Spatial Effects in a Common Trend Model of US City-Level CPI

Peter Burridge[a]

[a] Department of Economics, University of York,
York, YO10 5DD, UK (peter.burridge@york.ac.uk)

Fabrizio Iacone*[b]

[b] Department of Economics, University of York,
York, YO10 5DD, UK (fabrizio.iacone@york.ac.uk)

Štěpána Lazarová[c]

[c] School of Economics and Finance, Queen Mary University of London,
London E1 4NA, UK (s.lazarova@qmul.ac.uk)

2 July 2015

Abstract

This paper studies relative movements in price indices of 17 US cities. We employ an unobserved common trend model where the trend can be stochastic or deterministic with possible breaks or other nonlinearities. To accommodate the spatial nature of the data we allow for spatially correlated short-run shocks. In this way, the speed of convergence to the equilibrium implied by the law of one price is estimated taking into account the effect of distances across cities. The parameters of the model are estimated using a generalized method of moments (GMM) method which incorporates moment conditions corresponding to a generalized least squares-like within estimator of regression parameters. We find a slow rate of convergence of the price levels and strong evidence of spatial effects.

Keywords: Common trend, general method of moments, law of one price, price index, spatial correlation
JEL classification: C33, E31, R12

*Corresponding author: Fabrizio Iacone, Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, UK.

1 Introduction

The law of one price (LOP), as generally understood, follows from the assumption that individuals and firms will not systematically ignore opportunities to profit from risk-free arbitrage. In the absence of transactions costs or institutional barriers, it should not be possible to buy a commodity at one price and immediately sell it for a higher price. On the contrary, so the argument goes, the very possibility of arbitrage will eliminate such price differences. Like many core ideas in economics the LOP is easy to state but by no means easy to verify empirically. To help account for the frequent rejection of the LOP, Pippenger and Phillips (2008, p.916) identify four confounding factors in studies of commodity prices: use of retail prices, ignoring transport costs, ignoring time, and pricing non-identical products. The first three factors directly affect potential arbitrage, which requires the goods being traded to be resaleable, while the fourth is obviously fundamental. Many studies that challenge the empirical validity of the LOP, it is argued, fail to attend adequately to one or more of these details. On the other hand, when the data employed are not contaminated in this way, support for the LOP improves, a good example being the analysis of data from various multi-national internet traders by Cavallo, Neiman and Rigobon (2014). At any given time, there is always some observed price dispersion; consequently, many studies investigate whether prices can be shown to be converging to the LOP, and if so, how rapidly. The picture here is complicated by the underlying price dynamics: in many markets prices are non-stationary, and so following Johansen and Juselius (1992), testing for the presence of cointegration between two or more price series has become routine, with rejection interpreted as evidence against PPP or the LOP.

Closer in scope to the present work are studies that evaluate the size of international or internal border effects, or rates of price convergence within countries. In the first case it is necessary to distinguish between cross-border distance effects, which may be magnified by political boundaries, on the one hand, and inter- and intra-jurisdictional price distributional differences which may confound these. Surveying numerous North American studies, from Engel and Rogers (1996) onwards, Gorodnichenko and Tesar (2009) argue that much of the US-Canada border impact identified may be a side-effect of the greater price dispersion within the US. This line of argument demonstrates that price dispersion, per se, is not taken as evidence against the LOP. Studies of price convergence at the sub-national scale typically suppose that systems of states, regions or cities exhibit movement around a common trend, the point being to establish convergence towards such a trend. In an influential paper, Cecchetti, Mark and Sonora (2002) "believe that studying the behavior of prices across U.S. cities will help us in understanding the likely nature of inflation convergence in the Euro area." They work with relative price indices, arguing that it is the behaviour of such aggregates that is of primary concern to monetary policy makers. Their headline result is that city relative price indices do not have unit roots, but that convergence is very slow, with a half-life of about 9 years, attributed to the difficulty in trading some goods. They found that relative prices between distant cities were significantly more dispersed than those between near neighbours, while convergence between cities that were closer together was faster, but not significantly so (op. cit. p.1090 Table 3). Earlier, Parsley and Wei (1996) had also shown that the variability of relative commodity prices between U.S. cities was related to the distance separating them, while a unit root in relative prices was similarly rejected. Noting that both Cecchetti et al, and Parsley and Wei, and others, could only secure rejection of the crucial unit root null hypothesis by adopting panel unit root tests, that gloss over any individual series that might be non-stationary, Sonora (2008) repeats the analysis using a new generation of more powerful univariate tests. He finds in favour of stationarity in a majority of cases, and detects faster convergence rates than in the previous studies.

The common finding that relative price dispersion observed over time at pairs of locations increases with their physical separation suggests to us that spatial effects should be incorporated into the model, rather than being investigated separately. Although the inflation convergence literature stimulated by the creation of the Eurozone has a vigorous regional strand, and there are a number of studies of price dispersion between U.S. cities, space is generally introduced at a second stage of the analysis. In this paper, therefore, dynamic and spatial interactions in U.S. city-level prices are integrated via a panel data model with explicit spatial dependence. There are currently at least two alternative approaches to the modeling of such panels, and so the next section describes these briefly to provide some context. Section 3 introduces the model in detail, and Section 4 presents the estimation method and the asymptotic properties of the estimates. Section 5 gives a description of the data and empirical results, and finally Section 6 comments on the implications. Proofs of the theorems are set out in a separate section.

2 Estimating dynamic spatial panel models

Kelejian and Prucha (1999) propose a generalised method of moments (GMM) estimator for a static cross-section model with spatially correlated errors. This set-up is further developed by Kapoor et al (2007) and Mutl (2006) who introduce GMM estimators for stationary dynamic panel models with temporal and spatial correlation in the disturbance handled via random effects. Baltagi et al (2014) propose a GMM estimator for a model that also includes a temporal and spatial lag of the dependent variable, while Mutl and Pfaffermayr (2011), develop a test of the random effects assumption in a static Cliff-Ord type model. Similarly, Baltagi and Liu (2011) propose generalized least squares (GLS) estimators for panel data with fixed or random effects for a generalized spatial error components panel data model and develop a Hausman specification test. Lee and Yu (2010) review both static and dynamic spatial panel data models, providing a concise guide to recent developments in this rapidly expanding field. Following Yu, de Jong and Lee (2012) (YJL) the dynamic spatial panel model underlying this strand of work can be written as

$$\mathbf{Y}_{n,t} = \lambda_0 \mathbf{W}_n \mathbf{Y}_{n,t} + \gamma_0 \mathbf{Y}_{n,t-1} + \rho_0 \mathbf{W}_n \mathbf{Y}_{n,t-1} + \mathbf{X}_{n,t}\beta_0 + \mathbf{c}_{n,0} + \alpha_{t,0}\mathbf{1}_n + \mathbf{V}_{n,t} \qquad (1)$$

in which $\mathbf{Y}_{n,t} = [y_{1,t}, ..., y_{n,t}]'$ is observed at the n locations for each time period, $\mathbf{X}_{n,t}$ is an $n \times k$ matrix of exogenous covariates, $\mathbf{c}_{n,0}$ a vector of location-specific fixed effects, $\alpha_{t,0}$ a panel-wide time effect, and $\mathbf{V}_{n,t}$ an independent, identically distributed (IID) disturbance. In this structure, the vector of current endogenous variables $\mathbf{Y}_{n,t}$ is seen to be influenced by its own past, and also by a contemporaneous spill-over effect via the vector of weighted neighbouring values, $\mathbf{W}_n\mathbf{Y}_{n,t}$. To discuss the dynamics implicit in (1), first assume that the matrix, $[\mathbf{I}_n - \lambda_0 \mathbf{W}_n] = \mathbf{S}_n$ is invertible, and then write, $\mathbf{A}_n = \mathbf{S}_n^{-1}[\gamma_0 \mathbf{I}_n + \rho_0 \mathbf{W}_n]$. With this notation the reduced form may be written,

$$\mathbf{Y}_{n,t} = \mathbf{A}_n \mathbf{Y}_{n,t-1} + \mathbf{S}_n^{-1}[\mathbf{X}_{n,t}\beta_0 + \mathbf{c}_{n,0} + \alpha_{t,0}\mathbf{1}_n + \mathbf{V}_{n,t}] \qquad (2)$$

from which we obtain the Error Correction Model (ECM) representation

$$\mathbf{\Delta Y}_{n,t} = [\mathbf{A}_n - \mathbf{I}_n]\mathbf{Y}_{n,t-1} + \mathbf{S}_n^{-1}[\mathbf{X}_{n,t}\beta_0 + \mathbf{c}_{n,0} + \alpha_{t,0}\mathbf{1}_n + \mathbf{V}_{n,t}].$$

It is now easy to see that the dynamics of $\mathbf{Y}_{n,t}$ are determined by the dynamics of $\mathbf{X}_{n,t}$, $\alpha_{t,0}$, and the eigenvalues of $\mathbf{A}_n$. If $\mathbf{W}_n$ is obtained from a symmetric matrix of non-negative constants by row-normalisation, the interesting cases identified by YJL are (i) if all the eigenvalues of $\mathbf{A}_n$ have magnitude smaller than 1 the process may be stationary, (ii) if all the eigenvalues of $\mathbf{A}_n$ are equal to 1 we may have a pure unit root process without cointegration, and (iii) if some of the eigenvalues of $\mathbf{A}_n$ are equal to 1 we may have the case of "spatial cointegration". We say "may" here, because YJL assume that $\mathbf{X}_{n,t}$ is non-stochastic, while as they note, various further possibilities arise according to how $\alpha_{t,0}$ evolves. However, with the specification (1), the common time effect may be eliminated by a simple transformation, as is the case for our model introduced in Section 3. After some manipulation, YJL (2012, p. 30) show that the endogenous variable may be expressed as the sum of three components:

$$\mathbf{Y}_{n,t} = \mathbf{Y}_{n,t}^{unit} + \mathbf{Y}_{n,t}^{sta} + \mathbf{Y}_{n,t}^{\alpha} \tag{3}$$

where $\mathbf{Y}_{n,t}^{unit}$ is a non-stationary vector process, $\mathbf{Y}_{n,t}^{sta}$ is a stationary component, and $\mathbf{Y}_{n,t}^{\alpha} = \frac{1}{1-\lambda_0}\mathbf{1}_n\sum_{h=0}^{t}\alpha_{t-h,0}$ is a common trend. Furthermore, in the "spatial cointegration" case that is of greatest interest, two of these components are eliminated by the transformation, $(\mathbf{W}_n - \mathbf{I}_n)$; it can be shown that both $(\mathbf{W}_n - \mathbf{I}_n)\mathbf{Y}_{n,t}^{unit} = \mathbf{0}$ and $(\mathbf{W}_n - \mathbf{I}_n)\mathbf{Y}_{n,t}^{\alpha} = \mathbf{0}$ so that $(\mathbf{W}_n - \mathbf{I}_n)\mathbf{Y}_{n,t}$ is stationary, revealing that the rows of $(\mathbf{W}_n - \mathbf{I}_n)$ are cointegrating vectors, and that the rank of this matrix is the cointegrating rank of the system of related sites, in the sense that these vectors define linear combinations of the $\mathbf{Y}$ values observed at different locations that are stationary.

A somewhat different approach that introduces dependence and dynamics via observed and unobserved common factors, building on the work of Pesaran (2006), is developed in recent papers by Kapetanios, Pesaran and Yamagata (2011), Chudik, Pesaran and Tosetti (2011), and Pesaran and Tosetti (2011), who introduce a model of the form,

$$y_{it} = \alpha_i^{'}\mathbf{d}_t + \beta_i^{'}\mathbf{x}_{it} + \gamma_i^{'}\mathbf{f}_t + e_{it} \tag{4}$$

in which $\mathbf{d}_t$ is an $m_d \times 1$ vector of observed common effects (such as time trends, or aggregate prices), $\mathbf{x}_{it}$ is a $k \times 1$ vector of observed regressors, for individual i at time t, $\mathbf{f}_t$ is an $m_f \times 1$ vector of unobservable common factors ($m_f < n$) and e_{it} is the i^{th} element of the disturbance vector, $\mathbf{e}_t$. The primary object to be estimated is the mean of the β_i coefficients. To allow for both spatial and serial autocorrelation in $\mathbf{e}_t$ the fixed matrix $\mathbf{R}_t$ is introduced, and the stationary process ε_t such that

$$\mathbf{e}_t = \mathbf{R}_t\varepsilon_t$$
$$\varepsilon_{it} = \sum_{s=0}^{\infty} a_{is}\epsilon_{i,t-s}$$

with $\epsilon_{is} \sim IID(0,1)$ with finite $4th$ moments. Evidently, the YJL and the Pesaran *et al.* models are different but related. Since (4) is a final form equation, their connections and differences can be seen by comparing it with the final form of YJL, (3). First consider the treatment of unobservables. In (4) both the disturbance, $\mathbf{e}_t$ and the m_f-dimensional dynamic factors, $\mathbf{f}_t$ are unobserved, and in practice, the latter are proxied by augmenting the right-hand-side with cross-section means of both y and $\mathbf{x}$ in order that the mean of the β_i may be estimated. Furthermore, there are two possible sources of spatial dependence in the unobservables: via the factor loadings, γ_i' and via $\mathbf{R}_t$. In YJL's treatment of (3) on the other hand, the common $\mathbf{Y}_{n,t}^{\alpha}$ sequence is eliminated by subtraction of cross-section means, and the spatial dependence is introduced via a cross-sectional autoregression in the observables. Because Pesaran and Tosetti's paper is mostly addressed to estimation of and inference about the mean of the β_i, the presence of possible spatial correlation in the unobservables is essentially an inconvenience. Non-stationary dynamics may appear in the errors (in the unobserved $\mathbf{f}_t$ which are merely proxied by cross-sectional means) or in the common observed effects, $\mathbf{d}_t$. In this set-up, cointegration across space, in the sense of possible existence of (linear) combinations of the y_{it} that are stationary even when the y_{it} themselves have unit roots, is of no particular interest because, "the nature of the factors does not matter for inferential analysis of the coefficients of the observed variables." (Kapetanios et al 2011, p. 327) and it is these coefficients that are the objects of interest. Indeed, as is clear from (4), if the exogenous regressors, $\mathbf{x}_{it}$ are assumed to be stable, (in Pesaran and Tosetti their sums of squares converge at rate T in time and n in space) then cross-sectional cointegration requires the existence of vectors, $\mathbf{g}$ such that $\mathbf{g}'\mathbf{Y}_t$ depends only on $\mathbf{X}_t$ and stable components of $\mathbf{d}_t$ and $\mathbf{f}_t$ which have been left unspecified.

However, potentially non-stationary dynamics arising from the combination of spatial and temporal dependence are centre-stage in the discussion of YJL, whose model is therefore necessarily more tightly structured. This is apparent from (3), in which the spatial weights matrix explicitly defines the cointegration space. Notice, however, that the dynamic structure in (2) is quite restrictive, being a first-order spatial VAR(X). In sum, then, it is not easy to compare these approaches as they have different purposes. However, when the β_i of (4) are homogeneous, $\beta_i = \beta$, and the $\gamma_i = \mathbf{0}$ for all i but the shocks are spatially correlated an interesting special case emerges. Pesaran and Tosetti (2011, p. 186, Theorems 3 and 4) give the asymptotic distributions of the mean group and pooled estimators of β under these conditions. These asymptotic distributions are of course affected by the presence of the spatial and serial correlation in the shocks, and the estimators that do not exploit the spatial structure will not be fully efficient. To construct valid inference, the covariance matrix of $\hat{\beta}$ must use the spatial heteroscedasticity and autocorrelation consistent (SHAC) estimator of Kelejian and Prucha (2007), or some other method that accounts for the spatial structure, as Pesaran and Tosetti observe. Thus it is not possible to avoid entirely the need to specify or estimate a spatial structure, and given this, when both cross-section and time dimension are small, it could be important not to lose efficiency. Our model to be introduced in Section 3 is formally a special case of that described by Pesaran and Tosetti. However, our treatment of the spatial structure has more in common with the approach of YJL. Rather than handling the stationary dynamics and spatial dependence non-parametrically, essentially only to permit inference about β, we are interested in estimating the dynamic and spatial dependence parameters themselves; in this respect our approach is similar in spirit to that of Moscone et al (2014) but for the fact that in our model the cross-section dimension is finite, leading to a different treatment of the asymptotics.

Our model is built around an unobserved common trend. To allow for uncertainty about the existence of a common stochastic trend in city-level CPI documented by Chen and Devereux (2003) and potential changes in the mean rate of inflation, as discussed for example by Bierens (2000), we employ a general specification that permits the trend to be stochastic or deterministic with possible breaks or other nonlinearities. Using a vector error correction model (VECM) representation for the deviation of city-specific inflation rates from the trend, the common trend may be removed from the model, in similar fashion to YJL. We allow for spatially correlated idiosyncratic shocks, and, in pursuit of efficiency, first describe an infeasible GLS procedure for estimating the slope parameters of the model, employing a within estimator. Since the GLS procedure is infeasible when the spatial correlation parameter is not known, we then incorporate the GLS moment conditions into a GMM framework and estimate all parameters simultaneously.

Our estimate of the spatial dependence parameter in the model is significantly different from zero, confirming that relative distance has a strong influence on short run price dynamics. Location therefore should not be neglected in an analysis of city-level price movements. We find that prices are slowly converging to an equilibrium, and shocks to city prices have half-life of approximately nine years, in agreement with Cecchetti, Mark and Sonora (2002).

3 The model

The model we propose is built around a logarithm of a price level, p_t^*, which is not directly observed but which has time-series properties that characterize the movements of the n observable series p_{it}, the logarithm of the price index for city i at time t ($n = 17$

for our data). This is described by the equation

$$p_{it} = p_t^* + c_i + z_{it}, \qquad i = 1, \ldots, n, \, t = 1, \ldots, T, \tag{5}$$

where z_{it} are $I(0)$ processes for which $E(z_{it}) = 0$ for all t and $i = 1, \ldots, n$ and such that, for $\mathbf{z}_t = (z_{1,t}, \ldots, z_{n,t})'$, the matrix $E(\mathbf{z}_t \mathbf{z}_t')$ is positive definite. We also assume that the fixed effects sum to zero, $\sum_{i=1}^n c_i = 0$.

Equation (5) implies that the expected growth rate of prices is shared across cities, $E(\Delta p_{it}) = E(\Delta p_t^*)$. The departures from the price level p_t^*, $p_{it} - p_t^*$, follow a set of stationary equilibrium-correction model equations,

$$\Delta (p_{it} - p_t^*) = \alpha_i + \beta_1 \Delta \left(p_{it-1} - p_{t-1}^*\right) + \ldots + \beta_{k-1} \Delta \left(p_{it-k+1} - p_{t-k+1}^*\right) - \gamma \left(p_{it-1} - p_{t-1}^*\right) + u_{it} \tag{6}$$

for $i = 1, \ldots, n$, where $\alpha_i = \gamma c_i$. Coefficient γ measures the speed of adjustment to the equilibrium $p_t^* + c_i$.

The vector of shocks in the VECM form (6), $\mathbf{u}_t = (u_{1t}, \ldots, u_{nt})'$ follows the first order spatial autoregression

$$\mathbf{u}_t = \rho \mathbf{W} \mathbf{u}_t + \mathbf{v}_t, \qquad t = 1, \ldots T, \tag{7}$$

where $\mathbf{W}$ is a known weighting matrix and $\mathbf{v}_t$ is an independent, identically distributed vector process with $E(\mathbf{v}_t) = 0$ and $E(\mathbf{v}_t \mathbf{v}_t') = \sigma^2 \mathbf{I}_n$. The weight matrix $\mathbf{W} = \{w_{ij}\}$ is obtained by row-normalizing a symmetric matrix with non-negative components and with zero diagonal, so that $\sum_{j=1}^n w_{ij} = 1$ for $i = 1, \ldots, n$, $w_{ij} \geq 0$ and $w_{ii} = 0$ for $i = 1, \ldots, n$.

The trend in (5) is generic. A leading example is the case of a common stochastic trend,

$$p_t^* = c + p_{t-1}^* + z_t^*, \qquad t = 1, \ldots, T,$$

say, where p_0^* is a finite random variable and where the increment, z_t^*, is a zero mean $I(0)$ process, that is a process whose spectral density is finite and strictly positive. In this case the elements of the $n+1$ dimensional vector $(p_{1t}, ..., p_{nt}, p_t^*)'$ are cointegrated and the cointegration rank is n and the ECM representation (6) can be replaced by

$$\Delta p_{it} = \alpha_i + \beta_1 \Delta p_{it-1} + \ldots + \beta_{k-1} \Delta p_{it-k+1} - \gamma \left(p_{it-1} - p_{t-1}^*\right) + u_{it}, \quad i = 1, \ldots, n.$$

A similar model has been employed by Hall et al. (1992) in their analysis of the term structure of US treasury bills.

The term p_t^* may also have a common linear trend, $p_t^* = c_0 + ct + z_t^*$, as discussed by Chapman and Ogaki (1993). In this case, the series $p_{1t}, ..., p_{nt}$ are cotrending. Other polynomial trends, such as quadratic trends, are possible. The price level p_t^* may also be subject to a break, $p_t^* = c_0 + c^* \mathbb{I}(t \geq T_a + 1) + z_t^*$, where $c^* \neq 0$, and $p_{1t}, ..., p_{nt}$ are then cobreaking, a concept examined by Hendry and Massmann (2007). More generally, p_t^* may be characterized as a non-linear, non-parametric trend, $p_t^* = c(t/T) + z_t^*$, a case analyzed by Bierens (2000).

All the examples cited above suggest that $(p_{1t}, ..., p_{nt}, p_t^*)'$ may have a common feature, as defined in Engle and Kozicki (1993). See Urga (2007) for further examples. In our case, however, it is not strictly necessary that p_t^* is a common feature. Price level p_t^* may be an unobserved common factor, as previously noted, and $p_t^* = 0$ is also admitted.

When p_t^* characterizes long term dynamics such as a stochastic or a deterministic trend, possibly with breaks or other nonlinearities, $p_t^* + c_i$ can be seen as an equilibrium level, and (5) implies that departures from long run equilibrium, z_{it}, are short-lived and that the long run dynamics of p_{it} are driven by the trend p_t^*.

Notice also that p_t^* cannot be a weighted average of the prices of the various cities. Had p_t^* been a weighted average of the prices of the various cities, $E(\mathbf{z}_t \mathbf{z}_t')$ would not be positive definite and this would conflict with the belief that each city has an idiosyncratic inflation component. In fact, we prefer to view p_t^* as a shared price trend. As Beck et al. (2009) argue, this trend could be determined by national monetary policy as well as by international factors such as oil price and exchange rate dynamics.

Equation (6) is a standard VECM, and coefficient γ can be used to compute the half-life of a shock which is defined as $-(\ln 2) / \ln(1+\gamma)$.

We include the fixed effects in (5) and then in (6) for two reasons. First, the price indices obtained from the Bureau of Labor Statistics measure relative city price levels, that is the CPI for each city has the same base year. This means the CPI series has been multiplied by an individual constant for each city. An additive constant in the model in logarithms controls for the arbitrary base year. Second, the fixed effects account for the heterogeneity among cities leading to long-term differences in relative prices.

In (7) we explicitly assume that the disturbances are spatially autocorrelated, further assumptions to ensure invertibility of $\mathbf{I} - \rho \mathbf{W}$ being in Assumption 3 below.

With the model in place, we wish to estimate the parameters in equations (6) and in (7). If p_t^* was observable, we could estimate the parameters directly. Since p_t^* is unobservable, in order to estimate the parameters in (6) and in (7) we must either approximate p_t^* or eliminate it. If the number of observational units is increasing, $n \rightarrow$

∞, it is possible to estimate p_t^* consistently by a cross-sectional average (as is done implicitly in the common factors approach). For example, we have from (5) that

$$\frac{1}{n}\sum_{i=1}^{n} p_{it} = p_t^* + \frac{1}{n}\sum_{i=1}^{n} z_{it}$$

in which $n^{-1}\sum\limits_{i=1}^{n} z_{it} \xrightarrow{p} 0$ by a law of large numbers. In this approach the condition that $n \to \infty$ is necessary. In general, panel data techniques for the treatment of a common factor also require $n \to \infty$.

On the other hand, and consistently with the dimension of our dataset, in which $n = 17$ and $T = 94$, we view n as fixed. Since p_t^* is present in each of the n equations in (6), it may be removed by subtracting a weighted average of the equations from each equation. Stacking all the equations in (6) as an equation for an n dimensional vector, we take an $n \times n$ matrix $\mathbf{M}$ of rank $n-1$ with the property that $\mathbf{M1} = \mathbf{0}$, where $\mathbf{1} = (1, \ldots, 1)'$, and premultiply both sides of (6) by $\mathbf{M}$, obtaining an estimable equation

$$\mathbf{M\Delta p}_t = \mathbf{M}\alpha + \beta_1 \mathbf{M\Delta p}_{t-1} + \ldots + \beta_{k-1}\mathbf{M\Delta p}_{t-k+1} - \gamma \mathbf{Mp}_{t-1} + \mathbf{Mu}_t \qquad (8)$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)'$, $\mathbf{p}_t = (p_{1t}, \ldots, p_{nt})'$.

Examples of eligible matrices are $\mathbf{M} = \mathbf{M_1} = \mathbf{I} - \frac{1}{n}\mathbf{11}'$, where the weights across cities are equal, and $\mathbf{M} = \mathbf{I} - \mathbf{W}$, where the weights are given by the weighting matrix $\mathbf{W}$.

In the next Section, we estimate our parameters of interest, β_1, ..., β_{k-1}, γ, ρ by GMM. For this purpose, we also assume:

Assumption 1 Let p_{it} be as in (5). The fixed effects are such that $\sum_{i=1}^{n} c_i = 0$, whereas the idiosyncratic shock, z_{it}, is an $I(0)$ process, with $E(\mathbf{z}_t) = \mathbf{0}$ and $E(\mathbf{z}_t\mathbf{z}_t')$ positive definite.

Assumption 2 The price deviations follow the model (6). For $\phi_1, \ldots, \phi_k$ defined as

$$\phi_1 = 1 + \beta_1 - \gamma, \qquad \phi_s = \beta_s - \beta_{s-1}, \quad s = 2, \ldots, k-1, \qquad \phi_k = -\beta_{k-1} \qquad (9)$$

the solutions of $1 - \phi_1 z - \ldots - \phi_k z^k = 0$ are outside the unit circle.

Assumption 3 The vector of shocks $\mathbf{u}_t$ follows the model (7), where $\mathbf{W}$ has elements w_{ij} so that $\sum_{j=1}^{n} w_{ij} = 1$ for $i = 1, \ldots, n$, $w_{ij} \geq 0$ and $w_{ii} = 0$ for $i = 1, \ldots, n$, and $-1/|\lambda(W)_{\min}| < \rho < 1$ where $\lambda(W)_{\min}$ is the smallest eigenvalue of $\mathbf{W}$. The innovations $\mathbf{v}_t$ follow an independent, identically distributed vector process with mutually independent components, $E(\mathbf{v}_t) = 0$ and $E(\mathbf{v}_t\mathbf{v}_t') = \sigma^2\mathbf{I}_n$, and cumulant

$$\operatorname{cum}(v_{it}, v_{jt}, v_{kt}, v_{lt}) = \begin{cases} \kappa < \infty & i = j = k = l, \\ 0 & \text{otherwise.} \end{cases}$$

(cumulants and moments are related, and note in particular that $\operatorname{cum}(v_{it}, v_{it}, v_{it}, v_{it}) = E(v_{it}^4) - 3E(v_{it}^2))$.

4 Estimation method

4.1 Infeasible fully efficient GLS

Since the matrix $\mathbf{M}$ is singular, premultiplying both sides of (6) by $\mathbf{M}$ induces a spatial moving average unit root which cannot be eliminated by inversion. This effect is akin to the effect of over-differencing a time series. As a consequence, while the parameters of (8) can be estimated consistently by OLS under $T \to \infty$, the OLS estimator is not efficient. Let $\mathbf{B} = (\mathbf{I} - \rho\mathbf{W})^{-1}$. The variance matrix of the error term in (8) is $\operatorname{var}\mathbf{M}\mathbf{u}_t = \sigma^2\mathbf{M}\mathbf{B}\mathbf{B}'\mathbf{M}' \neq \sigma^2\mathbf{I}$. Theorem 8 of Magnus and Neudecker (1999, p. 272–273) implies that the best unbiased linear estimator of parameters in (8) can be obtained by premultiplying (8) by the Moore-Penrose generalized inverse $(\mathbf{MB})^+$ of matrix $\mathbf{MB}$,

$$\begin{aligned}(\mathbf{MB})^+\mathbf{M}\boldsymbol{\Delta}\mathbf{p}_t = (\mathbf{MB})^+\mathbf{M}\alpha + \beta_1(\mathbf{MB})^+\mathbf{M}\boldsymbol{\Delta}\mathbf{p}_{t-1} + \ldots + \beta_{k-1}(\mathbf{MB})^+\mathbf{M}\boldsymbol{\Delta}\mathbf{p}_{t-k+1} \\ - \gamma(\mathbf{MB})^+\mathbf{M}\mathbf{p}_{t-1} + (\mathbf{MB})^+\mathbf{M}\mathbf{u}_t, \qquad (10)\end{aligned}$$

and estimating the parameters of (10) by OLS. For example, when $\mathbf{M} = \mathbf{M_1}\mathbf{B}^{-1}$ then $(\mathbf{MB})^+\mathbf{M} = \mathbf{M_1}\mathbf{B}^{-1}$, or when $\mathbf{M} = \mathbf{M_1}$ then $(\mathbf{MB})^+\mathbf{M} = (\mathbf{M_1}\mathbf{B})^+\mathbf{M_1}$.

The two steps of transforming equation (6) and estimating the transformed equation (10) by OLS can be seen as a generalized least squares (GLS) procedure in a model where the covariance matrix of errors is singular. The variance of the term $(\mathbf{MB})^+\mathbf{M}\mathbf{u}_t$ in (10) is $\operatorname{var}\left((\mathbf{MB})^+\mathbf{M}\mathbf{u}_t\right) = \sigma^2(\mathbf{MB})^+\mathbf{MB}(\mathbf{MB})'(\mathbf{MB})^{+\prime} = \sigma^2(\mathbf{MB})^+\mathbf{MB} = \sigma^2\mathbf{M_1}$ independently of matrix $\mathbf{M}$. The last equality can be seen by noting that the matrix $\mathbf{MB}$ is of rank $n-1$ and has a singular value decomposition $\mathbf{MB} = \mathbf{R}\boldsymbol{\Lambda}\mathbf{T}'$ where $\mathbf{R}$ and $\mathbf{T}$ are $n\times(n-1)$ matrices such that $\mathbf{R}'\mathbf{R} = \mathbf{T}'\mathbf{T} = \mathbf{I}_{n-1}$ and $\boldsymbol{\Lambda}$ is an $(n-1)\times(n-1)$ diagonal matrix with positive diagonal elements. The fact that $\mathbf{MB1} = \mathbf{0}$ implies that $\mathbf{T}'\mathbf{1} = \mathbf{0}$ and so $\mathbf{TT}' = \mathbf{M_1}$, therefore indeed $(\mathbf{MB})^+\mathbf{MB} = \mathbf{T}\boldsymbol{\Lambda}^{-1}\mathbf{R}'\mathbf{R}\boldsymbol{\Lambda}\mathbf{T}' = \mathbf{TT}' = \mathbf{M_1}$.

To implement this approach, let $\mathbf{G} = (\mathbf{MB})^+\mathbf{M}$, let $\mathbf{X}_t = \left(\boldsymbol{\Delta}\mathbf{p}_{t-1}, \ldots, \boldsymbol{\Delta}\mathbf{p}_{t-k+1}, -\mathbf{p}_{t-1}\right)$

and write $\beta = (\beta_1, \ldots, \beta_{k-1}, \gamma)'$ to write model (10) as

$$\mathbf{G\Delta p}_t = \mathbf{G}\alpha + \mathbf{GX}_t\beta + \mathbf{Gu}_t.$$

Vector α contains n parameters that are usually not of primary interest, so in order to obtain a within estimator of parameters $\beta_1, \ldots, \beta_{k-1}$ and γ in (10), we subtract time averages from both sides of this equation and obtain

$$\mathbf{\Delta G}\widetilde{\mathbf{p}}_t = \beta_1\mathbf{\Delta G}\widetilde{\mathbf{p}}_{t-1} + \ldots + \beta_{k-1}\mathbf{\Delta G}\widetilde{\mathbf{p}}_{t-k+1} - \gamma\mathbf{G}\widetilde{\mathbf{p}}_{t-1} + \mathbf{G}\widetilde{\mathbf{u}}_t, \qquad t = k+1, \ldots, T, \quad (11)$$

where here and in what follows, for generic $n \times m$ matrices $\mathbf{Y}_{k+1}, \ldots, \mathbf{Y}_T$, matrix $\widetilde{\mathbf{Y}}_t$ is defined as

$$\widetilde{\mathbf{Y}}_t = \mathbf{Y}_t - \frac{1}{T-k}\sum_{t=k+1}^{T}\mathbf{Y}_t, \qquad t = k+1, \ldots, T.$$

Denoting $\check{\mathbf{Y}}_t = \mathbf{G}\widetilde{\mathbf{Y}}_t$, equation (11) can be written as

$$\mathbf{\Delta}\check{\mathbf{p}}_t = \beta_1\mathbf{\Delta}\check{\mathbf{p}}_{t-1} + \ldots + \beta_{k-1}\mathbf{\Delta}\check{\mathbf{p}}_{t-k+1} - \gamma\check{\mathbf{p}}_{t-1} + \check{\mathbf{u}}_t, \qquad t = k+1, \ldots, T. \quad (12)$$

The OLS estimator of $\beta_1, \ldots, \beta_{k-1}$ and γ in (12) is consistent and efficient as long as n is finite and $T \to \infty$. However, the above estimator is infeasible because ρ is not known.

4.2 Feasible GMM estimation

We can however estimate parameters $\beta_1, \ldots, \beta_{k-1}, \gamma, \rho$ and σ^2 simultaneously using the generalized method of moments (GMM). Note that for any $n \times n$ matrix $\mathbf{\Omega}$ such that $\mathbf{1}'\mathbf{\Omega} = \mathbf{0}$ and $\mathbf{\Omega 1} = \mathbf{0}$, the following moment conditions hold,

$$E\left(\widetilde{\mathbf{X}}_t'\mathbf{G}'\mathbf{G}\widetilde{\mathbf{u}}_t\right) = O\left(T^{-1}\right),$$
$$E\left(\widetilde{\mathbf{u}}_t'\mathbf{\Omega}\widetilde{\mathbf{u}}_t\right) = \sigma^2 \operatorname{tr}\left(\mathbf{B}'\mathbf{\Omega B}\right)\left(1 - \frac{1}{T-k}\right).$$

Let $\theta_0 = (\beta_1, \ldots, \beta_{k-1}, \gamma, \rho, \sigma^2)'$ be the vector of true values of parameters in the model and let $\theta = (b_1, \ldots, b_{k-1}, g, r, s^2)'$. Define

$$\mathbf{u}_t(\theta) = \mathbf{\Delta p}_t - \alpha - b_1\mathbf{\Delta p}_{t-1} - \ldots - b_{k-1}\mathbf{\Delta p}_{t-k+1} + g(\mathbf{p}_{t-1} - p^*_{t-1}\mathbf{1})$$

so that

$$\mathbf{M}\widetilde{\mathbf{u}}_t(\theta) = \mathbf{M}\boldsymbol{\Delta}\widetilde{\mathbf{p}}_t - b_1\mathbf{M}\boldsymbol{\Delta}\widetilde{\mathbf{p}}_{t-1} - \ldots - b_{k-1}\mathbf{M}\boldsymbol{\Delta}\widetilde{\mathbf{p}}_{t-k+1} + g\mathbf{M}\widetilde{\mathbf{p}}_{t-1}$$

and $\mathbf{M}\widetilde{\mathbf{u}}_t(\theta_0) = \mathbf{M}\mathbf{B}\widetilde{\mathbf{v}}_t$ for any matrix $\mathbf{M}$ such that $\mathbf{M}\mathbf{1} = \mathbf{0}$. While errors $\mathbf{u}_t(\theta)$ are unobservable, errors $\mathbf{M}\widetilde{\mathbf{u}}_t(\theta)$ can be observed. Noting that $\mathbf{G}'\mathbf{G} = (\mathbf{B}^{-1})'\mathbf{M}_{\mathbf{1}}\mathbf{B}^{-1}$ for any $\mathbf{M}$ such that $\mathbf{M}\mathbf{1} = \mathbf{0}$, we let $\mathbf{B}(r) = (\mathbf{I} - r\mathbf{W})^{-1}$ and define moment function $\widetilde{\mathbf{m}}_t$

$$\widetilde{\mathbf{m}}_t(\theta) = \begin{pmatrix} \widetilde{\mathbf{X}}_t'(\mathbf{B}^{-1}(r))'\mathbf{M}_{\mathbf{1}}\mathbf{B}^{-1}(r)\widetilde{\mathbf{u}}_t(\theta) \\ \widetilde{\mathbf{u}}_t'(\theta)\boldsymbol{\Omega}_1\widetilde{\mathbf{u}}_t(\theta) - s^2\operatorname{tr}\left(\mathbf{B}(r)'\boldsymbol{\Omega}_1\mathbf{B}(r)\right)\left(1 - \frac{1}{T-k}\right) \\ \vdots \\ \widetilde{\mathbf{u}}_t'(\theta)\boldsymbol{\Omega}_q\widetilde{\mathbf{u}}_t(\theta) - s^2\operatorname{tr}\left(\mathbf{B}(r)'\boldsymbol{\Omega}_q\mathbf{B}(r)\right)\left(1 - \frac{1}{T-k}\right) \end{pmatrix}$$

for $t = k+1, \ldots, T$, where $q \geq 2$ and $\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_q$ are $n \times n$ matrices such that $\mathbf{1}'\boldsymbol{\Omega}_i = \mathbf{0}$ and $\boldsymbol{\Omega}_i\mathbf{1} = \mathbf{0}$, $i = 1, \ldots, q$.

Let

$$\widetilde{\mathbf{S}}(\theta) = \frac{1}{T-k}\sum_{t=k+1}^{T}\widetilde{\mathbf{m}}_t(\theta)\widetilde{\mathbf{m}}_t(\theta)',$$
$$\widetilde{\mathbf{D}}(\theta) = \frac{1}{T-k}\sum_{t=k+1}^{T}\frac{\partial\widetilde{\mathbf{m}}_t(\theta)}{\partial\theta'},$$

and denote $\mathbf{S} = \operatorname{plim}_{T\to\infty}\widetilde{\mathbf{S}}(\theta_0)$ and $\mathbf{D} = \operatorname{plim}_{T\to\infty}\widetilde{\mathbf{D}}(\theta_0)$. Let Θ be a set such that $\Theta \subset \mathbb{R}^{k+2}$. For $\theta \in \Theta$, we define loss function $\widetilde{q}_{\boldsymbol{\Sigma}}$ as

$$\widetilde{q}_{\boldsymbol{\Sigma}}(\theta) = \left(\frac{1}{T-k}\sum_{t=k+1}^{T}\widetilde{\mathbf{m}}_t(\theta)'\right)\boldsymbol{\Sigma}_T\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\widetilde{\mathbf{m}}_t(\theta)\right)$$

where $\boldsymbol{\Sigma}_T$ is a weighting matrix that may depend on data and on sample size T. We define the GMM estimator of θ_0 as

$$\hat{\theta} = \arg\min_{\theta\in\Theta}\widetilde{q}_{\boldsymbol{\Sigma}}(\theta) \tag{13}$$

We introduce the following assumption:

Assumption 4 $\mathbf{S}$ is positive definite and $\mathbf{D}$ has full column rank, θ_0 is interior to Θ, Θ is compact, $E\mathbf{m}_t(\theta) = 0$ only if $\theta = \theta_0$ and $\boldsymbol{\Sigma}_T \xrightarrow{p} \boldsymbol{\Sigma}$ where matrices $\boldsymbol{\Sigma}_T$ and $\boldsymbol{\Sigma}$ are positive definite.

Assumption 4 ensures that $\mathbf{S}^{-1}$ and $(\mathbf{D}'\boldsymbol{\Sigma}\mathbf{D})^{-1}$ exist, that function $\operatorname{plim}_{T\to\infty}\widetilde{q}_{\boldsymbol{\Sigma}}$ has a unique minimum, and that $\sqrt{T}\left(\hat{\theta}-\theta_0\right)$ has limit normal distribution. Then, as stated in Theorem 1 below, $\sqrt{T}\left(\hat{\theta}-\theta_0\right)$ is asymptotically normal with covariance matrix $(\mathbf{D}'\mathbf{\Sigma}\mathbf{D})^{-1}\mathbf{D}'\mathbf{\Sigma}\mathbf{S}\mathbf{\Sigma}\mathbf{D}(\mathbf{D}'\mathbf{\Sigma}\mathbf{D})^{-1}$. The asymptotic variance matrix of $\sqrt{T}\left(\hat{\theta}-\theta_0\right)$ is minimized when $\mathbf{\Sigma}=\mathbf{S}^{-1}$. In practice, optimal GMM estimation requires a preliminary consistent estimation of $\mathbf{S}$. We estimate $\mathbf{S}$ by $\hat{\mathbf{S}}=\widetilde{\mathbf{S}}\left(\widetilde{\theta}\right)$ where

$$\widetilde{\theta}=\arg\min_{\theta\in\Theta}\widetilde{q}_{\mathbf{I}}(\theta)=\arg\min_{\theta\in\Theta}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\widetilde{\mathbf{m}}_t(\theta)'\right)\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\widetilde{\mathbf{m}}_t(\theta)\right).$$

The asymptotic properties of estimator $\hat{\theta}$ are summarized in the following theorem.

Theorem 1 *Let $\hat{\theta}$ be the GMM estimator defined in (13). Under Assumptions 1–4, as $T\to\infty$,*

$$\sqrt{T}\left(\hat{\theta}-\theta_0\right)\xrightarrow{d}N\left(\mathbf{0},(\mathbf{D}'\mathbf{\Sigma}\mathbf{D})^{-1}\mathbf{D}'\mathbf{\Sigma}\mathbf{S}\mathbf{\Sigma}\mathbf{D}(\mathbf{D}'\mathbf{\Sigma}\mathbf{D})^{-1}\right).$$

When the weighting matrix is $\mathbf{\Sigma}_T=\widetilde{\mathbf{S}}^{-1}\left(\widetilde{\theta}\right)$ then

$$\sqrt{T}\left(\hat{\theta}-\theta_0\right)\xrightarrow{d}N\left(\mathbf{0},(\mathbf{D}'\mathbf{\Sigma}\mathbf{D})^{-1}\right).$$

In order to obtain critical values for the asymptotic distribution of $\hat{\theta}$, matrix $\mathbf{D}$ can be estimated by $\hat{\mathbf{D}}=\widetilde{\mathbf{D}}\left(\hat{\theta}\right)=(T-k)^{-1}\sum_{t=k+1}^{T}\frac{\partial\widetilde{\mathbf{m}}_t(\hat{\theta})}{\partial\theta'}$ where

$$\frac{\partial\widetilde{\mathbf{m}}_t(\theta)}{\partial\theta'}=-\begin{pmatrix}\widetilde{\mathbf{X}}_t'\left(\mathbf{B}^{-1}(r)\right)'\mathbf{M_1}\mathbf{B}^{-1}(r)\widetilde{\mathbf{X}}_t & \widetilde{\mathbf{X}}_t'\mathbf{H}(r)\widetilde{\mathbf{u}}_t(\theta) & \mathbf{0}\\ \widetilde{\mathbf{u}}_t'(\theta)\left(\mathbf{\Omega}_1+\mathbf{\Omega}_1'\right)\widetilde{\mathbf{X}}_t & s^2C_1(r) & E_1(r)\\ \vdots & \vdots & \vdots\\ \widetilde{\mathbf{u}}_t'(\theta)\left(\mathbf{\Omega}_q+\mathbf{\Omega}_q'\right)\widetilde{\mathbf{X}}_t & s^2C_q(r) & E_q(r)\end{pmatrix}$$

and

$$\begin{aligned}\mathbf{H}(r)&=-\frac{\partial}{\partial r}\left(\left(\mathbf{B}^{-1}(r)\right)'\mathbf{M_1}\mathbf{B}^{-1}(r)\right)=\mathbf{W}'\mathbf{M_1}(\mathbf{I}-r\mathbf{W})+(\mathbf{I}-r\mathbf{W}')\mathbf{M_1}\mathbf{W},\\ C_i(r)&=-\frac{\partial}{\partial r}\operatorname{tr}\left(\mathbf{B}(r)'\mathbf{\Omega}_i\mathbf{B}(r)\right)\left(1-\frac{1}{T-k}\right),\\ &=-\operatorname{tr}\left(\mathbf{B}(r)'\left(\mathbf{\Omega}_i+\mathbf{\Omega}_i'\right)\mathbf{B}(r)\mathbf{W}\mathbf{B}(r)\right)\left(1-\frac{1}{T-k}\right),\\ E_i(r)&=\operatorname{tr}\left(\mathbf{B}(r)'\mathbf{\Omega}_i\mathbf{B}(r)\right)\left(1-\frac{1}{T-k}\right),\qquad i=1,\ldots,q.\end{aligned}$$

The following theorem shows that matrices $\widetilde{\mathbf{S}}\left(\widetilde{\theta}\right)$ and $\widetilde{\mathbf{D}}\left(\widetilde{\theta}\right)$ are consistent estimators of $\mathbf{S}$ and $\mathbf{D}$.

Theorem 2 *Under Assumptions 1–4,*

$$\widetilde{\mathbf{S}}\left(\widetilde{\theta}\right) \xrightarrow{p} \mathbf{S} \qquad \textit{and} \qquad \widetilde{\mathbf{D}}\left(\widetilde{\theta}\right) \xrightarrow{p} \mathbf{D}.$$

Proofs of Theorems 1 and 2 can be found in Section 7.

When the number of moment conditions exceeds the number of parameters that are being estimated, the model is overidentified and it is possible to test whether the corresponding sample moments $(T-k)^{-1}\sum_{t=k+1}^{T}\widetilde{m}_t\left(\hat{\theta}\right)$ are statistically close to 0, see for example Newey and McFadden (1994, p. 2231). If there are ℓ overidentifying moments in the moment function $\widetilde{m}_t$, then under the null that $E\left(\widetilde{\mathbf{m}}_t\left(\theta_0\right)\right) = 0$ and Assumptions 1–4,

$$J = (T-k)\,\widetilde{q}_{\Sigma}\left(\hat{\theta}\right) \xrightarrow{d} \chi^2_{\ell}.$$

In the remainder of the paper, we refer to this test as the overidentification test.

5 Empirical results

The data employed are annual observations of CPI in 17 cities in the US: Atlanta, Boston, Chicago, Cincinnati, Cleveland, Detroit, Houston, Kansas City, Los Angeles, Minneapolis, New York, Philadelphia, Pittsburgh, Portland, San Francisco, Seattle, St. Louis. The data are from the Bureau of Labor Statistics (http://www.bls.gov) and span the years 1918-2011 inclusive, for a total of 94 observations of each series.

The time path of the logarithms of the series is shown in Figure 1.

Figure 1 here

Figure 2 here

The series appear to trend upwards. The slope of the trend may have changed over time and it may have been steeper during World War II and during the oil shock. In Figure 2, which plots changes of the logarithm as approximation of inflation, it can be seen that in some periods, for example the years of World War II and the oil price shock, inflation has been higher than usual.

The price level series seem to follow a single common trend. If prices share a common trend and departures from this trend are stationary, then centering the price series around that trend should render all series stationary. We estimate the trend as the average of logarithms of prices for each period and subtract the estimated trend from each price series. Figure 3 displays the centered series. The panel data unit root test of Im et al. (2003) applied to the panel of centered series strongly rejects the null hypothesis of unit root. The plot of the series and the result of the unit root test suggest that a model in which there is a single common trend underlying all price series is highly plausible.

Figure 3 here

We define the matrix $\mathbf{W}$ using distances in minutes between cities, taken from *googlemap*™. The table of distances between the cities is shown in Table 3. We model the spatial weights as declining with the inverse squared distance between cities. Denoting the distance in minutes between cities i and j as d_{ij}, we put

$$w_{ij} = \frac{\frac{1}{d_{ij}^2}}{\sum_{j=1}^{n} \frac{1}{d_{ij}^2}}, \qquad i \neq j, \quad i, j = 1, \ldots, n, \tag{14}$$

and $w_{ii} = 0$. This weighting scheme has been used by Ertur and Koch (2007).

We first estimate a given model with five lags of $\boldsymbol{\Delta}\breve{\mathbf{p}}_t$, then drop insignificant lags of $\boldsymbol{\Delta}\breve{\mathbf{p}}_t$ and reestimate the model. We set $q = 3$ and $\boldsymbol{\Omega}_1 = \mathbf{M_1}$, $\boldsymbol{\Omega}_2 = (\mathbf{I} - \mathbf{W})' \mathbf{M_1} (\mathbf{I} - \mathbf{W})$ and $\boldsymbol{\Omega}_3 = \mathbf{M_1} (\mathbf{I} - \mathbf{W})$. We chose $\boldsymbol{\Omega}_i$ primarily focussing on $\mathbf{M}_1$ and combining it with other matrices. The matrix $\mathbf{M}_1$ seems a natural choice, because it generates mean correction as in a within group regression. We considered three matrices $\boldsymbol{\Omega}_i$ to have an overidentified model and therefore to be able to assess the validity of the orthogonality conditions by means of the J test.

Our estimate of model (12) using GMM defined in (13) is

$$\begin{aligned} \boldsymbol{\Delta}\breve{\mathbf{p}}_t &= \underset{(0.029)}{0.252}\boldsymbol{\Delta}\breve{\mathbf{p}}_{t-1} + \underset{(0.028)}{0.066}\boldsymbol{\Delta}\breve{\mathbf{p}}_{t-3} - \underset{(0.009)}{0.067}\breve{\mathbf{p}}_{t-1} + \widehat{\mathbf{u}}_t, \\ \widehat{\mathbf{u}}_t &= \underset{(0.044)}{0.370}\mathbf{W}\ \widehat{\mathbf{u}}_t + \widehat{\mathbf{v}}_t, \qquad \widehat{\sigma^2} = \underset{(1.02\times10^{-6})}{11.06 \times 10^{-6}}, \quad J = 3.236, \end{aligned} \tag{15}$$

where for estimates $\widehat{\beta}_1, \ldots, \widehat{\beta}_{k-1}, \widehat{\gamma}$, we let $\widehat{\mathbf{u}}_t = \Delta\breve{\mathbf{p}}_t - \widehat{\beta}_1 \Delta\breve{\mathbf{p}}_{t-1} - \ldots - \widehat{\beta}_{k-1}\Delta\breve{\mathbf{p}}_{t-k+1} + \widehat{\gamma}\breve{\mathbf{p}}_{t-1}$. Here estimated standard errors of coefficient estimates are reported in parentheses and J is the overidentifying test statistic.

The estimated value of the coefficient γ of $\breve{\mathbf{p}}_{t-1}$ in model (15) is negative and small, indicating slow reversion to the equilibrium implied by the law of one price. The estimated half-life of a shock, computed as $-\ln 2/\ln(1+\widehat{\gamma})$, is therefore just above nine years. This is similar to the half-life of nearly nine years found by Cecchetti et al. (2002, page 1081), and larger than the half-life of almost five years estimated by Chen and Devereux (2003, page 220). The estimate of ρ is significantly different from 0, suggesting that there is a strong spatial effect in the short term dynamics of inflation, in the sense that idiosyncratic shocks tend to be correlated in cities that are closely located.

With $q = 3$, there is one overidentifying moment in the moment function. The asymptotic null distribution of J is therefore χ^2_1 and the null that $E\left(\widetilde{\mathbf{m}}_t\left(\theta\right)\right) = 0$ is not rejected on the 5% significance level.

Parameters $\beta_1, \ldots, \beta_{k-1}$ and γ in model (8) can be estimated by OLS: this is convenient, as no spatial assumption need be imposed in (7) and may thus give a qualitative feel of the reliability of the GMM estimate under that additional assumption. The within OLS estimates with $\mathbf{M} = \mathbf{M_1}$ and $\mathbf{M} = \mathbf{I} - \mathbf{W}$, respectively, are

$$\mathbf{\Delta}\breve{\mathbf{p}}_t = \underset{(0.024)}{0.271}\mathbf{\Delta}\breve{\mathbf{p}}_{t-1} + \underset{(0.023)}{0.084}\mathbf{\Delta}\breve{\mathbf{p}}_{t-3} - \underset{(0.008)}{0.066}\breve{\mathbf{p}}_{t-1} + \widehat{\mathbf{u}}_t, \qquad \widehat{\sigma^2} = 11.70 \times 10^{-6},$$

$$\mathbf{\Delta}\breve{\mathbf{p}}_t = \underset{(0.024)}{0.222}\mathbf{\Delta}\breve{\mathbf{p}}_{t-1} + \underset{(0.023)}{0.046}\mathbf{\Delta}\breve{\mathbf{p}}_{t-3} - \underset{(0.008)}{0.073}\breve{\mathbf{p}}_{t-1} + \widehat{\mathbf{u}}_t, \qquad \widehat{\sigma^2} = 11.79 \times 10^{-6},$$

where the figures in brackets are the standard errors of the estimated coefficients computed assuming mistakenly that there is no correlation of disturbances across cities. Not surprisingly, failing to account for the spatial dependence in the estimation of the standard errors of the estimates results in underestimated standard error of estimated coefficients on $\mathbf{\Delta}\breve{\mathbf{p}}_{t-1}$, $\mathbf{\Delta}\breve{\mathbf{p}}_{t-3}$ and $\breve{\mathbf{p}}_{t-1}$. It can also be seen that the parameter estimates are sensitive to the choice of averaging matrix $\mathbf{M}$ as should be expected.

In the definition of the weight matrix, we assume that the decline of the strength of interaction is proportional to the inverse of square distance. However, the choice of $\mathbf{W}$ is bound to be arbitrary to a degree. We therefore examine the robustness of our results to the choice of $\mathbf{W}$ by verifying that alternative specifications of $\mathbf{W}$ do not lead to substantially different conclusions.

We consider several types of weights employed by practitioners. First, we consider weights declining slower with distance than in (14), namely decreasing with the inverse of distance, $w_{ij} = d_{ij}^{-1} / \left(\sum_{j=1}^{n} d_{ij}^{-1}\right)$ for $i, j = 1, \ldots, n$ and $i \neq j$, as proposed by Robinson (2010). The result of this estimation is reported in column II of Table 1. For comparison, the estimated parameters of the preferred model (15) are summarized in column I of the table. The estimated parameters for β_1, β_3 and γ are very similar in both specifications, although the estimated standard errors are slightly larger when squared distances are used. The spatial correlation parameter $\hat{\rho}$ is larger in column II than in column I and strongly significant. We also examine exponentially decreasing weights considered

	I baseline model	II inverse distance	III exponential decrease	IV nn nearest neighbors	V cut-off distance
$\hat{\beta}_1$	0.252 (0.029)	0.251 (0.027)	0.262 (0.030)	0.261 (0.028)	0.276 (0.033)
$\hat{\beta}_3$	0.066 (0.028)	0.066 (0.025)	0.062 (0.028)	0.075 (0.026)	0.067 (0.026)
$\hat{\gamma}$	-0.067 (0.009)	-0.071 (0.008)	-0.066 (0.009)	-0.071 (0.009)	-0.071 (0.009)
$\hat{\rho}$	0.370 (0.044)	0.841 (0.079)	0.233 (0.031)	0.367 (0.039)	0.394 (0.052)
$\widehat{\sigma^2} \times 10^6$	11.06 (1.02)	12.05 (1.11)	10.89 (1.00)	11.08 (0.93)	11.36 (1.05)
J	3.236	0.283	3.606	1.152	2.435

Table 1: Check of robustness to various specifications of weighting matrix $\mathbf{W}$

by Ertur and Koch (2007), that is, $w_{ij} = \exp(-ad_{ij})/\left(\sum_{j=1}^{n}\exp(-ad_{ij})\right)$ for $i \neq j$. Estimates for the case of $a = 0.2$ are summarized in Column III of Table 1. Again the spatial correlation parameter is significant and estimates $\hat{\beta}_1$, $\hat{\beta}_3$ and $\hat{\gamma}$ are close to their counterparts in the baseline model estimate.

Further, we estimate the model using a weight matrix based on nearest neighbors, with $1/d_{ij}$ set to 1 only for nn nearest neighbors and 0 otherwise. This type of weighting matrix has been employed by Baltagi and Liu (2011), among others. We carry out estimation for $nn = 1, \ldots, 8$. The estimates for $nn = 3$ are reported in Column IV of Table 1. The results for other values of nn are qualitatively similar. In addition, we examine the weighting matrices where only cities within a certain cut-off distance are considered as neighbors. We set $1/d_{ij}$ to 0 if d_{ij} exceeds the cut-off value and to 1 otherwise. We allow the cut-off point to vary between 750 to 1250 minutes of distance. In Column V of Table 1 we report the estimates for the cut-off point set to 1000 minutes. The results here are consistent with the previous results in that the estimated values of parameters β_1, β_3 and γ are similar to the estimates from the baseline model (15) and $\widehat{\rho}$ is significantly different from zero.

In the majority of cases discussed above, the test of overidentifying restrictions cannot reject the null of moment restrictions being satisfied. For cases summarized in Table 1, the J statistic is reported in the last row of the table.

As a final check, since the distances in minutes are taken from googlemap where they are changing frequently, we analyze the robustness to moderate changes in the distances between cities. We generate an $n \times n$ matrix of independent random numbers distributed uniformly on $[0.95, 1.05]$ or $[0.90, 1.10]$ and multiply the matrix of distances d_{ij} by this matrix element by element. Matrix $\mathbf{W}$ is then constructed using (14). We generate 1000 replications of matrix $\mathbf{W}$, each time estimating parameters of the overidentified model (15). Table 2 reports intervals containing 90% and 95% values of parameter estimates, using 5% and 95%, and 2.5% and 97.5% sample quantiles, respectively. Estimated

	original	Sample quantile intervals $\pm 5\%$ perturbation		$\pm 10\%$ perturbation	
	estimates	$5\% - 95\%$	$2.5\% - 97.5\%$	$5\% - 95\%$	$2.5\% - 97.5\%$
$\hat{\beta}_1$	0.252	$(0.251, 0.253)$	$(0.251, 0.253)$	$(0.250, 0.254)$	$(0.250, 0.255)$
$\hat{\beta}_3$	0.066	$(0.065, 0.067)$	$(0.065, 0.068)$	$(0.064, 0.068)$	$(0.064, 0.069)$
$-\hat{\gamma}$	0.067	$(0.067, 0.068)$	$(0.067, 0.068)$	$(0.066, 0.068)$	$(0.066, 0.069)$
$\hat{\rho}$	0.370	$(0.363, 0.377)$	$(0.363, 0.379)$	$(0.357, 0.384)$	$(0.354, 0.386)$
$\hat{\sigma}^2 \times 10^6$	11.06	$(10.95, 11.14)$	$(10.93, 11.15)$	$(10.86, 11.23)$	$(10.81, 11.27)$

Table 2: Check of robustness to $\pm 5\%$ and $\pm 10\%$ perturbation of distances

values of parameters β_1, β_3 and γ tend to be concentrated around their estimates from the model (15). The intervals for $\hat{\rho}$ are slightly wider but still narrow.

Summarizing our results we conclude that the values of estimated parameters may change when variations in the weighting matrix are considered, but the main message remains unchanged. The reversion of city-level prices to an equilibrium is slow and spatial correlation in errors is present.

6 Conclusions and comments

An unobserved common trend model with spatially correlated idiosyncratic shocks was introduced and applied to study relative movements in the CPI of 17 US cities. The model was estimated by GMM. The estimated half-life of a shock is approximately 9 years which is at the upper end of the range that has been reported in the literature. Strong evidence of spatial effects was found. Our estimate of the spatial error autocorrelation parameter, $\hat{\rho} = 0.37$, is significantly different from zero. In cases like this, if spatial structure is ignored, the estimated standard errors routinely associated with OLS regression may underestimate the effective dispersion of the estimates.

The structure we adopt seems to be novel and may have a variety of potential applications not confined to the law of one price literature. The method could be applied in a range of practical situations, including the modelling of income or output in different regions or other cases in which the "distance" affecting the correlation between idiosyncratic shocks may be in a dimension other than space, such as relative maturities applied to a vector of interest rates. The proposed detrending therefore applies in many setups.

7 Proofs of technical results

Before proceeding to prove Theorems 1 and 2, we introduce some additional notation, and note some properties of the moment function $\mathbf{m}_t$.

For any matrix $\mathbf{M}$ such that $\mathbf{M1} = \mathbf{0}$, equation (8) can be rewritten as

$$\mathbf{M}\mathbf{p}_t = \mathbf{M}\alpha + \sum_{\ell=1}^{k} \phi_\ell \mathbf{M}\mathbf{p}_{t-\ell} + \mathbf{M}\mathbf{u}_t$$

where the coefficients $\phi_1, \ldots, \phi_k$ are defined in (9).

By Assumption 2, $z^k - \phi_1 z^{k-1} - \ldots - \phi_{k-1} z - \phi_k \neq 0$ if $|z| \geq 1$, and it follows from Theorem 4' of Hannan (1970, p. 14) that $\mathbf{M}\mathbf{p}_t$ is second order stationary and can be written as

$$\mathbf{M}\mathbf{p}_t = \frac{1}{1 - \phi_1 - \cdots \phi_k} \mathbf{M}\alpha + \sum_{j=0}^{\infty} \psi_j \mathbf{M}\mathbf{u}_{t-j} = \frac{1}{\gamma} \mathbf{M}\alpha + \Psi(L) \mathbf{M}\mathbf{u}_t$$

where $\psi_0 = 1$, $\psi_j = O(c^j)$ for $0 < c < 1$, L is the backshift operator and $\Psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \ldots$.

Let $\mathbf{P}_t$ and $\mathbf{V}_t$ be $n \times (k+1)$ matrices defined as $\mathbf{P}_t = (\mathbf{p}_t, \mathbf{p}_{t-1}, \ldots, \mathbf{p}_{t-k})$, $\mathbf{V}_t = (\mathbf{v}_t, \mathbf{v}_{t-1}, \ldots, \mathbf{v}_{t-k})$, $\mathbf{F}$ be a $(k+1) \times k$ matrix defined as

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & & \vdots & -1 \\ -1 & 1 & & \vdots & 0 \\ & -1 & \ddots & 0 & \vdots \\ & & \ddots & 1 & 0 \\ 0 & & & -1 & 0 \end{pmatrix},$$

and δ and $\mathbf{d}$ be $(k+1) \times 1$ vectors defined as $\delta = (1, -\phi_1, \ldots, -\phi_k)' = (1, -1 - \beta_1 + \gamma, \ldots, \beta_{k-1})'$ and $\mathbf{d} = (1, -1 - b_1 + g, \ldots, b_{k-1})'$.

For any matrix $\mathbf{M}$ such that $\mathbf{M1} = \mathbf{0}$, we write

$$\widetilde{\mathbf{X}}_t = \widetilde{\mathbf{P}}_t \mathbf{F}, \qquad \widetilde{\mathbf{u}}_t(\theta) = \widetilde{\mathbf{P}}_t \mathbf{d}, \qquad \mathbf{M}\widetilde{\mathbf{P}}_t = \mathbf{MB}\Psi(L)\widetilde{\mathbf{V}}_{\mathbf{t}}.$$

We further define

$$\begin{aligned}
\boldsymbol{\Omega}_0(r) &= \left(\mathbf{B}^{-1}(r)\right)' \mathbf{M}_{\mathbf{1}} \mathbf{B}^{-1}(r), \\
\mathbf{A}_0(r) &= \mathbf{B}' \boldsymbol{\Omega}_0(r) \mathbf{B}, \\
\mathbf{A}_i(r) &= \mathbf{B}(r)' \boldsymbol{\Omega}_i \mathbf{B}(r) \qquad\qquad i = 1, \ldots, q,
\end{aligned}$$

and denote $\mathbf{A}_i = \mathbf{A}_i(\rho)$. The moment function $\widetilde{\mathbf{m}}_t$ can now be written as

$$\widetilde{\mathbf{m}}_t(\theta) = \begin{pmatrix} \widetilde{\mathbf{m}}_{0t}(\theta) \\ \widetilde{m}_{1t}(\theta) \\ \widetilde{m}_{2t}(\theta) \\ \widetilde{m}_{3t}(\theta) \end{pmatrix} = \begin{pmatrix} \mathbf{F}' \left(\Psi(L)\widetilde{\mathbf{V}}_t\right)' \mathbf{A}_0(r) \left(\Psi(L)\widetilde{\mathbf{V}}_t\right) \mathbf{d} \\ \mathbf{d}' \left(\Psi(L)\widetilde{\mathbf{V}}_t\right)' \mathbf{A}_1 \left(\Psi(L)\widetilde{\mathbf{V}}_t\right) \mathbf{d} - s^2 \operatorname{tr} \mathbf{A}_1(r) \left(1 - \frac{1}{T-k}\right) \\ \vdots \\ \mathbf{d}' \left(\Psi(L)\widetilde{\mathbf{V}}_t\right)' \mathbf{A}_q \left(\Psi(L)\widetilde{\mathbf{V}}_t\right) \mathbf{d} - s^2 \operatorname{tr} \mathbf{A}_q(r) \left(1 - \frac{1}{T-k}\right) \end{pmatrix}.$$

We define vector function $\mathbf{m}_t$ as

$$\mathbf{m}_t(\theta) = \begin{pmatrix} \mathbf{m}_{0t}(\theta) \\ m_{1t}(\theta) \\ m_{2t}(\theta) \\ m_{3t}(\theta) \end{pmatrix} = \begin{pmatrix} \mathbf{F}' (\Psi(L)\mathbf{V}_t)' \mathbf{A}_0(r) (\Psi(L)\mathbf{V}_t) \mathbf{d} \\ \mathbf{d}' (\Psi(L)\mathbf{V}_t)' \mathbf{A}_1 (\Psi(L)\mathbf{V}_t) \mathbf{d} - s^2 \operatorname{tr} \mathbf{A}_1(r) \\ \vdots \\ \mathbf{d}' (\Psi(L)\mathbf{V}_t)' \mathbf{A}_q (\Psi(L)\mathbf{V}_t) \mathbf{d} - s^2 \operatorname{tr} \mathbf{A}_q(r) \end{pmatrix}$$

so that

$$\mathbf{m}_t(\theta_0) = \begin{pmatrix} \mathbf{F}' (\Psi(L)\mathbf{V}_t)' \mathbf{B}' \boldsymbol{\Omega}_0 \mathbf{u}_t \\ \mathbf{u}_t' \boldsymbol{\Omega}_1 \mathbf{u}_t - \sigma^2 \operatorname{tr} \mathbf{A}_1 \\ \vdots \\ \mathbf{u}_t' \boldsymbol{\Omega}_q \mathbf{u}_t - \sigma^2 \operatorname{tr} \mathbf{A}_q \end{pmatrix}.$$

Since $\mathbf{u}_t$ is an independent process and $\Psi(L)\mathbf{V}_t\mathbf{F}$ is independent of $\mathbf{u}_t$, $\mathbf{m}_t(\theta_0)$ is a martingale difference process.

Let $\psi(j) = \sum_{\ell=0}^{\infty} \psi_\ell \psi_{\ell+|j|}$ and let $\boldsymbol{\Psi}$ be a $(k+1) \times (k+1)$ matrix with elements $\boldsymbol{\Psi}_{ij} = \psi(i-j)$. From Theorems 2 and 3 of Hannan (1970, p. 203–204) we can deduce that for any $n \times n$ matrix $\mathbf{\Omega}$,

$$\frac{1}{T-k}\sum_{t=k+1}^{T}\Psi\left(L\right)\mathbf{V}_t \xrightarrow{p} \mathbf{0}, \tag{16}$$

$$\frac{1}{T-k}\sum_{t=k+1}^{T}\left(\Psi\left(L\right)\mathbf{V}_t'\right)\mathbf{\Omega}\left(\Psi\left(L\right)\mathbf{V}_t\right) \xrightarrow{p} \sigma^2\operatorname{tr}\left(\mathbf{\Omega}\right)\mathbf{\Psi}, \tag{17}$$

because $\mathbf{v}_t$ are independently and identically distributed, $E\mathbf{v}_t\mathbf{v}_t' < \infty$ and $\sum_{j=0}^{\infty}\psi_j^2 \leq C\sum_{j=0}^{\infty}c^{2j} < \infty$ where C is a finite positive constant.

For any $n \times n$ matrix $\mathbf{\Omega}$ such that $\mathbf{1}'\mathbf{\Omega} = \mathbf{0}$ and $\mathbf{\Omega}\mathbf{1} = \mathbf{0}$, we have

$$E\mathbf{X}_t'\mathbf{\Omega}\mathbf{u}_t\left(\theta\right) = \sigma^2\operatorname{tr}\left(\mathbf{B}'\mathbf{\Omega}\mathbf{B}\right)\mathbf{F}'\mathbf{\Psi}\mathbf{d} + \frac{1}{\gamma}\left(1-\frac{g}{\gamma}\right)\alpha'\mathbf{\Omega}\alpha\mathbf{e}_k$$
$$E\mathbf{u}_t'\left(\theta\right)\mathbf{\Omega}\mathbf{u}_t\left(\theta\right) = \sigma^2\operatorname{tr}\left(\mathbf{B}'\mathbf{\Omega}\mathbf{B}\right)\mathbf{d}'\mathbf{\Psi}\mathbf{d} + \left(1-\frac{g}{\gamma}\right)^2\alpha'\mathbf{\Omega}\alpha,$$

where $\mathbf{e}_k = (0,\ldots,0,1)'$ is a $k \times 1$ vector. This implies that

$$E\mathbf{X}_t'\mathbf{\Omega}\mathbf{u}_t\left(\theta_0\right) = \sigma^2\operatorname{tr}\left(\mathbf{B}'\mathbf{\Omega}\mathbf{B}\right)\mathbf{F}'\mathbf{\Psi}\delta,$$
$$E\mathbf{u}_t'\left(\theta_0\right)\mathbf{\Omega}\mathbf{u}_t\left(\theta_0\right) = \sigma^2\operatorname{tr}\left(\mathbf{B}'\mathbf{\Omega}\mathbf{B}\right)\delta'\mathbf{\Psi}\delta,$$

and since $E\mathbf{X}_t'\mathbf{\Omega}\mathbf{u}_t = 0$ and $E\mathbf{u}_t'\mathbf{\Omega}\mathbf{u}_t = \sigma^2\operatorname{tr}\left(\mathbf{B}'\mathbf{\Omega}\mathbf{B}\right)$ by Assumptions 1-3, it can be seen that

$$\mathbf{F}'\mathbf{\Psi}\delta = \mathbf{0}, \qquad \delta'\mathbf{\Psi}\delta = \mathbf{1}. \tag{18}$$

Proposition 1 *For any sequence $\{\theta^*\}_{T=1}^{\infty}$ satisfying $\theta^* \xrightarrow{p} \theta_0$, as $T \to \infty$,*

$$\widetilde{\mathbf{S}}\left(\theta^*\right) \xrightarrow{p} \mathbf{S},$$

where

$$\mathbf{S} = E\mathbf{m}_t\left(\theta_0\right)\mathbf{m}_t\left(\theta_0\right)' = \begin{pmatrix} \mathbf{S}_1 & 0 \\ 0 & \mathbf{S}_2 \end{pmatrix}$$

with $\mathbf{S}_1 = (n-1)\,\sigma^4\mathbf{F}'\mathbf{\Psi}\mathbf{F}$ and $(\mathbf{S}_2)_{ij} = \sigma^4\operatorname{tr}\left(\mathbf{A}_i\left(\mathbf{A}_j + \mathbf{A}_j'\right)\right) + \kappa\sum_{\ell=1}^{n}\left(\mathbf{A}_i\right)_{\ell\ell}\left(\mathbf{A}_j\right)_{\ell\ell}$ for $i,j = 1,\ldots,q$.

Proof. Let $\theta^* = \left(\beta_1^*,\ldots,\beta_{k-1}^*,\gamma^*,\rho^*,(\sigma^*)^2\right)'$. Denote $\widetilde{\mathbf{m}}_t^* = \widetilde{\mathbf{m}}_t(\theta^*)$, $\mathbf{m}_t^* = \mathbf{m}_t(\theta^*)$ and

$\mathbf{m}_t = \mathbf{m}_t(\theta_0)$. We first show that

$$\widetilde{\mathbf{S}}(\theta^*) = \frac{1}{T-k}\sum_{t=k+1}^{T} \mathbf{m}_t \mathbf{m}_t' + o_p(1).$$

Let the matrix norm be defined as $\|\mathbf{A}\| = (\operatorname{tr}\mathbf{A}'\mathbf{A})^{\frac{1}{2}}$. By the Schwarz and triangle inequalities,

$$\left\|\widetilde{\mathbf{S}}(\theta^*) - \frac{1}{T-k}\sum_{t=k+1}^{T} \mathbf{m}_t \mathbf{m}_t'\right\| \leq 2(a_T + b_T) + 2\left[a_T^{\frac{1}{2}} + b_T^{\frac{1}{2}}\right]\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\|\mathbf{m}_t\|^2\right)^{\frac{1}{2}}$$

where $a_T = (T-k)^{-1}\sum_{t=k+1}^{T}\|\widetilde{\mathbf{m}}_t^* - \mathbf{m}_t^*\|^2$ and $b_T = (T-k)^{-1}\sum_{t=k+1}^{T}\|\mathbf{m}_t^* - \mathbf{m}_t\|^2$. We denote $\delta^* = \left(1, -1-\beta_1^*+\gamma^*, \ldots, \beta_{k-1}^*\right)'$ and write $\widetilde{\mathbf{m}}_t^* - \mathbf{m}_t^* = \left(\widetilde{\mathbf{m}}_{0t}^{*\prime} - \mathbf{m}_{0t}^{*\prime}, \widetilde{m}_{1t}^* - m_{1t}^*, \ldots, \widetilde{m}_{qt}^* - m_{qt}^*\right)'$ where

$$\begin{aligned}
\widetilde{\mathbf{m}}_{0t}^* - \mathbf{m}_{0t}^* &= \mathbf{F}'\left(\Psi(L)\widetilde{\mathbf{V}}_t\right)'\mathbf{A}_0(\rho^*)\left(\Psi(L)\widetilde{\mathbf{V}}_t\right)\delta^* - \mathbf{F}'(\Psi(L)\mathbf{V}_t)'\mathbf{A}_0(\rho^*)(\Psi(L)\mathbf{V}_t)\delta^*,\\
\widetilde{m}_{it}^* - m_{it}^* &= \delta^{*\prime}\left(\Psi(L)\widetilde{\mathbf{V}}_t\right)'\mathbf{A}_i\left(\Psi(L)\widetilde{\mathbf{V}}_t\right)\delta^* - \delta^{*\prime}(\Psi(L)\mathbf{V}_t)'\mathbf{A}_i(\Psi(L)\mathbf{V}_t)\delta^* \qquad i = 1, \ldots, q.
\end{aligned}$$

By the triangle and Schwarz inequalities, the term a_T is bounded by

$$\begin{aligned}
&C\|\delta^*\|^2\left(\|\mathbf{F}\|^2\|\mathbf{A}_0(\rho^*)\|^2 + \sum_{i=1}^{q}\|\delta^*\|^2\|\mathbf{A}_i\|^2\right)\left\|\frac{1}{T-k}\sum_{t=1}^{T-k}\Psi(L)\mathbf{V}_t\right\|^2\\
&\times\left(\operatorname{tr}\left(\frac{1}{T-k}\sum_{t=1}^{T-k}(\Psi(L)\mathbf{V}_t)'(\Psi(L)\mathbf{V}_t)\right) + \left\|\frac{1}{T-k}\sum_{t=1}^{T-k}\Psi(L)\mathbf{V}_t\right\|^2\right).
\end{aligned}$$

Matrices $\mathbf{F}$, δ^*, $\mathbf{A}_0(\rho^*)$ and $\mathbf{A}_1, \ldots, \mathbf{A}_q$ have finite norms. It follows from (16) and (17) that $a_T = O_p(T^{-1})$.

We further write $\mathbf{m}_t^* - \mathbf{m}_t = \left(\mathbf{m}_{0t}^{*\prime} - \mathbf{m}_{0t}', m_{1t}^* - m_{1t}, \ldots, m_{qt}^* - m_{qt}\right)'$ and note that

$$\begin{aligned}
\mathbf{m}_{0t}^* - \mathbf{m}_{0t} &= \mathbf{F}'(\Psi(L)\mathbf{V}_t)'(\mathbf{A}_0(\rho^*) - \mathbf{A}_0(\rho))(\Psi(L)\mathbf{V}_t)(\delta^* - \delta)\\
&+ \mathbf{F}'(\Psi(L)\mathbf{V}_t)'(\mathbf{A}_0(\rho^*) - \mathbf{A}_0(\rho))(\Psi(L)\mathbf{V}_t)\delta\\
&+ \mathbf{F}'(\Psi(L)\mathbf{V}_t)'\mathbf{A}_0(\rho)(\Psi(L)\mathbf{V}_t)(\delta^* - \delta),
\end{aligned}$$

$$\begin{aligned}
m_{it}^* - m_{it} &= (\delta^* - \delta)'(\Psi(L)\mathbf{V}_t)'\mathbf{A}_i(\Psi(L)\mathbf{V}_t)(\delta^* - \delta)\\
&+ (\delta^* - \delta)'(\Psi(L)\mathbf{V}_t)'(\mathbf{A}_i + \mathbf{A}_i')(\Psi(L)\mathbf{V}_t)\delta\\
&- (\sigma^*)^2\operatorname{tr}\mathbf{A}_1(\rho^*) + \sigma^2\operatorname{tr}\mathbf{A}_1(\rho) \qquad i = 1, \ldots, q.
\end{aligned}$$

By the triangle and Schwarz inequalities, the term b_T is bounded by

$$C\left(\left\|\mathbf{A}_0\left(\rho^*\right)-\mathbf{A}_0\left(\rho\right)\right\|^2\left(\left\|\delta^*-\delta\right\|^2+\left\|\delta\right\|^2\right)+\left\|\mathbf{A}_0\left(\rho\right)\right\|^2\left\|\delta^*-\delta\right\|^2\right)$$
$$\times\left\|\mathbf{F}\right\|^2\frac{1}{T-k}\sum_{t=1}^{T}\left\|\Psi\left(L\right)\mathbf{V}_t\right\|^4$$

$$+C\left(\left\|\delta^*-\delta\right\|^4+\left\|\delta^*-\delta\right\|^2\left\|\delta\right\|^2\right)\sum_{i=1}^{q}\left\|\mathbf{A}_i\right\|^2\frac{1}{T-k}\sum_{t=1}^{T}\left\|\Psi\left(L\right)\mathbf{V}_t\right\|^4$$
$$+C\sum_{i=1}^{q}\left(\left(\sigma^*\right)^2\operatorname{tr}\mathbf{A}_i\left(\rho^*\right)-\sigma^2\operatorname{tr}\mathbf{A}_i\left(\rho\right)\right)^2. \qquad (19)$$

Matrices $\mathbf{F}$, $\mathbf{M_1}$ and $\mathbf{A}_1,\ldots,\mathbf{A}_q$ have finite norms. By assumption, $\delta^*-\delta=o_p\left(1\right)$. By the Schwarz inequality,

$$\begin{aligned}E\left\|\Psi\left(L\right)\mathbf{V}_t\right\|^4&=\sum_{i=1}^{n}\sum_{j=1}^{k+1}\sum_{\ell=1}^{n}\sum_{m=1}^{k+1}E\left(\Psi\left(L\right)v_{i,t-j+1}\right)^2\left(\Psi\left(L\right)v_{\ell,t-m+1}\right)^2\\&\leq n\left(k+1\right)\sum_{i=1}^{n}\sum_{j=1}^{k}E\left(\Psi\left(L\right)v_{i,t-j+1}\right)^4<\infty\end{aligned}$$

because

$$\begin{aligned}E\left(\Psi\left(L\right)v_{i,t-j+1}\right)^4&=\sum_{p,q,r,s=0}^{\infty}\psi_p\psi_q\psi_r\psi_s Ev_{i,t-j+1-p}v_{i,t-j+1-q}v_{i,t-j+1-r}v_{i,t-j+1-s}\\&=\kappa\sum_{p=0}^{\infty}\psi_p^4+3\sigma^4\sum_{p,q=0}^{\infty}\psi_p^2\psi_q^2<\infty\end{aligned}$$

where the last inequality is due to the fact that ψ_j is square summable and that the fourth moments of $\mathbf{v}_t$ are finite. Therefore $\left(T-k\right)^{-1}\sum_{t=k+1}^{T}E\left\|\Psi\left(L\right)\mathbf{V}_t\right\|^4<\infty$ and the first two terms in (19) are $o_p\left(1\right)$. Functions $\mathbf{A}_0\left(r\right)$ and $s^2\operatorname{tr}\mathbf{A}_i\left(r\right)$, $i=1,\ldots,q$, are continuous in s^2 and r, and $\left(\sigma^*\right)^2\xrightarrow{p}\sigma^2$ and $\rho^*\xrightarrow{p}\rho$ by assumption, therefore by the continuous mapping theorem all terms in (19) are $o_p\left(1\right)$. It follows that $b_T=o_p\left(1\right)$.

To complete the proof of the present proposition we show that $\left(T-k\right)^{-1}\sum_{t=k+1}^{T}\mathbf{m}_t\mathbf{m}_t{}'$ $\xrightarrow{p}\mathbf{S}$. To show this, it is sufficient by Theorem 2.19 of Hall and Heyde (1980) to demonstrate that

$$\frac{1}{T-k}\sum_{t=k+1}^{T}E\left(\mathbf{m}_t\mathbf{m}_t'|\mathcal{F}_{t-1}\right)\xrightarrow{p}\mathbf{S} \qquad (20)$$

because $\mathbf{m}_t\mathbf{m}_t'$ is a strictly stationary process. We have

$$\begin{aligned}
E\left(\mathbf{m}_{0t}\mathbf{m}_{0t}'|\mathcal{F}_{t-1}\right) &= \sigma^2\mathbf{F}'\left(\Psi\left(L\right)\mathbf{V}_t'\right)\mathbf{M_1}\left(\Psi\left(L\right)\mathbf{V}_t\right)\mathbf{F},\\
E\left(\mathbf{m}_{0t}m_{it}|\mathcal{F}_{t-1}\right) &= \mathbf{F}'\left(\Psi\left(L\right)\mathbf{V}_t'\right)\mathbf{M_1}E\left(\mathbf{v}_t\mathbf{v}_t'\mathbf{A}_i\mathbf{v}_t\right), \qquad i=1,\dots,q,\\
E\left(m_{it}m_{jt}|\mathcal{F}_{t-1}\right) &= \kappa\sum_{l=1}^{n}\left(\mathbf{A}_i\right)_{ll}\left(\mathbf{A}_j\right)_{ll}+\sigma^4\operatorname{tr}\left(\mathbf{A}_i\left(\mathbf{A}_j+\mathbf{A}_j'\right)\right), \qquad i,j=1,\dots,q.
\end{aligned}$$

Limit statements in (16) and (17) imply that the convergence in (20) holds and that therefore $\widetilde{\mathbf{S}}\left(\theta^*\right)\xrightarrow{p}\mathbf{S}$. ■

Proposition 2 *For any sequence $\{\theta^*\}_{T=1}^{\infty}$ satisfying $\theta^*\xrightarrow{p}\theta_0$, as $T\to\infty$,*

$$\widetilde{\mathbf{D}}\left(\theta^*\right)\xrightarrow{p}\mathbf{D}$$

where

$$\mathbf{D}=E\frac{\partial\mathbf{m}_t\left(\theta_0\right)}{\partial\theta'}=-\begin{pmatrix}\sigma^2\left(n-1\right)\mathbf{F}'\mathbf{\Psi}\mathbf{F} & \mathbf{0} & \mathbf{0}\\ 0 & \sigma^2C_1\left(\rho\right) & \operatorname{tr}\mathbf{A}_1\\ \vdots & \vdots & \vdots\\ 0 & \sigma^2C_q\left(\rho\right) & \operatorname{tr}\mathbf{A}_q\end{pmatrix}.$$

Proof. We define matrix $\mathbf{D}\left(\theta\right)$ as $\mathbf{D}\left(\theta\right)=\left(T-k\right)^{-1}\sum_{t=k+1}^{T}\frac{\partial\mathbf{m}_t(\theta)}{\partial\theta'}$, so that

$$\mathbf{D}\left(\theta^*\right)=-\begin{pmatrix}\mathbf{F}'\frac{1}{T-k}\sum_{t=k+1}^{T}\left(\Psi\left(L\right)\mathbf{V}_t\right)'\mathbf{A}_0\left(\rho^*\right)\left(\Psi\left(L\right)\mathbf{V}_t\right)\mathbf{F} & \mathbf{D}_H\left(\theta^*\right) & \mathbf{0}\\ \delta^{*\prime}\frac{1}{T-k}\sum_{t=k+1}^{T}\left(\Psi\left(L\right)\mathbf{V}_t\right)'\left(\mathbf{A}_1+\mathbf{A}_1'\right)\left(\Psi\left(L\right)\mathbf{V}_t\right)\mathbf{F} & \left(\sigma^*\right)^2C_1\left(\rho^*\right) & \operatorname{tr}\mathbf{A}_1\left(\rho^*\right)\\ & \vdots & \\ \delta^{*\prime}\frac{1}{T-k}\sum_{t=k+1}^{T}\left(\Psi\left(L\right)\mathbf{V}_t\right)'\left(\mathbf{A}_q+\mathbf{A}_q'\right)\left(\Psi\left(L\right)\mathbf{V}_t\right)\mathbf{F} & \left(\sigma^*\right)^2C_q\left(\rho^*\right) & \operatorname{tr}\mathbf{A}_q\left(\rho^*\right)\end{pmatrix}$$

where $\mathbf{D}_H=-\mathbf{F}'\frac{1}{T-k}\sum_{t=k+1}^{T}\left(\Psi\left(L\right)\mathbf{V}_t\right)'\mathbf{B}'\mathbf{H}\left(\rho^*\right)\mathbf{B}\left(\Psi\left(L\right)\mathbf{V}_t\right)\delta^*$. We have

$$\begin{aligned}
&\left\|\widetilde{\mathbf{D}}\left(\theta^*\right)-\mathbf{D}\left(\theta^*\right)\right\|^2\\
&\le C\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}\Psi\left(L\right)\mathbf{V}_t\right\|^4\left(\left\|\mathbf{F}\right\|^4\left\|\mathbf{A}_0\left(\rho^*\right)\right\|^2+\left\|\delta^*\right\|^2\left\|\mathbf{F}\right\|^2\left(\sum_{i=1}^{q}\left\|\mathbf{A}_i\right\|^2+\left\|\mathbf{B}\right\|^4\left\|\mathbf{H}\left(\rho^*\right)\right\|^2\right)\right).
\end{aligned}$$

Since matrices $\mathbf{F}$, $\mathbf{A}_0\left(\rho^*\right)$, δ^*, $\mathbf{A}_1,\dots\mathbf{A}_q$, $\mathbf{B}$ and $\mathbf{H}\left(\rho^*\right)$ have finite norms, it follows from (16) that $\widetilde{\mathbf{D}}\left(\theta^*\right)=\mathbf{D}\left(\theta^*\right)+o_p\left(1\right)$. Therefore to prove the statement of the proposition it is sufficient to show that $\mathbf{D}\left(\theta^*\right)\xrightarrow{p}\mathbf{D}$. By the continuous mapping theorem and (17), the bottom right $q\times 2$ submatrix of $\mathbf{D}\left(\theta^*\right)$ converges in probability to the corresponding submatrix of $\mathbf{D}$ because $\left(s^*\right)^2\xrightarrow{p}\sigma^2$ and $\rho^*\xrightarrow{p}\rho$ and functions $s^2C_i\left(r\right)$ and $\operatorname{tr}\mathbf{A}_i\left(r\right)$,

$i = 1, \ldots, q$, are continuous in s^2 and r. For the top left $1 \times (k+1)$ block of matrix $\mathbf{D}(\theta^*)$, we can write

$$\begin{aligned}
&\frac{1}{T-k} \sum_{t=k+1}^{T} \left(\Psi(L)\mathbf{V}_t\right)' \mathbf{\Omega}_0(\rho^*) \left(\Psi(L)\mathbf{V}_t\right) \\
&= \frac{1}{T-k} \sum_{t=k+1}^{T} \left(\Psi(L)\mathbf{V}_t\right)' \mathbf{\Omega}_0 \left(\Psi(L)\mathbf{V}_t\right) \\
&+ \frac{1}{T-k} \sum_{t=k+1}^{T} \left(\Psi(L)\mathbf{V}_t\right)' \left(\mathbf{\Omega}_0(\rho^*) - \mathbf{\Omega}_0\right) \left(\Psi(L)\mathbf{V}_t\right). \qquad (21)
\end{aligned}$$

The first term of (21) converges in probability to $\sigma^2(n-1)\mathbf{\Psi}$ by (17). The norm of the second term of (21) is bounded by $\|\mathbf{\Omega}_0(\rho^*) - \mathbf{\Omega}_0\| (T-k)^{-1} \sum_{t=k+1}^{T} \|\Psi(L)\mathbf{V}_t\|^2$ which is $o_p(1)$ because $\rho^* \xrightarrow{p} \rho$ and function $\mathbf{\Omega}_0(r)$ is continuous in r, and $E\|\Psi(L)\mathbf{V}_t\|^2 < \infty$. Using similar arguments, we can see that since $\delta^* \xrightarrow{p} \delta$ by assumption, the left column of blocks of $\mathbf{D}(\theta^*)$ converges in probability to

$$-\begin{pmatrix} \sigma^2(n-1)\mathbf{F}'\mathbf{\Psi}\mathbf{F} \\ \sigma^2 \operatorname{tr}(\mathbf{A}_1)\delta'\mathbf{\Psi}\mathbf{F} \\ \vdots \\ \sigma^2 \operatorname{tr}(\mathbf{A}_q)\delta'\mathbf{\Psi}\mathbf{F} \end{pmatrix} = -\begin{pmatrix} \sigma^2(n-1)\mathbf{F}'\mathbf{\Psi}\mathbf{F} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

because by (18), $\delta'\mathbf{\Psi}\mathbf{F} = \mathbf{0}$. The convergence of the remaining block $\mathbf{D}_H(\theta^*)$ to $\mathbf{0}$ can be shown in a similar way. We can conclude that $\widetilde{\mathbf{D}}(\theta^*) \xrightarrow{p} \mathbf{D}$. ■

Proposition 3 *Let $\hat{\theta}$ be the GMM estimator defined in (13). Under Assumptions 1-4, as $T \to \infty$,*

$$\hat{\theta} \xrightarrow{p} \theta_0.$$

Proof. We first prove that $(T-k)^{-1} \sum_{t=k+1}^{T} \widetilde{\mathbf{m}}_t(\theta) \xrightarrow{p} E\mathbf{m}_t(\theta)$ uniformly in $\theta \in \Theta$. We show that

$$\sup_{\theta \in \Theta} \left\| \frac{1}{T-k} \sum_{t=k+1}^{T} \left(\widetilde{\mathbf{m}}_t(\theta) - \mathbf{m}_t(\theta)\right) \right\| = O_p\left(T^{-1}\right) \qquad \text{and} \qquad (22)$$

$$\sup_{\theta \in \Theta} \left\| (T-k)^{-1} \sum_{t=k+1}^{T} \mathbf{m}_t(\theta) - E\mathbf{m}_t(\theta) \right\| = o_p(1). \qquad (23)$$

Expression $(T-k)^{-1}\sum_{t=k+1}^{T}\left(\widetilde{\mathbf{m}}_t(\theta)-\mathbf{m}_t(\theta)\right)$ is equal to

$$-\begin{pmatrix} \mathbf{F}'\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\Psi(L)\mathbf{V}_t\right)'\mathbf{A}_0(r)\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\Psi(L)\mathbf{V}_t\right)\mathbf{d} \\ \mathbf{d}'\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\Psi(L)\mathbf{V}_t\right)'\mathbf{A}_1\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\Psi(L)\mathbf{V}_t\right)\mathbf{d}+\frac{1}{T-k}s^2\operatorname{tr}\mathbf{A}_1(r) \\ \vdots \\ \mathbf{d}'\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\Psi(L)\mathbf{V}_t\right)'\mathbf{A}_q\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\Psi(L)\mathbf{V}_t\right)'\mathbf{d}+\frac{1}{T-k}s^2\operatorname{tr}\mathbf{A}_q(r) \end{pmatrix}.$$

Using the triangle and Schwarz inequalities, we bound $\left\|(T-k)^{-1}\sum_{t=k+1}^{T}\left(\widetilde{\mathbf{m}}_t(\theta)-\mathbf{m}_t(\theta)\right)\right\|^2$ by

$$C\left(\|\mathbf{F}\|^2\|\mathbf{d}\|^2\|\mathbf{A}_0(r)\|^2+\|\mathbf{d}\|^4\sum_{i=1}^{q}\|\mathbf{A}_i(r)\|^2\right)\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}\Psi(L)\mathbf{V}_t\right\|^4$$
$$+\frac{Cs^4}{(T-k)^2}\sum_{i=1}^{q}\left(\operatorname{tr}\mathbf{A}_i(r)\right)^2$$

where C is a finite positive constant. Since the parameter space Θ is compact, matrices $\mathbf{d}$ and $\mathbf{A}_i(r)$ have finite norms. By a central limit theorem, $(T-k)^{-1/2}\sum_{t=k+1}^{T}\Psi(L)\mathbf{v}_{t-j}=O_p(1)$ (see for example Eicker 1967), therefore (22) holds.

We now show that (23) holds. For any finite $n\times n$ matrix $\mathbf{\Omega}$ we have $E\left(\Psi(L)\mathbf{V}_t\right)'\mathbf{\Omega}\left(\Psi(L)\mathbf{V}_t\right)=\sigma^2\operatorname{tr}(\mathbf{\Omega})\mathbf{\Psi}$, so

$$E\mathbf{m}_t(\theta)=\begin{pmatrix} \sigma^2\operatorname{tr}(\mathbf{A}_0(r))\mathbf{F}'\mathbf{\Psi}\mathbf{d} \\ \sigma^2\operatorname{tr}\mathbf{A}_1\mathbf{d}'\mathbf{\Psi}\mathbf{d}-s^2\operatorname{tr}\mathbf{A}_1(r) \\ \vdots \\ \sigma^2\operatorname{tr}\mathbf{A}_q\mathbf{d}'\mathbf{\Psi}\mathbf{d}-s^2\operatorname{tr}\mathbf{A}_q(r) \end{pmatrix}. \tag{24}$$

The i-th element of vector $(T-k)^{-1}\sum_{t=k+1}^{T}\mathbf{m}_{0t}(\theta)-E\mathbf{m}_{0t}(\theta)$ is equal to

$$\sum_{p,s=1}^{k+1}\sum_{q,v=1}^{n}\mathbf{F}_{pi}\left(\mathbf{A}_0(r)\right)_{qv}\mathbf{d}_s\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\left(\Psi(L)v_{q,t-p+1}\right)\left(\Psi(L)v_{v,t-s+1}\right)-\sigma^2\mathbf{I}_{qv}\mathbf{\Psi}_{ps}\right)$$

where $\mathbf{I}$ is an $n\times n$ identity matrix. The expression in the parentheses is $o_p(1)$ by (17), therefore

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}\mathbf{m}_{0t}(\theta)-E\mathbf{m}_{0t}(\theta)\right\|\leq C\sup_{\theta\in\Theta}\|\mathbf{F}\|\,\|\mathbf{A}_0(r)\|\,\|\mathbf{d}\|\,o_p(1)=o_p(1)$$

because Θ is compact, function $\mathbf{A}_0$ is continuous and $\left(\sum_{i=1}^{n}a_i\right)^2\leq n\sum_{i=1}^{n}a_i^2$. In a similar way, it can be shown that $\sup_{\theta\in\Theta}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}\mathbf{m}_{it}(\theta)-E\mathbf{m}_{it}(\theta)\right\|=o_p(1)$ for $i=1,\ldots,q$. We have therefore that $\sup_{\theta\in\Theta}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}\mathbf{m}_{t}(\theta)-E\mathbf{m}_{t}(\theta)\right\|\xrightarrow{p}0$ and consequently

$$\sup_{\theta\in\Theta}\left\|q_{\mathbf{\Sigma}}(\theta)-E\mathbf{m}_t(\theta)'\,\mathbf{\Sigma}E\mathbf{m}_t(\theta)\right\|\xrightarrow{p}0$$

by the continuous mapping theorem.

Consistency now follows by standard arguments, see for example Theorem 2.6 of Newey and McFadden (1994). ■

It is worth checking that the parameters of the model are identified, that is that $E\mathbf{m}_t(\theta)=0$ if and only if $\theta=\theta_0$. Regarding the expression $\mathbf{F}'\mathbf{\Psi}\mathbf{d}$ in (24), matrix $\mathbf{F}'\mathbf{\Psi}$ is $k\times(k+1)$ and has rank k, and since by (18) $\mathbf{F}'\mathbf{\Psi}\delta=0$, the one-dimensional null space of $\mathbf{F}'\mathbf{\Psi}$ is spanned by vector δ. Vectors $\mathbf{d}$ are normalized to have the first component equal to 1, therefore vector δ is the unique vector $\mathbf{d}$ for which $\mathbf{F}'\mathbf{\Psi}\mathbf{d}=0$ and also for which $\sigma^2\operatorname{tr}(\mathbf{A}_0(r))\,\mathbf{F}'\mathbf{\Psi}\mathbf{d}=0$ because $\sigma^2\operatorname{tr}(\mathbf{A}_0(r))>0$ for $r\in(-1,1)$. The parameter vector $\delta=(1,-\phi_1,\ldots,-\phi_k)'=(1,-1-\beta_1+\gamma,\ldots,\beta_{k-1})'$ is therefore identified.

A sufficient condition for parameters ρ and σ^2 to be identified is that

$$Em_{1t}(\theta)=\cdots=Em_{qt}(\theta)=0 \tag{25}$$

if and only if $s^2=\sigma^2$ and $r=\rho$. Conditions (25) imply that

$$\frac{\operatorname{tr}(\mathbf{B}'\mathbf{\Omega}_i\mathbf{B})}{\operatorname{tr}(\mathbf{B}'\mathbf{\Omega}_j\mathbf{B})}=\frac{\operatorname{tr}\left(\mathbf{B}(r)'\,\mathbf{\Omega}_i\mathbf{B}(r)\right)}{\operatorname{tr}\left(\mathbf{B}(r)'\,\mathbf{\Omega}_j\mathbf{B}(r)\right)}\qquad\text{for any } i,j=1,\ldots,q \tag{26}$$

because $\delta'\mathbf{\Psi}\delta=1$ by (18). For $\mathbf{\Omega}_1=\mathbf{M_1}$ and $\mathbf{\Omega}_2=(\mathbf{I}-\mathbf{W})'\,\mathbf{M_1}(\mathbf{I}-\mathbf{W})$ equation (26) has a unique solution $r=\rho$, therefore parameter ρ is identified and from (24) it can be seen that parameter σ^2 is also identified.

Proof of Theorem 1. By the mean value theorem,

$$\frac{1}{T-k}\sum_{t=k+1}^{T}\widetilde{\mathbf{m}}_t\left(\hat{\theta}\right)=\frac{1}{T-k}\sum_{t=k+1}^{T}\widetilde{\mathbf{m}}_t(\theta)+\mathbf{\Delta}\left(\hat{\theta}\right)\left(\hat{\theta}-\theta_0\right)$$

where the i-th row of matrix $\mathbf{\Delta}\left(\hat{\theta}\right)$ is the i-th row of matrix $\widetilde{\mathbf{D}}(\theta_i^*)$ for some θ_i^*,

$|\theta_i^* - \theta_0| \leq \left|\hat{\theta} - \theta_0\right|$, $i = 1, \ldots, k+q$. We have

$$\frac{\partial \widetilde{q}_{\mathbf{\Sigma}}\left(\hat{\theta}\right)}{\partial \theta} = 2\widetilde{\mathbf{D}}'\left(\hat{\theta}\right)\mathbf{\Sigma}_T \frac{1}{T-k}\sum_{t=k+1}^{T}\widetilde{\mathbf{m}}_t\left(\hat{\theta}\right)$$
$$= 2\widetilde{\mathbf{D}}'\left(\hat{\theta}\right)\mathbf{\Sigma}_T \mathbf{\Delta}\left(\hat{\theta}\right)\left(\hat{\theta} - \theta_0\right) + 2\widetilde{\mathbf{D}}'\left(\hat{\theta}\right)\mathbf{\Sigma}_T \frac{1}{T-k}\sum_{t=k+1}^{T}\widetilde{\mathbf{m}}_t\left(\theta_0\right).$$

Proposition 3, bound (22) and continuity of $\partial \widetilde{q}_{\mathbf{\Sigma}}\left(\theta\right)/\partial\theta$ imply that $\partial \widetilde{q}_{\mathbf{\Sigma}}\left(\hat{\theta}\right)/\partial\theta = 0$ with probability approaching 1 as $T \to \infty$. Using (22) we obtain

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) = -\left(\widetilde{\mathbf{D}}'\left(\hat{\theta}\right)\mathbf{\Sigma}_T\mathbf{\Delta}\left(\hat{\theta}\right)\right)^{-1}\widetilde{\mathbf{D}}'\left(\hat{\theta}\right)\mathbf{\Sigma}_T\left(\frac{1}{\sqrt{T-k}}\sum_{t=k+1}^{T}\mathbf{m}_t\left(\theta_0\right) + O_p\left(T^{-1/2}\right)\right).$$

By Propositions 3 and 2, $\widetilde{\mathbf{D}}\left(\hat{\theta}\right) \xrightarrow{p} \mathbf{D}$ and $\widetilde{\mathbf{D}}\left(\theta_i^*\right) \xrightarrow{p} \mathbf{D}$, $i = 1, \ldots, k+q$, and by the continuous mapping theorem also $\mathbf{\Delta}\left(\hat{\theta}\right) \xrightarrow{p} \mathbf{D}$. Since $(T-k)^{-1/2}\sum_{t=k+1}^{T}\mathbf{m}_t\left(\theta_0\right) \xrightarrow{d} N\left(0, \mathbf{S}\right)$ as we show below, we can conclude that

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, \left(\mathbf{D}'\mathbf{\Sigma}\mathbf{D}\right)^{-1}\mathbf{D}'\mathbf{\Sigma}\mathbf{S}\mathbf{\Sigma}\mathbf{D}\left(\mathbf{D}'\mathbf{\Sigma}\mathbf{D}\right)^{-1}\right).$$

If we set $\mathbf{\Sigma}_T = \widetilde{\mathbf{S}}^{-1}\left(\widetilde{\theta}\right)$, then by Proposition 1, $\mathbf{\Sigma}_T \xrightarrow{p} \mathbf{S}^{-1}$ and $\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, \left(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D}\right)^{-1}\right)$.

We now show that $(T-k)^{-1/2}\sum_{t=k+1}^{T}\mathbf{m}_t\left(\theta_0\right) \xrightarrow{d} N\left(0, \mathbf{S}\right)$. Denote $\mathbf{m}_t = \mathbf{m}_t\left(\theta_0\right)$. Pick an $(k+q) \times 1$ vector λ such that $0 < \|\lambda\| < \infty$. Since $\mathbf{m}_t$ is a zero-mean martingale difference sequence and $E\left(\lambda'\mathbf{m}_t\right)^2 = \lambda'\mathbf{S}\lambda < \infty$, Theorem 1 of Scott (1973) implies that $(T-k)^{-1/2}\sum_{t=k+1}^{T}\lambda'\mathbf{m}_t$ converges in distribution to a $N\left(0, \lambda'\mathbf{S}\lambda\right)$ random variable if

$$\frac{1}{T-k}\sum_{t=k+1}^{T}\left(\lambda'\mathbf{m}_t\right)^2 \xrightarrow{p} \lambda'\mathbf{S}\lambda, \tag{27}$$

$$\frac{1}{T-k}\sum_{t=k+1}^{T}E\left(\left(\lambda'\mathbf{m}_t\right)^2 1\left(\left|\lambda'\mathbf{m}_t\right| \geq \varepsilon T^{1/2}\right)\right) \to 0 \qquad \text{for all } \varepsilon > 0, \tag{28}$$

where $1\left(\cdot\right)$ is the indicator function. Condition (27) is implied by the fact that $(T-k)^{-1}\sum_{t=k+1}^{T}\mathbf{m}_t\mathbf{m}_t' \xrightarrow{p} \mathbf{S}$ as shown in the proof of Proposition 1. Since $\mathbf{m}_t$ are identically distributed, condition (28) is equivalent to $E\left(\lambda'\mathbf{m}_t\right)^2 1\left(\left|\lambda'\mathbf{m}_t\right| \geq \varepsilon T^{1/2}\right) \to 0$ which holds because $E\left(\lambda'\mathbf{m}_t\right)^2 < \infty$. We have shown that $(T-k)^{-1/2}\sum_{t=k+1}^{T}\lambda'\mathbf{m}_t\left(\theta_0\right) \xrightarrow{d}$

$N(0, \lambda' \mathbf{S} \lambda)$ and, since λ is arbitrary, that $(T-k)^{-1/2} \sum_{t=k+1}^{T} \mathbf{m}_t(\theta_0) \xrightarrow{d} N(0, \mathbf{S})$ by the Cramér-Wold device. ■

Proof of Theorem 2. Theorem 2 follows from Propositions 1–3. ■

Acknowledgements

We would like to thank two anonymous referees, Emanuela Marrocu, Takashi Yamagata, Emmanuel Guerre and the participants of the 14th Brazilian Time Series and Econometrics School in Gramado, the Econometrics reading group at the School of Economics and Finance at Queen Mary University of London and of the 19th International Panel Data Conference in London for helpful comments and suggestions.

References

[1] Baltagi, Badi H., and Long Liu, "Random effects, fixed effects and Hausman's test for the generalized spatial panel data regression model," Working paper, Syracuse University (2011).

[2] Baltagi, Badi H., Bernard Fingleton, and Alain Pirotte, "Estimating and forecasting with a dynamic spatial panel data model", Oxford Bulletin of Economics and Statistics 76, (2014) 112-138 doi: 10.1111/obes.12011

[3] Beck, Guenter W., Kirstin Hubrich, and Massimiliano Marcellino, "Regional inflation dynamics within and across euro area countries and a comparison with the United States," *Economic Policy* 24 (2009), 141–184. DOI: 10.1111/j.1468-0327.2009.00214.x

[4] Bierens, Herman J., "Nonparametric nonlinear cotrending analysis, with an application to interest and inflation in the United States," *Journal of Business & Economic Statistics* 18 (2000), 323–337. DOI:10.1080/07350015.2000.10524874

[5] Cavallo, Alberto, Brent Neiman, and Roberto Rigobon, "Currency unions, product introductions and the real exchange rate," *The Quarterly Journal of Economics* (2014) 529-595. DOI: 10.1093/qje/qju008

[6] Cecchetti, Stephen G., Nelson C. Mark, and Robert J. Sonora, "Price index convergence among United States cities," *International Economic Review* 43 (2002), 1081–1099. DOI: 10.1111/1468-2354.t01-1-00049

[7] Chapman, David A., and Masao Ogaki, "Cotrending and the stationarity of the real interest rate," *Economics Letters* 42 (1993), 133–138. DOI: 10.1016/0165-1765(93)90050-M

[8] Chen, Lein-Lein, and John Devereux, "What can US city price data tell us about purchasing power parity?," *Journal of International Money and Finance* 22 (2003), 213–222. DOI: 10.1016/S0261-5606(02)00102-X

[9] Chudik, Alexander, M. Hashem Pesaran, and Elisa Tosetti, "Weak and strong cross-section dependence and estimation of large panels," *The Econometrics Journal* 14 (2011), C45–C90. DOI: 10.1111/j.1368-423X.2010.00330.x

[10] Eicker, Friedhelm, "Limit theorems for regressions with unequal and dependent errors," (Vol. 1, pp. 59–82) in L. M. LeCam and J. Neyman, eds., *Proceedings of the fifth Berkeley symposium* (Berkeley and Los Angeles: University of California Press, 1967).

[11] Engel, Charles, and John H. Rogers, "How wide is the border?," *The American Economic Review* 86 (1996), 1112–1125.

[12] Engle, Robert F., and Sharon Kozicki, "Testing for common features," *Journal of Business & Economic Statistics* 11 (1993), 369–395. DOI:10.1080/07350015.1993.10509966

[13] Ertur, Cem, and Wilfried Koch, "Growth, technological interdependence and spatial externalities: theory and evidence," *Journal of Applied Econometrics* 22 (2007), 1033–1062. DOI: 10.1002/jae.963

[14] Gorodnichenko, Yuriy, and Linda L. Tesar, "Border effect or country effect? Seattle may not be so far from Vancouver after all," *American Economic Journal: Macroeconomics* 1 (2009) 219-241. DOI: 10.1257/mac.1.1.219

[15] Hall, Anthony D., Heather M. Anderson, and Clive W. J. Granger, "A cointegration analysis of treasury bill yields," *The Review of Economics and Statistics* 74 (1992), 116–126.

[16] Hall, Peter, and Christopher Charles Heyde, *Martingale limit theory and its applications (*London: Academic Press, 1980).

[17] Hannan, Edward J., *Multiple time series (*New York: Wiley, 1970).

[18] Hendry, David F., and Michael Massmann, "Co-Breaking: recent advances and a synopsis of the literature," *Journal of Business & Economic Statistics* 25 (2007), 33–51. DOI: 10.1198/073500106000000422

[19] Im, Kyung So, M. Hashem Pesaran, and Yongcheol Shin, "Testing for unit roots in heterogeneous panels," *Journal of Econometrics* 115 (2003), 53–74. DOI:10.1016/S0304-4076(03)00092-7

[20] Johansen, Soren, and Juselius, Katarina, "Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK," *Journal of Econometrics* 53 (1992) 211-244.

[21] Kapetanios, George, M. Hashem Pesaran, and Takashi Yamagata, "Panels with non-stationary multifactor error structures," *Journal of Econometrics* 160 (2011), 326–348. DOI: 10.1016/j.jeconom.2010.10.001

[22] Kapoor, Mudit, Harry H. Kelejian, and Ingmar R. Prucha, "Panel data models with spatially correlated error components," *Journal of Econometrics* 140 (2007), 97–130. DOI:10.1016/j.jeconom.2006.09.004

[23] Kelejian, Harry H., and Ingmar R. Prucha, "A generalized moments estimator for the autoregressive parameter in a spatial model," *International Economic Review* 40 (1999), 509–533. DOI: 10.1111/1468-2354.00027

[24] Kelejian, Harry H., and Ingmar R. Prucha, "HAC estimation in a spatial framework," *Journal of Econometrics* 140 (2007) 131–154. DOI: 10.1016/j.jeconom.2006.09.005

[25] Lee, Lung-fei, and Jihai Yu, "Some recent developments in spatial panel data models." *Regional Science and Urban Economics* 40 (2010), 255–271. DOI: 10.1016/j.regsciurbeco.2009.09.002

[26] Magnus, Jan R., and Heinz Neudecker, *Matrix differential calculus with applications in statistics and econometrics* (Chichester: Wiley, 1999).

[27] Moscone, Francesco, Elisa Tosetti and Alessandra Canepa, "Real estate market and financial stability in US metropolitan areas: A dynamic model with spatial effects", *Regional Science and Urban Economics* 49 (2014) 129–146, DOI: dx.doi.org/10.1016/j.regsciurbeco.2014.08.003

[28] Mutl, Jan, *Dynamic panel data models with spatially correlated disturbances.* (Ph.D. dissertation, University of Maryland, 2006).

[29] Mutl, Jan and Michael Pfaffermayr, "The Hausman test in a Cliff and Ord panel model," *Econometrics Journal* 14 (2011), 48–76. DOI: 10.1111/j.1368-423X.2010.00325.x

[30] Newey, Whitney K., and Daniel McFadden, "Large sample estimation and hypothesis testing" (Vol. 4, pp. 2111–2245), in R. F. Engle and D. McFadden, eds., *Handbook of Econometrics* (Amsterdam: Elsevier, 1994). DOI: 10.1016/S1573-4412(05)80005-4

[31] Parsley, David C., and Shang-Jin Wei, "Convergence to the law of one price without trade barriers or currency fluctuations," *The Quarterly Journal of Economics* 111 (1996), 1211–1236. DOI:10.2307/2946713

[32] Pippenger, John, and Llad Phillips, "Some pitfalls in testing the law of one price in commodity markets," *Journal of International Money and Finance* 27 (2008) 915-925. DOI:10.1016/j.jimonfin.2008.05.003

[33] Pesaran, M. Hashem, "Estimation and inference in large heterogeneous panels with a multifactor error structure," *Econometrica* 74 (2006), 967–1012. DOI: 10.1111/j.1468-0262.2006.00692.x

[34] Pesaran, M. Hashem, and Elisa Tosetti, "Large panels with common factors and spatial correlation," *Journal of Econometrics* 161 (2011), 182–202. DOI: 10.1016/j.jeconom.2010.12.003

[35] Robinson, Peter M., "Efficient estimation of the semiparametric spatial autoregressive model," *Journal of Econometrics* 157 (2010), 6–17. DOI:10.1016/j.jeconom.2009.10.031

[36] Scott, David J., "Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach," *Advances in Applied Probability* 5 (1973), 119–137.

[37] Sonora, Robert J., "Bivariate relative city price convergence in the United States: 1918–1997," *Review of Financial Economics* 17 (2008), 92–111. DOI: 92–111. 10.1016/j.rfe.2007.02.010

[38] Urga, Giovanni, "Common Features in Economics and Finance, an overview of recent developments," *Journal of Business and Economics Statistics* 25 (2007), 2–11. DOI 10.1198/073500106000000602

[39] Yu, Jihai, Robert de Jong, and Lung-fei Lee, "Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration," *Journal of Econometrics* 167 (2012), 16–37. DOI: 10.1016/j.jeconom.2011.05.014

	At	Bo	Ch	Ci	Cl	De	Ho	KC	LA	Mi	NY	Ph	Pi	Po	SF	Se	SL
At	0	1023	667	433	663	671	736	758	2040	1046	821	754	661	2460	2340	2520	497
Bo	1023	0	940	801	612	698	1740	1349	2640	1312	230	330	551	2880	2700	2820	1125
Ch	667	940	0	282	331	270	1081	507	1740	407	766	734	430	1980	1800	1920	298
Ci	433	801	282	0	237	246	1046	565	2040	661	598	534	272	2220	2220	2280	334
Cl	663	612	331	237	0	169	1253	782	2040	722	456	428	128	2340	2100	2220	558
De	671	698	270	246	169	0	1264	749	1980	664	605	572	266	2220	2040	2160	531
Ho	736	1740	1081	1046	1253	1264	0	756	1299	1160	1500	1437	795	2220	1620	2280	869
KC	758	1349	507	565	782	749	756	0	1500	433	1147	1082	821	1740	1680	1800	249
LA	2040	2640	1740	2040	2040	1980	1299	1500	0	1800	2460	2400	2220	901	349	1057	1680
Mi	1046	1312	407	661	722	664	1160	433	1800	0	1143	1102	827	1680	1920	1620	580
NY	821	230	766	598	456	605	1500	1147	2460	1143	0	111	349	2700	2580	2640	923
Ph	754	330	734	534	428	572	1437	1082	2400	1102	111	0	284	2640	2520	2640	858
Pi	661	551	430	272	128	266	795	821	2220	827	349	284	0	2400	2400	2400	565
Po	2460	2880	1980	2220	2340	2220	2220	1740	901	1680	2700	2640	2400	0	591	166	1980
SF	2340	2700	1800	2220	2100	2040	1620	1680	349	1920	2580	2520	2400	591	0	746	1920
Se	2520	2820	1920	2280	2220	2160	2280	1800	1057	1620	2640	2640	2400	166	746	0	2040
SL	497	1125	298	334	558	531	869	249	1680	580	923	858	565	1980	1920	2040	0

Table 3: Distances between 17 US cities in minutes (At: Atlanta, Bo: Boston, Ch: Chicago, Ci: Cincinnati, Cl: Cleveland, De: Detroit, Ho: Houston, KC: Kansas City, LA: Los Angeles, Mi: Minneapolis, NY: New York, Ph: Philadelphia, Pi: Pittsburgh, Po: Portland, SF: San Francisco, Se: Seattle, SL: St. Louis. Source: googlemaps.com)