Forecasting house price growth rates with factor models and spatio-temporal clustering[☆]

Raffaele Mattera [a],[*], Philip Hans Franses [b]

[a] *Department of Social and Economic Sciences, Sapienza University of Rome, Rome, Italy*
[b] *Econometric Institute, Erasmus School of Economics, Rotterdam, The Netherlands*

ARTICLE INFO

ABSTRACT

This paper proposes to use factor models with cluster structure to forecast growth rates of house prices in the US. We assume the presence of global and cluster-specific factors and that the clustering structure is unknown. We adopt a computational procedure that automatically estimates the number of global factors, the clustering structure and the number of clustered factors. The procedure enhances spatial clustering so that the nature of clustered factors reflects the similarity of the time series in the time domain and their spatial proximity. Considering house prices in 1975–2023, we highlight the existence of four main clusters in the US. Moreover, we show that forecasting approaches incorporating global and cluster-specific factors provide more accurate forecasts than models using only global factors and models without factors.

1. Introduction

Fluctuations in house prices provide important signals for consumption, inflation and financial stability, making the housing sector a critical leading indicator for the economy. Given the strong interlinkages between the regional housing market and the national business cycle, house price forecasting is paramount for policymakers. Previous studies have employed various methodologies to predict house prices at the regional level, but those based on factor models are among the most common (*e.g.* see Das, Gupta, & Kabundi, 2011; Emiris, 2016; Moench & Ng, 2011). Factor models assume that house prices can be modelled as a sum of two components: a global component driven by an unobserved factor common to all the regions and an idiosyncratic component specific to each region.

We conjecture that it is restrictive to assume that all the regions are affected by the same latent factors when dealing with regional house prices. In the US, house prices are notably affected by spillovers from neighbouring states[1] (Brady, 2014; Kuethe & Pede, 2011) and are best described by clusters (Kim & Rous, 2012). The distinction into clusters also explains why previous studies found differences in the forecastability of housing price growth rates across the US states. For example, interior states appear easier to forecast than coastal states (Rapach & Strauss, 2009). Therefore, it seems useful from a forecasting perspective to use models based on global (*e.g.* country-level) and regional (or cluster-level) latent factors.

Regional factor models, which have become increasingly popular in empirical studies conducted by central banks (*e.g.* Aastveit, Bjørnland, & Thorsrud, 2016; Beck, Hubrich, & Marcellino, 2009; Breitung & Eickmeier, 2014)

[☆] The numerical results presented in this manuscript were reproduced by the Editor-in-Chief (up to minor discrepancies) on the 15th of September 2024.

[*] Corresponding author.
E-mail addresses: raffaele.mattera@uniroma1.it (R. Mattera), franses@ese.eur.nl (P.H. Franses).

---

[1] Similar findings are common also for other housing markets outside the US, such as Germany (Otto & Schmid, 2018), Italy (Cipollini & Parla, 2020), the Netherlands (Van Dijk, Franses, Paap, & Van Dijk, 2011) and the UK (Blatt, Chaudhuri, & Manner, 2023) among others.

are particular types of factor models with cluster structure, where the clustering structure is known and depends on spatial information. These methods allow for the estimation of both global and cluster-specific factors. For example, using a two-step procedure based on the principal components method, Beck et al. (2009) proposed to model the regional inflation dynamics in the Euro Area countries by distinguishing between national and regional (*i.e.* NUTS 2) factors. Eickmeier, Gambacorta, and Hofmann (2014), considering financial and macroeconomic variables from twenty-four economies, adopted a similar two-step procedure to study the determinants of global liquidity using global factors and two cluster-specific factors for developed and not developed countries. Aastveit et al. (2016) extended the global factor model of Mumtaz and Surico (2009), used for studying the international transmission of structural shocks, to include regional factors for thirty-two countries worldwide. Moench and Ng (2011) adopted a Factor-Augmented VAR (FAVAR) model, distinguishing between global and regional factors, for studying the effect of housing shocks on consumption in the US. Other papers have proposed alternative approaches for including the spatial dimension in factor models (*e.g.* see Aquaro, Bailey, & Pesaran, 2021; Bailey, Holly, & Pesaran, 2016; Ciccarelli & Elhorst, 2018; Yang, 2021), but these generally consider global factors only or assume known spatial cluster structures. Lopes, Salazar, and Gamerman (2008) and Ippoliti, Valentini, and Gamerman (2012) proposed using spatially structured factor loadings to identify groups of spatial sites with common factors. These methods have been mainly applied for environmental modelling (*e.g.* see Gamerman, Ippoliti, & Valentini, 2022), but also for housing market modelling (Valentini, Ippoliti, & Fontanella, 2013). However, these models have not been proposed with the aim of clustering and do not fully account for the role played by the clusters in the forecasting task.

In this paper, we study cluster-specific factor models' ability to forecast house price growth rates in the US under the assumption of unknown spatio-temporal clustering. We estimate the cluster structure of the US states directly from the data and assume that all the states in the same cluster share the cluster-specific latent factors. Moreover, we assume all the US states share the same global factor. We notice that the cluster structure depends on spatial information and the similarity of the temporal pattern in housing prices.

In doing so, we follow a recent strand of the literature proposing approaches for estimating global and clustered factors under the assumption of unknown cluster membership. Ando and Bai (2017) develop a model for asset pricing that includes observable factors and global and cluster-specific latent factors. The (Ando & Bai, 2017) procedure requires, as in *k*-means clustering, the ex-ante specification of the correct number of clusters and the knowledge of the number of global and cluster-specific factors within each cluster. This can be inconvenient. An additional model-searching algorithm is therefore used to tune these parameters. Alonso, Galeano, and Peña (2020) consider an alternative robust procedure for estimating global and cluster-specific factors. The authors overcome

the problem of deciding on the number of clusters ex-ante by adopting a hierarchical procedure. Taking inspiration from the time series clustering literature, the authors adopt a cross-correlation distance (Alonso & Peña, 2019) in determining the clustering structure of the units. However, neither study includes a spatial dimension while updating the clusters and determining the nature of the cluster-specific factors, which is relevant in the case of regional house prices.

In this paper, we, therefore, adopt a computational procedure for estimating global and cluster-specific factors that account for spatial clustering. We then estimate factor-augmented models, using the estimated factors, and predict future house price growth rates.

The contribution of our paper is threefold. First, we show that cluster-specific factors can achieve more accurate forecasts of house price growth rates. Forecasting models based on global factors only provide less accurate forecasts out-of-sample. This evidence is confirmed by considering the results of predictive accuracy tests. Second, we adopt an unsupervised learning approach to build the clusters. We indeed consider an iterative procedure similar to Alonso et al. (2020), based on hierarchical clustering, that estimates the number of clusters and the number of factors within the algorithm. Different from (Alonso et al., 2020), who propose a four-step procedure, we adopt an algorithm that iteratively solves the problem until convergence. Third, we introduce the spatial dimension in the problem to enhance spatial clustering in estimating the cluster-specific factors. The amount of spatial penalty introduced in the algorithm is also chosen in a data-driven manner, but it can also be set manually by the user.

The rest of the paper is structured as follows. Section 2 presents the data about house prices in the US states and the forecasting model based on a Panel VAR with factors. Section 3 discusses the computational procedure for estimating the clusters and global and cluster-level factors. In contrast, Section 4 shows the outcomes of simulation experiments used to support the validity of the proposed computational procedure. Section 5 shows the main results, distinguishing between in-sample goodness-of-fit and out-of-sample forecasting accuracy and some robustness checks. Section 6 concludes with final remarks and future research directions.

2. Forecasting house prices growth rates in the US states

We now present the approach adopted for forecasting house price growth in the US states. Section 2.1 briefly describes the data, while Section 2.2 discusses the forecasting methodology.

2.1. Data

We consider the average house prices in the 48 continental US states for the period 1975–2023. Alaska and Hawaii are not included in the sample. The data source is the Federal House Finance Agency website.[2] The time

---

[2] Data can be retrieved at the following link https://www.fhfa.gov/.
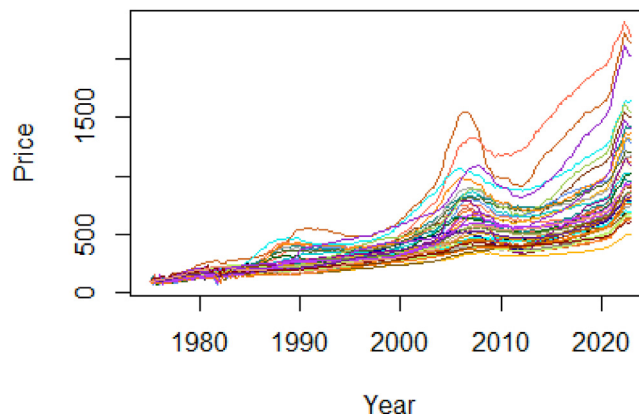
House prices in the U.S. (1975=100)

Fig. 1. Temporal evolution of regional house prices in the US (1975 = 100).

series are sampled quarterly, so we have $T = 193$ temporal observations for $N = 49$ cross-sectional units. The temporal evolution of the regional house prices in the US is shown in Fig. 1.

All the price indices show a clear positive trend, with a large downturn in correspondence with the 2007–2008 great financial crisis. Due to the nonstationarity of the prices time series, we consider the growth rates. The data we consider are georeferenced.

2.2. Forecasting model

In this paper, we adopt a suitably augmented Panel VAR model, which includes global and cluster-specific factors. A Panel VAR is a popular tool in macroeconomic forecasting (*e.g.* see Bai, Carriero, Clark, & Marcellino, 2022; Dées & Güntner, 2017; Feldkircher, Huber, & Pfarrhofer, 2020). This allows for the introduction of US state-specific fixed effects to capture the time-invariant state-specific factors not considered by the global and cluster-specific factors.

Let $\mathbf{y}_{i,t}$ be a P-dimensional vector including a set of P endogenous variables for the ith ($i = 1, \ldots, N$) US state at time t ($t = 1, \ldots, T$). The Panel VAR(L) can be written as follows

$$\mathbf{y}_{it} = \boldsymbol{\mu}_i + \sum_{l=1}^{L} \mathbf{A}_l \mathbf{y}_{it-l} + \boldsymbol{\varepsilon}_{it}, \qquad (1)$$

where $\boldsymbol{\mu}_i$ is an P-dimensional vector of time-invariant fixed effects, specific to each ith US State, $\mathbf{A}_l$ is the $P \times P$ matrix of coefficients for the lth lag, $\mathbf{y}_{it-l}$ is the P-dimensional vector associated with the lth lag of $\mathbf{y}_{it}$ and there are $\varepsilon_{i,t}$ i.i.d. disturbances. We assume that all roots of $\mathbf{A}$ are outside the unit circle, and hence, we deal with a stationary Panel VAR.

Let us denote x_{it} the house price growth rates in the ith US state at time t. We define $\mathbf{f}_{0t}$ the R-dimensional vector ($r = 1, \ldots, R$) with global factors at time t affecting all the series x_{it}, and $\mathbf{f}_{kt}$ the R_k dimensional vector ($r_k = 1, \ldots, R_k$) of latent factors at time t, affecting only the

units belonging to the kth cluster. We can estimate a Panel VAR given $\mathbf{y}_{it} = [x_{it}, \mathbf{f}'_{0t}, \mathbf{f}'_{kt}]'$.

The parameters are estimated using the fixed-effects approach, and the fixed effects are then estimated following (Sigmund & Ferstl, 2021). To choose the optimal number of lags to include, we rely on the minimization of the BIC criterion (*e.g.* see Han, Phillips, & Sul, 2017).

In the case of a Panel VAR(1), the forecasts for $t+1$ are given by

$$\widehat{\mathbf{y}}_{it+1} = \widehat{\boldsymbol{\mu}}_i + \widehat{\mathbf{A}}\mathbf{y}_{it}, \qquad (2)$$

where $\mathbf{y}_{i,t} = [x_{it}, \mathbf{f}'_{0t}, \mathbf{f}'_{kt}]'$. A similar equation can be used for larger Panel VAR orders. We then compare, from both in-sample fitting and out-of-sample forecasting, the following models

1. (m1): a Panel VAR with both global and cluster-specific factors $\mathbf{y}_{it} = [x_{it}, \mathbf{f}'_{0t}, \mathbf{f}'_{kt}]'$;
2. (m2): constrained version of (m1) where only the global latent factors are considered and the clustered factors are excluded $\mathbf{y}_{it} = [x_{it}, \mathbf{f}'_{0t}]'$;
3. (m3): baseline model where only the lags of house price growth rates are considered for forecasting.

We are interested in evaluating if the information included in the cluster-specific factors helps out-of-sample forecasting the house price growth rates. Therefore, we consider the m1 model as the main benchmark. We find evidence favouring cluster-specific factors in house price forecasting if the model m1 provides better forecasts out-of-sample than the alternatives, which are all restricted versions.

The cluster structure and the factors need to be estimated before forecasting the house price growth rates with the Panel VAR (1). The following section discusses the computational procedure for estimating the factors and the clustering structure. In particular, our computational procedure accounts for the spatial dimension while clustering US states to enhance the cluster-specific factors capturing spatial information.

3. Estimating latent factors with spatio-temporal clustering

Let us define with $\mathbf{X}$ a $T \times N$ a matrix of time series $(x_{i,t}; i = 1, \ldots, N; t = 1, \ldots, T)$. We assume that $\mathbf{X}$ can be written as a linear combination of R $(r = 1, \ldots, R)$ global factors, R_k $(r_k = 1, \ldots, R_k)$ specific factors that differ for each of the K $(k = 1, \ldots, K)$ clusters and an idiosyncratic noise $\boldsymbol{\eta}$, that is,

$$\mathbf{X} = \mathbf{F}_0 \boldsymbol{\Lambda}_0' + \sum_{k=1}^{K} \mathbf{F}_k \boldsymbol{\Lambda}_k' + \boldsymbol{\eta}. \tag{3}$$

$\mathbf{F}_0$ is the matrix of $T \times R$ global factors, while $\boldsymbol{\Lambda}_0 = \left[\boldsymbol{\Lambda}_{01}' \mid \cdots \mid \boldsymbol{\Lambda}_{0k}' \mid \cdots \mid \boldsymbol{\Lambda}_{0K}' \right]'$ is the $N \times R$ matrix of global factor loadings, with $\boldsymbol{\Lambda}_{0k}$ the $N_k \times R$ global factor loadings for the time series belonging to the kth cluster. Moreover, $\mathbf{F}_k$ is $T \times R_k$ matrix of the R_k factors specific to the kth cluster and $\boldsymbol{\Lambda}_k = \left[\mathbf{0}_{k1}' \mid \cdots \mid \boldsymbol{\Lambda}_{kk}' \mid \cdots \mid \mathbf{0}_{kK}' \right]'$ is the $N \times R_k$ matrix of cluster-specific factor loadings, with $\boldsymbol{\Lambda}_{kk}$ be the loading matrix for the factors affecting only the N_k time series in the kth cluster. Without loss of generality, we assume that the columns of $\mathbf{X}$ are ordered by cluster. We call (3) a Clustered Factor Model (CFM).

Following (Alonso et al., 2020), we consider the usual model assumptions required for the existence and identification of global and cluster-specific factors.[3]

First, we assume a fixed number of clusters $K > 1$. Then, we assume orthonormal loading matrices, that is, $\boldsymbol{\Lambda}_0' \boldsymbol{\Lambda}_0 = \mathbf{I}_R$, where $\mathbf{I}_R$ is the identity matrix of order R, $\boldsymbol{\Lambda}_k' \boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_{k,k}' \boldsymbol{\Lambda}_{k,k} = \mathbf{I}_{R_k}$ for $k = 1, \ldots, K$ and $\boldsymbol{\Lambda}_0' \boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_{0,k}' \boldsymbol{\Lambda}_{k,k} = \mathbf{0}_{R \times R_k}$, and diagonal covariance matrix of the factors. Furthermore, cross-loadings are not allowed, and thus $\boldsymbol{\Lambda}_k' \boldsymbol{\Lambda}_{k'} = \mathbf{0}_{R_k \times R_{k'}}$, for $k \neq k'$. We assume the idiosyncratic terms to have some weak forms of dependency, as discussed in Bai (2003) and Wang (2010). In particular, we allow for temporal and cross-sectional dependence and heteroskedasticity.

3.1. Computational procedure

Under the assumption that R global and R_k clustered factors exist, two issues need to be addressed. First, we must estimate the unknown parameters: the unobservable factor structure and corresponding loadings. Second, we must estimate the number of K clusters and the corresponding factors. These two problems are inevitably related. Indeed, the number of clustered factors depends on the global factors as they represent the latent variables explaining fluctuations of the time series in a cluster, which is not explained by the global factors. However, consistent estimation of a global factor cannot assume the absence of any clustered (local) factor. Indeed, we could get biased identification of the global factors if we ignore clustered factors in the estimation process. Thus, the two problems have to be solved iteratively. In what follows,

---

[3] Following (Alonso et al., 2020), we refer to the assumptions A, B, C, D and E of Wang (2010), which extends (Bai, 2003) to the case of group-specific factors.

we propose an iterative procedure for overcoming this issue.

First of all, we estimate a suitable initial number of global factors R in the absence of a clustering structure. According to Bai and Ng (2002), this can be obtained by minimizing

$$\min_R \ln (V_R) + R \left(\frac{N + T}{NT} \right) \ln \left(\frac{NT}{N + T} \right) \tag{4}$$

with

$$V_R = \min_{\boldsymbol{\Lambda}_0, \mathbf{F}_0} (NT)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} - \boldsymbol{\lambda}_{i0}' \mathbf{f}_{0t} \right)^2, \tag{5}$$

where $\boldsymbol{\lambda}_{i0}$ is the R-dimensional vector of global factor loadings for the ith unit and $\mathbf{f}_{0t}$ the R-dimensional vector of the global factors at time t. By assuming $T > N$, the R global factors can be consistently estimated with the Principal Component Estimator (PCE),

$$\widehat{\mathbf{F}}_0 = \mathbf{X} \widehat{\boldsymbol{\Lambda}}_0 / N \tag{6}$$

given that $\widehat{\boldsymbol{\Lambda}}_0$ is estimated by $\sqrt{N}$ times the eigenvectors of $\mathbf{X}'\mathbf{X}$ associated with the R largest eigenvalues. Note that (6) equals the solution of the minimization problem (4) given the number of factors R.

To estimate the R_k clustered factors, we consider the residuals $\widehat{\mathbf{E}} = \mathbf{X} - \widehat{\mathbf{F}}_0 \widehat{\boldsymbol{\Lambda}}_0'$. The matrix $\widehat{\mathbf{E}}$ contains the information of $\mathbf{X}$, which is not explained by the global factor structure. Thus, we estimate the clustering structure based on $\widehat{\mathbf{E}}$ and identify the factors within each cluster.

In principle, any clustering approach can be used to determine the clustering structure of the residuals $\widehat{\mathbf{E}}$. In our case, we suppose that $\widehat{\mathbf{E}}$ is a matrix of georeferenced time series, and thus either spatial, temporal or spatial–temporal approaches can be used. Therefore, we obtain a *spatially clustered factor model* if only the spatial dimension is considered, a *clustered factor model* if only the temporal dimension is considered and a *spatio-temporal clustered factor model* if both are included. The following sub-section provides more details on clustering. For each kth cluster, the R_k number of factors needs to be estimated. The cluster-specific counterpart of (6) can be used for this aim, and the corresponding R_k factors can be estimated as

$$\widehat{\mathbf{F}}_k = \widehat{\mathbf{E}} \widehat{\boldsymbol{\Lambda}}_k / N_k \tag{7}$$

with $\widehat{\boldsymbol{\Lambda}}_k = \left[\mathbf{0}_{k1}' \mid \cdots \mid \widehat{\boldsymbol{\Lambda}}_{kk}' \mid \cdots \mid \mathbf{0}_{kK}' \right]'$ and, by defining $\widehat{\mathbf{E}}_k$ the columns of $\widehat{\mathbf{E}}$ belonging to the kth cluster, we compute $\widehat{\boldsymbol{\Lambda}}_{kk}$ as $\sqrt{N_k}$ times the eigenvectors of the matrix $\widehat{\mathbf{E}}_k' \widehat{\mathbf{E}}_k$ associated with the R_k largest eigenvalues.

As previously highlighted, stopping at this stage would not guarantee an optimal solution. Indeed, we need to estimate the number of global factors under the presence of suitably identified cluster-specific factors. In other words, we need to remove the information of the clustered factors from the whole data matrix $\mathbf{X}$. A suitable operation is

$$\widehat{\mathbf{U}} = \mathbf{X} - \sum_{k=1}^{K} \widehat{\mathbf{F}}_k \widehat{\boldsymbol{\Lambda}}_k'. \tag{8}$$

The number of factors can be estimated by a modified version of (4), that is,

$$V_R = \min_{\Lambda_0, \mathbf{F}_0} (NT)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it} - \lambda'_{i0} \mathbf{f}_{0t} \right)^2. \tag{9}$$

The identification of the new global factors $\widetilde{\mathbf{F}}_0$ relies on the matrix $\widehat{\mathbf{U}}$ so we use the PCE in (6). Subsequently, the new cluster-specific factors $\widetilde{\mathbf{F}}_k$ are identified considering the updated residuals $\widetilde{\mathbf{E}}$. However, we first need to remove the effect of previously identified clustered factors, so we consider

$$\widetilde{\mathbf{E}} = \widehat{\mathbf{U}} - \widetilde{\mathbf{F}}_0 \widetilde{\Lambda}'_0 + \sum_{k=1}^{K} \widehat{\mathbf{F}}_k \widehat{\Lambda}'_k = \mathbf{X} - \widetilde{\mathbf{F}}_0 \widetilde{\Lambda}'_0, \tag{10}$$

and the estimation procedure continues until convergence.

A few remarks on the iterative procedure are needed. First, we note that the loss function adopted in the iterative procedure, consistent with (Ando & Bai, 2017) and Alonso et al. (2020), is least squares based and is similar to the Iterative PCA adopted in Wang (2010). Second, it relates to the EM algorithm in that it alternates two main steps, the cluster assignment (the E-step) and the parameter (*i.e.* factors and loadings) estimation (the M-step), iteratively until convergence. However, the clustering assignment in the E-step depends on a hierarchical procedure discussed in Section 3.2. Third, we intuitively set that if a kth cluster is a singleton, there is no local factor to estimate in that cluster. Even though we assume that a clustered factorial structure exists within each cluster, more sophisticated rules can be easily considered for testing whether a factorial structure at the cluster level exists. Moreover, alternative procedures can also be used for estimating the number of factors (Ahn & Horenstein, 2013). Fourth, we assume a static approximate factor structure (Bai & Ng, 2002; Stock & Watson, 2002). As shown in Wang (2010), including a finite number of lagged factors does not invalidate the properties of the Iterative PCA. However, other approaches for dynamic factors estimation, such as the dynamic principal component (DPCA, Brillinger, 2001; Forni, Hallin, Lippi, & Reichlin, 2000, 2005) and the generalized dynamic principal component (GDPCA, Peña & Yohai, 2016) may be used in this setting. Then, we notice that the proposed approach is computationally feasible if $N > T$ and $T > N$. As we intend to cluster housing prices of US states where $T > N$, the procedure highlighted so far is computationally very fast. When $N > T$, instead, a computationally suitable approach estimates the factor matrix as $\sqrt{T}$ times the eigenvectors corresponding to the R largest eigenvalues of the matrix $\mathbf{XX}'$ and $\Lambda_0 = \mathbf{F}'_0 \mathbf{X}$ the matrix of global factor loadings. The same applies to clustered factors. This second solution is recommended when dealing with, for example, housing prices in metropolitan areas where the cross-sectional dimension N is usually larger than T. The above-described procedure is better suited when (relatively long) georeferenced time series are considered.

Finally, we briefly remark that the main consequence of enhancing spatial clustering is an increase in the cross-sectional correlation of idiosyncratic errors within each cluster. As discussed in the introduction, this is a typical feature in regional factor models. Given that a very general form of cross-sectional dependence is allowed, introducing spatial clustering does not invalidate the properties of the principal component approach for global and cluster-specific factor estimation once the cluster structure has been identified.

3.2. Hierarchical spatio-temporal clustering

In what follows, we discuss the procedure while clustering residuals in the computational estimation of the CFM. Different approaches can be used for this task. The clustering approach adopted defines the nature of the clustered factor model considered, which is spatial, temporal, or spatio-temporal. Spatially clustered factor models are quite simple and define the partition based only on the spatial dimension. The CFM algorithm becomes a two-step procedure in this simple case because the clustering structure does not change with the iterations. Temporally clustered factor models, such as Ando and Bai (2017) and Alonso et al. (2020), do not consider the spatial dimension, which can be less appropriate in some empirical instances. Therefore, we propose a suitable spatio-temporal clustering approach in updating the CFM algorithm.

The use of partitional approaches is not recommended in this context to reduce the computational burden of the procedure. Indeed, these algorithms, such as k-means or k-medoids, require the apriori selection of the number of clusters G. Therefore, as in Alonso et al. (2020), we prefer hierarchical algorithms to reduce the degree of freedom of the problem.

Following Chavent, Kuentz-Simonet, Labenne, and Saracco (2018), Mattera and Franses (2023) and Bucci, Ippoliti, Valentini, et al. (2023), we consider a Ward-like spatio-temporal hierarchical clustering algorithm for updating the clustering structure at each iteration. Let us consider a set of N ($i = 1, \ldots, N$) statistical units and let $\mathbf{D} = [d_{ij}]$ be the $N \times N$ dissimilarity matrix associated with the N units, with d_{ij} being the dissimilarity measure between two units i and j. Let us define $\mathcal{P}_K = (\mathcal{C}_1, \ldots, \mathcal{C}_K)$ a partition of the dataset into K clusters. In the standard cross-sectional clustering, we can express the within-clusters inertia $\mathcal{C}_k$ ($k = 1, \ldots, K$) as follows:

$$I(\mathcal{C}_k) = \sum_{i \in \mathcal{C}_k} \sum_{j \in \mathcal{C}_k} \frac{w_i w_j}{2 \sum_{i \in \mathcal{C}_k} w_i} d_{ij}^2, \tag{11}$$

with w_i the weight associated to the ith statistical unit. Without any a priori information, it is common to set $w_i = 1/N$. The dissimilarity matrix $\mathbf{D}$ is computed in the attribute space, considering a set of P ($p = 1, \ldots, P$) variables describing the N statistical units. In spatio-temporal data, the attribute space is represented by observing an attribute over T ($t = 1, \ldots, T$) time periods and by the spatial dimension.

We define the spatio-temporal inertia of a $\mathcal{C}_k$ ($k = 1, \ldots, K$) cluster as the convex combination between the attribute inertia and the inertia of spatial clusters. In

other words, in the case of spatio-temporal clustering, (11) becomes:

$$I\left(\mathcal{C}_k\right) = (1-\alpha)\sum_{i\in\mathcal{C}_k}\sum_{j\in\mathcal{C}_k}\frac{w_i w_j}{2\sum_{i\in\mathcal{C}_k}w_i}d_{ij,t}^2$$
$$+ \alpha\sum_{i\in\mathcal{C}_k}\sum_{j\in\mathcal{C}_k}\frac{w_i w_j}{2\sum_{i\in\mathcal{C}_k}w_i}d_{ij,g}^2, \quad (12)$$

where $d_{ij,t}^2$ is the squared distance between units i and j over the (temporal) attribute space and $d_{ij,g}^2$ is the squared distance in the geographical space. The smaller the inertia $I\left(\mathcal{C}_k\right)$ is, the more homogenous are the observations in cluster $\mathcal{C}_k$. Note that the spatio-temporal inertia (12) is a convex combination of temporal inertia and spatial inertia. Moreover, spatio-temporal inertia depends on the so-called *mixing parameter* α, which in our case can be interpreted as the relevance of spatial information relative to the temporal information.

The resulting hierarchical clustering approach considers an initial partition with N clusters of singletons. Then, the algorithm aggregates two clusters at each step according to an objective function related to the within-cluster inertia. The procedure adopted in determining spatial weights α aims at finding a value representing the best compromise between loss in the temporal space and loss of geographical homogeneity. The procedure adopted in our paper is explained with more details in Chavent et al. (2018) and Mattera and Franses (2023). Let us define

$$W\left(\mathcal{P}_K^{(\alpha)}\right) = \sum_{k=1}^{K} I\left(\mathcal{C}_k^{(\alpha)}\right),$$

the spatio-temporal inertia of a kth cluster for a given value of the mixing parameter α. The total spatio-temporal inertia $W\left(\mathcal{P}_K^{(\alpha)}\right)$ equals the sum of the within-cluster inertia of its K clusters, $I\left(\mathcal{C}_k^{(\alpha)}\right)$. To choose α, we compare the proportion of the total inertia explained by the partition obtained with the temporal dimension

$$Q_t\left(\mathcal{P}_K^\alpha\right) = 1 - \frac{W_t\left(\mathcal{P}_K^\alpha\right)}{W_t\left(\mathcal{P}_1\right)}$$

with the proportion associated with the spatial dimension

$$Q_g\left(\mathcal{P}_K^\alpha\right) = 1 - \frac{W_g\left(\mathcal{P}_K^\alpha\right)}{W_g\left(\mathcal{P}_1\right)}$$

for different S values of α. For example, we can consider $\alpha = \{\alpha_1 = 0, \alpha_2 = 0.01, \ldots, \alpha_{S-1} = 0.99, \alpha_S = 1\}$. We notice that $W_t\left(\mathcal{P}_1\right)$ and $W_g\left(\mathcal{P}_1\right)$ are the total inertia using temporal and geographical distance only, i.e. $W\left(\mathcal{P}_K^{(\alpha)}\right)$ with $\alpha = 0$ and $\alpha = 1$, respectively.

The difference between the two criteria for all the S values of α provides a simple way for choosing the mixing parameter, that is, when the difference between the loss of temporal homogeneity and larger geographical cohesion, obtained through increasing values of α, is minimum.

We need only to define both temporal and spatial dissimilarity measures. A suitable choice, which is very common for clustering economic time series, is the correlation-based distance (Maharaj, D'Urso, & Caiado, 2019)

$$d_{ij,t} = \sqrt{2\left(1 - \hat{\rho}_{i,j}\right)^2}, \quad (13)$$

with $\hat{\rho}_{i,j}$ being the sample correlation coefficient between the two time series i and j. Notice that the correlation-based distance is meaningful if stationary time series are considered. Other dissimilarity measures for time series data can also be considered, such as cross-correlation-based (Alonso & Peña, 2019). For nonstationary series, suitable alternative distances are the ARIMA (Piccolo, 1990) or the feature-DTW (Franses & Wiemann, 2020) distances. Spatial clustering, instead, can be achieved in different ways, depending mainly on the measure used to quantify proximity. Here, we consider a spatial dissimilarity based on the geographical distances among the statistical units, measured in terms of latitude and longitude

$$d_{ij,g} = \sqrt{\left(x_{i,\text{lat}} - x_{j,\text{lat}}\right)^2 + \left(x_{i,\text{long}} - x_{j,\text{long}}\right)^2}. \quad (14)$$

Using dissimilarity (14) ensures that statistical units closer in the geographical space are clustered together. The distance (14) is not the only possible approach for enhancing spatial clustering (Bucci et al., 2023; Fouedjio, 2016; Mattera, 2022). Nevertheless, as shown in Chavent et al. (2018), if combined with another non-spatial dissimilarity, it can be successfully used to provide a soft spatial constraint to the final partition.

4. Simulation study

In what follows, we assess the performance of the proposed computational procedure through a simulation study. In particular, we assess the accuracy in clustering and forecasting the statistical units. Moreover, we evaluate the suitability of the Elbow criterion in correctly choosing the number of clusters K and shed light on the mechanism behind the selection of the mixing parameter α. The proposed approach is compared with the (Ando & Bai, 2017) which does not account for the spatial dimension in the clustering task. The clustering accuracy is evaluated in terms of the Rand index (RI, Hubert & Arabie, 1985), which lies between 0 and 1. The higher the value of RI is, the better the clustering accuracy. Given that the true labels of the units are known in the simulated environment, we have that the RI equals 1 in the case of perfect overlap in clustering obtained by the method and the true partition. The forecasting ability is evaluated by computing the loss $L_{\hat{\mathbf{X}},\tilde{\mathbf{X}}} = E\left[\left(\hat{\mathbf{X}}_{T+1|T} - \tilde{\mathbf{X}}_{T+1|T}\right)^2\right]$, that is the expectation of the squared difference between the one-step-ahead forecast obtained with the estimated factors $\hat{\mathbf{X}}_{T+1|T}$ and those obtained with the true values $\tilde{\mathbf{X}}_{T+1|T}$. Hence, the lower $L_{\hat{\mathbf{X}},\tilde{\mathbf{X}}}$, the better is the forecasting accuracy. To obtain one-step-ahead forecasts, we use the Panel VAR(1) model (2) with the fixed-effects estimator as discussed in Section 2.2. We simulate 1000 datasets, averaging the results over the simulations.
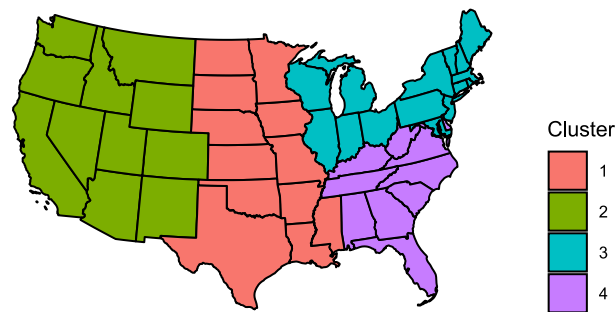
To include a spatial dimension in the simulation study consistent with our empirical application to house prices,

Simulated scenario with K=3
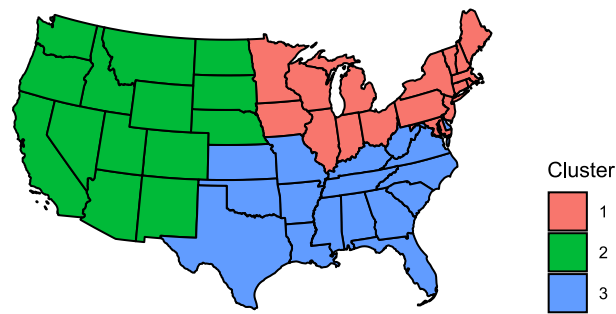
(a) Clustering structure with $K = 3$.

Simulated scenario with K=4

(b) Clustering structure with $K = 4$.

Fig. 2. Clustering structures considered for the simulation experiments.

we simulate $N = 49$ statistical units representing the US states with their spatial coordinates. We assume the presence of $K = 3$ and $K = 4$ clusters, respectively, which are shown in Fig. 2

Given the $N = 49$ units, we consider one global factor and one local factor for each cluster. Following (Ando & Bai, 2017), under the DGP as in (3), we simulate the global factor from a uniform $[0, 1]$ distribution and the factor loadings as uniform $[-2, 2]$ variables. Moreover, each element of the cluster-specific factors is simulated from standard Normal variables $N(0, 1)$ and each element of the factor-loading matrix follows also $N(0, 1)$. Then, we consider two alternative specifications for the error term. In the first scenario (Scenario I), we allow some cross-sectional correlation between the errors within each cluster, which are simulated from a multivariate normal distribution with zero mean and covariance matrix with constant correlation $\rho_{ij} = 0.3$ if i and j belong to the

same cluster, and $\rho_{ij} = 0$ otherwise. The errors associated with statistical units placed in different clusters are independent.

In the second scenario (Scenario II), we assume that correlations depend on geographical proximity. In particular, we assume that the correlation between two units i and j increases with smaller geographical distances. To be more precise, let us define $\widetilde{d}_{ij,g}$ the normalized geographical distance, that is $\widetilde{d}_{ij,g} = d_{ij,g}/\max(d_{ij,g})$, such that all the values of $\widetilde{d}_{ij,g}$ fall in the interval $[0, 1]$. We assume that the correlation between units i and j equals $\widetilde{\rho}_{ij} = \rho_{ij}^{\widetilde{d}_{ij,g}}$. In this way, we simulate a covariance matrix where the actual correlations $\widetilde{\rho}_{ij}$ depend on the spatial correlation (ρ_{ij}), which increases faster with decreasing geographical distance. Indeed, large values of $\widetilde{d}_{ij,g}$ indicate that the time series associated with the states i and j are very far in the geographical space. For $\widetilde{d}_{ij,g} \to 1$, so for very large distances, we have that $\widetilde{\rho}_{ij} \to \rho_{ij}$ while for $\widetilde{d}_{ij,g} \to 0$, so

Table 1

Simulation study results under Scenario I. Rand Index is the accuracy of the clustering resulting from the approach on the rows, while $L_{\hat{\mathbf{x}},\tilde{\mathbf{x}}}$ is a measure of forecasting accuracy. The results are obtained with an average of 1000 simulations; standard deviations are shown in parentheses.

Panel A: $T = 100$				
	Rand Index		$L_{\hat{\mathbf{x}},\tilde{\mathbf{x}}}$	
	$K = 3$	$K = 4$	$K = 3$	$K = 4$
Ando and Bai (2017)	0.539	0.596	5.646	5.200
	(0.03)	(0.03)	(3.719)	(3.270)
Our method	0.674	0.731	5.534	5.197
	(0.03)	(0.03)	(3.702)	(3.316)
Panel B: $T = 200$				
	Rand Index		$L_{\hat{\mathbf{x}},\tilde{\mathbf{x}}}$	
	$K = 3$	$K = 4$	$K = 3$	$K = 4$
Ando and Bai (2017)	0.533	0.587	4.025	3.731
	(0.03)	(0.04)	(2.687)	(2.209)
Our method	0.675	0.732	3.915	3.567
	(0.03)	(0.03)	(2.660)	(2.223)
Panel C: $T = 400$				
	Rand Index		$L_{\hat{\mathbf{x}},\tilde{\mathbf{x}}}$	
	$K = 3$	$K = 4$	$K = 3$	$K = 4$
Ando and Bai (2017)	0.529	0.576	2.799	2.608
	(0.02)	(0.04)	(1.868)	(1.568)
Our method	0.677	0.733	2.673	2.521
	(0.03)	(0.03)	(1.796)	(1.617)

Table 2

Simulation study results under Scenario II. Rand Index is the accuracy of the clustering resulting from the approach on the rows, while $L_{\hat{\mathbf{x}},\tilde{\mathbf{x}}}$ is a measure of forecasting accuracy. The results are obtained with an average of 1000 simulations; standard deviations are shown in parentheses.

Panel A: $T = 100$				
	Rand Index		$L_{\hat{\mathbf{x}},\tilde{\mathbf{x}}}$	
	$K = 3$	$K = 4$	$K = 3$	$K = 4$
Ando and Bai (2017)	0.551	0.616	10.327	10.155
	(0.02)	(0.02)	(6.239)	(5.907)
Our method	0.696	0.772	10.222	10.181
	(0.03)	(0.03)	(6.469)	(6.210)
Panel B: $T = 200$				
	Rand Index		$L_{\hat{\mathbf{x}},\tilde{\mathbf{x}}}$	
	$K = 3$	$K = 4$	$K = 3$	$K = 4$
Ando and Bai (2017)	0.551	0.614	7.086	7.244
	(0.02)	(0.02)	(4.173)	(4.669)
Our method	0.696	0.772	6.981	7.236
	(0.03)	(0.03)	(4.359)	(4.781)
Panel C: $T = 400$				
	Rand Index		$L_{\hat{\mathbf{x}},\tilde{\mathbf{x}}}$	
	$K = 3$	$K = 4$	$K = 3$	$K = 4$
Ando and Bai (2017)	0.548	0.612	5.189	5.221
	(0.02)	(0.02)	(3.365)	(3.114)
Our method	0.697	0.771	5.070	5.201
	(0.03)	(0.03)	(3.339)	(3.321)

for small distances, we have that $\widetilde{\rho}_{ij} \to 1$. In what follows, we show the results assuming $\rho_{ij} = 0.1$.[4]

Given the two scenarios, we consider different time series lengths, that is, $T = 100, 200, 400$, and we keep $N = 49$ fixed. The results of the simulations under Scenario I are shown in Table 1, while those under Scenario II are in Table 2. The column Rand Index measures the (average) accuracy of the clustering, while $L_{\hat{\mathbf{x}},\tilde{\mathbf{x}}}$ is a measure of forecasting accuracy. DGP1 assumes independence among error terms, while DGP2 assumes some cross-sectional correlation exists. The results are averaged over 1000 simulated datasets, while the standard deviations of the results are in parentheses.

Considering the results of both simulated scenarios, we note that the clustering accuracy does not vary much when increasing the sample size from $T = 100$ to $T = 400$. The proposed spatio-temporal approach shows better clustering accuracy than the temporal approach proposed by Ando and Bai (2017). Indeed, the rate of correct classification is 0.7 for our spatio-temporal approach, while it is about 0.5 for the (Ando & Bai, 2017) method. Moreover, in the case of Scenario II, the clustering accuracy improves compared to the case of Scenario I due to the increasing importance of the spatial dimension. In terms of forecasting accuracy, both Tables 1 and 2 show that our proposed spatio-temporal approach provides better results. Moreover, forecast accuracy improves in both simulated scenarios with increasing T, while the standard deviation decreases with increasing temporal observations.

Finally, we evaluate the accuracy of the Elbow criterion in selecting the number of clusters K, and the (Chavent

et al., 2018) procedure in choosing the mixing parameter α. Fig. 3 summarizes the results on the choice of the number of clusters K. We consider the values $T = 100$, $T = 200$ and $T = 400$, and both $K = 3$ and $K = 4$ for both the considered scenarios.

Fig. 3 indicates that the median of the boxplot equals the correct number of factors using the Elbow method. This holds for all the considered values of T. The variability of the results reduces for larger T. Regarding the mixing parameter α choice, the boxplots with the results of the simulations for the different scenarios are shown in Fig. 4.

We find that α is lower when the correlations, used for computing the distance in the temporal domain $d_{ij,t}$, do not depend on the spatial information. This is the case of Scenario I. When the correlations lead to some spatial clustering, for instance, because they depend on the spatial dimension as in Scenario II, we find the mixing parameter α to be more important. These results confirm that the procedure adopted in the paper for selecting α is appropriate. It also suggests that the selected value of α provides relevant information on the importance of spatial dimension in a given dataset. The larger the α, the more relevant the spatial dimension for determining the clustering structure of the data.

5. Forecasting US house price growth rates: results

We now empirically evaluate the usefulness of cluster-specific factors in forecasting the house price growth rates in the US states. For the out-of-sample forecasting experiment, a rolling-window procedure is adopted. The sample is divided into a train and a test sample. We leave the last eight years of observations for testing. A fixed window of forty years ($M = 160$) quarters is adopted, and the

---

[4] The results assuming $\rho_{ij} = 0.05$ and $\rho_{ij} = 0.01$ are available upon request, and lead to qualitatively similar findings.
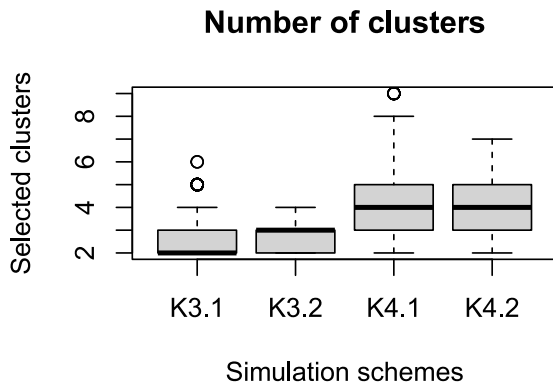
Number of clusters

(a) Results for $T = 100$
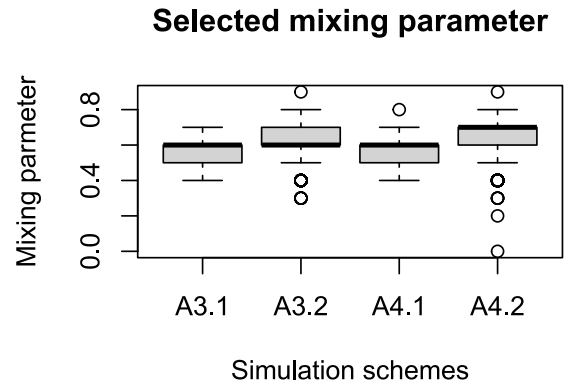
Number of clusters

(b) Results for $T = 200$

Number of clusters
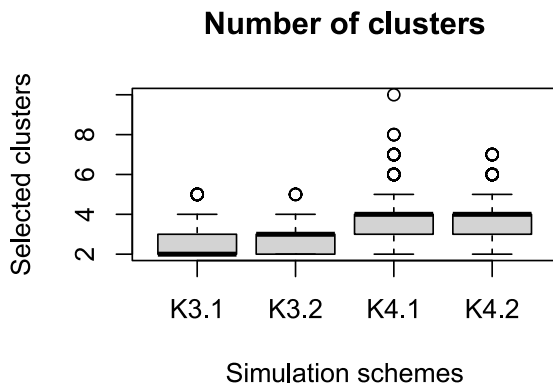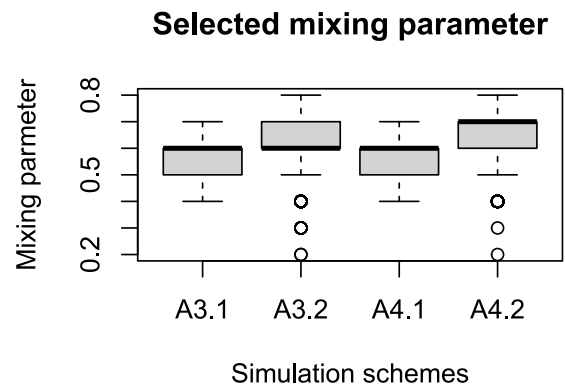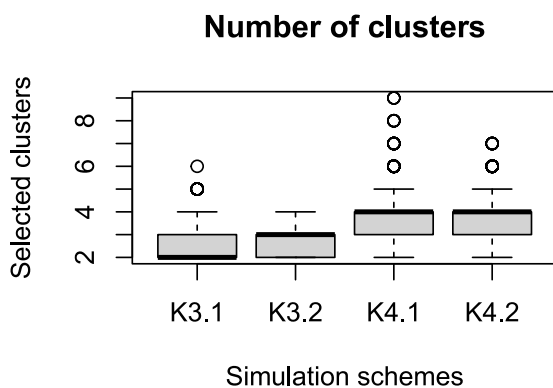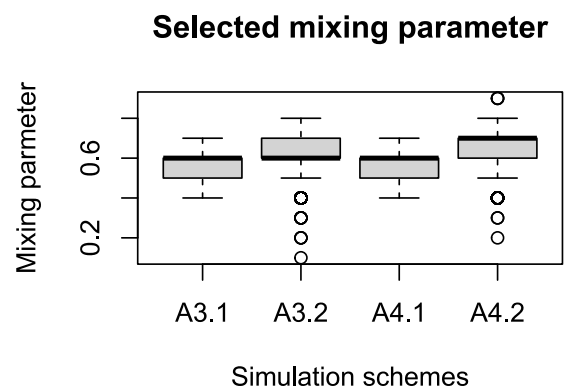
(c) Results for $T = 400$

Fig. 3. Optimal K selection with the Elbow criterion (our procedure), for different Scenarios ($T = 100, 200, 400$). "K3.1" means $K = 3$ under Scenario I, while "K3.2" means $K = 3$ under Scenario II. "K4.1" means $K = 4$ under Scenario I, and "K4.2" means $K = 4$ under Scenario II.

Selected mixing parameter

(a) Results for $T = 100$

Selected mixing parameter

(b) Results for $T = 200$

Selected mixing parameter

(c) Results for $T = 400$

Fig. 4. Optimal α selection for different Scenarios ($T = 100, 200, 400$). "A3.1" means $K = 3$ under Scenario I, while "A3.2" means $K = 3$ under Scenario II. "A4.1" means $K = 4$ under Scenario I, and "A4.2" means $K = 4$ under Scenario II.
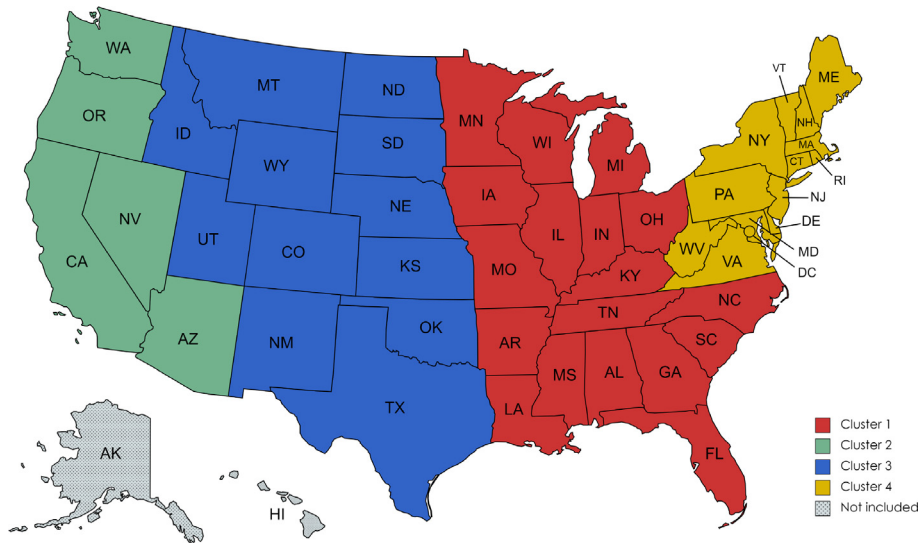
Fig. 5. Spatio-temporal clustering of the US house prices growth.

automatic model selection is performed to determine the optimal number of lags to include in each recursion. The number and the composition of clusters are allowed to vary during the rolling window procedure, and the global and clustered factors are re-estimated at each step. Given the estimated parameters, a one-step-ahead forecast is produced. Then, a new observation is included in the sample while the oldest one is removed. The procedure is repeated until no new observation is available.

To evaluate forecasting accuracy, we rely on two commonly employed accuracy metrics, namely the Root Mean Squared Error (RMSE), that is

$$\text{RMSE}_i = \sqrt{\frac{1}{T-M} \sum_{t=M+1}^{T} \left(\widehat{x}_{i,t} - x_{i,t}\right)^2}, \tag{15}$$

where $\widehat{x}_{i,t}$ represents the forecast at time t for the ith US state's house price growth rate and $x_{i,t}$ the realized value, and the Mean Absolute Error (MAE), that is,

$$\text{MAE}_i = \frac{1}{T-M} \sum_{t=M+1}^{T} \left|\widehat{x}_{i,t} - x_{i,t}\right|, \tag{16}$$

Notice that the models (m2) and (m3) are nested into (m1). It is interesting, therefore, to test if the additional information considered in the full models has a statistically significant role in more accurately forecasting house price growth rates. We conjecture that the model with global and cluster-specific factors provides more accurate out-of-sample forecasts than the nested models. Put differently, the cluster-specific factors are useful for forecasting purposes. As highlighted by Clark and West (2006), the parsimonious models should have smaller prediction errors than more complex ones in which they are nested. Indeed, parsimonious models gain accuracy by setting the parameters that are zero in the population to zero, while the more complex models introduce noise into

the forecasting process. Therefore, we adopt the (Clark & West, 2007) test based on adjusted errors for accurate testing. Given a generic ith unit (we omit the ith subscript for easy reading), let us define $\hat{e}_{1,t}$ and $\hat{e}_{2,t}$, $t = M + 1, \ldots, T$, the one-step-ahead forecast errors of the parsimonious model and the alternative one, respectively. Let $\hat{f}_t = \hat{e}_{1,t}^2 - \left[\hat{e}_{2,t}^2 - \left(\hat{x}_{1,t} - \hat{x}_{2,t}\right)^2\right]$ be the so-called adjusted-MSE (Clark & West, 2006). Let $\bar{f}$ be the corresponding sample average, that is, $\bar{f} = (T - M)^{-1} \sum_{t=M+1}^{T} \hat{f}_t$. The (Clark & West, 2007) test statistic is

$$\sqrt{(T-M)}\bar{f} / \sqrt{V\left[\hat{f}_t - \bar{f}\right]}. \tag{17}$$

with $V[\cdot]$ be the variance. Under the null hypothesis, the two models provide equally good forecasts. Clark and West (2007) recommend constructing the usual one-tailed t-test to assess if the adjusted difference in mean squared errors is zero. For the sake of robustness, we also consider the (Diebold & Mariano, 2002) predictive accuracy test.

5.1. Clustering results

The number of clusters is estimated through the Elbow criterion, and the resulting partition into $K = 4$ clusters is shown in Fig. 5.

The spatial weight computed in the spatio-temporal clustering algorithm is $\alpha = 0.6$, thus suggesting that the spatial dimension is more relevant than the temporal one in discriminating the data and determining the nature of the clustered factors. At the same time, the temporal dimension is not negligible. Thus, our spatio-temporal approach is recommended instead of a pure spatial or pure temporal approach. Moreover, we find that $R = 1$ for the global factor and $R_k = 1$ for all four clusters. The partition in Fig. 5 highlights that local factors arise in the
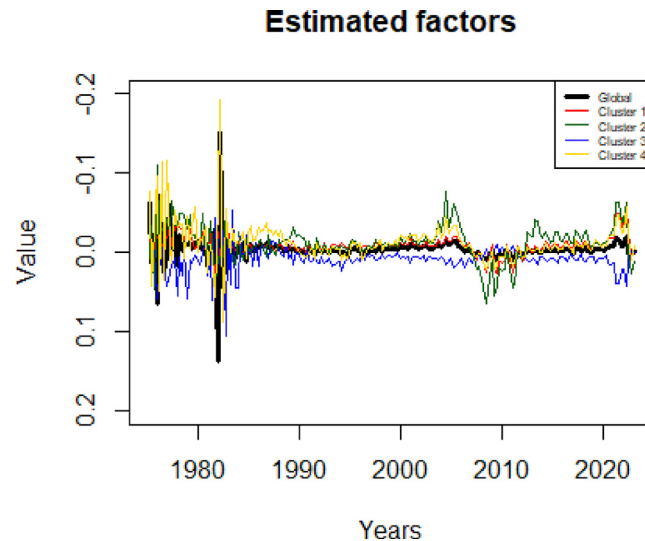
Estimated factors

Fig. 6. Temporal evolution of the estimated factors: both global and local.

west (Cluster 2) and in the north-east of the US (Cluster 4), as well as two distinct clustering factorial structures that exist in the centre of the US territory. In particular, it is interesting to notice that in Cluster 4 we have urban states with higher population densities, like New York, Massachusetts, and New Jersey, which tend to have more competitive housing markets with higher property prices.

Cluster 2, instead, includes Pacific states with higher housing price growth rates than the national average. California, in particular, is known for its high-priced housing market, with cities like San Francisco and Los Angeles experiencing some of the highest home price changes in the country. Oregon, Washington, and Nevada also have areas with relatively high housing price growth rates, especially in major metropolitan cities such as Portland and Seattle. Cluster 1 includes most coastal states compared with Cluster 3, but differences between the two clusters can be better exploited considering the estimated factors in Fig. 6. The solid black line in Fig. 6 represents the global factor, while the local factors have the same colours as in Fig. 5. The local factor of Cluster 1 is characterized by large volatility in the first years of the sample compared to other clusters. In contrast, Clusters 2 and 3 local factors show larger variability for most of the considered time span. Instead, the local factor of Cluster 1 shows the lowest level of variability compared to the others. Moreover, while most of the clusters show an increase in the local factors since 2020, Cluster 3 is the only one with a local factor with negative values.

5.2. In sample fit

Given the estimated latent (global and clustered) factors, we estimate the optimal lag length of the Panel VAR. Considering the values of the BIC-type criterion (*e.g.* Han et al., 2017) for different values of $L \in 1, \ldots, 5$, we choose $L = 1$ as it minimizes the BIC. We compare the Panel

VAR(1) models (*i.e.* m1–m3) in terms of in-sample RMSE and MAE, given the cluster structure into $K = 4$ groups. Fig. 7(a) shows the distribution of the $N = 49$ US states in terms of in-sample fit under RMSE loss, while Fig. 7(b) in terms of MAE loss. Both figures consider the relative accuracy, given by the ratio between the in-sample fit of the competing models to the main benchmark, that is, the m1 model. Values larger than 1, highlighted with the red dashed line, indicate that model m1 provides a better in-sample fit (*i.e.* lower RMSE or MAE) than the competing model.

Fig. 7 shows that both boxplots, comparing the different models m2 and m3 with the full model m1, are above the dashed red line, and as such, indicating a better in-sample fit of the full Panel VAR model including both global and cluster-specific factors. Moreover, we consider the in-sample fit obtained in the rolling window procedure. We estimate the optimal lag order of the Panel VAR models in each iteration. Figs. 7(c) and 7(d) show the average distributions, that is, the average RMSE and MAE for each US state across all the rolling windows. The results are similar to those obtained considering the full sample, thus suggesting that using both global and cluster-specific factors improves the in-sample fit of the Panel VAR model.

5.3. Out of sample results

Next, we investigate the accuracy of the model's out-of-sample forecasting. Figs. 8(a) and 8(b) show the distributions of the RMSE and MAE for each US state in the testing set. Also, in this case, the results are reported in terms of relative accuracy, where the accuracy of the full model, m1, is compared with those of the competing restricted Panel VAR models m2 and m3. Values larger than 1, indicated with a dashed red line, highlight that the

(a) Relative (RMSE) accuracy of the competing models with m1 for the full-sample experiment.

(b) Relative (MAE) accuracy of the competing models with m1 for the full-sample experiment.

(c) Relative (RSME) accuracy of the competing models with m1 for the rolling-window experiment.

(d) Relative (MAE) accuracy of the competing models with m1 for the rolling-window experiment.
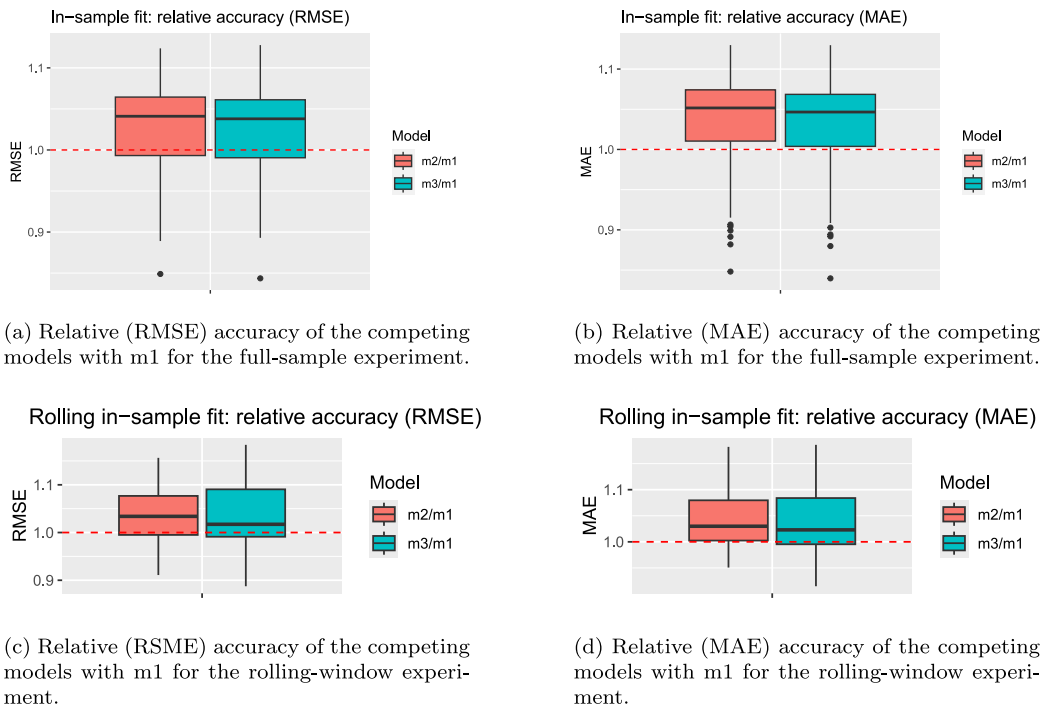
Fig. 7. Relative accuracy measures (RMSE and MAE) of the competing models compared with the model m1. Both full-sample and (averaged) rolling-window experiments are considered. Values larger than 1 (the red dashed line) indicate that m1 provides a better in-sample fit compared to the competing model.

model, including cluster-specific factors, provides more accurate forecasts.

In both figures, the boxplots are above the dashed red line, and therefore, the full model provides the most accurate out-of-sample forecasts for most US states. In particular, the median of the boxplots in terms of both accuracy measures is about 1.1. The cluster-specific factors are, however, not equally relevant across the US states. While these are very important in some states in predictive terms, they are not in others. Table 3 allows us to evaluate the differences at the state level.

According to Table 3, we highlight that the cluster-specific factors provide better accuracy for the US states in all the clusters, except for those in Cluster 3. For these states, *i.e.*, the mountain states in the US centre, we find that the model with global factors only performs worse than the baseline model. However, the differences in terms of average losses are not very large. In fact, the average loss for the states in Cluster 3 is equal to 0.0221 for the full model (m1), 0.0211 for the model with global factors only (m2) and 0.0212 for the baseline model (m3). The cluster-specific factors allow for considerable improvements in out-of-sample forecasting accuracy for the remaining states. For instance, the RMSE for the full model in Cluster 1 equals 0.0159, while it takes values 0.0187 and 0.0188 for the restricted models m2 and m3, respectively. Therefore, we find that global factors alone perform worse than the baseline model for US states in the east of the US. Including cluster-specific factors significantly improves the forecasting performance. Similar

findings are obtained for the US states in Cluster 2 and Cluster 4, for which the model with cluster-specific factors provides an average RMSE of 0.0199 for Cluster 2 and 0.0150 for Cluster 4,. In contrast, the model m2 performs worse than the baseline in both clusters. Interestingly, such regional differences in forecastability have also been acknowledged in other previous studies (*e.g.* see Rapach & Strauss, 2009).

Finally, we test the equality in the predictive accuracy of the full model m1 compared to both m2 and m3. In this setting, we also consider differences across the states. Table 4 shows the results of both the (Clark & West, 2007) test for nested models, and the (Diebold & Mariano, 2002) test. The results confirm that the differences in terms of forecasting accuracy described in Table 3 are statistically significant. Indeed, we reject the null hypothesis of the predictive accuracy tests for the US states in Clusters 1, 2 and 4. For the states in Cluster 3, however, we do not reject the null hypothesis of equal predictive accuracy.

5.4. Results with observable factors

The results obtained in previous sub-sections consider lagged values of the house price growth rates and lags of latent factors, estimated with the computational procedure discussed in Section 3. However, the accuracy of the forecasts can be improved, including observable factors. For instance, Holly, Pesaran, and Yamagata (2010) found that real per capita income, the long-term interest rates

Rolling out–of–sample: relative accuracy (RMSE)
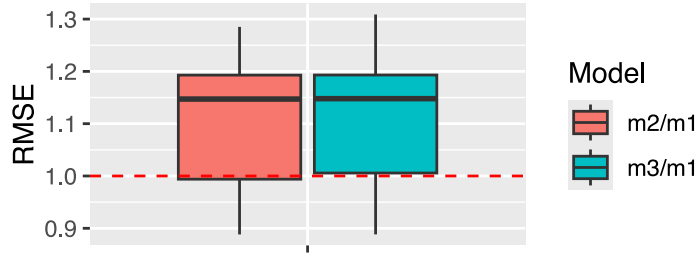
(a) Relative (MAE) accuracy of the competing models with m1. Values larger than 1 (the red dashed line) indicate that m1 provides a better in-sample fit compared to the competing model.

Rolling out–of–sample: relative accuracy (MAE)

(b) Relative (MAE) accuracy of the competing models with m1. Values larger than 1 (the red dashed line) indicate that m1 provides a better in-sample fit compared to the competing model.

Fig. 8. Relative accuracy of the competing models with m1.

and the global inflation rate are relevant predictors for house prices. In what follows, we include this source of additional information in the Panel VAR (1) and evaluate if cluster-specific factors are still useful even if observable factors are included in the forecasting model.

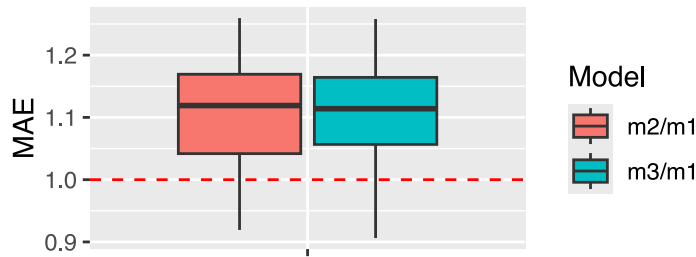Let us define by $\mathbf{z}_{i,t}$ the vector of P observable factors for the ith state at time t. We consider the real per capita income, the long-term interest rates and the global inflation rate, and we estimate the following additional models

1. (m4): Panel VAR where $\mathbf{y}_{i,t} = [x_{i,t}, \mathbf{z}'_{i,t}, \mathbf{f}'_{0,t}, \mathbf{f}'_{k,t}]'$;
2. (m5): constrained version of m4 without clustered factors, that is, $\mathbf{y}_{i,t} = [x_{i,t}, \mathbf{z}'_{i,t}, \mathbf{f}'_{0,t}]'$
3. (m6): constrained version of m4 without factors, where $\mathbf{y}_{i,t} = [x_{i,t}, \mathbf{z}'_{i,t}]'$. This is the standard Panel VAR;

The difference between models m1 and m4 is that the last also includes observable factors, while the first does not. Fig. 9 shows the comparison in terms of relative accuracy measures between the models m5 and m6 with model m4. Both in-sample (Figs. 9(a) and 9(b)) and out-of-sample (Figs. 9(c) and 9(d)) results from the

rolling-window experiments are considered. Values larger than 1 (the red dashed line) indicate that m4 provides better results than the competing model.

The results shown in Fig. 9 align with those discussed in the previous subsection. The medians of the boxplots are consistently above the threshold of 1, suggesting that model m4, including cluster-specific factors, provides better results both in-sample and out-of-sample. This result holds for both RMSE and MAE accuracy measures. Table 5 shows the out-of-sample accuracy results for each state, while Table 6 shows the results of predictive accuracy tests. Regarding regional differences, we find that model m4, including cluster-specific factors, provides more accurate out-of-sample forecasts than models m5 and m6 for all the states in Clusters 1, 2 and 4. For states in these clusters, the relative accuracy measures (RMSE and MAE) are larger than 1, while for states placed in Cluster 3, we find relative accuracy measures lower than 1. The best results can be found for the states in Cluster 1, like in the previous experiment, without observable factors. Also, regarding statistical significance, we reject the null hypothesis of equal predictive accuracy for both CW and DM tests while comparing model m4 with m5 and m6 in

Rolling in–sample fit: relative accuracy (RMSE)

Rolling in–sample fit: relative accuracy (MAE)

(a) Relative RMSE (in-sample) of the Panel VAR model with observable factors, including cluster-specific factors compared to the restricted models.

(b) Relative MAE (in-sample) of the Panel VAR model with observable factors, including cluster-specific factors compared to the restricted models.

Rolling out–of–sample: relative accuracy (RMSE)

Rolling out–of–sample: relative accuracy (MAE)

(c) Relative RMSE (out-of-sample) of the Panel VAR model with observable factors, including cluster-specific factors compared to the restricted models.

(d) Relative MAE (out-of-sample) of the Panel VAR model with observable factors, including cluster-specific factors compared to the restricted models.

Fig. 9. Relative accuracy measures (RMSE and MAE) of the models m5 and m6 compared to m4. Both in-sample and out-of-sample results from the rolling-window experiments are considered. Values larger than 1 (the red dashed line) indicate that m4 provides better results compared to the competing model.

the case of states in Clusters 1, 2 and 4. In contrast, we do not reject the null hypothesis for the states in Cluster 3.

Finally, we compare the forecasting accuracy of the models, including observable factors, with model m1. The results are summarized by Fig. 10. Interestingly, we find that, while in-sample the model including observable factors m4 provides the best results, out-of-sample, we find that the model m1 has the lowest RMSE and MAE. Therefore, the model m1 is the most accurate out-of-sample. The results thus confirm that including global and cluster-specific factors allows for good forecasting performance. Moreover, Fig. 10 also suggests that model m4 provides better results than models m5 and m6. This highlights again that including both cluster factors and observable factors in the model significantly enhances its accuracy.
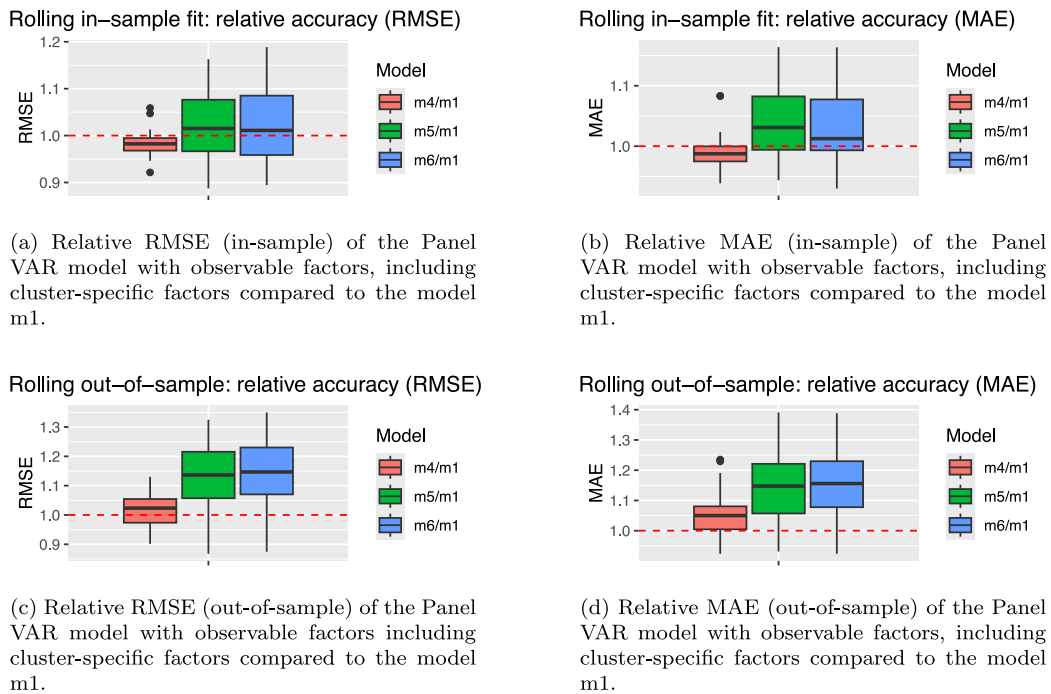
6. Final remarks

In this paper, we study cluster-specific factor models' ability to forecast house price growth rates in the US, assuming unknown spatio-temporal clustering. We estimate the cluster structure of the US states directly from the data and assume that all the states in the same cluster share the cluster-specific latent factors. Moreover, we assume all the US states share the same global factor. In doing so, we follow the idea behind regional factor models, which have become increasingly popular in empirical studies conducted by central banks. Regional factor models are a particular type of factor model with a known

clustering structure based on spatial information. Differently, however, we notice that the cluster structure depends not only on spatial information but also on the similarity of the temporal pattern of the housing market and propose a novel computational procedure.

Our method provides a structured way of estimating the factors and the clusters. This is accomplished by estimating a factor model, boosting along the residuals, estimating a type of guided hierarchical clustering model, and estimating additional "local" clusters. The described process, which automatically estimates the number of global factors, the clustering structure and the number of clustered factors, is then iterated. In doing so, we combine the principal components approach for estimating factors and guided hierarchical clustering. The proposed dimension reduction implied by the algorithm is sensible, particularly when the spatial indexes can be clustered geographically, and geographic clusters have corresponding factors that predict the outcome.

The novelties can be summarized as follows. First, we show that more accurate forecasts of house price growth rates can be achieved using cluster-specific factors. Forecasting models based on global factors only and models without factors provide less accurate forecasts out-of-sample. This evidence is confirmed by considering the results of predictive accuracy tests. Second, we adopt an unsupervised learning approach to build the clusters and propose a computational procedure that enhances spatial clustering. The main strength of the proposed method is that it does not require prior knowledge of the number

(a) Relative RMSE (in-sample) of the Panel VAR model with observable factors, including cluster-specific factors compared to the model m1.

(b) Relative MAE (in-sample) of the Panel VAR model with observable factors, including cluster-specific factors compared to the model m1.

(c) Relative RMSE (out-of-sample) of the Panel VAR model with observable factors including cluster-specific factors compared to the model m1.

(d) Relative MAE (out-of-sample) of the Panel VAR model with observable factors, including cluster-specific factors compared to the model m1.

Fig. 10. Relative accuracy measures (RMSE and MAE) of the Panel VAR models with observable factors (m4–m6) compared to the Panel VAR with latent factors only (m1). Both in-sample and out-of-sample results from the rolling-window experiments are considered. Values larger than 1 (the red dashed line) indicate that m1 provides better results than competing models, including observable factors.

of clusters in the dataset, as it is based on a hierarchical procedure. Moreover, we estimate the number of global and clustered factors within the algorithm, so the user does not need to specify these quantities too.

Considering the full sample of quarterly data on house prices in the period 1975–2023, we identify four clusters of US states characterized by their own clustered factor structure, given the presence of a single latent factor affecting the whole set of US states. Moreover, we show the results obtained using a Panel VAR approach for both in-sample and out-of-sample predictions. Still, we also study how the results change if separate models for each US state are estimated.[5] The conclusions we get for the relevance of cluster-specific factors are the same. Indeed, we find that the inclusion of cluster factors, estimated with the computational procedure discussed in Section 3, allows for more accurate forecasting of house price growth rates, even if state-specific univariate models are considered rather than the Panel VAR.

Several future research directions can be highlighted. First, the spatial structure considered in this paper could be extended to include an alternative distance across variables (Elhorst, 2013). Within this framework, a different definition of the distance (14) should also be considered. For instance, the geographical distance (14) can be replaced with a distance between multivariate time series (*e.g.* see, D'Urso, 2000). Future research may investigate if cluster-specific factors estimated using distances based on socio-economic variables have a larger predictive power than the geographical distance. Second, the proposed approach is effective if global factors dominate over cluster-specific factors, typically when there are a few large clusters among many smaller ones. However, in the opposite case, there is a risk of confounding global and cluster-specific factors, particularly when the variance of the global factors is relatively small compared to that of the cluster factors. Further investigation in this direction is needed, and methods for the principal component approach for factor estimation should be developed to address this issue. Third, the methods discussed in this paper rely on the stationarity assumption. While stationarity is commonly studied in the literature, nonstationary factor models present a promising topic for future research, especially when considering clustering structures. A possible solution to estimate factors in this context is the Generalized PCA (Peña & Yohai, 2016), which is suitable for nonstationary time series. The properties related to factor reconstruction have been studied by Smucler (2019), but further research seems needed in this direction. Finally, we admit that the proposed approach provides a limited understanding of how uncertainty in estimating groups, loadings, and latent factors propagates in the temporal forecasting model. These limitations could potentially be mitigated by using Bayesian spatial dynamic factor models. Gamerman et al. (2022), Ippoliti et al. (2012), Lopes et al. (2008).

---

[5] Results are available upon request.

Table 3

Relative accuracy measures for each US state (out-of-sample). Values larger than 1 indicate that the full model, including global and cluster-specific factors, provides more accurate forecasts.

States	RMSE		MAE	
	m2/m1	m3/m1	m2/m1	m3/m1
AL	1.2541	1.2480	1.1773	1.1643
AR	1.2187	1.2079	1.1693	1.1502
AZ	1.1946	1.2055	1.1879	1.2030
CA	1.2552	1.2592	1.2243	1.2344
CO	0.9307	0.9352	0.9594	0.9613
CT	1.1318	1.1295	1.1042	1.0925
DC	0.9939	1.0057	1.0067	1.0208
DE	1.1650	1.1424	1.1369	1.1122
FL	1.1248	1.1376	1.0963	1.1206
GA	1.1505	1.1598	1.0893	1.0913
IA	1.2534	1.2460	1.2595	1.2361
ID	0.9661	0.9746	1.0084	1.0184
IL	1.2175	1.2127	1.1954	1.1773
IN	1.1850	1.1873	1.1178	1.1121
KS	1.1460	1.1478	1.0867	1.0819
KY	1.2722	1.2657	1.2029	1.1798
LA	0.8882	0.8883	0.9251	0.9305
MA	1.1471	1.1409	1.1831	1.1694
MD	1.0902	1.0868	1.0740	1.0624
ME	1.1333	1.1338	1.1270	1.1134
MI	1.1639	1.1651	1.1297	1.1235
MN	1.1193	1.1156	1.0917	1.0818
MO	1.1611	1.1555	1.1637	1.1514
MS	1.2851	1.3088	1.1920	1.2249
MT	0.9272	0.9378	0.9650	0.9860
NC	1.1441	1.1498	1.0811	1.0771
ND	0.9340	0.9352	0.9873	1.0124
NE	0.9633	0.9560	0.9787	0.9694
NH	1.1333	1.1303	1.0907	1.0812
NJ	1.1707	1.1687	1.1562	1.1394
NM	0.9503	0.9543	1.0295	1.0524
NV	1.1929	1.2043	1.1591	1.1773
NY	1.2424	1.2435	1.2401	1.2580
OH	1.1704	1.1676	1.1380	1.1307
OK	0.9305	0.9323	0.9389	0.9394
OR	1.1849	1.1892	1.1740	1.1803
PA	1.1980	1.1950	1.1490	1.1353
RI	1.1218	1.1180	1.1317	1.1257
SC	1.2108	1.2161	1.1854	1.1740
SD	0.9324	0.9337	0.9548	0.9569
TN	1.1687	1.1794	1.1115	1.1261
TX	0.9321	0.9403	0.9439	0.9544
UT	0.9730	0.9808	1.0418	1.0566
VA	1.1170	1.1123	1.0823	1.0646
VT	1.1187	1.1189	1.1190	1.1140
WA	1.1688	1.1753	1.1882	1.1934
WI	1.1908	1.1894	1.1604	1.1467
WV	1.2222	1.1931	1.1563	1.1340
WY	0.9186	0.9155	0.9193	0.9062

Table 4

Predictive accuracy tests results: (Clark & West, 2007) (CW) and Diebold and Mariano (2002) (DM) with squared errors. The p-values are shown, and the full model m1 is compared with the models m2 and m3 for each US state.

States	CW test		DM test	
	m1 vs m2	m1 vs m3	m1 vs m2	m1 vs m3
AL	0.03	0.03	0.05	0.05
AR	0.04	0.04	0.05	0.05
AZ	0.02	0.01	0.03	0.03
CA	0.02	0.02	0.05	0.05
CO	0.83	0.79	0.96	0.96
CT	0.02	0.01	0.03	0.03
DC	0.14	0.12	0.59	0.38
DE	0.03	0.04	0.04	0.05
FL	0.05	0.03	0.05	0.05
GA	0.05	0.03	0.06	0.05
IA	0.02	0.03	0.05	0.05
ID	0.70	0.63	0.93	0.89
IL	0.02	0.02	0.05	0.05
IN	0.03	0.03	0.06	0.06
KS	0.03	0.03	0.06	0.06
KY	0.03	0.03	0.05	0.05
LA	0.60	0.57	0.98	0.98
MA	0.03	0.03	0.03	0.03
MD	0.07	0.06	0.09	0.08
ME	0.02	0.01	0.03	0.04
MI	0.03	0.03	0.04	0.03
MN	0.07	0.07	0.05	0.04
MO	0.03	0.03	0.04	0.04
MS	0.04	0.03	0.05	0.05
MT	0.84	0.78	0.95	0.95
NC	0.04	0.04	0.05	0.05
ND	0.58	0.53	0.98	0.97
NE	0.55	0.56	0.92	0.92
NH	0.02	0.02	0.05	0.05
NJ	0.01	0.01	0.03	0.03
NM	0.58	0.52	0.94	0.92
NV	0.02	0.01	0.02	0.02
NY	0.02	0.01	0.02	0.03
OH	0.03	0.03	0.04	0.04
OK	0.65	0.61	0.95	0.94
OR	0.03	0.02	0.03	0.03
PA	0.02	0.01	0.03	0.03
RI	0.02	0.02	0.04	0.04
SC	0.03	0.02	0.04	0.04
SD	0.73	0.70	0.95	0.95
TN	0.03	0.02	0.05	0.04
TX	0.78	0.72	0.96	0.95
UT	0.61	0.55	0.86	0.80
VA	0.04	0.04	0.07	0.07
VT	0.02	0.01	0.06	0.06
WA	0.02	0.02	0.03	0.02
WI	0.03	0.03	0.04	0.04
WV	0.04	0.05	0.05	0.05
WY	0.75	0.73	1.00	1.00

Table 5
Relative accuracy measures for each US state (out-of-sample). Values larger than 1 indicate that the full model, including global and cluster-specific factors, provides more accurate forecasts. The compared models include observable factors.

States	RMSE		MAE	
	m5/m4	m6/m4	m5/m4	m6/m4
AL	1.2111	1.2188	1.1372	1.1332
AR	1.1650	1.1684	1.1064	1.0982
AZ	1.1825	1.1999	1.1886	1.2122
CA	1.2073	1.2206	1.2452	1.2682
CO	0.9556	0.9661	0.9714	0.9766
CT	1.1311	1.1413	1.0912	1.0792
DC	0.9635	0.9824	0.9455	0.9667
DE	1.1198	1.1110	1.0950	1.0767
FL	1.1094	1.1318	1.0781	1.1113
GA	1.1276	1.1448	1.1014	1.1107
IA	1.1748	1.1885	1.1318	1.1296
ID	0.9851	0.9986	1.0098	1.0232
IL	1.1530	1.1629	1.0950	1.0830
IN	1.1655	1.1798	1.1056	1.1094
KS	1.1134	1.1293	1.1206	1.1245
KY	1.2015	1.2089	1.1217	1.1142
LA	1.0041	1.0166	0.9744	0.9830
MA	1.1250	1.1353	1.1423	1.1381
MD	1.0701	1.0784	1.0348	1.0374
ME	1.1268	1.1420	1.0923	1.0849
MI	1.1249	1.1378	1.1506	1.1536
MN	1.0801	1.0906	1.1261	1.1255
MO	1.1333	1.1406	1.1339	1.1360
MS	1.1820	1.2166	1.1271	1.1659
MT	0.9530	0.9708	0.9577	0.9822
NC	1.1320	1.1484	1.1027	1.1088
ND	1.0058	1.0259	1.0182	1.0466
NE	0.9987	1.0052	0.9852	0.9853
NH	1.1239	1.1338	1.0661	1.0654
NJ	1.1585	1.1712	1.1232	1.1084
NM	1.0373	1.0544	0.9981	1.0296
NV	1.1671	1.1867	1.1543	1.1763
NY	1.1967	1.2183	1.1977	1.2236
OH	1.1460	1.1559	1.1387	1.1401
OK	0.9875	1.0016	0.9506	0.9644
OR	1.1557	1.1683	1.2066	1.2184
PA	1.1662	1.1819	1.1325	1.1386
RI	1.1081	1.1168	1.1003	1.1041
SC	1.1761	1.1915	1.1432	1.1463
SD	0.9665	0.9829	0.9510	0.9721
TN	1.1509	1.1711	1.1096	1.1327
TX	0.9674	0.9845	0.9731	0.9945
UT	1.0015	1.0155	1.0449	1.0654
VA	1.1028	1.1131	1.0744	1.0644
VT	1.1211	1.1328	1.0956	1.0910
WA	1.1403	1.1534	1.1812	1.1909
WI	1.1789	1.1944	1.2085	1.2053
WV	1.1573	1.1445	1.0621	1.0464
WY	0.9625	0.9705	0.9263	0.9189

Table 6
Predictive accuracy tests results: (Clark & West, 2007) (CW) and Diebold and Mariano (2002) (DM) with squared errors. The p-values are shown, and the full model m1 is compared with the models m2 and m3 for each US state. The compared models include observable factors.

States	CW test		DM test	
	m5/m4	m6/m4	m5/m4	m6/m4
AL	0.05	0.04	0.05	0.05
AR	0.07	0.07	0.05	0.05
AZ	0.01	0.01	0.03	0.03
CA	0.03	0.02	0.04	0.04
CO	0.74	0.65	0.94	0.92
CT	0.02	0.01	0.04	0.04
DC	0.33	0.24	0.86	0.77
DE	0.08	0.08	0.06	0.07
FL	0.05	0.02	0.05	0.04
GA	0.06	0.04	0.05	0.05
IA	0.03	0.02	0.07	0.06
ID	0.54	0.40	0.87	0.55
IL	0.04	0.03	0.07	0.07
IN	0.05	0.04	0.07	0.06
KS	0.03	0.02	0.05	0.04
KY	0.06	0.05	0.05	0.04
LA	0.25	0.21	0.44	0.27
MA	0.03	0.03	0.03	0.03
MD	0.12	0.10	0.11	0.10
ME	0.02	0.01	0.05	0.05
MI	0.03	0.02	0.03	0.03
MN	0.07	0.06	0.03	0.03
MO	0.04	0.04	0.04	0.04
MS	0.07	0.04	0.05	0.06
MT	0.74	0.61	0.94	0.94
NC	0.05	0.03	0.05	0.04
ND	0.24	0.16	0.40	0.25
NE	0.35	0.33	0.54	0.36
NH	0.02	0.02	0.06	0.06
NJ	0.03	0.02	0.04	0.05
NM	0.19	0.13	0.08	0.08
NV	0.01	0.01	0.02	0.02
NY	0.02	0.01	0.04	0.04
OH	0.03	0.03	0.04	0.03
OK	0.36	0.28	0.73	0.46
OR	0.02	0.02	0.03	0.03
PA	0.03	0.02	0.04	0.04
RI	0.02	0.02	0.05	0.05
SC	0.05	0.03	0.04	0.04
SD	0.55	0.44	0.94	0.87
TN	0.04	0.03	0.04	0.04
TX	0.61	0.48	0.91	0.82
UT	0.37	0.26	0.45	0.09
VA	0.06	0.05	0.08	0.08
VT	0.02	0.02	0.08	0.08
WA	0.01	0.01	0.02	0.02
WI	0.02	0.02	0.04	0.04
WV	0.07	0.08	0.08	0.08
WY	0.49	0.44	0.99	0.98

Algorithm 1 Clustered factor model

Fix the maximum iterations *max.iter* and a convergence rule ε;
Estimate a suitable initial R;
Estimate R factors $\widehat{\mathbf{F}}_0$ with PCE;
Take the residuals $\widehat{\mathbf{E}}$ and provide an initial suitable partition;
for k=1,...,K **do**
   if $N_k > 1$ **then**
      Estimate the number of within-cluster factors R_k;
      Estimate R_k factors $\widehat{\mathbf{F}}_k$ with cluster-wise PCE;
   else
      No local factor;
   end if
end for
Store the initial loss:

$$\ell_0 = \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left(x_{it} - \widehat{\lambda}'_{i0}\widehat{f}_{0t} - \widehat{\lambda}'_{ik}\widehat{f}_{kt} \right)^2 \tag{18}$$

repeat
   Compute the residuals based on cluster-wise factors $\widehat{\mathbf{U}}$;
   Update R;
   Estimate the new R factors $\widetilde{\mathbf{F}}_0$ with PCE;
   Take the newly adjusted residuals $\widetilde{\mathbf{E}}$ and update the clusters with Algorithm 2;
   for k=1,...,K **do**
      if $N_k > 1$ **then**
         Estimate the number of within-cluster factors R_k;
         Estimate R_k factors $\widetilde{\mathbf{F}}_k$ with cluster-wise PCE;
      else
         No local factor;
      end if
   end for
   Store the loss:

$$\ell_1 = \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left(x_{it} - \widetilde{\lambda}'_{i0}\widetilde{f}_{0t} - \widetilde{\lambda}'_{ik}\widetilde{f}_{kt} \right)^2$$

until $| \left(\ell_i - \ell_{i-1} \right) / \ell_{i-1} | \leq \varepsilon$ or *max.iter* is reached.

Algorithm 2 Hierarchical spatio-temporal clustering

Consider a grid of S values for $\alpha \in [0, 1]$:

$$\mathcal{U} = \{\alpha_1 = 0, \ldots, \alpha_s, \ldots, \alpha_S = 1\}.$$

for i=1,...,N **do**
   for j=1,...,N **do**

$$d_{ij,t} = \sqrt{2(1 - \hat{\rho}_{ij})},$$

$$d_{ij,g} = \sqrt{\left(x_{i,\text{lat}} - x_{j,\text{lat}} \right)^2 + \left(x_{i,\text{long}} - x_{j,\text{long}} \right)^2}.$$

   end for
end for
for s=1,...,S **do**

$$Q_t\left(\mathcal{P}_G^{\alpha=\alpha_s} \right) = 1 - \frac{W_t\left(\mathcal{P}_G^{\alpha=\alpha_s} \right)}{W_t\left(\mathcal{P}_1 \right)}$$

$$Q_g\left(\mathcal{P}_G^{\alpha=\alpha_s} \right) = 1 - \frac{W_g\left(\mathcal{P}_G^{\alpha=\alpha_s} \right)}{W_g\left(\mathcal{P}_1 \right)}$$

end for
Choose $\alpha^* \in \mathcal{U}$ (Chavent et al., 2018):

$$\alpha^* = \min_{\alpha} \left[Q_t(\alpha) - Q_g(\alpha) \right]$$

Given $K = N$, start with an initial partition with singletons $\mathcal{P}_N$;
for $K = N - 1, \ldots, 2$ **do**
   Aggregate two merged clusters $\mathcal{A} \cup \mathcal{B}$ with a cluster $\mathcal{C}$ by means of the modified (Lance & Williams, 1967) equation

$$\begin{aligned}\delta(\mathcal{A} \cup \mathcal{B}, \mathcal{C}) =\ & \frac{N_{\mathcal{A}} + N_{\mathcal{C}}}{N_{\mathcal{A}} + N_{\mathcal{B}} + N_{\mathcal{C}}} \delta(\mathcal{A}, \mathcal{C})^{\alpha^*} \\ & + \frac{N_{\mathcal{B}} + N_{\mathcal{C}}}{N_{\mathcal{A}} + N_{\mathcal{B}} + N_{\mathcal{C}}} \delta(\mathcal{B}, \mathcal{C})^{\alpha^*} \\ & - \frac{N_{\mathcal{C}}}{N_{\mathcal{A}} + N_{\mathcal{B}} + N_{\mathcal{C}}} \delta(\mathcal{A}, \mathcal{B})^{\alpha^*}\end{aligned}$$

end for
Stop when $K = 1$, thus the partition is $\mathcal{P}_1$.

CRediT authorship contribution statement

Raffaele Mattera: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Philip Hans Franses:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The replication files (data and R codes) can be accessed by the GitHub repository at the following link: https://github.com/raffmattera/stcfm/.

Acknowledgments

References

Aastveit, K. A., Bjørnland, H. C., & Thorsrud, L. A. (2016). The world is not enough! small open economies and regional dependence. *Scandinavian Journal of Economics*, *118*(1), 168–195.

Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, *81*(3), 1203–1227.

Alonso, A. M., Galeano, P., & Peña, D. (2020). A robust procedure to build dynamic factor models with cluster structure. *Journal of Econometrics*, *216*(1), 35–52.

Alonso, A. M., & Peña, D. (2019). Clustering time series by linear dependency. *Statistics and Computing*, *29*(4), 655–676.

Ando, T., & Bai, J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, *112*(519), 1182–1198.

Aquaro, M., Bailey, N., & Pesaran, M. H. (2021). Estimation and inference for spatial models with heterogeneous coefficients: an application to US house prices. *Journal of Applied Econometrics*, *36*(1), 18–44.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, *71*(1), 135–171.

Bai, Y., Carriero, A., Clark, T. E., & Marcellino, M. (2022). Macroeconomic forecasting in a multi-country context. *Journal of Applied Econometrics*, *37*(6), 1230–1255.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Bailey, N., Holly, S., & Pesaran, M. H. (2016). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*, *31*(1), 249–280.

Beck, G. W., Hubrich, K., & Marcellino, M. (2009). Regional inflation dynamics within and across euro area countries and a comparison with the united states. *Economic Policy*, *24*(57), 142–184.

Blatt, D., Chaudhuri, K., & Manner, H. (2023). A changepoint analysis of UK house price spillovers. *Regional Studies*, *57*(7), 1223–1238.

Brady, R. R. (2014). The spatial diffusion of regional housing prices across US states. *Regional Science and Urban Economics*, *46*, 150–166.

Breitung, J., & Eickmeier, S. (2014). Analyzing business and financial cycles using multi-level factor models. Deutsche Bundesbank Discussion Paper.

Brillinger, D. (2001). Time series: Data analysis and theory. In *Classics in applied mathematics*. USA, Philadelphia: Society for Industrial and Applied Mathematics.

Bucci, A., Ippoliti, L., Valentini, P., et al. (2023). Analysing spatiotemporal patterns of covid-19 confirmed deaths at the nuts-2 regional level. *Regional Statistics*, *13*(2), 214–239.

Chavent, M., Kuentz-Simonet, V., Labenne, A., & Saracco, J. (2018). Clustgeo: an r package for hierarchical clustering with spatial constraints. *Computational Statistics*, *33*(4), 1799–1822.

Ciccarelli, C., & Elhorst, J. P. (2018). A dynamic spatial econometric diffusion model with common factors: The rise and spread of cigarette consumption in Italy. *Regional Science and Urban Economics*, *72*, 131–142.

Cipollini, A., & Parla, F. (2020). Housing market shocks in Italy: A gvar approach. *Journal of Housing Economics*, *50*, Article 101707.

Clark, T. E., & West, K. D. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, *135*(1–2), 155–186.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291–311.

Das, S., Gupta, R., & Kabundi, A. (2011). Forecasting regional house price inflation: a comparison between dynamic factor models and vector autoregressive models. *Journal of Forecasting*, *30*(2), 288–302.

Dées, S., & Güntner, J. (2017). Forecasting inflation across euro area countries and sectors: A panel var approach. *Journal of Forecasting*, *36*(4), 431–453.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.

D'Urso, P. (2000). Dissimilarity measures for time trajectories. *Journal of the Italian Statistical Society*, *9*, 53–83.

Eickmeier, S., Gambacorta, L., & Hofmann, B. (2014). Understanding global liquidity. *European Economic Review*, *68*, 1–18.

Elhorst, J. P. (2013). *Spatial econometrics: from cross-sectional data to spatial panels*. Germany, Heidelberg: Springer.

Emiris, M. (2016). A dynamic factor model for forecasting house prices in belgium. National Bank of Belgium Working Paper (313).

Feldkircher, M., Huber, F., & Pfarrhofer, M. (2020). Factor augmented vector autoregressions, panel vars, and global vars. (pp. 65–93). Macroeconomic Forecasting in the Era of Big Data: Theory and Practice.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, *82*(4), 540–554.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*(471), 830–840.

Fouedjio, F. (2016). A hierarchical clustering method for multivariate geostatistical data. *Spatial Statistics*, *18*, 333–351.

Franses, P. H., & Wiemann, T. (2020). Intertemporal similarity of economic time series: An application of dynamic time warping. *Computational Economics*, *56*(1), 59–75.

Gamerman, D., Ippoliti, L., & Valentini, P. (2022). A dynamic structural equation approach to estimate the short-term effects of air pollution on human health. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *71*(3), 739–769.

Han, C., Phillips, P. C., & Sul, D. (2017). Lag length selection in panel autoregression. *Econometric Reviews*, *36*(1–3), 225–240.

Holly, S., Pesaran, M. H., & Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, *158*(1), 160–173.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

Ippoliti, L., Valentini, P., & Gamerman, D. (2012). Space–time modelling of coupled spatiotemporal environmental variables. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *61*(2), 175–200.

Kim, Y. S., & Rous, J. J. (2012). House price convergence: Evidence from us state and metropolitan area panels. *Journal of Housing Economics*, *21*(2), 169–186.

Kuethe, T. H., & Pede, V. O. (2011). Regional housing price cycles: A spatio-temporal analysis using us state-level data. *Regional Studies*, *45*(5), 563–574.

Lopes, H. F., Salazar, E., & Gamerman, D. (2008). Spatial dynamic factor analysis. *Bayesian Analysis*, *3*(4), 759–792.

Maharaj, E. A., D'Urso, P., & Caiado, J. (2019). *Time series clustering and classification*. USA, Boca Raton: Chapman and Hall/CRC.

Mattera, R. (2022). A weighted approach for spatio-temporal clustering of covid-19 spread in Italy. *Spatial and Spatio-temporal Epidemiology*, *41*, Article 100500.

Mattera, R., & Franses, P. H. (2023). Are african business cycles synchronized? evidence from spatio-temporal modeling. *Economic Modelling*, *128*, Article 106485.

Moench, E., & Ng, S. (2011). A factor analysis of housing market dynamics in the US and the regions. *The Econometrics Journal*, *14*, C1–C24.

Mumtaz, H., & Surico, P. (2009). The transmission of international shocks: a factor-augmented var approach. *Journal of Money, Credit and Banking*, *41*, 71–100.

Otto, P., & Schmid, W. (2018). Spatiotemporal analysis of german real-estate prices. *The Annals of Regional Science*, *60*(1), 41–72.

Peña, D., & Yohai, V. J. (2016). Generalized dynamic principal components. *Journal of the American Statistical Association*, *111*(515), 1121–1131.

Piccolo, D. (1990). A distance measure for classifying arima models. *Journal of Time Series Analysis*, *11*(2), 153–164.

Rapach, D. E., & Strauss, J. K. (2009). Differences in housing price forecastability across us states. *International Journal of Forecasting*, *25*(2), 351–372.

Sigmund, M., & Ferstl, R. (2021). Panel vector autoregression in r with the package panelvar. *The Quarterly Review of Economics and Finance*, *80*, 693–720.

Smucler, E. (2019). Consistency of generalized dynamic principal components in dynamic factor models. *Statistics & Probability Letters*, *154*, Article 108536.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Valentini, P., Ippoliti, L., & Fontanella, L. (2013). Modeling us housing prices by spatial dynamic structural equation models. *The Annals of Applied Statistics*, *7*(2), 763–798.

Van Dijk, B., Franses, P. H., Paap, R., & Van Dijk, D. (2011). Modelling regional house prices. *Applied Economics*, *43*(17), 2097–2110.

Wang, P. (2010). *Large dimensional factor models with a multi-level factor structure: identification, estimation and inference*: Working paper, Department of Economics, HKUST.

Yang, C. F. (2021). Common factors and spatial dependence: An application to US house prices. *Econometric Reviews*, *40*(1), 14–50.

Further reading

Lance, G. N., & Williams, W. T. (1967). A general theory of classificatory sorting strategies: 1. hierarchical systems. *The Computer Journal*, *9*(4), 373–380.