ELSEVIER

Journal of Econometrics 105 (2001) 59–83

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

# A new semiparametric spatial model for panel time series

Xiaoheng Chen[a], Timothy G. Conley[b,*]

[a]*Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK*

[b]*Department of Economics, Northwestern University, 2003 Sheridan Road, Evanston, IL 60208, USA*

**Abstract**

This paper presents a semiparametric model for large-dimension vector time series whose elements correspond to economic agents. Dependence between agents' variables is characterized using a spatial model. Functions of agents' *economic distances* provide restrictions that enable estimation of a vector autoregressive specification. We present sufficient conditions for our model to generate stationary, $\beta$-mixing series with finite higher-order moments. We estimate the model using a simple two-step sieve least-squares procedure, where the sieve estimators are constructed to preserve shape restrictions on the functions of economic distance, e.g., positive definiteness of a covariance function. We provide rates of convergence for the sieve estimators, $\sqrt{T}$ limiting distributions for the model's finite-dimensional parameters, and a bootstrap method for inference. In an illustrative application, we use this model to characterize how the comovement in output growth across US industrial sectors depends on the similarity of sectors' technologies. We also present a small Monte Carlo evaluation of our estimators. 

*JEL classification:* C3; C4; C5

*Keywords*: Space–time model; Sieve estimator; Time-varying coefficient VAR; Isotropic conditional covariances

* Corresponding author. Tel.: +1-847-491-8266; fax: +1-847-491-7001.

*E-mail addresses:* x.chen1@lse.ac.uk (X. Chen), tconley@nwu.edu (T.G. Conley).

0304-4076/01/$ - see front matter 
PII: S0304-4076(01)00070-7

## 1. Introduction

This paper presents an econometric model for large-dimension vector time series with a panel structure where there is dependence across variables as well as over time. By panel structure, we mean repeated observations of a large but fixed number of agents over time, rather than a vector containing a large number of aggregate attributes. Examples of this type of data include quarterly observations on sector-specific variables and weekly price data for many retail firms in a region. In situations like these, there are too few degrees of freedom to permit unrestricted time series estimation; restrictions are needed to make progress. Restrictions that are typically imposed are those that limit interrelationships such as exclusion restrictions in vector autoregressions, or those that assume there are a few underlying factors responsible for interrelationships. [1] However, the panel structure of these series presents the possibility that there may be some replication across agents' series that can be used to mitigate the problem of having a relatively short time series. [2] Our idea is to use additional information on the relationships between agents to model this replication across agents.

We model agents as residing in a Euclidean space where distances between them are determined by an 'economic metric'. For example, if agents correspond to sectors they may be close if they use inputs in nearly the same proportions, far if they use inputs in very different proportions. [3] If agents correspond to firms, measures of the overlap in their retail markets might be plausible measures of economic distances. [4] If agents are countries, measures of trade volumes or transport costs between countries may be suitable economic distance metrics. [5] Our data generating model allows economic distances between agents to be jointly determined along with the variables of interest and to change over time. [6] However, we require that these economic distances be observable to the econometrician.

The data-generating process is assumed to be characterized by economic distances between agents. The basic idea is that if there are two groups of agents with the same positions relative to each other, then there is replication in the cross section that can be exploited. Our data generation model is a

[1] See, for example, Horvath and Verbrugge (1997) for a VAR with many exclusion restrictions and Quah and Sargent (1993) for a dynamic factor model.

[2] Obviously, replication across series is what enables consistent (large $N$) estimation of large-cross-section panel data models, including those with temporal dynamic structure like, e.g., Holz-Eakin et al. (1988).

[3] Conley and Dupor (1999, 2000) construct such an economic distance for US industrial sectors from a series of benchmark input–output tables. We use this measure in Section 4.

[4] For other examples of economic distance in firm panels see Pulvino (1998).

[5] See, e.g., Conley and Ligon (2000) for measures of economic distances between countries meant to reflect transport costs.

[6] In our empirical example distances vary slowly over time, and are treated as exogenous.

vector autoregression (VAR) whose coefficient matrix and shock covariance matrix are functions of economic distances between agents. The impact of other agents' variables on the conditional mean of a given agent's variable is a function of their economic distances from this agent. Similarly, covariances of VAR shocks are functions of distances between agents in the previous period, a property we refer to as being isotropic.

We take a semiparametric approach to estimate our VAR coefficient matrix and shock covariance matrix. We estimate both using sieve estimators whose large sample properties are studied by Chen and Shen (1998).[7] By expanding the order of the sieves, we can approximate VAR coefficients that are arbitrary (subject to smoothness conditions) functions of distance and we can approximate any valid shock covariance matrix. We use representations of isotropic covariance functions borrowed from the statistics literature on random fields (vector-indexed variables) to choose basis functions for our sieves that will guarantee that our estimates of shock covariance matrices are positive definite.[8] The same sieve basis also allows us to estimate the VAR coefficient matrix imposing shape restrictions in a straightforward manner.

We present an empirical example that demonstrates how the paper's model can be used to characterize the dynamic correlation properties of sector-level variables and their relation to input–output structures. Several important models of business cycles rely on the presence of strong intersectoral linkages to account for fluctuations.[9] Quantifying these linkages using the paper's econometric method can provide a useful assessment of the plausibility of these models. Following Conley and Dupor (1999, 2000), we construct economic distances between sectors that measure linkages due to similarity of input technology. We present evidence suggesting that there is substantial comovement in sectoral output growth rates that is related to this metric.

There are many applications in other fields where large-dimension time series are of interest and our model or its potential extensions should be useful. Industrial organization applications include modeling commodity price dynamics for many locations, as in Slade (1986), and estimating demand elasticities using state-level series, as in Sumner (1981). Finance applications include characterizing distributions of many asset returns, as in Connor and Koracjczk (1993), and pricing of many macroeconomic factors, as in Chen et al. (1986). International applications include studies of purchasing power parity using series from many countries, see, for example, Taylor (2000).

Our spatial model puts different restrictions on comovement across agents than typical factor models. Covariance across variables in factor models is

[7] Sieve estimators use flexible approximating basis functions such as orthogonal series, polynomials, splines, and neural networks, see e.g., Gallant and Nychka (1987), Newey et al. (1999).

[8] See, e.g., Yaglom (1987) for a discussion of isotropic random fields.

[9] For example, see Long and Plosser (1983), Horvath (1998), or Dupor (1997).

mediated by some relatively low-dimensional set of factors, as in Quah and Sargent (1993) or Forni and Reichlin (1998). Our model does not require the existence of a low-dimensional source of covariation, instead it relies upon replications in covariance patterns as functions of economic distances. We also impose a certain symmetry in order to estimate conditional covariances. Still, our model can result in very general unconditional covariance patterns because we allow agents to move (potentially endogenously) over time. [10] Of course, we are limited to situations where measurements of economic distances are available and afford a sensible model of variables' relationships. Our spatial model may not be useful for large-dimension vector time series without a panel structure. For example, if the analysis is concerned with forecasting given a large number of aggregate indicators, as in Stock and Watson (1997), and these indicators are a large number of aggregate statistics, a spatial model may not be suitable to describe their relationships.

The remainder of the paper is organized as follows. Section 2 describes our model of data generation, and presents sufficient conditions for this model to generate stationary, $\beta$-mixing time series with finite higher-order moments. Section 3 presents our estimation strategy for the VAR coefficient and shock covariance matrices. We obtain consistency with rates of convergence for our sieve estimators, provide $\sqrt{T}$ asymptotic normality for the model's finite-dimensional parameters, and present a bootstrap method for inference. Section 4 presents an empirical example in which our approach is used to study output comovement across US industrial sectors. Section 5 presents a small Monte Carlo experiment evaluating the performance of our estimators. Finally, Section 6 discusses possible extensions and concludes.

## 2. Data generation

We use a model that characterizes the relationships between agents' variables by the economic distances between them. For simplicity, we use economic distance configurations that can be represented by a set of points (one for each agent at each time period) in $\Re^k$. Thus we model agents as residing at locations in $\Re^k$, with agent $i$ at time $t$ located at a point $s_{i,t}$. Each agent also has a variable of interest $X_{i,t}$ which we take to be a scalar for simplicity.

[10] Our modeling moving locations and flexible dynamic interactions are the main differences between our model and related work in the spatial–temporal statistics literature. In this literature, typical data sets consist of repeated observations at a fixed number of physical locations. Dependence over time is usually treated in a symmetric manner with dependence over space, see Carroll et al. (1997), limiting the ability of these models to capture dynamic interactions (see Cressie, 1997). There are a few spatial statistics models that do allow a linear dynamic structure, for example Huang and Cressie (1997) and Meiring et al. (1998). However, they still restrict attention to fixed, exogenous locations.

The econometrician's sample of $N$ agents' variables consists of the realizations of $X_{i,t}$ at a collection of locations $\{s_{i,t}\}_{i=1}^N$ for each of $T$ time periods: $\{X_{i,t}\colon i=1,\dots,N,\ t=1,\dots,T\}$. We denote $Y_t=(X_{1,t},X_{2,t},\dots,X_{N,t})'\in\mathfrak{R}^N$ as a vector collecting $\{X_{i,t}\}_{i=1}^N$ and $D_t$ as a vector of distances between the $\{s_{i,t}\}_{i=1}^N$. Referring to the distance between points $i$ and $j$ as $D_t(i,j)=\|s_{i,t}-s_{j,t}\|$, we stack the distances as

$$D_t=(D_t(1,2),\dots,D_t(1,N),D_t(2,3),\dots,D_t(2,N),\dots,D_t(N-1,N))' \in\mathfrak{R}^{N(N-1)/2}.$$

We assume that $D_t(i,j)$ has common support $(0,\bar{a}]$ for all $t$, $i\neq j$.

We assume that the joint process $\{(Y_t,D_t)\colon t=1,\dots,T\}$ is a first-order Markov process, [11] and that $Y_t$ evolves according to the following nonlinear VAR model:

$$Y_{t+1}=A(D_t)Y_t+e_{t+1},\quad e_{t+1}\equiv Q(D_t)u_{t+1}, \tag{2.1}$$

where $u_{t+1}$ is an IID sequence with $Eu_{t+1}=0$ and $Eu_{t+1}u_{t+1}'=I_N$. The $N\times N$ matrix $A(D_t)$ consists of elements that are functions of economic distances between agents. $\Sigma(D_t)\equiv Q(D_t)Q(D_t)'$ is a conditional covariance matrix that is also a function of distances between agents.

## 2.1. Structure on conditional means

We model the conditional mean of $X_{i,t+1}$ given $\{(Y_{t-l},D_{t-l}),l\geqslant 0\}$ as follows:

$$E[X_{i,t+1}|\{(Y_{t-l},D_{t-l}),l\geqslant 0\}]=\alpha_i X_{i,t}+\sum_{j\neq i} g_i(D_t(i,j))X_{j,t},$$

where the $g_i(\cdot)$ are continuous functions from $(0,\infty)$ to $\mathfrak{R}^1$. Hence the conditional mean of $Y_{t+1}$ given $\{(Y_{t-l},D_{t-l}),l\geqslant 0\}$ is $A(D_t)Y_t$,

$$A(D_t)=\begin{bmatrix} \alpha_1 & g_1(D_t(1,2)) & \dots & g_1(D_t(1,N)) \\ g_2(D_t(2,1)) & \alpha_2 & \dots & g_2(D_t(2,N)) \\ \cdots & \cdots & \cdots & \cdots \\ g_N(D_t(N,1)) & g_N(D_t(N,2)) & \dots & \alpha_N \end{bmatrix}. \tag{2.2}$$

In the following theoretical development, we model the $g_i$ functions in a general way, letting each be distinct. However, in practice it may be quite desirable to model $g_i$ as having considerable features in common across $i$. For example, $g_i$ might differ across $i$ only in scale: $g_i=\gamma_i g$ with $g$ being a monotone function.

[11] Note that we only model the joint process of $Y_t$ and the economic distance $D_t$ as a Markov process, without imposing any structure directly on locations $\{s_{i,t}\}_{i=1}^N$ which may include physical locations as parts of the elements in $s_{i,t}$.

### 2.2. Structure on conditional variances

We model the conditional variance of $Y_{t+1}$ given $\{(Y_{t-l}, D_{t-l}), l \geqslant 0\}$ as

$$\Sigma(D_t) = \begin{bmatrix} \sigma_1^2 + C(0) & C(D_t(1,2)) & \dots & C(D_t(1,N)) \\ C(D_t(2,1)) & \sigma_2^2 + C(0) & \dots & C(D_t(2,N)) \\ \dots & \dots & \dots & \dots \\ C(D_t(N,1)) & C(D_t(N,2)) & \dots & \sigma_N^2 + C(0) \end{bmatrix}, \tag{2.3}$$

where $C(\cdot)$ is assumed to be continuous at zero and is a $k$-dimensional isotropic covariance function. In other words, $C$ is assumed to be covariance function for stationary random fields with indices in $\Re^k$ whose covariance depends only on distance, not direction. This choice of $C$ implies that $\Sigma(D_t)$ will be positive definite for any set of interpoint distances $D_t$ and any values of the $\sigma_i^2 \geqslant 0$.

We use a representation of $C$ that is analogous to the spectral representation of a time-series covariance function. An isotropic covariance function has a representation as an integral of a generalized Bessel function against a positive finite measure, see e.g. Yaglom (1987, Section 22, pp. 353–354). That is, for any $\tau \in \Re^k$ with $||\tau|| = \sqrt{\sum_{i=1}^k \tau_i^2}$,

$$C(||\tau||) = \int_0^\infty h(y||\tau||)\, \mathrm{d}\Phi(y), \tag{2.4}$$

$$h(y||\tau||) \equiv 2^{(k-2)/2}\Gamma\left(\frac{k}{2}\right) \frac{J_{(k-2)/2}(y||\tau||)}{(y||\tau||)^{(k-2)/2}} \tag{2.5}$$

where $\Phi$: $[0,\infty) \to [0, \Phi(\infty)]$ is a bounded nondecreasing function, and $J_{(k-2)/2}(\cdot)$ is a Bessel function. The function $\Phi$ is analogous to spectral measure for covariance stationary time series. The smoother the function $C(||\tau||)$ is, the more moments the measure $\Phi$ possesses. This representation allows us to easily approximate any isotropic covariance function by approximating $\Phi$.

For many values of $k$, $h(y||\tau||)$ simplifies. For example, when $k = 3$,

$$h(y||\tau||) = \frac{\sin(y||\tau||)}{y||\tau||}$$

and when $k = \infty$,

$$h(y||\tau||) = \exp(-[y||\tau||]^2). \tag{2.6}$$

If a function is an isotropic covariance function in $k$ dimensions then it is also a valid isotropic covariance function in any lower number of dimensions. Therefore, the limiting expression (2.6) for $h$ will generate a set of covariance functions that are valid for locations $\{s_{i,t}\}$ in $\Re^k$, for any $k$.

### 2.3. Some probabilistic properties of the model

In this subsection, we provide two sets of sufficient conditions to ensure the asymptotic stability of the dynamic system. We are particularly interested in properties such as stationarity, ergodicity, finiteness of higher-order moments, and temporal dependence structure described in terms of *β-mixing.*[12] Either of the following two Propositions (2.1) or (2.2) can be used in Section 3 to establish large sample properties of our estimators:

*Case* I. Exogenous $D_t$

AI.1. $\{D_t\colon t \geqslant 0\}$ is IID with continuous marginal density (w.r.t. Lebesgue measure restricted to $[0,\bar{a}]^{N(N-1)/2}$), and is independent of $\{u_{t+1}\}$ and $Y_0$. The density of $D_t$ is positive over $(0,\bar{a}]^{N(N-1)/2}$.

AI.2. $\{u_{t+1}\colon t \geqslant 0\}$ is IID and independent of $Y_0$. $Eu_{t+1} = 0$, $Eu_{t+1}u'_{t+1} = I_N$ and $E(||u_{t+1}||^s) < \infty$ for some $s \geqslant 2$. The marginal distribution is absolutely continuous w.r.t. Lebesgue measure on $\Re^N$, and has positive continuous marginal density on $\Re^N$.

AI.3. $\max\{|\alpha_1|,\ldots,|\alpha_N|\} < 1$, and $g_i(\cdot)$, $i=1,\ldots,N$ are continuously differentiable over $(0,+\infty)$.

AI.4. $0 < \min\{\sigma_1^2 + C(0),\ldots,\sigma_N^2 + C(0)\} < \infty$, and $C(\cdot)$ is continuously differentiable over $[0,+\infty)$.

AI.5. $E([\rho(A(D_t))]^s) < 1$ for some integer $s \geqslant 2$, where $\rho(\cdot)$ is the spectral radius of $A(D_t)$ (i.e., the largest eigenvalue in absolute value of $A(D_t)$).

*Proposition 2.1. Under AI.1–AI.5, $\{Y_t\}$ is geometrically ergodic Markov and $E[||Y_t||^s] < \infty$. If in addition $Y_0$ is initialized from the invariant distribution, then $\{Y_t\}$ is strictly stationary, β-mixing with an exponential decay rate.*

*Proof.* See the appendix. □

*Case* II. (Possibly) Endogenous $D_t$

We assume that $\{D_t\}$ evolves according to

$$\ln(D_{t+1}) = \Psi_1(Y_t, \ln(D_t)) + \Psi_2(\ln(D_t))v_{t+1},$$

where $\{v_t\}$ is an IID shock with zero mean and identity covariance matrix. We concentrate on estimation of objects that do not require estimation of $\Psi_1$ and $\Psi_2$. Therefore, we try to be agnostic about this process and make

[12] Recall that a time series $\{Z_t\colon t \geqslant 0\}$ is *β-mixing* (or *absolutely regular*) if $\beta_n \equiv \sup_t E[\sup_{B\in\mathscr{F}_{t+n}^{\infty}} |P(B|\mathscr{F}_0^t) - P(B)|] \to 0$ as $n \to \infty$, where $\mathscr{F}_0^t$ and $\mathscr{F}_{t+n}^{\infty}$ denote the sigma-fields generated by $(Z_0,\ldots,Z_t)$ and $(Z_{t+n},\ldots,Z_\infty)$ respectively. The process is called *β-mixing with an exponential decay rate* if $\beta_n = O(\rho^n)$ for some $\rho \in (0,1)$.

minimal assumptions about $\Psi_1$ and $\Psi_2$ that ensure stationarity and $\beta$-mixing. Obviously, to solve problems such as forecasting $Y$ more than a single period in advance, we will require more structure on $\Psi_1$ and $\Psi_2$ than we currently impose. We look at the joint process

$$Z_{t+1} = \begin{bmatrix} Y_{t+1} \\ \ln(D_{t+1}) \end{bmatrix} = \begin{bmatrix} A(D_t)Y_t \\ \Psi_1(Y_t, \ln(D_t)) \end{bmatrix} + \begin{bmatrix} Q(D_t) & 0 \\ 0 & \Psi_2(\ln(D_t)) \end{bmatrix} \begin{bmatrix} u_{t+1} \\ v_{t+1} \end{bmatrix}.$$

AII.1. $\{(u_{t+1}, v_{t+1}): t \geqslant 0\}$ is IID with $E(u_t, v_t) = (0, 0)$, $E(u_t, v_t)(u_t, v_t)' = I$ and independent of $(Y_0, D_0)$. The marginal distribution is absolutely continuous w.r.t. Lebesgue measure, and equivalent to Lebesgue measure in a neighborhood of zero, and has continuous marginal density, and finite $s \geqslant 2$ order moments.

AII.2. There exist two nonsingular, nonrandom matrices $\Omega_1$ and $\Omega_2$ s.t.

$$\Omega_1 \leqslant \begin{bmatrix} Q(D) & 0 \\ 0 & \Psi_2(\ln(D)) \end{bmatrix} \leqslant \Omega_2 \quad \text{almost surely.}$$

AII.3. (i) AI.3 holds and $\sup_{D(i,j)} |g_i(D(i,j))| \leqslant \bar{g} < \infty$ all $i = 1, \ldots, N, i < j$.
(ii) $\Psi_1(Y, \ln(D))$ is bounded over any compact set; and there exist constant matrices $B_1, B_2$ s.t. $\Psi_1(Y, \ln(D)) \approx B_1 Y + B_2 \ln(D)$ as $||(Y, \ln(D))|| \to \infty$.

AII.4.

$$\rho\left(\begin{bmatrix} \bar{A} & 0 \\ B_1 & B_2 \end{bmatrix}\right) < 1,$$

where the matrix $\bar{A}$ is the matrix $A(D)$ with all the $g_i(D(i,j))$ elements replaced by $\bar{g}$.

*Proposition 2.2. Under AII.1–AII.4, $\{(Y_t, \ln(D_t))\}$ is geometrically ergodic Markov with finite s-th moments. If in addition $(Y_0, \ln(D_0))$ is initialized from the invariant distribution, then $\{(Y_t, \ln(D_t))\}$ is strictly stationary and $\beta$-mixing with an exponential decay rate.*

*Proof.* See the appendix. □

For simplicity, we assume in the following sections that the stationary (marginal) probability distribution of $D_t(i,j)$ is absolutely continuous with respect to Lebesgue measure and has a bounded density on $[0, \bar{a}]$. We note that formally this rules out time-invariant distances. If distances were fixed then, strictly speaking, the functions $g_i$ and $C$ are only identified at a finite number of points and the estimation problem can be thought of as a completely parametric one. However, the number of distinct distances in many applications is likely to be very large relative to sample size. In such cases, we argue that proceeding as if the distances distribution were in fact absolutely continuous w.r.t. Lebesgue measure provides a very useful approximation.

## 3. Estimation and inference

We use the method of sieves (Grenander, 1981) to construct estimators for the $g_i$ and $\Phi$. Essentially, this approach is to use a flexible sequence of parametric families, e.g. Fourier series, orthogonal polynomials, splines, wavelets or neural networks, to approximate the unknown functions. The number of terms in the approximating sequence is allowed to grow as the sample size increases to improve the accuracy of the approximation. In this section, we first describe our choices of sieves to approximate the $g_i$ and $\Phi$. We then present our estimators, their large sample properties, and a bootstrap-based method for conducting inference.

### *3.1. Shape-preserving cardinal B-spline wavelet sieve*

Our choice of an approximating sequence is motivated by our need to obtain a nondecreasing estimate of $\Phi$ in (2.4) to ensure that our estimated $C$ is valid, implying a positive definite estimate of $\Sigma$. In addition, shape restrictions on $g_i$ in (2.2) will also be of interest. Thus, we choose a basis that makes it easy to impose shape restrictions; following Chen et al. (1997) we use a shape-preserving cardinal B-spline sieve. Let $B_m$ denote the cardinal B-spline of order $m$, a piecewise polynomial of highest degree $m-1$. The pieces are tied together so that the resulting function is $m-1$ times differentiable, nonnegative, has support on $[0,m]$ and is symmetric about the center of its support

$$B_m(x) = \frac{1}{(m-1)!}\sum_{k=0}^{m}(-1)^k \binom{m}{k} [\max(0, x-k)]^{m-1}. \tag{3.1}$$

Hence a first-order spline is the unit function with support $[0,1]$, and a second-order spline is a tent with support $[0,2]$ and a peak at 1, see Chui (1992). Thus the functions $g_i$ and $\Phi$ of interest are approximated by

$$g_i(y) \approx \sum_{j=-\infty}^{+\infty} a_j B_{m_i}(2^n y - j), \tag{3.2}$$

$$\Phi(y) \approx \sum_{j=-\infty}^{+\infty} b_j B_m(2^n y - j) \tag{3.3}$$

for some $n$. The index $n$ and coefficients $a_j$ are allowed to differ across these approximations of course, however we suppress this in our notation. [13]

[13] We also suppress the dependence in $n$ in the coefficients $a_j, b_j$. For instance, in Eq. (3.3) one often writes $b_{n,j}2^{n/2}B_m(2^n y - j)$ in the more standard wavelets approximation. Here we effectively use spline approximation.

This is a two index depiction of a family of basis functions. The index $j$ is a translation and the index $n$ provides a scale refinement. Since $B_m$ has compact support, the summation over $j$ is effectively summing over a finite number of terms, say $j \in [\underline{J}, \bar{J}]$, determined by the support of the data. The right sides of (3.2) and (3.3) show the approximation for given scale refinement (choice of $n$). As $n$ gets larger, more $B_m(2^n y - j)$ terms are included and the approximation improves. A discussion of the denseness of this approximation class and the shape-preserving property can be found in Anastassiou and Yu (1992); and the approximation accuracy can be found in Chen et al. (1997). Since $B_m$ is nonnegative, to obtain a nondecreasing and nonnegative approximator for $\Phi$, we can simply restrict the coefficient sequence $\{b_j\}$ to be nondecreasing and nonnegative.

### 3.1.1. Sieve for $g_i$

We suppose that the true function $g_i \in \mathscr{G}_i$, where

$$\mathscr{G}_i = \left\{ g \in C^{r_i}([0,\bar{a}]): \sup_{x,y\in[0,\bar{a}],x\neq y} \frac{|g^{(r_i)}(x) - g^{(r_i)}(y)|}{|x-y|^{\varsigma_i}} < \infty \right\} \tag{3.4}$$

with $r_i$ is some nonnegative integer and $0 < \varsigma_i \leqslant 1$. We assume that the total smoothness is $m_i \equiv r_i + \varsigma_i > 1/2$.

We consider the following sieve:

$$\mathscr{G}_{i,T} = \left\{ \begin{array}{c} g \in \mathscr{G}_i\text{: } g(y) = \sum_{j=\underline{J}_i}^{\bar{J}_i} a_j B_{m_i}(2^{n_i(T)} y - j), \\ \int [g^{(m_i)}(x)]^2 \, \mathrm{d}x \leqslant c_i(T) \end{array} \right\}, \tag{3.5}$$

where $\underline{J}_i = -m_i$ and $\bar{J}_i = 2^{n_i(T)}\bar{a}$, $c_i(T) \to \infty$ very slowly at the order of $\log(T)$ and $2^{n_i(T)} \to \infty$ at the order of $T^{1/(2m_i+1)}$. Thus, smooth functions (with larger $m_i$) allow slower growth of $n_i(T)$.

*Remark 1.* Monotonicity restrictions are easily enforced in this framework. For example, if $g$ is nonincreasing we can then impose $g^{(1)} \leqslant 0$ in the population parameter space (3.4), and $\cdots \geqslant a_j \geqslant a_{j+1} \geqslant \cdots$ in sieve (3.5).

### 3.1.2. Sieve for $C$

We assume that the true $C \in \mathscr{C}$, where

$$\mathscr{C} = \left\{ \begin{array}{c} C(\|\tau\|) = \int_0^\infty h(y\|\tau\|) \, \mathrm{d}\Phi(y)\text{: } \Phi \text{ bounded, positive nondecr.} \\ C \text{ continuous at 0, and has } m_c \text{ derivatives} \end{array} \right\}. \tag{3.6}$$

Our sieve for $C$ can be viewed as plugging the approximation for $\Phi$ in (3.3) into the representation in (2.4) or equivalently as an approximation of $C$ using a new sieve basis obtained from $B_m$:

$$H_{n,j}(x)=2^n\int_0^{\infty} h(y||x||)B_m^{(1)}(2^n y-j)\,\mathrm{d}y \quad \text{for } x\in[0,\infty).$$

Thus we use the following sieve:

$$\mathscr{C}_T=\left\{\begin{array}{c} C\colon C(||\tau||)=\sum_{j=\underline{J}}^{\bar{J}} b_j H_{n(T),j}(||\tau||),\ C \text{ continuous at } 0 \\ \cdots\leqslant b_j\leqslant b_{j+1}\leqslant\cdots,\ \int [C^{(m_c)}(z)]^2\,\mathrm{d}z\leqslant c(T) \end{array}\right\}, \tag{3.7}$$

where $\underline{J}$ and $\bar{J}$ are determined by the support of $\{D_t(i,j)\}$ and $c(T)$ grows at $\log(T)$ order. Note that the constraint $\cdots\leqslant b_j\leqslant b_{j+1}\leqslant\cdots$ implies that the corresponding estimates of $\Phi$ is nondecreasing and $C$ is positive definite.

### 3.2. Estimation: Two-step sieve least squares

In this subsection, we introduce an additional subscript to denote true functions and parameters to facilitate exposition of our estimators' properties. Thus, we denote the true values of $\alpha_i, g_i(\cdot), \sigma_i^2$, $i=1,\ldots,N$ and $C(\cdot)$ by $\alpha_{i,o}$, $g_{i,o}(\cdot)$, $\sigma_{i,o}^2$, $i=1,\ldots,N$ and $C_o(\cdot)$. We can consistently estimate all the parameters and functions of interest by the following simple two-step sieve least-squares (LS) procedure [14]:

*Step* 1: LS estimation of $\alpha_i$ and $g_i(\cdot)$ using the conditional mean specification (2.2) and sieve (3.5) for $g_i$.

We assume the true $\alpha_o=(\alpha_{1,o},\ldots,\alpha_{N,o})'\in\Re^N$ and $g_o=(g_{1,o},\ldots,g_{N,o})\in\prod_{i=1}^N\mathscr{G}_i$, and estimate $\alpha_o$ and $g_o$ by least squares:

$$(\hat{\alpha}_T,\hat{g}_T)=\arg\min_{(\alpha,g)\in\Re^N\times\Pi_{i=1}^N\mathscr{G}_{i,T}}\frac{1}{T}\sum_{t=1}^{T}[Y_{t+1}-A(D_t)Y_t]'[Y_{t+1}-A(D_t)Y_t]. \tag{3.8}$$

With the level of generality presented here, there are no cross-equation restrictions imposed among the true parameters $(\alpha_{1,o},\ldots,\alpha_{N,o}),(g_{1,o},\ldots,g_{N,o})$ and this will reduce to equation by equation estimation of the system

$$X_{i,t+1}=\alpha_{i,o}X_{i,t}+\sum_{j\neq i} g_{i,o}(D_t(i,j))X_{j,t}+e_{i,t+1}, \tag{3.9}$$

$$E[e_{i,t+1}|\{(Y_{t-l},D_{t-l}),l\geqslant 0\}]=0, \tag{3.10}$$

$$E[e_{i,t+1}^2|\{(Y_{t-l},D_{t-l}),l\geqslant 0\}]=\sigma_{i,o}^2+C_o(0). \tag{3.11}$$

[14] Weighted LS may, of course, provide more efficient estimates of some components.

Estimates for $\alpha_i$, $g_i$, $i=1,\dots,N$ can be obtained by performing $N$ individual sieve LS minimizations using

$$(\hat{\alpha}_{i,T},\hat{g}_{i,T})=\arg\min_{(\alpha_i,g_i)\in\Re\times\mathcal{G}_{i,T}}\frac{1}{T}\sum_{t=1}^{T}\left[X_{i,t+1}-\left[\alpha_i X_{i,t}+\sum_{j\neq i}g_i(D_t(i,j))X_{j,t}\right]\right]^2. \tag{3.12}$$

We denote $\hat{e}_{t+1}=(\hat{e}_{1,t+1},\dots,\hat{e}_{N,t+1})$ as the LS residual, that is for $i=1,\dots,N$,

$$\hat{e}_{i,t+1}=X_{i,t+1}-\left[\hat{\alpha}_{i,T}X_{i,t}+\sum_{j\neq i}\hat{g}_{i,T}(D_t(i,j))X_{j,t}\right]. \tag{3.13}$$

*Step* 2: Sieve estimation of $\sigma^2$ and $C(\cdot)$ using the conditional variance specification (2.3) and (2.4) and sieve (3.7) for $C$, and the fitted residuals $\hat{e}_{i,t+1}$ from step 1.

We assume the true $\sigma_o^2=(\sigma_{1,o}^2,\dots,\sigma_{N,o}^2)\in(0,\infty)^N$ and $C_o(\cdot)\in\mathcal{C}$. Their sieve estimates $(\hat{\sigma}_T^2,\hat{C}_T(\cdot))$ solve:

$$\min_{(\sigma^2,C)\in(0,\infty)^N\times\mathcal{C}_T}\sum_{t=1}^{T-1}\left\{\sum_i(\hat{e}_{i,t+1}^2-[\sigma_i^2+C(0)])^2\right.$$
$$\left.+\sum_i\sum_{i\neq j}(\hat{e}_{i,t+1}\hat{e}_{j,t+1}-C(D_t(i,j)))^2\right\}. \tag{3.14}$$

Simple calculation yields: $\hat{\sigma}_{i,T}^2=\max\{(1/(T-1))\sum_{t=1}^{T-1}\hat{e}_{i,t+1}^2-\hat{C}_T(0),0\}$, and $\hat{C}_T(0)=\sum_{l=\underline{J}}^{\bar{J}}\hat{b}_{l,T}H_{n(T),l}(0)$. Coefficient estimates $\{\hat{b}_{l,T}\}$ can be obtained from the first-order conditions for $l=\underline{J},\dots,\bar{J}$,

$$\sum_{t=1}^{T-1}\sum_i\sum_{i\neq j}\left(\hat{e}_{i,t+1}\hat{e}_{j,t+1}-\sum_{k=\underline{J}}^{\bar{J}}\hat{b}_{k,T}H_{n(T),k}(D_t(i,j))\right)H_{n(T),l}(D_t(i,j))=0.$$

*Remark 2.* If we allow $C(\cdot)$ to be discontinuous at zero then we can only identify and estimate $\sigma_i^2+C(0)$, $i=1,\dots,N$ jointly.

*Theorem 3.1. Suppose either Proposition 2.1 or 2.2 holds with* $s>2$, *and that all the distances* $D_t(i,j)$, $i\neq j$ *have positive continuous densities with support* $(0,\bar{a}]$. *Let the true* $g_{i,o}\in\mathcal{G}_i$ *with* $m_i>1/2$, $c_i(T)=O(\log(T))$, *and* $2^{n_i(T)}=O(T^{1/(2m_i+1)})$. *Then for* $i=1,\dots,N$,

(1) $\sqrt{\int[\hat{g}_{i,T}(x)-g_{i,o}(x)]^2\,\mathrm{d}x}=O_P(T^{-m_i/(2m_i+1)})$;

(2) *If* $X_{i,t} \neq E[X_{i,t} \mid D_t(i,j), X_{j,t}, j \neq i]$ *with positive probability, then* $\sqrt{T}(\hat{\alpha}_{i,T} - \alpha_{i,o}) \Rightarrow \mathcal{N}(0, \sigma^2_{\alpha_i})$ *where*

$$\sigma^2_{\alpha_i} = [\sigma^2_{i,o} + C_o(0)] \times \left\{ \inf_{w_i \neq 0, w_i \in \bar{\mathscr{G}}_i} E\left[ X_{i,t} - \sum_{j \neq i} w_i(D_t(i,j)) X_{j,t} \right]^2 \right\}^{-1} > 0, \tag{3.15}$$

*where* $\bar{\mathscr{G}}_i$ *denotes the linear completion of* $\mathscr{G}_i - \{g_{i,o}\}$.

(3) *If either Proposition* 2.1 *or* 2.2 *holds with* $s > 4$, *then* $\sqrt{T}([\hat{\sigma}^2_{i,T} + \hat{C}_T(0)] - [\sigma^2_{i,o} + C_o(0)]) \Rightarrow \mathcal{N}(0, V_{e_i})$ *where*

$$V_{e_i} = E[e^2_{i,t+1} - E(e^2_{i,t+1})]^2.$$

(4) *Let either Proposition* 2.1 *or* 2.2 *hold with* $s > 4$, *and the true* $C_o(\cdot) \in \mathscr{C}$ *with* $m_c \in (0.5, 2]$. *If* $c(T) = O(\log(T))$ *and* $2^{n(T)} = O(T^{1/(2m_c+1)})$, *then:* $\sqrt{\int_0^{\bar{a}} [\hat{C}_T(z) - C_o(z)]^2 \, \mathrm{d}z} = O_P(T^{-m_c/(2m_c+1)})$.

*Proof.* See the appendix. □

Under the assumptions for Theorem 3.1, by applying the results in Chen and Shen (1998), we actually obtain multivariate asymptotic normality of $\sqrt{T}(\hat{\alpha}_{1,T} - \alpha_{1,o}, \ldots, \hat{\alpha}_{N,T} - \alpha_{N,o})$ and $\sqrt{T}([\hat{\sigma}^2_{1,T} + \hat{C}_T(0)] - [\sigma^2_{1,o} + C_o(0)], \ldots, [\hat{\sigma}^2_{N,T} + \hat{C}_T(0)] - [\sigma^2_{N,o} + C_o(0)])$, see the appendix. However, the asymptotic covariance expressions are very complicated, so we do not report them here.

*Remark 3.* We can state (2) as: If $E[X_{i,t} - E[X_{i,t} \mid D_t(i,j), X_{j,t}, j \neq i]]^2 > 0$, and if there is a nonzero function $w_i^* \in \bar{\mathscr{G}}_i$ such that

$$E[X_{i,t} \mid D_t(i,j), X_{j,t}, j \neq i] = \sum_{j \neq i} w_i^*(D_t(i,j)) X_{j,t}, \tag{3.16}$$

then $\sqrt{T}(\hat{\alpha}_{i,T} - \alpha_{i,o}) \Rightarrow \mathcal{N}(0, \sigma^2_{\alpha_i})$ with

$$\sigma^2_{\alpha_i} = [\sigma^2_{i,o} + C_o(0)] \times \{E[X_{i,t} - E[X_{i,t} \mid D_t(i,j), X_{j,t}, j \neq i]]^2\}^{-1}.$$

Notice that when $(X_{1,t}, \ldots, X_{N,t})$ has a multivariate normal distribution conditional on $\{D_t(i,j)\}$, then (3.16) is satisfied. However without making specific distributional assumptions, it is in general difficult to check whether or not the conditional expectation $E[X_{i,t} \mid D_t(i,j), X_{j,t}, j \neq i]$ has the form $\sum_{j \neq i} w_i^*(D_t(i,j)) X_{j,t}$ for some function $w_i^* \in \bar{\mathscr{G}}_i$. Therefore, we express the asymptotic variance as (3.15) in Theorem 3.1. To conduct inference about $\alpha_i$, we need a consistent estimator of $\sigma^2_{\alpha_i}$. We could approximate the solution to $\inf_{w_i \neq 0, w_i \in \bar{\mathscr{G}}_i} E[X_{i,t} - \sum_{j \neq i} w_i(D_t(i,j)) X_{j,t}]^2 > 0$ by sieves which are dense in $\bar{\mathscr{G}}_i$. Alternately, we can conduct bootstrap-based inference as described in Section 3.3.

### 3.3. Inference

This section describes a heuristic bootstrap method for conducting inference for the special case of our model where distances are exogenous. Our primary motivation for using a bootstrap is the fact that we do not have a pointwise distribution result for our nonparametric estimators $\hat{g}$ and $\hat{C}$. In addition, although we have $\sqrt{T}$ asymptotic normality results for the parametric portions of our model, the asymptotic variance of $\hat{\alpha}$ may be difficult to estimate (see Remark 3).

In presenting the bootstrap method, we abstract from consideration of the distance generating process, conditioning on the observed sequence of distances $\{D_t\}$ and keeping them fixed in constructing our bootstrap samples. This is motivated by applications such as that in Section 4 where there is neither an obvious specification for a distance evolution process, nor sufficient data to estimate one were it available.[15] In this application, distances vary slowly relative to $Y_t$, and this 'fixed distance-sample-path bootstrap' should still give useful approximations. Of course, given a model of exogenous distance evolution and enough data, it would be straightforward to generalize this bootstrap to include this source of variation.

The initial step is to use our consistent estimates of the model's parameters, $\hat{A}(D_t)$ and $\hat{\Sigma}(D_t)$ to construct estimates of the shocks in the model. Defining $\hat{Q}(D_t)$ as the Cholesky factor of $\hat{\Sigma}(D_t)$, estimates of the underlying shocks can be obtained from residuals $\hat{e}_{t+1}$, by premultiplication by $\hat{Q}(D_t)^{-1}$ : $\hat{u}_{t+1} \equiv [\hat{Q}(D_t)]^{-1}\hat{e}_{t+1}$. We generate a sequence of $u$ shocks by drawing independently from the empirical distribution of $\hat{u}_{t+1}$. Then, using the first observation in our data, $Y_1$, as a starting value we generate bootstrap samples of length $T$ using the actual sequence of distances $\{D_1, D_2, \ldots\}$ (for all bootstrap samples) and the autoregression

$$Y_{t+1} = \hat{A}(D_t)Y_t + \hat{Q}(D_t)\hat{u}_{t+1}.$$

Once the bootstrap samples are generated, we can estimate $\{\alpha_i\}$, $\{\sigma_i^2\}$, $\{g_i\}$ and $C$ for each bootstrap sample using the same estimation method as was used for the point estimates. The resulting collection of estimates allows us to recover the bootstrap distribution of $\{\alpha_i\}$, $\{\sigma_i^2\}$ and $g_i(d)$ and $C(d)$ for each distance $d$ which we use to construct approximate confidence intervals and standard errors.

[15] The economic distances in Section 4 are formed using benchmark input–output matrix data from four years: 1972, 1977, 1982, and 1987. So there are not enough observations to estimate a dynamic model of their evolution.

## 4. An empirical example

This section presents an example application of our estimation method to the study of output comovement across US industrial sectors to illustrate our methods. Several important models of business cycles rely on the presence of strong intersectoral linkages to account for fluctuations. This motivates an empirical investigation of whether measures of such linkages, captured in economic metrics, are systematically related to comovements in sectoral output. Therefore, we estimate our model for 20 industrial sectors' output growth rates.

Our economic distance metric is motivated by the idea that comovements in output growth may be due to innovations in common technologies. Therefore, we construct an economic metric of similarity of technologies across sectors using input requirements. [16] Our metric implies sectors are close if they use inputs in similar proportions and far if they do not. Specifically, sector locations $s_{i,t}$ are identified with input shares calculated for 22 inputs including the 20 industrial sectors, labor, and an aggregate of all other goods. These input shares are calculated using benchmark use tables prepared by the Bureau of Economic Analysis in 1972, 1977, 1982, and 1987. To construct quarterly economic distances, we identify the benchmark input–output matrices with configurations of sectors in the first quarters of 1972, 1977, 1982 and 1987. Observations for quarters prior to 1972:1 and after 1987:1 are the 1972 and 1987 benchmarks, respectively. Configurations for intermediate quarters are imputed as linear interpolations of the benchmarks. Economic distances in the matrix $D_t$ are taken to be the Euclidean distances between sectors' input share vectors. The nonzero distances range from 0.09 to 0.81 with a median of 0.36. Our $Y_t$ observations are de-meaned quarterly growth rates of the Federal Reserve Board monthly industrial production index for the 20 two-digit SIC manufacturing sectors from 1947:2 to 1992:4.

We estimate our model defined by (2.1), (2.2), and (2.3) with $g_i$ specified to be common across sectors, using least squares (LS). We approximate $g$ as a linear combination of six third-order B-splines (bell-shaped curves given by (3.1) with $m = 3$) scaled to be evenly spaced over the support of the distance distribution. Estimation proceeds as though this approximation for $g$ and the $\alpha_i$ coefficients on the main diagonal of $A(D_t)$ are a completely parametric model of the conditional mean of $Y_{t+1}$ given $Y_t$ and $D_t$. The parameters in the approximation for $g$ and the $\alpha_i$ parameters are then estimated using LS. Residuals from this LS procedure are used to estimate $C$.

Our estimates of $C$ in (2.3) are obtained by minimizing the sum of squared differences between an approximation for $C$ and these residuals. Although our

[16] See Conley and Dupor (1999) for more detail on the construction of this and other economic distance metrics between sectors.

Table 1

| $\alpha_i$ Point estimates and bootstrap SE in ( ) | | | |
|---|---|---|---|
| SIC20 food | −0.42 (0.060) | SIC30 plastics | 0.03 (0.069) |
| SIC21 tobacco | −0.43 (0.071) | SIC31 leather | 0.15 (0.066) |
| SIC22 textile | 0.23 (0.069) | SIC32 stone, glass | 0.17 (0.063) |
| SIC23 apparel | 0.07 (0.070) | SIC33 prim. metals | −0.08 (0.075) |
| SIC24 lumber | 0.01 (0.068) | SIC34 fab. metals | 0.16 (0.060) |
| SIC25 furniture | 0.22 (0.060) | SIC35 nonelec. mach. | 0.53 (0.045) |
| SIC26 paper | 0.16 (0.066) | SIC36 elec. mach. | 0.28 (0.058) |
| SIC27 printing | −0.03 (0.050) | SIC37 transportation | 0.01 (0.071) |
| SIC28 chemicals | 0.19 (0.082) | SIC38 instruments | 0.31 (0.052) |
| SIC29 oil | −0.26 (0.069) | SIC39 miscellaneous | 0.25 (0.066) |

economic distances correspond to distances between share vectors in $\mathfrak{R}^{22}$, we do not use the specification given by (2.5) for $k = 22$. Instead, for computational simplicity we use the representation in (2.6) for $k = \infty$. Specifically, we approximate the measure $\Phi$ in (2.4) as a linear combination of six third-order B-splines evenly spaced from frequency 0 to 60. [17] The coefficients are constrained to be nondecreasing and $\Phi$ specified to be constant after a frequency of 60 to guarantee that this estimate is a valid measure. We plug our $\Phi$ approximation into (2.4) and numerically approximate this integral over the same 0 to 60 range to produce an approximation for $C$. The resulting approximation for $C$ is a linear combination of functions of distances with a monotonicity constraint on their coefficients (see (3.7)). We estimate $C$ by regressing products of residuals from distinct agents upon these functions of distances, subject to the monotonicity constraint. We estimate $\sigma_i^2 + C(0)$ in (2.3) for each sector $i$ by the maximum of either the sample second moment of its LS residuals or the implied estimate of $C(0)$, in order to respect the nonnegativity constraint on $\sigma_i^2$. [18]

The coefficient matrix $A(D_t)$ is summarized by the $\alpha$ coefficients on each sector's own lagged output growth rate and the function $g$ that governs the impact of other sectors' output growth. Estimates of $\alpha_i$ presented in Table 1 are not particularly large in absolute magnitude, though the majority are significantly different from zero. Our estimate of $g$ is presented in Fig. 1. The solid line is our point estimate of $g$ plotted over the distances in our sample. The pluses are a bootstrap 95% confidence interval (200 draws)

[17] We did not use any formal data-driven methods such as cross-validation to choose the number of splines. In limited experimentation, we obtained similar results using five and seven splines for $g$ and $C$.

[18] In other words, our $\sigma_i^2 + C(0)$ estimate is $\max\{(1/(T-1))\sum_{t=1}^{T-1}\hat{e}_{i,t+1}^2, \hat{C}_T(0)\}$, see (3.14). This nonnegativity constraint on $\sigma_i^2$ binds for four sectors.

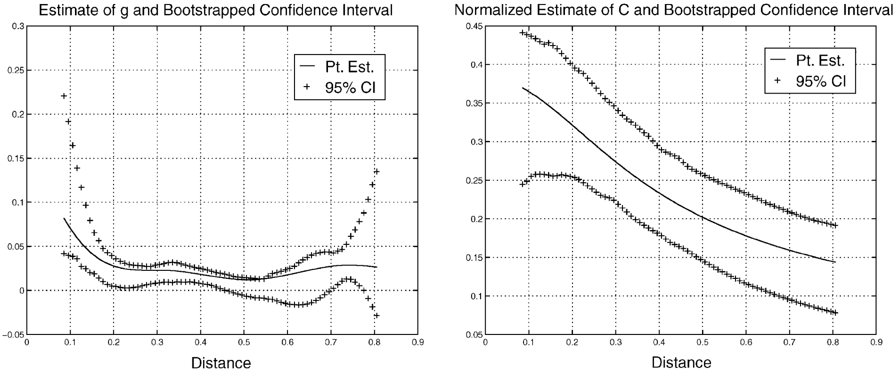


Fig. 1. The solid lines are point estimates of $g$ and normalized $C$, pluses indicate bootstrapped 95% confidence intervals (200 draws).

constructed as described in Section 3.3.[19] Our point estimates are small in magnitude but are statistically significant—the pointwise confidence intervals do not include zero—for distances less than 0.45 which is approximately the 80th percentile of nonzero distances. The point estimate is also decreasing for distances up to around 0.5 which is approximately the 90th percentile of the distribution. Thus there is some evidence of significant (but small) dynamic spatial correlation for most distances and that it declines with distance. However, $g$ is not precisely estimated, perhaps partially due to lack of variation in the preliminary distance measurements used here.

The conditional variances are described by idiosyncratic components $\sigma_i^2$ and the function $C$ that governs covariances. Sector specific variance estimates, $\hat{C}(0) + \hat{\sigma}_i^2$, presented in Table 2 differ across sectors with the largest being about 19 times the smallest, evidence of the importance of idiosyncratic shocks. Our estimate of $C$ divided by the average of the sector variance estimates is presented in Fig. 1. This normalized estimate of $C$ would be the sample spatial correlation if sector variances were identical, and in this case where they are unequal it still provides a rough measure of whether $C$ is large relative to sector variances. Pluses indicate a 95% confidence interval (200 draws) obtained using the bootstrap described in Section 3.3. The magnitude of the estimates of $C$ is quite large relative to the estimated sector-specific variances, even when only the lower endpoints of the confidence intervals are considered. There is strong evidence that the shocks in our VAR are spatially correlated as a function of this economic distance.

[19] Note that there is a bias in our residuals sampled in the bootstrap. They are not mean zero even though we use demeaned $Y_t$ because we have time-varying distances. This bias will be small where distances vary little, the only situation were one would expect the bootstrap to work well, and could be removed by, e.g., including constants in the regression.

Table 2

$\sigma_i^2 + C(0)$ Point estimates and bootstrap SE in ( )

| | | | |
|---|---|---|---|
| SIC20 food | 0.0005 (0.0001) | SIC30 plastics | 0.0016 (0.0002) |
| SIC21 tobacco | 0.0010 (0.0001) | SIC31 leather | 0.0010 (0.0001) |
| SIC22 textile | 0.0012 (0.0002) | SIC32 stone, glass | 0.0006 (0.0001) |
| SIC23 apparel | 0.0006 (0.0001) | SIC33 prim. metals | 0.0093 (0.0024) |
| SIC24 lumber | 0.0018 (0.0004) | SIC34 fab. metals | 0.0007 (0.0001) |
| SIC25 furniture | 0.0007 (0.0001) | SIC35 nonelec. mach. | 0.0005 (0.0001) |
| SIC26 paper | 0.0006 (0.0001) | SIC36 elec. mach. | 0.0009 (0.0001) |
| SIC27 printing | 0.0005 (0.0001) | SIC37 transportation | 0.0027 (0.0004) |
| SIC28 chemicals | 0.0005 (0.0001) | SIC38 instruments | 0.0005 (0.0001) |
| SIC29 oil | 0.0006 (0.0001) | SIC39 miscellaneous | 0.0007 (0.0001) |

## 5. Monte Carlo exercise

This section presents the results from a small Monte Carlo (MC) exercise that examines the finite sample properties of our estimators. The first experiment we consider uses the simplest possible data generating process (DGP) where locations and hence distances are fixed over time. We characterize the performance of our point estimators and bootstrap inferences with this DGP. One of the main issues in our empirical application is that distances vary over time but slowly relative to output growth rates. Effectively, we have too few distance observations to estimate a model of their evolution for use in our bootstrap inference procedures. This motivates our second set of experiments that assess the quality of inference using our bootstrap that conditions on a single sample path of distances when the DGP has moving distances.

Our DGP specifications are as follows. In each case, we have 20 agents with locations in $\mathfrak{R}^3$. Our fixed location DGP is calibrated to match the sector-level input distances by choosing the configuration that corresponds to the projection of the sector locations from our empirical example into $\mathfrak{R}^3$.[20] Our moving location DGPs take this projection as mean locations that are perturbed by expectation zero AR(1) shocks (with varying serial correlation) to each coordinate. The specification of $A$ and $\Sigma$ in (2.2) and (2.3) is common to both sets of experiments. We take the conditional mean components $\alpha_i$ and $g_i$ to be common across all agents with $\alpha_i = 0.4$ and $g_i$ to be linear from an intercept of 0.06 to zero at a distance of 0.35, and zero for all larger distances. The innovations are Gaussian with the conditional variance parameters $\sigma_i^2 = 1$ for all $i$ and $C(d) = \exp(-2d)$.

[20] See Mardia et al. (1979) for a description of this projection procedure, called multidimensional scaling, and an algorithm for its implementation. The deciles (0.1 to 0.9) of inter-agent distances in this configuration are: 0.08, 0.13, 0.17, 0.22, 0.26, 0.32, 0.37, 0.42, 0.49.

We examine the performance of approximations for both $g$ and $C$ that use third-order B-splines as in the empirical example. We approximate $g$ as a linear combination of six splines scaled to be evenly spaced over the support of the distance distribution. Our approximation for $C$ corresponds to approximating the spectral measure $\Phi$ using eight splines, evenly spaced from frequency 0 to 60, with $\Phi$ specified to be constant for higher frequencies. [21] The $\{\alpha_i\}$ parameters and the function $g$ are estimated by least squares and then the resulting residuals are used to estimate the $\{\sigma_i^2\}$ parameters and the function $C$ via least squares with a coefficient constraint. Bootstrap confidence intervals are then calculated using 100 bootstrap draws.

Table 3 summarizes 500 MC simulations from our fixed location DGP for sample lengths of $T=200$ and 500. Statistics are reported for $\alpha$, $\sigma^2$, and $g$ and $C$ evaluated at distances 0.12, 0.27, and 0.49, approximately the 10th, 50th, and 90th percentiles. The columns labeled percentiles report the 10th, 30th, 50th, 70th, and 90th percentiles of the various parameters across MC draws, pooling all agents' estimates for $\alpha_i$ and $\sigma_i^2$. The columns labeled Bias and Root MSE contain the resulting bias and root mean squared error across MC replications. Finally, the last column labeled 96% CI Cover Pr reports the fraction of MC draws where our bootstrap 96% confidence interval contains the true parameter value. [22]

Table 4 presents a comparison of coverage probabilities for our estimators of $\alpha_i$ and $\sigma_i^2$ and $g$ and $C$ for moving distance DGPs. The entries in each row are the proportion of 500 MC samples with $T=200$ where the bootstrap 96% confidence intervals contain the true value ($\alpha$ and $\sigma^2$ estimates pool across all agents). The differences across rows reflect different specifications for the distances: DGPs with location autocorrelation of zero, 0.2, and 0.9 are considered (mean locations are the same as in the fixed location DGP). [23]

Overall, our estimators perform well in these experiments. For the fixed distance DGP, typical biases are a few percent of the true parameter values and the 10th–90th percentile spread is small relative to the true parameter value. The poorest performance in relative bias (for nonzero parameters) is for $g(0.27)$. While 0.27 is near the median, it is also near the kink at 0.35 where we expect $g$ to be hard to estimate. Across parameters in Table 3, the 96% confidence interval coverage probabilities are below but generally close to 96%. The improvement in precision with $T=500$ over $T=200$ is

[21] These choices of numbers of splines are admittedly ad hoc. Six splines for $g$ was the only number examined as it appeared to fit well. Five, eight, ten, and 15 splines were tried for $C$. The fit appeared much better with more than five splines and eight was chosen as it was fastest to compute.

[22] We use 96% rather than 95% confidence intervals due to the need to limit the number of bootstrap draws (100) for each MC replication for computational tractability.

[23] The innovation variance for all these processes is 0.0001. Thus the processes with serial correlation have larger second moments.

Table 3
Results for 500 Monte Carlo simulations: fixed location DGP

| Param. | Truth | Percentiles of Monte Carlo distribution | | | | | Bias | Root | 96% Cl |
|---|---|---|---|---|---|---|---|---|---|
| | | 10 | 30 | 50 | 70 | 90 | | MSE | Cover Pr |
| $T=200$ | | | | | | | | | |
| $\alpha$ | 0.4 | 0.3179 | 0.3600 | 0.3894 | 0.4167 | 0.4541 | 0.0130 | 0.0557 | 0.940 |
| $\sigma^2$ | 1 | 0.7007 | 0.8616 | 0.9720 | 1.0813 | 1.2369 | 0.0287 | 0.6094 | 0.933 |
| $g$ (0.12) | 0.040 | 0.0313 | 0.0369 | 0.0400 | 0.0428 | 0.0466 | 0.0004 | 0.00027 | 0.917 |
| $g$ (0.27) | 0.014 | 0.0057 | 0.0098 | 0.0127 | 0.0150 | 0.0187 | 0.0018 | 0.00025 | 0.929 |
| $g$ (0.49) | 0 | −0.0092 | −0.0046 | −0.0016 | 0.0016 | 0.0054 | 0.0015 | 0.00027 | 0.925 |
| $C$(0.12) | 0.78 | 0.6677 | 0.7221 | 0.7644 | 0.8058 | 0.8634 | 0.0188 | 0.0037 | 0.908 |
| $C$(0.27) | 0.59 | 0.4810 | 0.5362 | 0.5720 | 0.6015 | 0.6581 | 0.0142 | 0.0032 | 0.913 |
| $C$(0.49) | 0.38 | 0.3064 | 0.3454 | 0.3730 | 0.3988 | 0.4433 | 0.0039 | 0.0024 | 0.931 |
| $T=500$ | | | | | | | | | |
| $\alpha$ | 0.4 | 0.3521 | 0.3786 | 0.3964 | 0.4133 | 0.4371 | 0.0046 | 0.0343 | 0.934 |
| $\sigma^2$ | 1 | 0.8320 | 0.9344 | 1.0022 | 1.0724 | 1.1771 | 0.0031 | 0.6180 | 0.932 |
| $g$ (0.12) | 0.040 | 0.0356 | 0.0385 | 0.0400 | 0.0417 | 0.0443 | 0.0008 | 0.00016 | 0.918 |
| $g$ (0.27) | 0.014 | 0.0091 | 0.0113 | 0.0130 | 0.0145 | 0.0172 | 0.0013 | 0.00016 | 0.944 |
| $g$ (0.49) | 0 | −0.0060 | −0.0034 | −0.0016 | 0.0002 | 0.0027 | 0.0016 | 0.00016 | 0.916 |
| $C$(0.12) | 0.78 | 0.7085 | 0.7466 | 0.7744 | 0.8008 | 0.8407 | 0.0088 | 0.0024 | 0.918 |
| $C$(0.27) | 0.59 | 0.5214 | 0.5499 | 0.5770 | 0.5993 | 0.6331 | 0.0087 | 0.0020 | 0.926 |
| $C$(0.49) | 0.38 | 0.3337 | 0.3580 | 0.3772 | 0.3963 | 0.4215 | 0.0005 | 0.0015 | 0.936 |

Table 4
96% Cl coverage probabilities with moving locations DGPs 500 Monte Carlo simulations, $T = 200$

| DGP location autocorrelation | Parameters | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | $\sigma^2$ | $g$ (0.12) | $g$ (0.27) | $g$ (0.49) | $C$(0.12) | $C$(0.27) | $C$(0.49) |
| 0 | 0.935 | 0.919 | 0.926 | 0.902 | 0.940 | 0.916 | 0.916 | 0.878 |
| 0.2 | 0.935 | 0.933 | 0.926 | 0.954 | 0.940 | 0.902 | 0.904 | 0.912 |
| 0.9 | 0.936 | 0.932 | 0.938 | 0.946 | 0.940 | 0.922 | 0.924 | 0.920 |

noticeable, however not overwhelming. The coverage probabilities in Table 4 are below, but close to 96% for all parameters and across the moving distance DGPs as well. Conducting bootstrap inference with a fixed distance sample path when the true DGP does have moving distances appears to give fairly reliable interval estimates for at least these types of moving distances.

## 6. Conclusion and extensions

This paper presented a tractable model for large-dimension vector time series that utilizes a spatial model of dependence across series, made operational by data on economic distances between agents. Spatial and temporal dependence between agents' variables is modeled using flexible functions of their economic distances. We present sufficient conditions for our model to generate stationary, $\beta$-mixing series with finite higher-order moments. Sieve estimators of functions of economic distance are constructed to preserve shape restrictions, in particular those required for $C$ to be a valid covariance function. We provide rates of convergence for the sieve estimators $\hat{g}_i$ and $\hat{C}$, $\sqrt{T}$ asymptotic normality for the parametric portions of the model, and a bootstrap method for inference. These estimators provide a good method of characterizing dynamics of agents' series as demonstrated in our example application where comovement in sector-specific output growth is seen to be systematically related to an economic distance measuring the similarity of technology between sectors. Short-term forecasts holding distances fixed are also feasible using the paper's results. Finally, our Monte Carlo evidence suggests that our estimators work reasonably well for the set of experiments we considered.

There are many ways to extend our current results. Perhaps the most interesting and important extensions will be to develop specific, estimable models of distance evolution. Clearly this will be necessary for long-term forecasts. Models of distance evolution are likely to be tied to a particular application so that the exact nature of economic configurations can be well understood, particularly if distances are endogenous. Developing a model of distance evolution for the intersector distances used here will be problematic. The time series of benchmark input–output data is likely not to have sufficient variation to estimate a model, even if one were suggested by economic theory. However, there is considerable promise for developing a model of endogenous economic distances in applications with firm-level data. For example, a model for the evolution of a measure of economic distance based on cross-price demand elasticities should be estimable in connection with a model describing price response functions for competing firms.

There are several other extensions that we feel will be worth pursuing. A more thorough Monte Carlo evaluation would help address questions like how different order sieve approximations perform and how our model's forecasting performance compares to that of alternatives like dynamic factor models. An extension of great empirical relevance is testing dimension-reducing restrictions on the $g_i$ functions. Another useful extension would be to allow agents' series to be cointegrated. Finally, the paper's basic model of dependence across series is well suited to extensions where cross-sectional dependence is allowed for in panel data when taking limits as both the cross-section and time-series dimensions increase.

## Acknowledgements

The authors would like to thank Ken West and two anonymous referees for valuable suggestions. They are also grateful to T. Andersen, I. Domowitz, J. Dubé, W. Dupor, G. Elliott, J. Hamilton, B. Hansen, C. Hsiao, E. Leeper, F. Molinari, M. Ogaki, P. Pedroni, I. Perrigne, S. Stern, R. Vigfusson, Q. Vuong, J. Wooldridge and seminar participants at UCSD, UCR, USC, Boston College, Cornell, Penn, Virginia and Wisconsin for helpful comments. F. Molinari and R. Vigfusson provided excellent research assistance. We of course remain responsible for any mistakes. Chen acknowledges support from the Social Science Division 1998/99 Dean's Grant at the University of Chicago. Conley acknowledges support from National Science Foundation grant no. SES-9905720.

## Appendix [24]

*Proofs of Propositions 2.1 and 2.2.* To establish *β-mixing* with exponential decay rates for $\{Z_t' = (Y_t, \ln(D_t))\}$, we need to show (e.g. Doukhan, 1994):

[24] The full proofs are available upon request from the authors.

(i) $\phi$-irreducible and aperiodic; (ii) Lyapunov criteria or "drift condition". Either AI.1–AI.5 or AII.1–AII.4 are sufficient to imply (i) and (ii).

*Proof of Theorem 3.1.* We establish the results by applying the Theorems 1–3 of Chen and Shen (1998, CS), here we outline steps needed to verify their conditions.

(1) For convergence rate of $\hat{g}_{i,T} - g_{i,o}$, we apply CS Theorem 1 by letting their $\theta = (\alpha_i, g_i)$, true $\theta_o = (\alpha_{i,o}, g_{i,o})$, their sieve estimator $\hat{\theta}_T = (\hat{\alpha}_{i,T}, \hat{g}_{i,T})$ solves $\sup_{\theta \in (-1,1)\times \mathscr{G}_{i,T}} (1/T)\sum_{t=1}^{T} l(\theta, Y_t, D_t)$ where $l(\theta, Y_t, D_t) = -\frac{1}{2}(X_{i,t+1} - [\alpha_i X_{i,t} + \sum_{j\neq i} g_i(D_t(i,j))X_{j,t}])^2$, and their norm $||\theta - \theta_o||^2 \equiv E[X_{i,t}(\alpha_i - \alpha_{i,o}) + \sum_{j\neq i} X_{j,t}[g_i(D_t(i,j)) - g_{i,o}(D_t(i,j))]]^2$. Their condition A.1 of *β-mixing* is satisfied by our Propositions 2.1 or 2.2. Their condition A.2 of local variance behavior, and their condition A.4 of global Lipschitz continuous property of $\{l(\theta, Y_t, D_t) - l(\theta_o, Y_t, D_t)\colon \theta - \theta_o\}$ are satisfied by the same proof as CS for their Proposition 1 on nonparametric regression. Now CS Theorem 1 implies $\sqrt{\int [\hat{g}_{i,T}(x) - g_{i,o}(x)]^2\,dx} = O_P(\max\{||\Pi_T g_{i,o} - g_{i,o}||, \delta_T\})$ where $\delta_T = const. \times \sqrt{2^{n_i(T)} \times T^{-1}}$ is obtained by solving CS condition A.3 on metric entropy, and $||\Pi_T g_{i,o} - g_{i,o}|| = const. \times 2^{-n_i(T)\times m_i}$ is the deterministic approximation error rates. The optimal convergence rate is obtained by setting $||\Pi_T g_{i,o} - g_{i,o}|| = \delta_T$, hence $\sqrt{\int [\hat{g}_{i,T}(x) - g_{i,o}(x)]^2\,dx} = O_P(T^{-m_i/(2m_i+1)})$.

(2) To establish $\sqrt{T}(\hat{\alpha}_{i,T} - \alpha_{i,o}) \Rightarrow \mathscr{N}(0, \sigma^2_{\alpha_i})$, we apply CS Theorem 2 with their smooth functional $f(\theta) = \alpha_i$. In particular, we use the same proof as for CS Proposition 2 on partial linear regressions to verify conditions B.1–B.5 for their Theorem 2. Hence we obtain: $\sqrt{T}(\hat{\alpha}_{i,T} - \alpha_{i,o}) \Rightarrow \mathscr{N}(0, \sigma^2_{\alpha_i})$, where

$$\sigma^2_{\alpha_i} = Var(l'_{\theta_o}[v^*, Y_t, D_t]) + 2\sum_{k=1}^{\infty} Cov(l'_{\theta_o}[v^*, Y_0, D_0], l'_{\theta_o}[v^*, Y_k, D_k]),$$

and $v^* \in \mathfrak{R} \times \bar{\mathscr{G}}_i$ is the Riesz representor for $f(\theta) = \alpha_i$ with

$||v^*||^2 = \{\inf_{w_i \neq 0, w_i \in \bar{\mathscr{G}}_i} E[X_{i,t} - \sum_{j\neq i} w_i(D_t(i,j))X_{j,t}]^2\}^{-1}$. Clearly $||v^*|| < \infty$ under our assumption that $X_{i,t} \neq E[X_{i,t} \mid D_t(i,j), X_{j,t}, j\neq i]$ with positive probability. Simple computation yields:

$\sigma^2_{\alpha_i} = [\sigma^2_{i,o} + C_o(0)] \times \{\inf_{w_i \neq 0, w_i \in \bar{\mathscr{G}}_i} E[X_{i,t} - \sum_{j\neq i} w_i(D_t(i,j))X_{j,t}]^2\}^{-1}$.

(3) To establish $\sqrt{T}[(\hat{\sigma}^2_{i,T} + \hat{C}_T(0)) - (\sigma^2_{i,o} + C_o(0))] \Rightarrow \mathscr{N}(0, \delta^2_{e_i})$, we apply CS Theorem 3 (p. 307). In particular, we have with $\varepsilon_T = O(T^{-m_i/(2m_i+1)})$:

$\sup (1/T)\sum_{t=1}^{T} [l(\theta, Y_t, D_t) - l(\theta_o, Y_t, D_t) - E(l(\theta, Y_t, D_t) - l(\theta_o, Y_t, D_t))] = O_p(\varepsilon^2_T)$, with sup taken over $\theta \in \mathfrak{R} \times \mathscr{G}_{i,T}, ||\theta - \theta_o|| \leqslant \varepsilon_T$. This and $m_i > 1/2$ imply that: $\frac{1}{T}\sum_{t=1}^{T} [l(\hat{\theta}_T, Y_t, D_t) - l(\theta_o, Y_t, D_t)] = O_p(\varepsilon^2_T) = o_p(T^{-1/2})$. Hence, $-\sqrt{T}[(\hat{\sigma}^2_{i,T} + \hat{C}_T(0)) - (\sigma^2_{i,o} + C_o(0))] = o_p(1) + (2/\sqrt{T})\sum_{t=1}^{T} [l(\theta_o, Y_t, D_t) - E[l(\theta_o, Y_t, D_t)]]$. Now by the standard CLT for *β-mixing* process, we obtain

$\frac{2}{\sqrt{T}}\sum_{t=1}^{T}[l(\theta_o,Y_t,D_t) - E[l(\theta_o,Y_t,D_t)]] \Rightarrow \mathcal{N}(0,V_{e_i}),\quad V_{e_i}=E[e_{i,t+1}^2 - E(e_{i,t+1}^2)]^2$.

(4) For convergence rate of $\hat{C}_T(\cdot)-C_o(\cdot)$: Recall $\hat{C}_T(\cdot)$ solves:

$$\min_{C\in\mathscr{C}_T}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=2,j>i}^{N}(\hat{e}_{i,t+1}\hat{e}_{j,t+1}-C(D_t(i,j)))^2,$$

where $\hat{e}_{i,t+1}=X_{i,t+1}-[\hat{\alpha}_{i,T}X_{i,t}+\sum_{j\neq i}\hat{g}_{i,T}(D_t(i,j))X_{j,t}]$.

First under $m_g>1/2$ and by the same proofs as those for CS Propositions 1 and 2, the conclusions of (1) and (2) are valid uniformly over $i=1,\ldots,N$, and $\sqrt{T}(\hat{\alpha}_{N,T}-\alpha_{N,o},\ldots,\hat{\alpha}_{N,T}-\alpha_{N,o})$ is asymptotically jointly normal. These and

$$\hat{e}_{i,t+1}\hat{e}_{j,t+1}-C(D_t(i,j)) = (\hat{e}_{i,t+1}\hat{e}_{j,t+1}-e_{i,t+1}e_{j,t+1}) + (e_{i,t+1}e_{j,t+1}-C(D_t(i,j))),$$

$$(\hat{e}_{i,t+1}\hat{e}_{j,t+1}-e_{i,t+1}e_{j,t+1}) = \hat{e}_{i,t+1}(\hat{e}_{j,t+1}-e_{j,t+1})+e_{j,t+1}(\hat{e}_{i,t+1}-e_{i,t+1}),$$

imply that $\hat{C}_T(\cdot)$ is an approximate sieve solution to:

$$\min_{C\in\mathscr{C}_T}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=2,j>i}^{N}(e_{i,t+1}e_{j,t+1}-C(D_t(i,j)))^2+O_p(T^{-2m_g/(2m_g+1)}).$$

We apply CS Theorem 1 again, by letting their $\theta=C$, their $l(\theta,Y_t,D_t)$ to be $-\sum_{i=1}^{N}\sum_{j=2,j>i}^{N}(e_{i,t+1}e_{j,t+1}-C(D_t(i,j)))^2$, and their norm $||\theta-\theta_o||^2$ to be $E[C(D_t(i,j))-C_o(D_t(i,j))]^2$, hence $\sqrt{\int_0^{\bar{a}}[\hat{C}_T(z)-C_o(z)]^2\,\mathrm{d}z}=O_P(T^{-m_c/(2m_c+1)})$.

## References


Anastassiou, G.A., Yu, X.M., 1992. Monotone and probabilistic wavelet approximation. Stochastic Analysis and Applications 10, 251–264.

Carroll, R.J, Chen, R., George, E.I., Li, T.H., Newton, H.J., Schmiediche, H., Wang, N., 1997. Ozone exposure and population density in Harris County Texas. Journal of the American Statistical Association 92, 392–415.

Chen, N., Roll, R., Ross, S.A., 1986. Economic forces and the stock market. Journal of Business 59 (3), 383–403.

Chen, X., Shen, X., 1998. Sieve Extremum estimates for weakly dependent data. Econometrica 66, 289–314.

Chen, X., Hansen, L.P., Scheinkman, J., 1997. Shape-preserving estimation of diffusions. Working Paper, University of Chicago.

Conley, T.G., Ligon, E., 2000. Economic distance and cross-country spillovers. Working Paper, UC Berkeley.

Conley, T.G., Dupor, W.D., 1999. A spatial analysis of sectoral complementarity. Working Paper, Northwestern University.

Conley, T.G., Dupor, W.D., 2000. Sectoral business cycle dynamics using spatial methods. Working Paper, Northwestern University.

Connor, G., Koracjczk, R.A., 1993. A test for the number of factors in an approximate factor model. Journal of Finance 48 (4), 1263–1291.
Cressie, N., 1997. Comment on Carroll et al. Journal of the American Statistical Association 92, 411–412.
Chui, C.K., 1992. An Introduction to Wavelets. Academic Press, San Diego.
Doukhan, P., 1994. Mixing: Properties and Examples. Springer, New York.
Dupor, W.D., 1997. Aggregate fluctuations and production complementarities. Working Paper, Wharton School.
Forni, M., Reichlin, L., 1998. Let's get real: a factor analytic approach to disaggregated business cycle dynamics. Review of Economic Studies 65, 453–473.
Gallant, A.R., Nychka, D., 1987. Semi-non-parametric Maximum Likelihood Estimation. Econometrica 55, 363–390.
Grenander, U., 1981. Abstract Inference. Wiley Series, New York.
Holz-Eakin, D., Newey, W., Rosen, H.S., 1988. Estimating vector autoregressions with panel data. Econometrica 56 (6), 1371–1395.
Horvath, M., 1998. Cyclicality and sectoral linkages: aggregate fluctuations from independent sectoral shocks. Review of Economic Dynamics 1, 781–808.
Horvath, M.T.K., Verbrugge, R., 1997. Shocks and sectoral interactions: in empirical investigation. Working Paper, Stanford University.
Huang, H., Cressie, N., 1997. Spatio-temporal prediction of snow water equivalent using the Kalman filter. Working Paper, Iowa State University.
Long, J., Plosser, C., 1983. Real business cycles. Journal of Political Economy 91, 39–69.
Mardia, K., Kent, J., Bibby, J., 1979. Multivariate Analysis. Academic Press, London.
Meiring, W., Guttorp, P., Sampson, P.D., 1998. Space–time estimation of grid-cell hourly ozone levels for assessment of a deterministic model. Environmental and Ecological Statistics 5(3), 197–222.
Newey, W., Powell, J., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models Econometrica 67, 565–603.
Pulvino, T., 1998. Do asset fire sales exist: an empirical investigation of commercial aircraft transactions. Journal of Finance 53(3), 939–978.
Quah, D., Sargent, T.J., 1993. A dynamic index model for large cross sections. In: Stock, J.H., Watson, W.H. (Eds.), Business Cycles, Indicators, and Forecasting. University of Chicago Press, Chicago, pp. 285–306 (Chapter 7).
Slade, M.E., 1986. Exogeneity tests of market boundaries applied to petroleum products. Journal of Industrial Economics 34, 291–303.
Stock, J.H., Watson, M.W., 1997. Diffusion indices. Working Paper, Harvard University.
Sumner, D.A., 1981. Measurement of monopoly behavior: an application to the cigarette industry. Journal of Political Economy 89 (5), 1010–1019.
Taylor, A.M., 2000. A century of purchasing power parity. Working Paper, UC Davis.
Yaglom, A.M., 1987. Correlation Theory for Stationary and Related Random Functions, Vols. I and II. Springer, New York.