covariance matrix. For contributions to this field see Butte, Tamayo, Slonim, Golub and Kohane (2000), Meinshausen and Buhlmann (2006), Chaudhuri, Drton and Richardson (2007), Peng, Wang, Zhou and Zhu (2009), Rothman, Bickel, Levina and Zhu (2010), Khare and Rajaratnam (2011), and Bien and Tibshirani (2011) among others.

where CV_{iT} is some suitably chosen critical value of the test, and p_{iT} is the observed p value for H_{0i}.

Consider now the family-wise error rate (FWER) defined by

$$FWER_T = \Pr\left[\cup_{i=1}^{m}\left(|Z_{iT}| > CV_{iT}\,|H_{0i}\right)\right],$$

and suppose that we wish to control $FWER_T$ to lie below a pre-determined value, p. Bonferroni (1935, 1936) provides a general solution, which holds for all possible degrees of dependence across the separate tests. By Boole's inequality we have

$$\begin{aligned}\Pr\left[\cup_{i=1}^{m}\left(|Z_{iT}| > CV_{iT}\,|H_{0i}\right)\right] &\leq \sum_{i=1}^{m}\Pr\left(|Z_{iT}| > CV_{iT}\,|H_{0i}\right)\\ &\leq \sum_{i=1}^{m} p_{iT}.\end{aligned}$$

Hence, to achieve $FWER_T \leq p$, it is sufficient to set $p_{iT} \leq p/m$. However, Bonferroni's procedure can be quite conservative, particularly when the tests are highly correlated. This means that the procedure does not reject as often as it should and therefore lacks power. A step-down procedure is proposed by Holm (1979) which is more powerful than Bonferroni's procedure, without imposing any further restrictions on the degree to which the underlying tests depend on each other.

If we abstract from the T subscript and order the p-values of the tests, so that

$$p_{(1)} \leq p_{(2)} \leq .... \leq p_{(m)},$$

are associated with the null hypotheses, $H_{(01)}, H_{(02)}, ..., H_{(0m)}$, respectively, Holm's procedure rejects $H_{(01)}$ if $p_{(1)} \leq p/m$, rejects $H_{(01)}$ and $H_{(02)}$ if $p_{(2)} \leq p/(m-1)$, rejects $H_{(01)}, H_{(02)}$ and $H_{(03)}$ if $p_{(3)} \leq p/(m-2)$, and so on.[8]

In our application, we apply multiple testing procedures to distinct non-diagonal elements of the sample estimate of $\mathbf{R} = (\rho_{ij})$, namely $\hat{\mathbf{R}} = (\hat{\rho}_{ij})$, where $\hat{\rho}_{ij}$ is the correlation of the de-factored price changes between i and j MSAs. Bailey, Pesaran and Smith (2013) show that the Holm approach applied to $\hat{\mathbf{R}}$ provides a regularised version of the matrix that converges to $\mathbf{R}$ at a faster rate and has superior support recovery than when applying Bonferroni as $N \to \infty$. To apply the Holm procedure to $\hat{\mathbf{R}} = (\hat{\rho}_{ij})$, we first observe that under the null i and j are unconnected, and $\hat{\rho}_{ij}$ is approximately distributed as $N\left(0, T^{-1}\right)$. Therefore, the p-values of the individual tests are (approximately) given by $p_{ij} = 2\left[1 - \Phi\left(\sqrt{T}\left|\hat{\rho}_{ij}\right|\right)\right]$ for $i = 1, 2, ..., N-1, j = i+1, ..., N$, with the total number of tests being carried out given by $m = N(N-1)/2$. To apply the Holm procedure we need to order these p-values in an ascending manner, which is equivalent to ordering $\left|\hat{\rho}_{ij}\right|$ in a descending manner. Denote the largest value of $\left|\hat{\rho}_{ij}\right|$ over all $i \neq j$, by $\left|\hat{\rho}_{(1)}\right|$, the second largest value by $\left|\hat{\rho}_{(2)}\right|$, and so on, to obtain the ordered sequence $\left|\hat{\rho}_{(s)}\right|$, for $s = 1, 2, ..., m$. Then the (i, j) pair associated with $\left|\hat{\rho}_{(s)}\right|$ are connected if $\left|\hat{\rho}_{(s)}\right| > \Phi^{-1}\left(1 - \frac{p/2}{m-s+1}\right)$, otherwise disconnected, for $s = 1, 2, ..., m$, where p is the pre-specified overall size of the test (which we set to 5% in the empirical application), and $\Phi^{-1}(.)$ is the inverse of the

[8] Other multiple testing procedures can also be considered and Efron (2010) provides a recent review. But most of these procedures tend to place undue prior restrictions on the dependence of the underlying test statistics while the Holm method is not subject to this problem.

standard normal distribution function. The resultant connection matrix will be denoted by $\hat{\mathbf{W}} = (\hat{w}_{ij})$, where $\hat{w}_{ij} = 1$ if the (i, j) pair are connected according to the Holm procedure, otherwise $\hat{w}_{ij} = 0$. Connections can also be classified as positive ($\hat{w}_{ij}^{+}$) if $\hat{\rho}_{ij} > 0$, and negative ($\hat{w}_{ij}^{-}$) if $\hat{\rho}_{ij} < 0$.

4 Application: US house prices

The two stage procedure developed in this paper can be applied to different types of panel data sets so long as the time series dimension of the panel is reasonably large such that reliable estimates of pair-wise correlations, ρ_{ij}, can be obtained. There are many such panels, covering regions or countries, that can be considered. Regional data in the United States have been studied by many including Cromwell (1992), Pollakowski and Ray (1997), Carlino and DeFina (1998, 2004), Carlino and Sill (2001), Del Negro (2002), Owyang, Piger and Wall (2005), and Partridge and Rickman (2005). The cross country data sets used in global modelling provide another example. Here we opt to study house price changes at the level of Metropolitan Statistical Areas (MSAs) in the US, where we have quarterly time series data on 363 MSAs over the period 1975Q1-2010Q4 ($T = 144$). There already exists a large literature on the spatial dimension of house price changes, partly because of the availability of spatially disaggregated data, but also because of the role that housing plays in household wealth and in the transmission of monetary policy shocks, and more recently as a conduit for transmission of global shocks. Recent contributions include Rapach and Strauss (2007, 2009), Kadiyala and Bhattacharya (2009), Gupta and Das (2010), Gupta, Kabundi and Miller (2011a,b), Kuethe and Pede (2011) and Gupta and Miller (2012). In these a number of models are considered ranging from ARDL, STAR, BVAR, FAVAR, FABVAR, Bayesian shrinkage LBVAR and the so-called SpVAR specification which are applied to house prices data directly without prior assessment of their degree of cross-sectional dependence which is required when implementing these model specifications. Moreover, house price shocks spill over into adjacent geographical areas and tend to ripple across the economy. See, for example, Meen (1999), and Holly, Pesaran and Yamagata (2011a,b).

MSAs are large urban concentrations.[9] They range in size, as measured by population in 2008, from the smallest - Carson City - with a population of 55,000, to New York and its environs with a population of 18.97 million. Moreover, there can be considerable distances between MSAs. The pair-wise average distance is 1,156 miles, though of course this is exaggerated by the relative sparseness of the distribution of MSAs in the Midwest. Indeed, by comparison, the study of regional house prices in the UK by Holly, Pesaran and Yamagata (2011b) deals with distances of a much smaller magnitude. Distance is, therefore, likely to be an important factor for the spatial distribution of house prices, though size could play a role as well.

Our choice of house prices is also motivated by the role that housing plays in spatial equilibrium models (Glaeser, Gyourko and Saiz (2008), Glaeser and Gottlieb (2009)). The standard approach in urban and regional economics is to assume a spatial equi-

[9]Metropolitan Statistical Areas are geographic entities delineated by the Office of Management and Budget and are used by Federal statistical agencies when collecting, tabulating, and publishing Federal statistics for spatial units in the USA. The MSAs are defined by a core area with a large population concentration, together with adjacent areas that have a high degree of economic and social integration with that core through commuting and transport links.

librium. At the margin firms and households have to be indifferent between different locations. Firms employ labour up to the point at which the wage is equal to marginal product; construction companies supply housing up to the point at which marginal cost is equal to marginal product. Finally, households have to be indifferent about where they are located, taking into account wages, the price of houses and the local availability of amenities (proximity to schools, sea, mountains, temperature, etc.). The combination of the labour supply curve, the supply curve for housing and the labour demand determines simultaneously the population of say a locality, wages and the price of housing. Idiosyncratic differences in space in terms of productivity, particular characteristics of an area and the construction sector determine differences across space in population density, household incomes and the price of houses. There are a number of equilibrating processes at work. Households tend to move across geographical areas in response to differences in wages, house prices and area characteristics. There can also be agglomeration effects due to economies of scale in relation to size and population density of cities. But, it should be clear that such equilibrating tendencies are likely to operate fully only in the long run, over a number of years rather than quarters. It takes time for households to relocate in response to changing economic circumstances. It also takes time (time to build) for construction companies to increase the supply of housing. Any model of house price diffusion across MSAs must also adequately deal with dynamics, both within and across MSAs.

4.1 Spatial weights matrices based on distance

We start our analysis with a standard specification of $\mathbf{W}$ based on contiguity measures, which we also use as a benchmark to examine the estimates of $\mathbf{W}$ that are based on pair-wise correlations, $\hat{\rho}_{ij}$. As noted above, MSAs are deliberately defined as one or more large cities with their core having a substantial influence over the surrounding region, but not distance MSAs, due to commuting or other travel costs. Hence, it can be argued that geographical distance can play an important role in the determination of connections across different MSAs. As noted earlier, we consider 363 MSAs in total, excluding three MSAs located in Alaska and Hawaii.[10] We denote a weights matrix based on physical distance by $\mathbf{W}_d$, and make use of data for geodesic distance (d) by applying the Haversine formula to data on the Latitude-Longitude of zip codes, cross referenced to each of the $N = 363$ MSAs.[11] We regard $\mathbf{W}_d$ (of dimension $N \times N$) to be symmetric. We identify as neighbours for each MSA, i ($i = 1, ..., N$), all MSAs that lie within a radius of d miles. This pattern translates into a value of 1 for elements (i, j) and (j, i) of $\mathbf{W}_d$ if MSA i is a neighbour (falling within the given radius) of MSA j, or a value of 0 otherwise. Diagonal entries (i, i) take a value of 0, indicating that MSA i cannot be a neighbour of itself. Also under this specification all non-zero elements of $\mathbf{W}_d$ are viewed as representing a positive connection, which should be contrasted with connections that are based on economic factors that could lead to negative as well as positive connections.

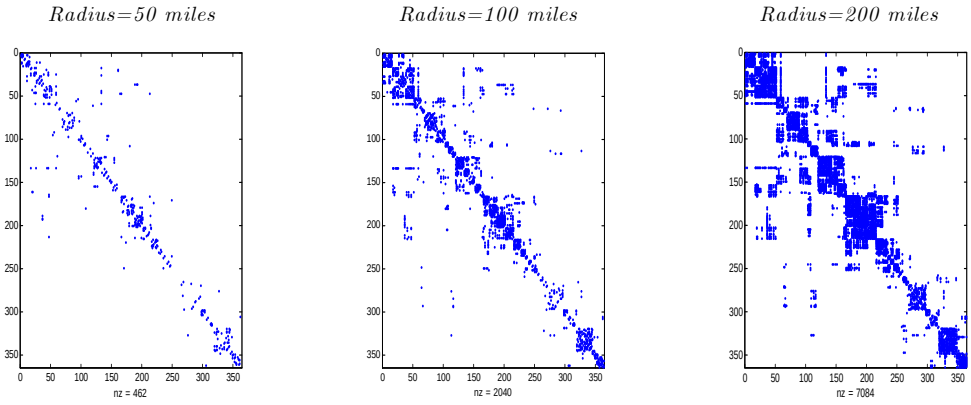
We study three cases: (i) MSAs within a radius of $d = 50$ miles, (ii) MSAs within a radius of $d = 100$ miles, and (iii) MSAs within a radius of $d = 200$ miles. These give rise to three $\mathbf{W}_d$ matrices, namely (i) $\mathbf{W}_{50m}$, (ii) $\mathbf{W}_{100m}$, and (iii) $\mathbf{W}_{200m}$, which are sparse by nature, but of different degree depending on the cut-off point set by the radius,

[10] Note that the District of Columbia is treated as a single MSA.

[11] See Appendix I for details of this formula.

d. We compare the degree of sparseness of $\mathbf{W}_{50m}$, $\mathbf{W}_{100m}$, and $\mathbf{W}_{200m}$ in terms of their percentage of non-zero elements (excluding the diagonal elements). This percentage is 0.35% for $\mathbf{W}_{50m}$, 1.55% for $\mathbf{W}_{100m}$, and 5.39% for $\mathbf{W}_{200m}$. As expected, the number of non-zero elements increases when the radius within which MSAs are considered to be neighbours rises. Figure 1 displays all three $\mathbf{W}_d$ matrices. In this figure we have ordered the MSAs by States starting at the East Coast and moving towards the West Coast, following the list provided in Table A of Appendix II, from top to bottom and from left to right. The sparseness of the $\mathbf{W}_d$ matrices is captured by white areas in the graph when the relevant entries are equal to zero. As to be expected there is considerable clustering along the diagonal, but because we are using a line to depict a plane, sometimes an MSA may lie at the edge of a State (or region) and fall within the radius of another State or region. Clearly, as the radius is increased from 50 to 200 miles the degree of leaching increases.

Figure 1: Spatial weights matrices specified by distance



4.2 Spatial weights matrices formed from de-factored house price correlations

As an alternative to $\mathbf{W}_d$, we now consider the problem of estimating $\mathbf{W}$ using pair-wise correlations of house price changes across MSAs. We consider the same 363 MSAs, over the period 1975Q1 to 2010Q4. We denote the level of house prices in MSA i, located in State s, in quarter t, by P_{ist}, for $i = 1, ..., N_s$, $s = 1, ..., S$, and $t = 1, ..., T$, where $\sum_{s=1}^{S} N_s = N = 363$, $S = 49$ (comprised of 48 contiguous States and the district of Columbia), and $T = 144$ quarters. Then we compute real house prices as:

$$p_{ist} = \ln\left(\frac{P_{ist}}{CPI_{st}}\right), \text{ for } i = 1, 2, ..., N_s; \; s = 1, ..., S; \; t = 1, 2, ..., T,$$

where CPI_{st} is the Consumer Price Index of State s in quarter t. Details on the sources of these data can be found in Appendix III. We have ordered the MSAs by State, as described in Section 4.1. Finally, we obtain seasonally adjusted changes in real house prices, π_{ist}, as residuals from regressing $p_{ist} - p_{is,t-1}$, the seasonally unadjusted rate of change in real house prices, on an intercept and three quarterly seasonal dummies. Before modelling the spatial dimension of the price changes, π_{ist}, we first need to examine the extent to which these price changes are cross sectionally strongly correlated and then de-factor such effects if necessary. (Section 2.1.3)

4.2.1 Stage 1(a,b) : Cross sectional dependence in house price changes and de-factoring of observations

To examine the degree of cross section dependence in house price changes, we computed the CD statistic of Pesaran (2013) for the price series, π_{ist}, without any de-factoring. (See (10) and (11)). We obtained $CD_\pi = 640.46$ as compared to a critical value of 1.96 at the 5% significance level. The test is clearly highly significant and suggests a very high degree of cross-sectional dependence in house price changes, which could be due to common national and regional effects. Applying the method proposed by BKP we calculate the exponent of cross sectional dependence (standard error in parenthesis) for the house price changes and obtain $\mathring{\alpha}_\pi = 0.989\ (0.03)$. This suggests the existence of cross-sectional strong or semi-strong dependence in real house price changes across MSAs. Therefore, it would be inappropriate to apply standard spatial modelling techniques directly to π_{ist}, as suggested in step 1a of the two-stage procedure set out in Section 2.1.3.

The strong cross-sectional dependence in house price changes can be modelled using observed (national/regional income, unemployment and interest rates), or unobserved common factors (using principal components). Alternatively, as argued in Pesaran (2006), we can use cross-sectional averages at the national and regional level as in (13).[12] We identify a total of $R = 8$ regions in the US containing an average of approximately 45 MSAs each. These are: (i) New England, (ii) Mid East, (iii) South East, (iv) Great Lakes, (v) Plains, (vi) South West, (vii) Rocky Mountains, and (viii) Far West (see Table A of Appendix II for more details). Accordingly, let π_{irt} denote the rate of change of real house prices (after seasonal adjustments) in the i^{th} MSA located in region $r = 1, 2, ..., R$, at time t, and consider the following hierarchical factor model

$$\begin{aligned} \pi_{irt} &= a_{ir} + \beta_{ir}\bar{\pi}_{rt} + \gamma_{ir}\bar{\pi}_t + \xi_{irt}, \\ i &= 1, 2, ..., N_r; r = 1, 2, ..., R;\ t = 2, ..., T, \end{aligned} \tag{12}$$

where , $\bar{\pi}_{rt} = N_r^{-1}\sum_{i=1}^{N_r}\pi_{irt}$, and $\bar{\pi}_t = N^{-1}\sum_{r=1}^{R}\sum_{i=1}^{N_r}\pi_{irt}$, with $N = \sum_{r=1}^{R} N_r$. Write the above model more compactly as

$$\boldsymbol{\pi}_t = \mathbf{a} + \mathbf{B}\mathbf{Q}_N\boldsymbol{\pi}_t + \boldsymbol{\Gamma}\mathbf{P}_N\boldsymbol{\pi}_t + \boldsymbol{\xi}_t, \tag{13}$$

where $\boldsymbol{\pi}_t$ is an $N \times 1$ vector of house price changes partitioned by regions, namely

$$\boldsymbol{\pi}_t = (\pi_{11t}, \pi_{21t}, ...\pi_{N_1 1t}; \pi_{12t}, \pi_{22t}, ...\pi_{N_2 2t}; .....; \pi_{1Rt}, \pi_{2Rt}, ..., \pi_{N_R Rt})'.$$

Similarly

$$\mathbf{a} = (a_{11}, a_{21}, ...a_{N_1 1}; a_{12}, a_{22}, ...a_{N_2 2}; .....; a_{1R}, a_{2R}, ..., a_{N_R R})'.$$

$\mathbf{B}$ and $\boldsymbol{\Gamma}$ are $N \times N$ diagonal matrices with their ordered elements given by

$$\beta_{11}, \beta_{21}, ...\beta_{N_1 1}; \beta_{12}, \beta_{22}, ...\beta_{N_2 2}; .....; \beta_{1R}, \beta_{2R}, ..., \beta_{N_R R},$$

and

$$\gamma_{11}, \gamma_{21}, ...\gamma_{N_1 1}; \gamma_{12}, \gamma_{22}, ...\gamma_{N_2 2}; .....; \gamma_{1R}, \gamma_{2R}, ..., \gamma_{N_R R},$$

respectively. Finally, $\mathbf{Q}_N$ and $\mathbf{P}_N$ are $N \times N$ projection matrices such that $\mathbf{Q}_N\boldsymbol{\pi}_t$ give the regional means and $\mathbf{P}_N\boldsymbol{\pi}_t$ the national mean of the local feature. More specifically, let $\boldsymbol{\tau}_{N_r}$ be an $N_r \times 1$ vector of ones, and $\boldsymbol{\tau}_N$ an $N \times 1$ vector of ones, then

$$\mathbf{P}_N = \boldsymbol{\tau}_N(\boldsymbol{\tau}'_N\boldsymbol{\tau}_N)^{-1}\boldsymbol{\tau}'_N,$$

[12] We also considered using State level averages, but there were only a few MSAs in some States.

and

$$\mathbf{Q}_N = \begin{pmatrix} \mathbf{P}_{N_1} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{N_2} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}_{N_{R-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{P}_{N_R} \end{pmatrix},$$

where $\mathbf{P}_{N_r} = \boldsymbol{\tau}_{N_r}(\boldsymbol{\tau}'_{N_r}\boldsymbol{\tau}_{N_r})^{-1}\boldsymbol{\tau}'_{N_r}$. It is assumed that R is fixed, and for each r, N_r/N tends to a non-zero constant as $N \to \infty$. $\mathbf{P}_{N_r}\boldsymbol{\pi}_t$, for $r = 1, 2, ..., R$, and $\mathbf{P}_N\boldsymbol{\pi}_t$ can be viewed as regional and national factors that are consistently estimated by simple averages. They also represent the strong form of cross-sectional dependence in the real house price changes across MSAs.

The de-factored real house price changes are then given by residuals from (13), namely

$$\hat{\boldsymbol{\xi}}_t = \boldsymbol{\pi}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{Q}_N\boldsymbol{\pi}_t - \hat{\boldsymbol{\Gamma}}\mathbf{P}_N\boldsymbol{\pi}_t, \ t = 2, ..., T. \tag{14}$$

Then, in accordance with step 1b of the two-stage procedure of Section 2.1.3, we apply the CD test of Pesaran (2013) to the vector of de-factored price changes, $\hat{\boldsymbol{\xi}}_t$. The resulting CD statistic is much reduced, falling from 640.45 to -6.05, showing that the simple hierarchical de-factoring procedure has managed to eliminate almost all of the strong cross-sectional dependence that had existed in the house price changes, and what remains could be due to the local dependencies that need to be modelled using spatial techniques. Also, the estimate of the exponent of cross sectional dependence, α, which stood at $\mathring{\alpha}_\pi = 0.989$ (0.03) is now reduced to $\mathring{\alpha}_\xi = 0.637$ (0.03) for the de-factored price changes which is close to the borderline between strong and weak dependence of 1/2.

For comparison we repeat the de-factoring analysis by applying the method of principal components developed for large panels in Bai (2003) to price changes. This entails running the following regressions

$$\pi_{it} = a_i + \boldsymbol{\gamma}'_i\hat{\mathbf{f}}_t + \xi_{it}, \ i = 1, 2, ..., N; \ t = 2, ..., T, \tag{15}$$

where $\hat{\mathbf{f}}_t$ is an $\ell \times 1$ vector of principal components (PC) of house price changes (without grouping by regions or States), and $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{i\ell})'$ is the associated vector of factor loadings. To select the number of PCs we applied the six information criteria proposed in Bai and Ng (2002), specifying 8 as the maximum number of factors to match the number of cross-sectional averages used in the hierarchical factor model. All six IC ended up selecting 8 as the optimal number of factors. We increased the maximum number of factors in the procedure, but still ended up selecting the maximum as the optimal. In view of the failure of the IC to lead to any meaningful outcome we decided to conduct the de-factoring analysis using $\ell = 2, 3, ..., 8$ principal components. We then computed CD statistics for the de-factored residuals for all 7 choices of ℓ, and obtained the values of 53.39, 10.21, 2.73, 3.27, 2.31, -1.96 and -4.42 respectively, for $\ell = 2, 3, ..., 8$. The corresponding exponents of cross-sectional dependence are $\mathring{\alpha}_{2pc} = 0.932$ (0.04), $\mathring{\alpha}_{3pc} = 0.799$ (0.04), $\mathring{\alpha}_{4pc} = 0.793$ (0.03), $\mathring{\alpha}_{5pc} = 0.785$ (0.03), $\mathring{\alpha}_{6pc} = 0.831$ (0.02), $\mathring{\alpha}_{7pc} = 0.718$ (0.02) and $\mathring{\alpha}_{8pc} = 0.622$ (0.02), respectively. In view of these results, and to strike a balance between purging the house price changes from common effects and still leaving a sufficient degree of spatial dependence in the de-factored observations we decided to opt for the mid value of ℓ at 4.

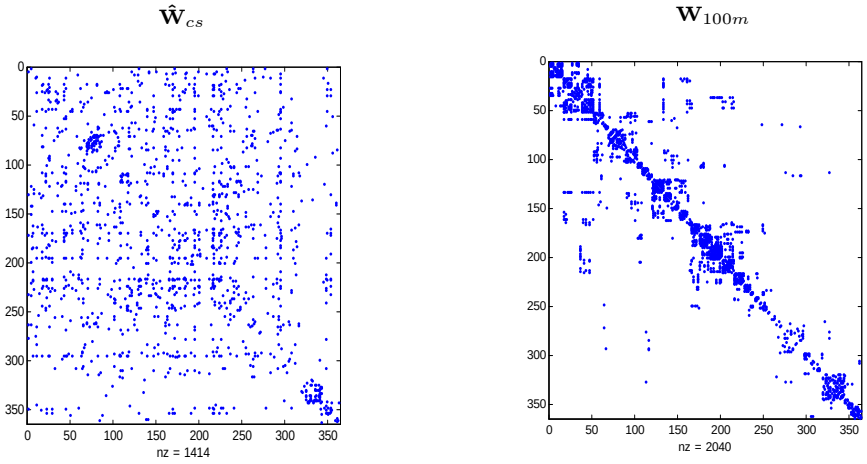
It is clear that both methods (cross-sectional averages or PCs with $\ell = 4$) do reasonably well in purging the price changes from the common effects, although the use of regional and national averages have a clearer economic interpretation as factors than do the statistically generated principal components. We conclude that once de-factoring has been implemented using either (14) or (15), the resulting residuals are weakly enough cross-correlated to consider them amenable to spatial modelling using existing econometric methods or a more elaborate specification, as shown in (18).

4.2.2 Step 2: Estimation of spatial connections

Having computed de-factored price changes, $\hat{\xi}_{it}$, and $\hat{\xi}_{it,PC_4}$, we are now in a position to apply the methodology developed in Section 3 to estimate the matrix of connections using pair-wise correlations of $\hat{\xi}_{it}$ (or those of $\hat{\xi}_{it,PC_4}$). To do this, first we obtain the sample correlation matrix of $\hat{\boldsymbol{\xi}}_t = (\hat{\xi}_{it})$, $\hat{\mathbf{R}}_\xi = (\hat{\rho}_{\hat{\xi},ij})$ from the residuals of regression (13), where $\hat{\rho}_{\hat{\xi},ij} = \hat{\rho}_{\hat{\xi},ji} = \hat{\sigma}_{\hat{\xi},ij}/\sqrt{\hat{\sigma}_{\hat{\xi},ii}\hat{\sigma}_{\hat{\xi},jj}}$, and $\hat{\sigma}_{\hat{\xi},ij} = T^{-1}\sum_{t=1}^{T}\hat{\xi}_{it}\hat{\xi}_{jt}$. Next, we apply Holm's multiple testing to the $N(N-1)/2$ pair-wise correlation coefficients, $\hat{\rho}_{\hat{\xi},ij}$, for $i = 1, 2, ..., N-1, j = i+1, ..., N$, as described in Section 3. We denote the resultant connection matrix by $\hat{\mathbf{W}}_{cs} = \left(\hat{w}_{cs_{ij}}\right)$. Here cs stands for multiple testing applied to residuals extracted from de-factoring using the cross-sectional averages approach.

As in Section 4.1, measuring the degree of sparseness of $\hat{\mathbf{W}}_{cs}$ by the percentage of its non-zero elements we obtain the figure of 1.08% which is comparable to the 1.55% we obtained for $\boldsymbol{W}_{100m}$, although as can be see from Figure 2 the pattern of sparseness of the two matrices, $\mathbf{W}_{100m}$ and $\hat{\mathbf{W}}_{cs}$, are quite different. In fact it is best to view the non-zero elements of $\hat{\mathbf{W}}_{cs}$ as connections rather than as neighbours (in a physical sense). According to $\hat{\mathbf{W}}_{cs}$, the connections extend well beyond geographical boundaries, though distinct clusters are evident especially in the West Coast and parts of the East Coast regions.[13] Divisions of the connections into the East, the Middle and West of the country are also visible.

Figure 2: Spatial weights matrix using multiple testing and $\mathbf{W}_{100m}$
Residuals from defactoring using cross-sectional averages



[13] Recall that in these graphs MSAs are ordered by State, moving from East to West, from top to bottom and left to right.

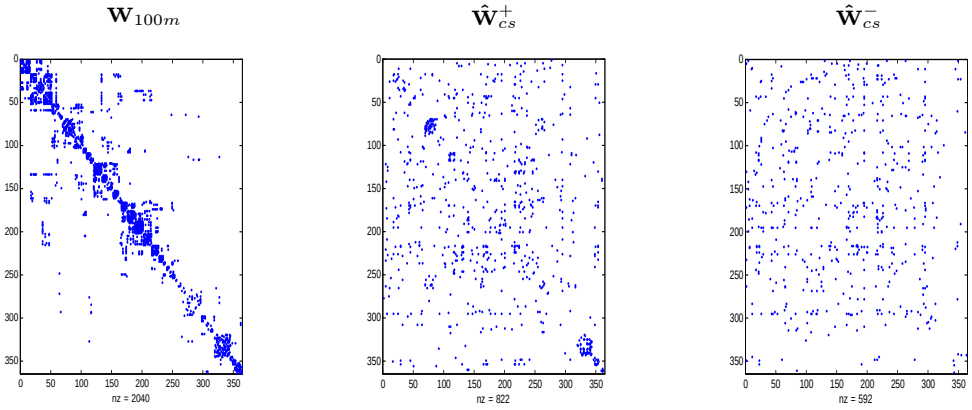
4.2.3 Positive and negative connections

Unlike the distance based $\mathbf{W}_d$ weights matrices, the connections identified by the correlation based approach can be readily distinguished into positive and negative ones. This can be done by associating positive connections with statistically significant evidence of a positive correlation, and the negative connections with the evidence of statistically pairwise negative correlations. Accordingly, we can now define the positively and negatively connected weights matrices, $\hat{\mathbf{W}}_{cs}^{+} = (\ \hat{w}_{cs_{ij}}^{+})$ and $\hat{\mathbf{W}}_{cs}^{-} = \left(\hat{w}_{cs_{ij}}^{-}\right)$, respectively, by

$$\hat{w}_{cs_{ij}}^{+} = \hat{w}_{cs_{ij}} I\left(\hat{\rho}_{\hat{\xi},ij} > 0\right), \text{ and } \hat{w}_{cs_{ij}}^{-} = \hat{w}_{cs_{ij}} I\left(\hat{\rho}_{\hat{\xi},ij} < 0\right).$$

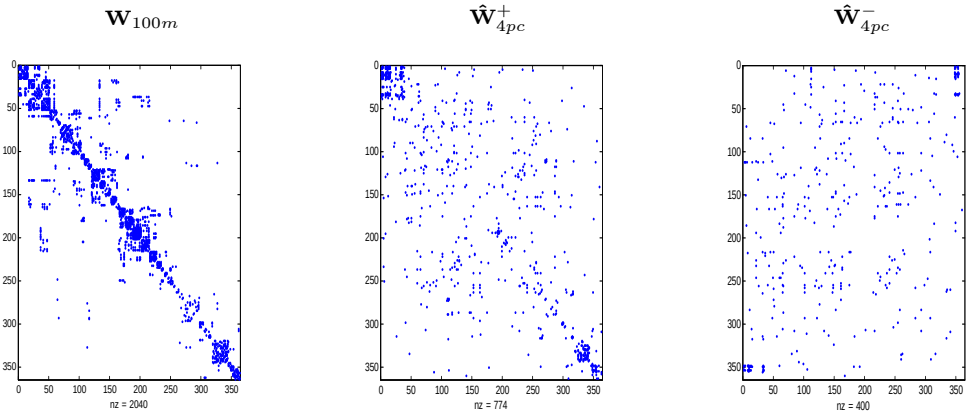
We note that $\hat{\mathbf{W}} = \hat{\mathbf{W}}_{cs}^{+} + \hat{\mathbf{W}}_{cs}^{-}$. Comparing these with the distance based weights matrix, $\mathbf{W}_{100m}$, in Figure 3, at first glance we notice that $\hat{\mathbf{W}}_{cs}^{+}$ is more closely related to $\mathbf{W}_{100m}$ than is $\hat{\mathbf{W}}_{cs}^{-}$. Further, it is evident that geographical proximity is not the only factor driving spatial connections between MSAs. There are significant correlations (positive or negative) well away from the diagonal, with a number of clusters suggesting connections at considerable distances.

Figure 3: Spatial weights matrices - distance and correlation-based connections (CS)



We also applied the multiple testing procedure to the de-factored house price changes using the PCs, as in (15), and obtained the weights matrices $\hat{\mathbf{W}}_{jpc}$, corresponding to $2, 3$, and 4, PCs, respectively. The degree of sparseness of these matrices, as measured by the percentage of their non-zero elements, at 1.64%, 1.15% and 0.89%, respectively, tend to rise with the number of PCs used in the de-factoring process, which is to be expected. Comparing $\hat{\mathbf{W}}_{jpc}$, $j = 2, 3, 4$, with the distance-based matrices $\mathbf{W}_{50m}, \mathbf{W}_{100m}$, and $\mathbf{W}_{200m}$, it appears that we need at least 3 PCs to match the degree of sparseness of $\mathbf{W}_{100m}$. We also constructed $\hat{\mathbf{W}}_{jpc}^{+}$ and $\hat{\mathbf{W}}_{jpc}^{-}$, $j = 2, 3, 4$ in line with the procedure described earlier. $\hat{\mathbf{W}}_{4pc}^{+}$, $\hat{\mathbf{W}}_{4pc}^{-}$ and $\mathbf{W}_{100m}$ are plotted in Figure 4 below. As with the cross-sectional averages approach, the PC method suggests that positively correlated connections match more closely the distance-based connections than do the negatively correlated connections.

Figure 4: Spatial weights matrices - distance and correlation-based connections (4PCs)



4.2.4 Statistical associations of different connection weights matrices

We assess the closeness of the correlation-based estimates, $\hat{\mathbf{W}}^{+}$ and $\hat{\mathbf{W}}^{-}$ (using either *cs* or *pc* regressions for de-factoring) with the distance-based weight matrix, $\mathbf{W}_d$, more formally by quantifying the statistical association of the two types of weights matrices. The analysis is complicated by the fact that these matrices are by nature sparse, and hence the probability of a zero realisation in both adjacency matrices $\hat{\mathbf{W}}^{+}$ (or $\hat{\mathbf{W}}^{-}$) and $\mathbf{W}_d$ is higher than obtaining a common entry of 1.[14] Given the symmetry of the weights matrices in our application, we focus on the upper triangular elements. We create contingency tables from these upper-triangular elements of the form

$$\begin{bmatrix} n_{11} & n_{10} \\ n_{01} & n_{00} \end{bmatrix},$$

where:

- n_{11} equals the number of times $\hat{\mathbf{W}}^{+}$ (or $\hat{\mathbf{W}}^{-}$) displays entry of 1 when $\mathbf{W}_d$ displays 1.
- n_{00} equals the number of times $\hat{\mathbf{W}}^{+}$ (or $\hat{\mathbf{W}}^{-}$) displays entry of 0 when $\mathbf{W}_d$ displays 0.
- n_{01} equals the number of times $\hat{\mathbf{W}}^{+}$ (or $\hat{\mathbf{W}}^{-}$) displays entry of 0 when $\mathbf{W}_d$ displays 1.
- n_{10} equals the number of times $\hat{\mathbf{W}}^{+}$ (or $\hat{\mathbf{W}}^{-}$) displays entry of 1 when $\mathbf{W}_d$ displays 0.

Then, $n_{11} + n_{00} + n_{01} + n_{10} = N(N-1)/2$, and Pearson's chi-squared statistic - Pearson (1900) - is given by

$$\chi^2 = \frac{1}{2} N(N-1) \left[\sum_{i,j=0}^{1} \frac{n_{ij}^2}{(n_{i.} + n_{.j})} - 1 \right].$$

[14] For more details regarding testing the dependence among multicategory variables see Pesaran and Timmermann (2009).

We set the significance level at 5%. We compare $\hat{\mathbf{W}}^{+}$ (or $\hat{\mathbf{W}}^{-}$) with the three versions of $\mathbf{W}_d$, namely $\mathbf{W}_{50m}, \mathbf{W}_{100m}$, and $\mathbf{W}_{200m}$. For brevity of exposition we present the contingency tables for $\hat{\mathbf{W}}^{+}_{cs}$ and $\hat{\mathbf{W}}^{-}_{cs}$ versus $\mathbf{W}_{100}$ only:

Table 1: Contingency tables - $\hat{\mathbf{W}}^{+}_{cs}$ and $\hat{\mathbf{W}}^{-}_{cs}$ versus $\mathbf{W}_{100m}$ spatial weights matrices

		$\mathbf{W}_{100m}$		
		1	**0**	$\sum_{rows}$
$\hat{\mathbf{W}}^{+}_{cs}$	**1**	54	357	**411**
	0	966	64326	**65292**
	$\sum_{cols}$	**1020**	**64683**	**65703**

		$\mathbf{W}_{100m}$		
		1	**0**	$\sum_{rows}$
$\hat{\mathbf{W}}^{-}_{cs}$	**1**	8	288	**296**
	0	1012	64395	**65407**
	$\sum_{cols}$	**1020**	**64683**	**65703**

It is clear that $\hat{\mathbf{W}}^{+}_{cs}$ has more elements in common with $\mathbf{W}_d$ than does $\hat{\mathbf{W}}^{-}_{cs}$. The $\chi^2_{5\%}$ statistics for $\hat{\mathbf{W}}^{+}_{cs}$ and $\hat{\mathbf{W}}^{-}_{cs}$ versus $\mathbf{W}_{50m}, \mathbf{W}_{100m}$, and $\mathbf{W}_{200m}$ respectively are shown in Table 2 below (to be compared with a critical value of 3.84):

Table 2: Pearson's $\chi^2_{5\%}$ test statistics
$\mathbf{W}^{+}_{cs}$ and $\mathbf{W}^{-}_{cs}$ versus $\mathbf{W}_d$, $d = 50, 100, 200m$

	$\mathbf{W}_{50m}$	$\mathbf{W}_{100m}$	$\mathbf{W}_{200m}$
$\hat{\mathbf{W}}^{+}_{cs}$	267.24	**363.27**	298.41
$\hat{\mathbf{W}}^{-}_{cs}$	0.89	2.57	4.30

The chi squared test statistics are highly significant especially when $\hat{\mathbf{W}}^{+}_{cs}$ is considered. Elements of $\hat{\mathbf{W}}^{+}_{cs}$ are much more closely associated with the spatial weights, $\mathbf{W}_d$, than the elements of $\hat{\mathbf{W}}^{-}_{cs}$. The association between $\hat{\mathbf{W}}^{+}_{cs}$ and $\mathbf{W}_d$ is the largest when $d = 100$.

Finally, we repeat these comparisons with weights based on PC de-factored price changes, and obtain similar results when the number of PCs is set to 4. See Tables 3 and 4 where $\hat{\mathbf{W}}^{+}_{4pc}$ and $\hat{\mathbf{W}}^{-}_{4pc}$ are compared with $\mathbf{W}_d$, $\hat{\mathbf{W}}^{+}_{cs}$ and $\hat{\mathbf{W}}^{-}_{cs}$. The association between $\hat{\mathbf{W}}^{+}_{cs}$ and $\hat{\mathbf{W}}^{+}_{4pc}$ is particularly high, and gives a chi-squared statistic of 7573.2 (compared with a critical value of 3.84).

Table 3: Pearson's $\chi^2_{5\%}$ test statistics
$\mathbf{W}^{+}_{pc}$ and $\mathbf{W}^{-}_{pc}$ versus $\mathbf{W}_d$, $d = 50, 100, 200m$

	$\mathbf{W}_{50m}$	$\mathbf{W}_{100m}$	$\mathbf{W}_{200m}$
$\hat{\mathbf{W}}^{+}_{pc}$	608.53	**1034.51**	876.42
$\hat{\mathbf{W}}^{-}_{pc}$	0.71	0.40	4.52

Table 4: Contingency tables - $\mathbf{W}_{cs}^{+/-}$ versus $\mathbf{W}_{pc}^{+/-}$ spatial weights matrices

		$\mathbf{W}_{pc}^{+}$		
		1	**0**	$\sum_{rows}$
$\hat{\mathbf{W}}_{cs}^{+}$	**1**	137	250	**387**
	0	274	65042	**65316**
	$\sum_{cols}$	**411**	**65292**	**65703**

		$\mathbf{W}_{pc}^{-}$		
		1	**0**	$\sum_{rows}$
$\hat{\mathbf{W}}_{cs}^{-}$	**1**	54	146	**200**
	0	242	65261	**65503**
	$\sum_{cols}$	**296**	**65407**	**65703**

4.3 A heterogeneous spatiotemporal model of US house price changes

We are now in a position to illustrate the utility of separate identification of positive and negative connections for the spatial analysis of house price changes. The de-factored house price changes, $\hat{\xi}_{it}$, can be modelled using the following spatiotemporal model

$$\hat{\xi}_{it} = a_{i\xi} + \sum_{j=1}^{h_{\lambda i}} \lambda_{ij} \hat{\xi}_{i,t-j} + \sum_{j=0}^{h_{\xi i}} \psi_{ij} \hat{\xi}^{*}_{i,t-j} + \zeta_{it}, \text{for } i = 1, 2, ..., N, \ t = 2, ..., T, \tag{16}$$

where

$$\begin{aligned} \hat{\xi}^{*}_{it} &= \frac{\mathbf{w}_i \hat{\boldsymbol{\xi}}_t}{\mathbf{w}_i \boldsymbol{\tau}_N}, \text{ if } \mathbf{w}_i \boldsymbol{\tau}_N > 0, \\ &= 0 \text{ if } \mathbf{w}_i \boldsymbol{\tau}_N = 0, \end{aligned}$$

and $\mathbf{w}_i$ denotes the i^{th} row of the $N \times N$ spatial matrix $\mathbf{W}$, which we take as given. Writing the above model in matrix notation we have

$$\hat{\boldsymbol{\xi}}_t = \mathbf{a}_\xi + \sum_{j=1}^{h_\lambda} \boldsymbol{\Lambda}_j \hat{\boldsymbol{\xi}}_{t-j} + \sum_{j=0}^{h_\xi} \boldsymbol{\Psi}_j \mathbf{W} \hat{\boldsymbol{\xi}}_{t-j} + \boldsymbol{\zeta}_t, \tag{17}$$

where $h_\lambda = \max(h_{\lambda 1}, h_{\lambda 2}, ..., h_{\lambda N})$, $h_\xi = \max(h_{\xi 1}, h_{\xi 2}, ..., h_{\xi N})$, $\boldsymbol{\Lambda}_j$ and $\boldsymbol{\Psi}_j$ are $N \times N$ diagonal matrices with λ_{ij} and ψ_{ij} over i as their diagonal elements, and $\boldsymbol{\zeta}_t = (\zeta_{1t}, \zeta_{2t}, ..., \zeta_{Nt})'$. This model provides a generalisation of the spatiotemporal models analysed in the literature to the case where the slope coefficients, λ_{ij} and ψ_{ij}, and the error variances, $\sigma^2_{\zeta_i} = var(\zeta_{it})$ are allowed to differ across i. An econometric analysis of this model is provided by Aquaro, Bailey and Pesaran (2013).

To accommodate negative as well as positive connections, (17) can be further generalised to

$$\hat{\boldsymbol{\xi}}_t = \mathbf{a}_\xi + \sum_{j=1}^{h_\lambda} \boldsymbol{\Lambda}_j \hat{\boldsymbol{\xi}}_{t-j} + \sum_{j=0}^{h_\xi^+} \boldsymbol{\Psi}_j^+ \mathbf{W}^+ \hat{\boldsymbol{\xi}}_{t-j} + \sum_{j=0}^{h_\xi^-} \boldsymbol{\Psi}_j^- \mathbf{W}^- \hat{\boldsymbol{\xi}}_{t-j} + \boldsymbol{\zeta}_t, \tag{18}$$

where $\mathbf{W}^+$ and $\mathbf{W}^-$ are $N \times N$ network matrices for positive and negative connections, respectively. Since $\mathbf{W} = \mathbf{W}^+ + \mathbf{W}^-$, the above specification reduces to (17) if we impose the restrictions $h_\xi^+ = h_\xi^- = h_\xi$, and that $\boldsymbol{\Psi}_j^+ = \boldsymbol{\Psi}_j^-$, for all $j = 1, 2, ..., h_\xi$, which can be tested.

We now use the estimates of the correlation-based connection matrices, $\hat{\mathbf{W}}^{+}$ and $\hat{\mathbf{W}}^{-}$, computed based on residuals from (13) or (15), to obtain estimates of $\mathbf{\Psi}_j^{+}$ and $\mathbf{\Psi}_j^{-}$. To simplify the exposition and given the desirable properties of the de-factoring based on cross- section averages we use the residuals from (13). Therefore (18) becomes

$$\hat{\boldsymbol{\xi}}_t = \mathbf{a}_\xi + \sum_{j=1}^{h} \mathbf{\Lambda}_j \hat{\boldsymbol{\xi}}_{t-j} + \sum_{j=0}^{h_\xi^{+}} \mathbf{\Psi}_j^{+} \tilde{\mathbf{W}}_{cs}^{+} \hat{\boldsymbol{\xi}}_{t-j} + \sum_{j=0}^{h_\xi^{-}} \mathbf{\Psi}_j^{-} \tilde{\mathbf{W}}_{cs}^{-} \hat{\boldsymbol{\xi}}_{t-j} + \boldsymbol{\zeta}_t,$$

where $\tilde{\mathbf{W}}_{cs}^{+}$ and $\tilde{\mathbf{W}}_{cs}^{-}$ are the scaled (row-standardised when applicable) versions of $\hat{\mathbf{W}}_{cs}^{+}$ and $\hat{\mathbf{W}}_{cs}^{-}$.[15] More precisely, for the positively correlated connections we compute the local averages as

$$\begin{aligned} \hat{\xi}_{it}^{+} &= \frac{\hat{\mathbf{w}}_{cs,i}^{+} \hat{\boldsymbol{\xi}}_t}{\hat{\mathbf{w}}_{cs,i}^{+} \boldsymbol{\tau}_N} = \tilde{\mathbf{w}}_{cs,i}^{+} \hat{\boldsymbol{\xi}}_t, \text{ if } \hat{\mathbf{w}}_{cs,i}^{+} \boldsymbol{\tau}_N > 0 \\ &= 0, \text{ if } \hat{\mathbf{w}}_{cs,i}^{+} \boldsymbol{\tau}_N = 0, \text{ for } i = 1, ..., N, \end{aligned}$$

where $\hat{\mathbf{w}}_{cs,i}^{+}$ and $\tilde{\mathbf{w}}_{cs,i}^{+}$ are the i^{th} row of $\hat{\mathbf{W}}_{cs}^{+}$ and $\tilde{\mathbf{W}}_{cs}^{+}$ respectively, while $\boldsymbol{\tau}_N$ is an $N \times 1$ vector of ones. Analogous expressions hold for $\hat{\mathbf{W}}_{cs}^{-}$. Setting h_λ, h_ξ^{+} and h_ξ^{-} equal to unity (18) becomes

$$\hat{\boldsymbol{\xi}}_t = \mathbf{a}_\xi + \mathbf{\Lambda}_1 \hat{\boldsymbol{\xi}}_{t-1} + \mathbf{\Psi}_0^{+} \tilde{\mathbf{W}}_{cs}^{+} \hat{\boldsymbol{\xi}}_t + \mathbf{\Psi}_0^{-} \tilde{\mathbf{W}}_{cs}^{-} \hat{\boldsymbol{\xi}}_t + \mathbf{\Psi}_1^{+} \tilde{\mathbf{W}}_{cs}^{+} \hat{\boldsymbol{\xi}}_{t-1} + \mathbf{\Psi}_1^{-} \tilde{\mathbf{W}}_{cs}^{-} \hat{\boldsymbol{\xi}}_{t-1} + \boldsymbol{\zeta}_t, \ t = 3, ..., 144. \tag{19}$$

Here $\mathbf{\Lambda}_1 = diag\left(\boldsymbol{\lambda}_1\right)$, $\mathbf{\Psi}_0^{+} = diag\left(\boldsymbol{\psi}_0^{+}\right)$, $\mathbf{\Psi}_0^{-} = diag\left(\boldsymbol{\psi}_0^{-}\right)$, $\mathbf{\Psi}_1^{+} = diag\left(\boldsymbol{\psi}_1^{+}\right)$, and $\mathbf{\Psi}_1^{-} = diag\left(\boldsymbol{\psi}_1^{-}\right)$. Also, $\boldsymbol{\lambda}_1, \boldsymbol{\psi}_0^{+}, \boldsymbol{\psi}_0^{-}, \boldsymbol{\psi}_1^{+}$ and $\boldsymbol{\psi}_1^{-}$ are $N \times 1$ vectors of parameters for the $N = 363$ MSAs. Finally, for quasi maximum likelihood (QML) estimation of the parameters we assume that $\zeta_{it} \sim IIDN\left(0, \sigma_{\zeta_i}^2\right)$, for $i = 1, ..., N$.

The model specification in (19) allows for a high degree of heterogeneity in dynamics and spatial dependence across the 363 MSAs. By comparison, the traditional spatial setting assumes that the spatial coefficients in $\boldsymbol{\psi}_0 = (\psi_{10}, ..., \psi_{N0})'$ are homogeneous. This is a strong assumption that we relax in (19). Furthermore, existing spatial literature assumes that all units in $\mathbf{W}_d$ have at least one (positive) neighbour - see Kelejian and Prucha (1999, 2010) among others. This need not always hold either. When applying the multiple testing procedure to the residuals from (13) we find a relatively small number of units, $N_0 = 76$ in total, that are not connected with the remaining MSAs. There are also a number of MSAs with only negative connections, $N_- = 34$, and a number with only positive connections, $N_+ = 90$, with the remaining $N_{+/-} = 163$ MSAs having both positive and negative connections, so that $N = N_{+/-} + N_- + N_+ + N_0 = 363$. The distribution of connections by the eight regions are given in Table 5.

It is clear that, for the most part, MSAs in all regions have both positive and negative connections. Also, more MSAs have exclusively positive connections than only negative connections across all regions, the most polarised regions being the South East and the Far West. On the other hand, a more balanced distribution of MSAs across N_- and N_+ can be seen for the Plains, South West and Rocky Mountains regions, with the latter two also having a proportionately larger number of MSAs with no connections at all.

[15] Here we abstract from sampling variations in the second stage of our modelling strategy. Potentially one could estimate both strong and weak forms of dependence simultaneously but this is outside the scope of the present paper.

Table 5: Distribution of MSAs by connections across 8 regions in the US

Region\No. of MSAs	$N_{+/-}$	N_{-}	N_{+}	N_0	$\sum_{row}$
New England	9	1	1	4	15
Mid East	17	2	9	8	36
South East	63	10	25	16	114
Great Lakes	28	8	13	12	61
Plains	16	5	8	3	32
South West	14	3	7	14	38
Rocky Mountains	7	3	3	9	22
Far West	9	2	24	10	45
$\sum_{col}$	163	34	90	76	363
	Proportional to total no. of MSAs per region				
New England	60.0%	6.7%	6.7%	26.7%	100.0%
Mid East	47.2%	5.6%	25.0%	22.2%	100.0%
South East	55.3%	8.8%	21.9%	14.0%	100.0%
Great Lakes	45.9%	13.1%	21.3%	19.7%	100.0%
Plains	50.0%	15.6%	25.0%	9.4%	100.0%
South West	36.8%	7.9%	18.4%	36.8%	100.0%
Rocky Mountains	31.8%	13.6%	13.6%	40.9%	100.0%
Far West	20.0%	4.4%	53.3%	22.2%	100.0%

$N_{+/-}$ denotes the number of MSAs with both positive and negative connections; N_{-} the no. of MSAs with only negative connections; N_{+} the no. of MSAs with only positive connections; finally N_0 the no. of MSAs with no connections.

Given the existence of contemporaneous effects we cannot estimate ψ_{i0}^{+} and ψ_{i0}^{-} consistently using OLS. Instead we use a QML method developed for given spatial weight matrices in Aquaro, Bailey and Pesaran (2013). For computational convenience we concentrate out the intercept and lagged effects and work with the concentrated log-likelihood function of (19) after stacking over T

$$\ell\left(\boldsymbol{\psi}_0^{+}, \boldsymbol{\psi}_0^{-}\right) \propto T \ln \left| I_N - \boldsymbol{\Psi}_0^{+} \tilde{\mathbf{W}}_{cs}^{+} - \boldsymbol{\Psi}_0^{-} \tilde{\mathbf{W}}_{cs}^{-} \right| - \frac{T}{2} \sum_{i=1}^{N} \ln \left(\frac{1}{T} \widetilde{\hat{\boldsymbol{\xi}}}_i' \mathbf{M}_i \widetilde{\hat{\boldsymbol{\xi}}}_i \right). \tag{20}$$

Here, $\widetilde{\hat{\boldsymbol{\xi}}}_i = \hat{\boldsymbol{\xi}}_i - \psi_{i0}^{+} \hat{\boldsymbol{\xi}}_i^{+} - \psi_{i0}^{-} \hat{\boldsymbol{\xi}}_i^{-}$, $\mathbf{M}_i = I_T - \mathbf{Z}_i \left(\mathbf{Z}_i' \mathbf{Z}_i\right)^{-1} \mathbf{Z}_i'$, $\mathbf{Z}_i = \left(\boldsymbol{\tau}_T, \hat{\boldsymbol{\xi}}_{i,-1}, \hat{\boldsymbol{\xi}}_{i,-1}^{+}, \hat{\boldsymbol{\xi}}_{i,-1}^{-}\right)$ for $\boldsymbol{\tau}_T$ a $T \times 1$ vector of ones, $\boldsymbol{\psi}_0^{+} = (\psi_{10}^{+}, \ldots, \psi_{N0}^{+})'$ and $\boldsymbol{\psi}_0^{-} = (\psi_{10}^{-}, \ldots, \psi_{N0}^{-})'$. The intercepts, $a_{i\xi}$, and the parameters of the lagged values, $\boldsymbol{\lambda}_1, \boldsymbol{\psi}_1^{+}$ and $\boldsymbol{\psi}_1^{-}$ can be estimated via OLS applied to the equations for individual MSAs conditional on ψ_{i0}^{+} and ψ_{i0}^{-}.

We note also that for the N_0 units with no connections, we set $\psi_{i0}^{+} = \psi_{i0}^{-} = \psi_{i1}^{+} = \psi_{i1}^{-} = 0$, and estimate the only remaining parameters, intercepts and λ_{i1} by OLS. In contrast for the $N_{+/-}$ MSAs with both positive and negative connections estimation of ψ_{i0}^{+} and ψ_{i0}^{-} needs to be performed as shown in (20). For MSAs with only negative or positive connections we impose the restriction that the corresponding ψ_{i0}^{+}, ψ_{i1}^{+}, ψ_{i1}^{-}, and ψ_{i0}^{-} coefficients are set to zero. This restriction is needed for identification purposes due to the simultaneity problem that arises in this case. Clearly these coefficients can be set to other values as well, such as the average of each coefficient within the region they belong to or even to the national average of each coefficient.

With respect to inference, it is important to compute standard errors using the second cross derivatives of the original likelihood function of (19) with respect to the full

vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_N)'$, where $\boldsymbol{\theta}_i = \left(\psi_{i0}^+, \psi_{i0}^-, \psi_{i1}^+, \psi_{i1}^-, \alpha_{i\xi}, \lambda_{i1}, \sigma_{\varsigma_i}^2\right)'$. The variance-covariance matrix of $\hat{\boldsymbol{\theta}}_{ML}$ is computed as

$$\hat{\Sigma}_{\boldsymbol{\theta}_{ML}} = \left[-\frac{1}{T} \frac{\partial^2 \ell \left(\hat{\boldsymbol{\theta}}_{ML} \right)}{\partial \hat{\boldsymbol{\theta}}_{ML} \partial \hat{\boldsymbol{\theta}}'_{ML}} \right]^{-1}.$$

For further details, see Aquaro, Bailey and Pesaran (2013).

QML estimates for individual MSAs are available on request. In what follows we give median and mean estimates, and the proportion of MSAs with statistically significant parameters (at the 5% level). The summary estimates for all MSAs are given in Table 6, which shows the median and mean (also known as mean group estimates, MGE) of $\hat{\lambda}_{i1}, \hat{\psi}_{i0}^+, \hat{\psi}_{i0}^-, \hat{\psi}_{i1}^+, \hat{\psi}_{i1}^-$ and $\hat{\sigma}_{i\varsigma}$ over the parameter coefficients that are not restricted to zero, together with the standard errors (in parenthesis) of the MGE.[16] For simplicity of exposition we refer to these summary estimates in the table as $\lambda_1, \psi_0^+, \psi_0^-, \psi_1^+, \psi_1^-$ and σ_ς.

A number of general conclusions readily emerge from an examination of the results in Table 6. The mean and median estimates are very close suggesting that the estimates across the MSAs are approximately symmetrically distributed. All the mean estimates are statistically significant at the 5% level, with the mean lagged spatial effects (ψ_1^+ and ψ_1^-) being less precisely estimated as compared to the other mean effects, λ_1, ψ_0^+ and ψ_0^-.[17] The size of the mean temporal effect, λ_1, at 0.392 (0.009) is reasonably large considering that de-factoring is likely to have removed some of the common dynamics in the house price changes. With regard to the cross section dynamics, contemporaneous positive spill-over effects are more sizeable than their equivalent negative effects with ψ_0^+ and ψ_0^- estimated at 0.345 (0.017) and -0.2763 (0.021), respectively. The estimates in both cases are correctly signed and clearly reject the hypothesis that $\boldsymbol{\Psi}_j^+ = \boldsymbol{\Psi}_j^-$, for $j = 0$ (and for $j = 1$ as discussed below). Therefore, it would appear inappropriate to estimate model specification (17) instead. Also, the lagged spatial effects show a slight reversal of direction of the contemporaneous effects as seen from the size and the magnitudes of ψ_0^+ and ψ_0^-. Indeed, ψ_1^+ and ψ_1^- average at -0.040 (0.015) and 0.071 (0.016). We also note that the mean spatial effects of positive connections at 0.345 is somewhat higher than the mean effects of negative connections at 0.276 (in absolute terms).

With regard to the statistical significance of the estimates for individual MSAs (abstracting from multiple testing issues) we note that $\hat{\lambda}_{i1}$ is statistically significant in 90% of the MSAs, whilst the contemporaneous spatial effects is significant in 65% of MSAs with positive connections, and significant in 62% of MSAs with negative connections. In contrast, the lagged spatial effects turned out to be much weaker, with only 28 present of positive connections and 26% of negative connections being statistically significant. Overall, the estimates suggest there exists a reasonably rich temporal and cross sectional dependence in US house price changes even after stripping them of strong, pervasive national and regional factors.

[16] The MG estimator is defined as the simple average of the estimates across the MSAs with non-zero coefficients. For example, the MGE of $E\left(\psi_{i0}^+\right) = \psi_0^+$, is given by $\psi_{0,MGE}^+ = (1/N_+^*) \sum_{i=1}^{N_+^*} \hat{\psi}_{i0}^+$, where N_+^* denotes the number of MSAs with positive connections ($N_+^* = N_{+/-} + N_+$), and $\hat{\psi}_{i0}^+$ is the QMLE of ψ_{i0}^+. The non-parametric estimator of the variance of $\psi_{0,MGE}^+$ is given by: $\widehat{Var}\left(\psi_{0,MGE}^+\right) = \frac{1}{N_+^*\left(N_+^*-1\right)} \sum_{i=1}^{N_+^*} \left(\hat{\psi}_{i0}^+ - \psi_{0,MGE}^+\right)^2$. For more details see Pesaran and Smith (1995).

[17] MGE standard errors are reported in parenthesis.

Table 6: Quasi-ML estimates of spatiotemporal model (19)

Applied to residual house price changes of 363 MSAs in the United States

	λ_1	ψ_0^+	ψ_0^-	ψ_1^+	ψ_1^-	σ_ζ
	Computed over non-zero parameter coefficients					
Median	0.3986	0.3124	-0.2493	-0.0430	0.0608	1.2416
Mean Group Estimates	0.3921	0.3454	-0.2763	-0.0398	0.0706	1.3056
	(0.0086)	(0.0168)	(0.0209)	(0.0147)	(0.0156)	(0.0181)
% significant (at 5% level)	89.8%	64.8%	61.9%	28.1%	26.4%	-
Number of non-zero coef.	363	253	197	253	197	363

[1] *Restricted parameter coefficients are set to zero.* $\hat{\psi}_{i0}^+ = 0$ *and* $\hat{\psi}_{i1}^+ = 0$ *if MSA* i *has no positive connections;* $\hat{\psi}_{i0}^- = 0$ *and* $\hat{\psi}_{i1}^- = 0$ *if MSA* i *has no negative connections;* $\hat{\psi}_{i0}^+ = 0$, $\hat{\psi}_{i1}^+ = 0$, $\hat{\psi}_{i0}^- = 0$ *and* $\hat{\psi}_{i1}^- = 0$ *if MSA* i *has no positive or negative connections, for* $i = 1, ..., 363$.

[2] *MGE standard errors are in brackets.*

To give an idea of the extent of parameter heterogeneity across MSAs, in Table 7 we provide median and mean estimates by regions. Interestingly enough the regional differences are not very pronounced, particularly if we focus on the more precisely estimated mean parameters, λ_1 and ψ_0^+. The regional estimates of λ_1 range from the low value of 0.316 (0.021) for Great Lakes to the high value of 0.458 (0.025) for the Rocky Mountains. The regional differences in the mean estimates of ψ_0^+ are even slightly lower and range from 0.264 (0.082) in South West to 0.374 (0.042) in the Plains. In contrast, the estimates of the negative connections, ψ_0^- are less precisely estimated and range from -0.078 (0.099) for New England to -0.370 (0.053) in South West). But one should consider such comparisons with care since in the case of some regions the number of non-zero estimates were quite small. Nevertheless, one of our main conclusions that positive and negative connections have opposite effects seems to be robust to the regional disaggregation. The estimates of ψ_0^+ and ψ_0^- are respectively positive and negative across all the regions. The results in Table 7 also support our conclusion that lagged spatial effects are generally not that important and tend to be statistically insignificant in a number of regions. But once again we need to bear in mind that some of the regional estimates are based on a rather small number of non-zero estimates.

Table 7: Quasi-ML estimates of spatiotemporal model (19) summarised by region

Applied to residual house price changes of 363 MSAs in the United States

	Computed over non-zero parameter coefficients					
	λ_1	ψ_0^+	ψ_0^-	ψ_1^+	ψ_1^-	σ_ζ
	New England					
Median	0.4064	0.2762	-0.0843	-0.0514	0.0209	1.1684
Mean Group Estimates	0.3944	0.3563	-0.0781	-0.0050	-0.0412	1.2704
	(0.0303)	(0.0996)	(0.0991)	(0.0430)	(0.0784)	(0.0966)
% significant (5% level)	86.7%	60.0%	50.0%	0.0%	30.0%	-
Number of non-zero coef.	15	10	10	10	10	10
	Mid East					
Median	0.4278	0.3439	-0.1904	-0.0096	0.0625	1.3977
Mean Group Estimates	0.3990	0.3603	-0.1938	-0.0755	0.1129	1.4368
	(0.0319)	(0.0465)	(0.1163)	(0.0487)	(0.0747)	(0.0634)
% significant (5% level)	91.7%	65.4%	68.4%	30.8%	26.3%	-
Number of non-zero coef.	36	26	19	26	19	36
	South East					
Median	0.4013	0.3242	-0.2686	-0.0538	0.0847	1.2384
Mean Group Estimates	0.4001	0.3563	-0.3062	-0.0596	0.0977	1.3469
	(0.0162)	(0.0262)	(0.0326)	(0.0234)	(0.0242)	(0.0427)
% significant (5% level)	90.4%	64.8%	61.6%	27.3%	31.5%	-
Number of non-zero coef.	114	88	73	88	73	114
	Great Lakes					
Median	0.3176	0.2660	-0.2227	0.0149	0.0361	1.2492
Mean Group Estimates	0.3160	0.3304	-0.2846	0.0229	0.0407	1.3142
	(0.0209)	(0.0463)	(0.0383)	(0.0435)	(0.0351)	(0.0392)
% significant (5% level)	78.7%	63.4%	50.0%	31.7%	13.9%	-
Number of non-zero coef.	61	41	36	41	36	61
	Plains					
Median	0.3808	0.3015	-0.2491	-0.1573	0.0597	1.1128
Mean Group Estimates	0.3751	0.3744	-0.2409	-0.1280	0.0825	1.1254
	(0.0243)	(0.0427)	(0.0540)	(0.0290)	(0.0421)	(0.0324)
% significant (5% level)	93.8%	75.0%	57.1%	29.2%	28.6%	-
Number of non-zero coef.	32	24	21	24	21	32
	South West					
Median	0.3935	0.2944	-0.3053	-0.1023	0.0077	1.2877
Mean Group Estimates	0.4024	0.2642	-0.3695	-0.0576	0.0377	1.3385
	(0.0209)	(0.0823)	(0.0525)	(0.0630)	(0.0560)	(0.0301)
% significant (5% level)	94.7%	57.1%	82.4%	38.1%	23.5%	-
Number of non-zero coef.	38	21	17	21	17	38
	Rocky Mountains					
Median	0.4435	0.3486	-0.2756	0.0155	0.1396	1.1618
Mean Group Estimates	0.4581	0.3177	-0.3086	0.0083	0.1033	1.2096
	(0.0253)	(0.0667)	(0.0542)	(0.0430)	(0.0557)	(0.0409)
% significant (5% level)	100.0%	70.0%	80.0%	10.0%	40.0%	-
Number of non-zero coef.	22	10	10	10	10	22
	Far West					
Median	0.4672	0.3667	-0.2438	0.0330	0.0480	1.2158
Mean Group Estimates	0.4400	0.3591	-0.2673	0.0137	0.0155	1.2437
	(0.0237)	(0.0488)	(0.0898)	(0.0428)	(0.0293)	(0.0336)
% significant (5% level)	91.1%	63.6%	63.6%	30.3%	18.2%	-
Number of non-zero coef.	45	33	11	33	11	45

[1] *Restricted parameter coefficients are set to zero.* $\hat{\psi}_{i0}^+ = 0$ *and* $\hat{\psi}_{i1}^+ = 0$ *if MSA* i *has no positive connections;* $\hat{\psi}_{i0}^- = 0$ *and* $\hat{\psi}_{i1}^- = 0$ *if MSA* i *has no negative connections;* $\hat{\psi}_{i0}^+ = 0$, $\hat{\psi}_{i1}^+ = 0$, $\hat{\psi}_{i0}^- = 0$ *and* $\hat{\psi}_{i1}^- = 0$ *if MSA* i *has no positive or negative connections, for* $i = 1, ..., 363$.

[2] *MGE standard errors are in brackets below their respective Mean Group Estimates.*

Finally, to assess the importance of de-factoring of house price changes we also estimated the connection matrices $\hat{\boldsymbol{W}}^{+}$ and $\hat{\boldsymbol{W}}^{-}$ without de-factoring, using the hierarchical factor model (12). Not surprisingly we found $\hat{\boldsymbol{W}}^{+}$ to be much denser as compared to the estimates obtained based on de-factored price changes, and $\hat{\boldsymbol{W}}^{-}$ to be less dense. The many more connections that we are finding when using price changes without de-factoring reflect the presence of common factors rather than genuine spatial effects. In line with this result we also find an estimate of spatial effects, ψ_0^{+}, which is very close to unity when we use $\boldsymbol{W}^{+}$ estimated based on non-defactored price changes. Details of these results are available upon request.

5 Conclusions

An understanding of the spatial dimension of economic and social activity requires methods that can separate out the relationship between spatial units that is due to the effect of common factors from that which is purely spatial, even in an abstract sense. We are able to distinguish between cross-sectional strong dependence and weak or spatial dependence. Strong dependence in turn suggests that there are common factors. We have proposed the use of cross unit averages to extract common factors and contrast this to a principal components approach widely used in the literature. We then use multiple testing to determine significant bilateral correlations (signifying connections) between spatial units and compare this to an approach that just uses distance to determine units that are neighbours. In a very data rich environment with observations on many spatial units over long periods of time a way of filtering the data to uncover spatial connections is crucial. We have applied these methods to real house price changes in the US at the level of the Metropolitan Statistical Area. Although there is considerable overlap between neighbours determined by distance and those by multiple testing, there is also considerable correlation between MSAs across the United States that suggests that other forces at work.

We also find that our analysis of connections based on pair-wise correlations of de-factored house price changes clearly points to the existence of negative as well as positive connections. This feature is absent if we base the spatial analysis exclusively on contiguity. It is common in the literature to think of spatial relationships as involving spillover from one area to another with the (implicit) assumption that the spillovers are positive. But this need not be the case. Migration across space could raise/lower wages or house prices in one locality and lower/raise them into another locality.

Further, we verify that basing the spatial analysis on house price changes without de-factoring ignores the possibility that there may be common national and regional factors that account for these correlations and failing to condition on the common factors may bias the inferences that can be drawn. Our analysis strips out such common effects and allows us to focus on spillover effects (positive or negative) which is of primary interest in spatial analysis. Although proximity measured by distance is a useful metric for constructing a weighting matrix, our analysis suggests that correlation analysis, once applied to de-factored price changes with appropriate application of multiple testing techniques can lead to important new insights as to the nature of spatial connections.

Appendices

Appendix I: Calculation of distance

The original data used were Latitude-Longitude of zip codes, cross referenced with each of the 366 Metropolitan Statistical Areas (MSAs). Any missing Latitude-Longitude coordinates were coded manually from Google searches. The geodesic distance between a pair of latitude/longitude coordinates was then calculated using the Haversine formula:

$$\begin{aligned} a &= \sin^2\left(\frac{\Delta lat}{2}\right) + \cos\left(lat1\right)\cos\left(lat2\right)\sin^2\left(\frac{\Delta long}{2}\right), \\ c &= 2a\tan 2\left(\sqrt{a}, \sqrt{1-a}\right), \\ d &= Rc, \end{aligned}$$

where R is the radius of the earth in miles and d is the distance. $\Delta lat = lat2 - lat1$, and $\Delta long = long2 - long1$.

Appendix II: Geographical classification of the United States

Table A provides the broader geographical breakdown used in our analysis of US house prices. We identify 8 regions which contain an average of 6 States, each of which contains an approximate average of 45 Metropolitan Statistical Areas. The classifications are shown in Table A together with the number of MSAs included in each State. Details on the exact MSAs used are available upon request.

Appendix III: Data sources

Monthly data for US house prices from January 1975 to December 2010 are taken from Freddie Mac. These data are available at:

http://www.freddiemac.com/finance/cmhpi

The quarterly figures are arithmetic averages of monthly figures.

Annual CPI data at State level are obtained from the Bureau of Labor Statistics:

http://www.bls.gov/cpi/

The quarterly figures are interpolated using the interpolation technique described in the appendix of GVAR toolbox 1.1 user guide.

The annual population data at MSA level are obtained from the Regional Economic Information System, Bureau of Economic Analysis, U.S. Department of Commerce:

http://www.bea.gov/regional/docs/footnotes.cfm?tablename=CA1-3

Again the quarterly figures are interpolated using the interpolation technique described in the appendix of GVAR toolbox 1.1 user guide.

Table A: Geographical Classification of Regions, States and MSAs in the United States

Regions	States	No of MSAs	Regions	States	No of MSAs	Regions	States	No of MSAs
New England	Connecticut (CT)	4	**Great Lakes**	Illinois (IL)	9	**South West**	Arizona (AZ)	6
	Maine (ME)	3		Indiana (IN)	13		New Mexico (NM)	4
	Massachussetts (MA)	5		Michigan (MI)	14		Oklahoma (OK)	3
	New Hampshire (NH)	1		Ohio (OH)	13		Texas (TX)	25
	Rhode Island (RI)	1		Wisconsin (WI)	12			
	Vermont (VT)	1						
Mid-East	Delaware (DE)	1	**Plains**	Iowa (IA)	8	**Rocky Mountains**	Colorado (CO)	7
	District of Columbia (DC)	1		Kansas (KS)	4		Idaho (ID)	5
	Maryland (MD)	4		Minnesota (MN)	5		Montana (MT)	3
	New Jersey (NJ)	4		Missouri (MO)	8		Utah (UT)	5
	New York (NY)	12		Nebraska (NE)	2		Wyoming (WY)	2
	Pennsylvania (PA)	14		North Dakota (ND)	3			
				South Dakota (SD)	2			
South East	Alabama (AL)	11				**Far West**	California (CA)	26
	Arkansas (AR)	6					Nevada (NV)	3
	Florida (FL)	20					Oregon (OR)	6
	Georgia (GA)	14					Washington (WA)	10
	Kentucky (KY)	5						
	Louisiana (LA)	8						
	Mississippi (MS)	4						
	North Carolina (NC)	14						
	South Carolina (SC)	8						
	Tennessee (TN)	10						
	Virginia (VA)	9						
	West Virginia (WV)	5						

References

[1] Anselin, L. (1988). Spatial Econometrics: Methods and Models. Boston: Kluwer Academic Publishers.

[2] Anselin, L. (2001). Spatial Econometrics. In B. H. Baltagi (Ed.), A Companion to Theoretical Econometrics. Blackwell, Oxford.

[3] Aquaro, M., Bailey, N. and Pesaran, M. H. (2013). MLE Type Estimation of Spatiotemporal Models. Unpublished Manuscript.

[4] Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. Econometrica, 71, 1, 135–171.

[5] Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. Econometrica, 70, 191-221.

[6] Bailey, N., Kapetanios, G. and Pesaran, H. M. (2013). Exponent of Cross-sectional Dependence: Estimation and Inference. Jan 2012, revised Nov 2013. Cambridge Working Paper 1206.

[7] Bailey, N., Pesaran, H. M. and Smith, L.V. (2013). A multiple Testing Approach to the Regularisation of Sample Correlation Matrices. Unpublished Manuscript.

[8] Barigozzi, M. and Brownlees, C. (2013). NETS: Network Estimation for Time Series. SSRN Working Paper. Available at SSRN: http://ssrn.com/abstract=2249909 or http://dx.doi.org/10.2139/ssrn.2249909.

[9] Bester, C. A., Conley,T. G. and Hansen, C. B. (2011). Inference with Dependent Data Using Cluster Covariance Estimators. Journal of Econometrics, 165, 2, 137-151.

[10] Bhattachajee, A. and Holly, S. (2013). Understanding Interactions in Social Networks and Committees. Spatial Economic Analysis, 8, 1, 23-53.

[11] Bien, J. and Tibshirani, R. J. (2011). Sparse Estimation of the Covariance Matrix. Biometrika, 98, 4, 807-820.

[12] Bonferroni, C. (1935). Il Calcolo delle Assicurazioni su Gruppi di Teste. Studi in Onore del Professore Salvatore Ortu Carboni, Rome, Italy, 13–60.

[13] Bonferroni, C. (1936). Teoria Statistica delle Classi e Calcolo delle Probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 8, 3–62.

[14] Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R. and Kohane, I. S. (2000). Discovering Functional Relationships between RNA Expression and Chemotherapeutic Susceptibility using Relevance Networks. Proc. Nat. Acad. Sci. U.S.A., 97, 12182-12186.

[15] Carlino, G. A. and DeFina, R. H. (1998). The Differential Regional Effects of Monetary Policy. Review of Economics and Statistics, 80, 4, 572-587.

[16] Carlino, G. A. and DeFina, R. H. (2004). How Strong Is Co-movement in Employment over the Business Cycle? Evidence from State/Sector Data. Journal of Urban Economics, 55, 2, 298-315.

[17] Carlino, G. and Sill, K. (2001). Regional Income Fluctuations: Common Trends and Common Cycles. Review of Economics and Statistics, 83, 3, 446-456.

[18] Chamberlain, G. and Rothschild, M. (1983). Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets. Econometrica, 51, 1305–1324.

[19] Chaudhuri, S., Drton, M. and Richardson, T. S. (2007). Estimation of a Covariance Matrix with Zeros. Biometrika, 94, 199-216.

[20] Chudik, A., Pesaran, M. H. and Tosetti, E. (2011). Weak and Strong Cross-section Dependence and Estimation of Large Panels. The Econometrics Journal, 14, C45-C90.

[21] Chudik, A. and Pesaran, M. H. (2013). Large Panel Data Models with Cross-Sectional Dependence: A Survey. Center for Applied Financial Economics (CAFE) research paper no. 13.15. In B. H. Baltagi (Ed.), forthcoming in The Oxford Handbook on Panel Data. Oxford University Press.

[22] Conley, T. G. (1999). GMM Estimation with Cross-sectional Dependence. Journal of Econometrics, 92, 1-45.

[23] Conley, T. G. and Dupor, B. (2003). A Spatial Analysis of Sectoral Complementarity. Journal of Political Economy, 111, 311-352.

[24] Conley, T. G. and Topa, G. (2003). Identification of Local Interaction Models with Imperfect Location Data. Journal of Applied Econometrics, 18, 5, 605-618.

[25] Connor, G. and Korajczyk, R. (1993). A Test for the Number of Factors in an Approximate Factor Model. The Journal of Finance XLVIII, 4, 1263-1291.

[26] Corrado, L. and Fingleton, B. (2012). Where is the Economics in Spatial Econometrics? Journal of Regional Science, 52, 2, 210-239.

[27] Cromwell B. (1992). Does California Drive the West? An Econometric Investigation of Regional Spillovers. Economic Review of the Federal Reserve Bank of San Francisco, 2, 13–23.

[28] Dees, S., di Mauro, F., Pesaran, M. H. and Smith, L.V. (2007). Exploring the International Linkages of the Euro Area: A Global VAR Analysis. Journal of Applied Econometrics, 22, 1-38.

[29] Del Negro, M. (2002). Asymmetric Shocks among U.S. States. Journal of International Economics, 56, 2, 273-297.

[30] Dempster, A. P. (1972). Covariance Selection. Biometrics, 28, 157-175.

[31] Efron, B. (2010). Large-scale Inference: Empirical Bayes Methods for Estimation, Testing and Prediction, Cambridge.

[32] Forni, M. and Reichlin, L. (1998). Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics. Review of Economic Studies, 65, 453-473.

[33] Glaeser, E. L. and Gottlieb, J. D. (2009). The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. Journal of Economic Literature, 47, 983–1028.

[34] Glaeser, E. L., Gyourko, J. and Saiz, A. (2008). Housing Supply and Housing Bubbles. Journal of Urban Economics, 64, 2, 198-217.

[35] Gupta, R. and Das, S. (2010). Predicting Downturns in the US Housing Market. Journal of Real Estate Economics and Finance 41, 3, 294-319.

[36] Gupta, R., Kabundi, A. and Miller, S. M. (2011a). Using Large Data Sets to Forecast Housing Prices: A Case Study of twenty US States. Journal of Housing Research 20, 2, 161-190.

[37] Gupta, R., Kabundi, A. and Miller, S. M. (2011b). Forecasting the US Real House Price Index: Structural and Non-structural Models with and without Fundamentals. Economic Modelling, 28, 4, 2013-2021.

[38] Gupta R. and Miller, S. M. (2012). The Time-series Properties on Housing Prices: A Case Study of the Southern California Market. Journal of Real Estate Finance and Economics, 44, 3, 2012.

[39] Holly, S., Pesaran, H. M. and Yamagata, T. (2011a). A Spatiotemporal Model of House Prices in the US. Journal of Econometrics, 158, 160-173.

[40] Holly, S., Pesaran, H. M. and Yamagata, T. (2011b). The Spatial and Temporal Diffusion of House Prices in the UK. Journal of Urban Economics, 69, 1, 2-23.

[41] Holly, S. and Petrella, I. (2012). Factor Demand Linkages, Technology Shocks and the Business Cycle. Review of Economics and Statistics, 94, 4, 948-963.

[42] Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics, 6, 2, 65–70.

[43] Hotelling, H. (1933). Analysis of a Complex of Statistical Variables with Principal Components. Journal of Educational Psychology, 24, 6, 417-441.

[44] Kadiyala, K.R. and Bhattacharya, R. (2009). Regional Housing Prices in the USA: An Empirical Investigation of Nonlinearity. The Journal of Real Estate Finance and Economics, 38, 4, 443–460.

[45] Kapetanios, G. and Pesaran, M. H. (2007). Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Return. In Garry Phillips and Elias Tzavalis (eds.), The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis, Cambridge University Press, Cambridge.

[46] Kapoor, M., Kelejian, H. H. and Prucha, I. R. (2007). Panel Data Models with Spatially Correlated Error Components. Journal of Econometrics, 140, 97-130.

[47] Khare, K. and Rajaratnam, B. (2011). Wishart Distributions for Decomposable Covariance Graph Models. Annals of Statistics, 30, 514-555.

[48] Kelejian, H. H. and Prucha, I. R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. International Economic Review, 40, 509-533.

[49] Kelejian, H. H. and Prucha, I. R. (2007). HAC Estimation in a Spatial Framework. Journal of Econometrics, 140, 131-154.

[50] Kelejian, H. H. and Prucha, I. R. (2010). Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances. Journal of Econometrics, 157, 53–67.

[51] Krugman, P. R. (1991). Increasing Returns and Economic Geography. Journal of Political Economy, 99, 483-499.

[52] Kuethe, T. and Pede, V. (2011). Regional Housing Price Cycles: A Spatiotemporal Analysis Using US State-level Data. Regional Studies, 45, 5, 563-574.

[53] Lee, L.-F. (2002). Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models. Econometric Theory, 18, 2, 252-277.

[54] Lee, L.-F. (2004). Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. Econometrica, 72, 6, 1899-1925.

[55] Lee, L.-F. and Yu, J. (2010). Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects. Journal of Econometrics, 154, 165-185.

[56] Lee, K. and Pesaran, M. H. (1993). Persistence Profiles and Business Cycle Fluctuations in a Disaggregated Model of UK Output Growth. Richerche Economiche, 47, 293-322.

[57] Lin, X. and Lee, L-F (2010). GMM Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity. Journal of Econometrics, 157, 1, 34-52.

[58] Meen, G. (1999). Regional House Prices and the Ripple Effect: A New Interpretation. Housing Studies, 14, 733–753.

[59] Meinshausen, N. and Buhlmann, P. (2006). High-Dimensional Graphs and Variable Selection with the Lasso. The Annals of Statistics, 34, 3, 1436-1462.

[60] Owyang, M. T., Piger, J. and Wall, H. J. (2005). Business Cycle Phases in U.S. States. Review of Economics and Statistics, 87, 4, 604-616.

[61] Partridge, M. D. and Rickman, D. S. (2005). Regional Cyclical Asymmetries in an Optimal Currency Area: An Analysis Using US State Data. Oxford Economic Papers, July, 57, 3, 373-397.

[62] Pearson, K. (1900). On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling. Philosophical Magazine Series, 5, 50 (302), 157–175.

[63] Peng, J., Wang, W., Zhou, N. and Zhu, J. (2009). Partial Correlation Estimation by Joint Sparse Regression Models. Journal of the American Statistical Association, June, 104, 486, 735-746.

[64] Pesaran, H. M. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. Econometrica, 74, 4, 967-1012.

[65] Pesaran, M. H. (2013). Testing Weak Cross-sectional Dependence in Large Panels. Forthcoming in Econometric Reviews.

[66] Pesaran, M. H., Schuermann, T. and Weiner, S. M. (2004). Modeling Regional Interdependencies using a Global Error-Correcting Macroeconomic Model. Journal of Business and Economics Statistics, 22, 2, 129-162.

[67] Pesaran, M. H. and Smith, R. P. (1995). Estimating Long-run Relationships from Dynamic Heterogeneous Panels. Journal of Econometrics, 68, 79-113.

[68] Pesaran, M. H. and Timmermann, A. (2009). Testing Dependence Among Serially Correlated Multicategory Variables. Journal of American Statistical Association, 104, 485, 325-337.

[69] Pollakowski H. O. and Ray T. S. (1997). Housing Price Diffusion Patterns at Different Aggregation Levels: An Examination of Housing Market Efficiency. Journal of Housing Research, 8, 1, 107–124.

[70] Rapach, D. E. and Strauss, J. K. (2007). Forecasting Real Housing Price Growth in the Eighth District States. Federal Reserve Bank of St. Louis. Regional Economic Development, 3, 2, 33–42.

[71] Rapach, D. E. and Strauss, J. K. (2009). Differences in Housing Price Forecast Ability Across U.S. States. International Journal of Forecasting, 25, 2, 351-372.

[72] Robertson, D. and Symons, J. (2007). Maximum Likelihood Factor Analysis with Rank-deficient Sample Covariance Matrices. Journal of Multivariate Analysis, 98, 4, 813-828.

[73] Rothman, A. J., Bickel, P. J., Levina, E. and Zhu, J. (2008). Sparse Permutation Invariant Covariance Estimation. Electron. J. Stat., 2, 494-515.

[74] Rothman, A. J., Bickel, P. J., Levina, E. and Zhu, J. (2010). Sparse Multivariate Regression with Covariance Estimation. Journal of Computational and Graphical Statistics, 19, 4, 947-962.

[75] Stock, J. H. and Watson, M. W. (1998). Diffusion Indexes. NBER Working Papers 6702, National Bureau of Economic Research, Inc.

[76] Stock, J. H. and Watson, M. W. (2002a). Forecasting Using Principal Components from a Large Number of Predictors. Journal of the American Statistical Association, 97, 460, 147–162.

[77] Stock, J. H. and Watson, M. W. (2002b). Macroeconomics Forecasting Using Diffusion Indexes. Journal of Business and Economic Statistics, 20, 2, 147–162.

[78] Stone, R. (1947). On the Interdependence of Blocks of Transactions. Journal of the Royal Statistical Society (Supplement), 9, 1, 1–45.

[79] Whittle, P. (1954). On Stationary Processes on the Plane. Biometrika, 41, 434–449.

[80] Yu, J., de Jong, R. M. and Lee, L-F (2008). Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed Effects when Both N and T Are Large. Journal of Econometrics, 146, 1, 118-134.