Journal of Econometrics 158 (2010) 160–173

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A spatio-temporal model of house prices in the USA[☆]

Sean Holly[a], M. Hashem Pesaran[a,b,*], Takashi Yamagata[c]

[a] Faculty of Economics and CIMF, University of Cambridge, United Kingdom

[b] University of Southern California, United States

[c] Department of Economics and Related Studies, University of York, United Kingdom

ARTICLE INFO

Article history:
Available online 17 March 2010

JEL classification:
C21
C23

Keywords:
House prices
Cointegration
Cross-sectional dependence
Spatial dependence

ABSTRACT

This paper provides an empirical analysis of changes in real house prices in the USA using State level data. It examines the extent to which real house prices at the State level are driven by fundamentals such as real per capita disposable income, as well as by common shocks, and determines the speed of adjustment of real house prices to macroeconomic and local disturbances. We take explicit account of both cross-sectional dependence and heterogeneity.This allows us to find a cointegrating relationship between real house prices and real per capita incomes with coefficients $(1, -1)$, as predicted by the theory. We are also able to identify a significant negative effect for a net borrowing cost variable, and a significant positive effect for the State level population growth on changes in real house prices. Using this model we then examine the role of spatial factors, in particular, the effect of contiguous states by use of a weighting matrix. We are able to identify a significant spatial effect, even after controlling for State specific real incomes, and allowing for a number of unobserved common factors. We do, however, find evidence of departures from long run equilibrium in the housing markets in a number of States notably California, New York, Massachusetts, and to a lesser extent Connecticut, Rhode Island, Oregon and Washington State.

1. Introduction

Recent developments in the housing markets in the USA and elsewhere have once again highlighted the importance of large changes in house prices for the functioning of credit and money markets (International Monetary Fund, 2004). Changes in housing wealth also play an important role in household behaviour with real implications for output and employment. There is also the possibility of bubbles in house prices with prices moving well away from their fundamental drivers, such as household disposable income (Case and Shiller, 2003; McCarthy and Peach, 2004). This in turn raises the issue of whether there is cointegration between real house prices and real per capita disposable incomes. The evidence on this is mixed.

Using US national-level data, Meen (2002) and Gallin (2006) do not find strong evidence of a cointegrating relationship, possibly because of the short time span of the data they consider. To cope with this problem, other studies use panel data. Malpezzi (1999) uses panel data on 133 metropolitan areas in the USA over 18 years from 1979 to 1996 and he is able to reject the null of a unit root in the residuals of the regressions of real house prices on real per capita incomes, using the panel unit root test of Levin et al. (2002, LLC). However, the LLC test does not take account of possible cross-sectional dependence of house prices and this could bias the test results. Capozza et al. (2002) recognize this problem and try to control for cross-sectional dependence by adding time dummies to their error correction specifications. However, as Gallin (2006) points out, local housing market shocks are likely to be correlated in ways that are not captured by simple time effects. To allow for more general error cross-sectional dependence, Gallin (2006) adopts a bootstrap version of Pedroni's 1999 residual-based cointegration test procedure, originally advanced in Maddala and Wu (1999), but fails to reject the hypothesis of no cointegration. However, his bootstrap approach is likely to be biased when the cross section dimension (N) is much larger than the time series dimension (T), as in Gallin's application.

In this paper, using recently developed econometric techniques for the analysis of heterogeneous dynamic panels subject to cross-sectional dependence, we study the determination of real house prices in a panel of 49 US States over 29 years. We examine the extent to which real house prices at the State level are driven by fundamentals such as real per capita disposable income, as well as by common shocks, and determine the speed of adjustment of real house prices to macroeconomic and local disturbances.

☆ We would like to thank Jörg Breitung, Peter Burridge, Bernie Fingleton, David Hendry, Andreas Pick, Ron Smith, Elisa Tosetti and Joakim Westerlund for helpful comments. We have also benefited from constructive comments and suggestions from two anonymous referees and the editors. Financial support from the ESRC under Grant No. RES-000-23-0135 is gratefully acknowledged.

* Corresponding address: Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, United Kingdom.

E-mail address: mhp1@econ.cam.ac.uk (M.H. Pesaran).

0304-4076/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2010.03.040

There are considerable differences across US States in both the level and rates of growth of real incomes.[1] This heterogeneity should in turn be reflected in real house prices. The importance of heterogeneity in spatially distributed housing markets has been highlighted by Fratantoni and Schuh (2003). They quantify the importance of spatial heterogeneity in US housing markets for the efficacy of monetary policy. Depending on local conditions monetary policy can have differing effects on particular US regions (Carlino and DeFina, 1998). However, there are significant dependences in house prices and real incomes across States that cannot be captured by spatial effects alone. For example, Pollakowski and Ray (1997), using vector autoregressive (VAR) models, show that at the national level (dividing the USA into nine regions) there are significant non-spatial diffusion patterns.

Real house prices can vary between States because real incomes differ, but they can also differ because of scarcity of land or other idiosyncratic factors. The effects of common shocks on house prices whether observed, such as changes in interest rates and oil prices, or unobserved, such as technological change, could also differ across States. We take account of these influences by making use of the common correlated effects (CCE) estimator of Pesaran (2006) which is consistent under heterogeneity and cross-sectional dependence. The CCE estimator can be computed by ordinary least squares (OLS), applied to an auxiliary regression where the observed regressors are augmented by (weighted) cross-sectional averages of the dependent variable and the individual specific regressors. Notably, CCE estimation allows for unobserved common factors to be possibly correlated with exogenously given State-specific regressors and it is invariant to the (unknown but fixed) number of unobserved common factors as N and T tend to infinity (jointly). These features are not shared with other approaches in Banerjee and Carrion-i-Silvestre (2006), Groen and Kleibergen (2003), Nelson et al. (2005), Westerlund (2005), Pedroni and Vogelsang (2005), Chang (2005), and Bai and Kao (2006).

The CCE procedure also copes with the presence of spatial effects (Pesaran and Tosetti, 2010). This is because spatial dependence is dominated by the common factor error structure that underlies the CCE estimators. But once the model parameters are estimated, the importance of spatial effects can be ascertained by fitting a spatial model to the residuals from the panel. It is also worth noting that the CCE procedure is robust to the choice of the spatial model, although for estimation of spatial effects some parametric formulation would be needed. In this paper we estimate a spatial autoregressive (SAR) error model and show that spatial effects are indeed statistically highly significant in the analysis of house prices in the USA.

Finally, to test for cointegration between real house prices and real disposable incomes, we apply the panel unit root tests of Moon and Perron (2004) and Pesaran (2007) to the log price–income ratio which allow for cross-sectional dependence.

The remainder of the paper is organized as follows. Section 2 discusses the theory underlying house price determination. Section 3 provides a review of the panel data model and estimation methods. Section 4 provides a preliminary data analysis. Section 5 reports the estimation results. Section 6 provides some concluding remarks.

2. Modelling house prices

It is now standard to see the determination of house prices as the outcome of a market for the services of the housing stock and as an asset. Demand for housing can be met either through rental of a residential property or by owner occupation. The expected net benefit from owner occupation needs to be set against the rental cost of the same property. Denote the real house price at the start of period t by P_t, and the real rental cost of the same house over the period t by R_t. Then the net benefit of owning the house over the period t to $t+1$ is given by $P_{t+1} - P_t(1 + r_t) + S_t$, where $r_t > 0$ is the real rate of interest, and S_t is the real value of housing services. This expression abstracts from transaction costs, depreciation, and other costs of home ownership. These can be readily incorporated into the analysis without affecting the long run relationship between real house prices and incomes that is the focus of our empirical analysis.

For a risk neutral household the one period arbitrage condition for the asset market equilibrium in real house prices is given by[2]

$$E\left(P_{t+1}|\mathcal{F}_t\right) - P_t(1 + r_t) + S_t = R_t,$$

or

$$P_t = \left(\frac{1}{1 + r_t}\right)\left[E\left(P_{t+1}|\mathcal{F}_t\right) + S_t - R_t\right],$$

where $\mathcal{F}_t$ is the information set available at time t.[3] To complete the model we shall assume that R_t cannot exceed household's real disposable income, Y_t, and represent this relationship by

$$R_t = \alpha_t Y_t, \quad 0 < \alpha_t < 1,$$

where α_t is assumed to follow a stationary process. We shall also assume that

$$S_t = \beta_t^{-1} R_t, \quad 0 < \beta_t < 1,$$

which ensures positive real house prices in all periods.[4] Under these assumptions

$$P_t = \left(\frac{1}{1 + r_t}\right)\left[E\left(P_{t+1}|\mathcal{F}_t\right) + \theta_t Y_t\right], \tag{2.1}$$

where

$$\theta_t = \frac{\alpha_t(1 - \beta_t)}{\beta_t} > 0.$$

It is also reasonable to assume that θ_t, the fraction of income allocated to net housing services, $S_t - R_t$, is stationary.

Accordingly, under rational expectations and assuming that r_t is sufficiently large relative to the growth of real disposable income, $g_t = \Delta \ln(Y_t)$, bubble-free real house prices will be given as the discounted stream of future net housing services, $S_t - R_t = \theta_t Y_t$. The solution simplifies considerably under $r_t = r$,

$$P_t = \sum_{j=0}^{\infty}\left(\frac{1}{1 + r}\right)^j E\left(\theta_{t+j} Y_{t+j}|\mathcal{F}_t\right),$$

which can be written equivalently as

$$\frac{P_t}{Y_t} = \sum_{j=0}^{\infty} E\left(\theta_{t+j}\prod_{s=1}^{j}\left(\frac{1 + g_{t+s}}{1 + r}\right)\middle|\,\mathcal{F}_t\right).$$

Therefore, under fairly general assumptions regarding the processes generating g_t and θ_t, the price–income ratio, P_t/Y_t, would also be stationary. In particular, $p_t = \ln(P_t)$ will be cointegrated with $y_t = \ln(Y_t)$ with the cointegrating vector given by $(1, -1)$, if y_t is an integrated variable of order 1. For example, if θ_t and g_t are

[1] For a recent review of the USA housing market see Green and Malpezzi (2003).

[2] Feldstein et al. (1978), Hendershott and Hu (1981) and Buckley and Ermisch (1982).

[3] Alternatively, expectations can be taken under the risk-neutral measure which does not necessarily require households to be individually risk neutral, although it does imply that at the aggregate households are treated as if they are risk neutral.

[4] The condition $0 < \beta_t < 1$ is sufficient but not necessary for $P_t > 0$ for all t.

independently distributed with $g_t = g + \varepsilon_{gt}$, $\varepsilon_{gt} \sim$ i.i.d. $(0, \sigma_g^2)$, $g < r$, and $\theta_t = \theta + \varepsilon_{\theta t}$, $\varepsilon_{\theta t} \sim (0, \sigma_\theta^2)$, we have

$$\frac{P_t}{Y_t} = \frac{\theta(1+r)}{r-g}. \tag{2.2}$$

In this simple case, the price–income ratio, also known as the affordability index, is non-stochastic and rises with θ and g, and falls with r. However, the elasticity of real house prices to real income does not depend on these parameters.[5]

The long run relationship between y_t and p_t holds more generally even if r_t is time varying, so long as it is sufficiently large relative to g_t such that the possibility of real house price bubbles can be ruled out *a priori*. Note that (2.1) can be written equivalently as

$$\frac{P_t}{Y_t} = \frac{\theta_t}{r_t - E\left(\Delta \ln P_{t+1} | \mathcal{F}_t\right)}. \tag{2.3}$$

It is clear that for a meaningful solution, the expected rate of real house price appreciation, $E(\Delta \ln P_{t+1}|\mathcal{F}_t)$, must be less than r_t. Also by the Cauchy–Schwarz inequality we have

$$E\left|\frac{P_t}{Y_t}\right| < \left[E(\theta_t^2)\right]^{1/2} \left\{E\left[\frac{1}{\left[r_t - E\left(\Delta \ln P_{t+1}|\mathcal{F}_t\right)\right]^2}\right]\right\}^{1/2}.$$

Since θ_t is a stationary process then $E(\theta_t^2) < K < \infty$. Therefore, P_t/Y_t will be uniformly integrable if

$$E\left[\frac{1}{\left[r_t - E\left(\Delta \ln P_{t+1}|\mathcal{F}_t\right)\right]^2}\right] < K < \infty.$$

Namely, so long as expected real house price inflation is sufficiently small relative to the real interest rate, the price income ratio cannot contain a unit root even though P_t and Y_t might.

In addition to real incomes and real interest rates, other factors such as the State level unemployment rate, changes in demographics, taxation and credit conditions are also likely to play a role in the determination of real house prices at the State level. Unfortunately many of these factors are either unobserved or are rather difficult to measure accurately. Fixed effects for each State will pick up State specific factors such as climate, location and culture. In this paper we also attempt to account for other, often short-term influences, by unobserved factors which we then proxy by cross State averages of the observables. Amongst observable short-term factors, we shall also consider the possible effect of State level population growth rates on real house prices. In aggregate time series analysis, it is difficult to identify the effects of slowly moving variables such as population growth on real house prices. But in a panel context the cross section dimension can be used to identify such effects. For a given level of real per capita disposable income, we would expect real house prices to be higher in States with a higher rate of population growth.

2.1. The spatial dimension

The model of the previous discussion abstracts from the spatial dimension. However, it is possible that States that are contiguous may influence each other's house prices. High prices in metropolitan areas may persuade people to commute from neighbouring States. Labour mobility compared to Europe is much higher in the USA and lower house prices may provide an incentive to migrate.

The development of spatial econometrics (Paelinck and Klaasen, 1979; Anselin, 1988; Krugman, 1998) has been spurred by a new interest in the role of space in economic processes, with a particular emphasis placed on interactions in space (spatial autocorrelation) and spatial structures (spatial heterogeneity). In spatial econometrics, the degree of cross-sectional dependence is typically calibrated by means of a weighting matrix. For example the (i, j) elements of a weighting matrix, s_{ij}, could take a value of 1 if the ith and jth areas/regions/countries are contiguous and zero otherwise.[6] Weights can also be based on physical distance, or even on other types of metrics, such as economic (Conley, 1999), or social distance (Conley and Topa, 2002). The class of cross-sectional dependence that typically is assumed in spatial econometrics, such as spatial moving average and/or spatial autoregressive models as introduced by Whittle (1954), Cliff and Ord (1973, 1981), and Haining (1978), all represent examples of weak forms of cross-sectional dependence, in the sense that the degree of dependence decreases sufficiently quickly as the distance between units increases.

Another possible source of cross-sectional dependence would be due to economy-wide common shocks that affect all cross section units. Changes in interest rates, oil prices, and technology are examples of such common shocks that may affect house prices, but with different degrees across States. This strong form of cross-sectional dependence is pervasive and unlike spatial dependence does not diminish along the space dimension.[7]

It is clear that for the analysis of house prices across States, we need a sufficiently general and flexible econometric model where both forms of cross-sectional dependence can be accommodated. This issue is considered next.

3. The econometric model and tests

From the model of Section 2 that suggested a cointegrating and proportional relationship between real house prices and incomes, as well as a role for the real interest rate and demographic shifts, we can write a panel model for US States as a long-run relation compatible with the long run theory in the following log-linear form:

$$p_{it} = \alpha_i + \beta_{i1} y_{it} + \beta_{i2} g_{i,t-1} + \beta_{i3} c_{i,t-1} + u_{it},$$
$$i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T, \tag{3.1}$$

where p_{it} is the logarithm of the real price of housing in the ith State during year t, and y_{it} is real per capita personal disposable income. We follow the macroeconomic literature and assume that real per capita income is best characterized by a unit root process with a drift, and based on the general theory developed in the previous section, we would expect p_{it} and y_{it} to be cointegrated with coefficients $(1, -1)$. The short-run effects, such as price dynamics, and their adjustments to the long-run equilibrium across States can be accommodated through the error terms, u_{it}. It is also possible to augment the real house price equation with observable short-term effects of variables such as changes in demographic factors, g_{it}, and the net cost of borrowing defined by $c_{it} = r_{it} - \Delta p_{it}$, where r_{it} represents the long-term real interest rate and g_{it} represents the rate of change of population in the ith State. The net borrowing cost variable can be viewed as a proxy for the denominator of (2.3), which is included in (3.1) with a lag to avoid simultaneity. *A priori* we would expect a rise in c_{it} to

[5] For some recent evidence on the relationship between real house prices and real incomes for a wide range of countries, see Almeida et al. (2006).

[6] See Moran (1948), Cliff and Ord (1973, 1981), Anselin (1988, 2001) and Haining (2003).

[7] Different forms of cross-sectional dependence are discussed and formally defined in Pesaran and Tosetti (2010).

be associated with a fall in the price–income ratio, and hence a negative coefficient for $c_{i,t-1}$ in (3.1). The effect of population growth on real house prices is expected to be positive ($\beta_{i3} > 0$).

As is clear from Eq. (3.1), the parameter vector of the slope coefficients, $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \beta_{i3})'$, is allowed to be heterogeneous across states. In order to assess the overall effects of covariates, in this paper we shall focus mainly on the estimation of the average value of $\boldsymbol{\beta}_i$, namely $E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$, assuming a random coefficient model, $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\varpi}_i$, where $\boldsymbol{\varpi}_i \sim IID(\mathbf{0}, \mathbf{V}_{\varpi})$.

We shall assume that u_{it} has the following multi-factor structure

$$u_{it} = \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it}, \tag{3.2}$$

in which $\mathbf{f}_t$ is an $m \times 1$ vector of unobserved common shocks (or factors), and ε_{it} are the individual-specific (idiosyncratic) errors assumed to be distributed independently of y_{it} and $\mathbf{f}_t$. However, we allow ε_{it} to be weakly dependent across i, and serially correlated over time. The pattern of serial correlation in ε_{it} could vary across i, (Pesaran, 2006). The common factors, $\mathbf{f}_t$, can also be serially correlated and possibly correlated with y_{it}, c_{it} and g_{it}. Furthermore, $\mathbf{f}_t$ is allowed to be stationary or nonstationary (Kapetanios et al., 2009).

Despite its simplicity the above specification is reasonably general and flexible and allows us to consider a number of different factors that drive house prices. In particular, some of the supply factors that are difficult to measure accurately can be captured through the unobserved common components of u_{it}.[8] The spatial aspect of house price formation can be accommodated through the assumed weak dependence of the idiosyncratic errors, ε_{it}.

In this way we are able to test for cointegration between real house prices and real disposable income, whilst allowing for both forms of cross dependence, weak and strong, at the State level.

3.1. The common correlated effects (CCE) estimator

We use the common correlated effects (CCE) type estimator, which asymptotically eliminates strong as well as weak forms of cross section dependence in large panels (Pesaran, 2006). To illustrate, suppose the $k \times 1$ ($k = 3$) vector $\mathbf{x}_{it} = (y_{it}, c_{i,t-1}, g_{i,t-1})'$ is generated as

$$\mathbf{x}_{it} = \mathbf{a}_i + \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it}, \tag{3.3}$$

where $\mathbf{a}_i$ is a $k \times 1$ vector of individual effects, $\boldsymbol{\Gamma}_i$ are $m \times k$ factor loading matrices with fixed components, $\mathbf{v}_{it}$ are the specific components of $\mathbf{x}_{it}$ distributed independently of the common effects and across i, but assumed to follow general covariance stationary processes. Combining (3.1)–(3.3), we now have

$$\underset{(k+1)\times 1}{\mathbf{z}_{it}} = \begin{pmatrix} p_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \underset{(k+1)\times 1}{\mathbf{d}_i} + \underset{(k+1)\times m}{\mathbf{C}_i'} \underset{m\times 1}{\mathbf{f}_t} + \underset{(k+1)\times 1}{\boldsymbol{\upsilon}_{it}}, \tag{3.4}$$

where

$$\boldsymbol{\upsilon}_{it} = \begin{pmatrix} \varepsilon_{it} + \boldsymbol{\beta}_i'\mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix}, \tag{3.5}$$

$$\mathbf{d}_i = \begin{pmatrix} 1 & \boldsymbol{\beta}_i' \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \begin{pmatrix} \alpha_i \\ \mathbf{a}_i \end{pmatrix}, \qquad \mathbf{C}_i = \begin{pmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Gamma}_i \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \tag{3.6}$$

$\mathbf{I}_k$ is an identity matrix of order k, and the rank of $\mathbf{C}_i$ is determined by the rank of the $m \times (k + 1)$ matrix of the unobserved factor loadings $\tilde{\boldsymbol{\Gamma}}_i = (\boldsymbol{\gamma}_i \quad \boldsymbol{\Gamma}_i)$. Pesaran (2006) has suggested using cross section averages of p_{it} and $\mathbf{x}_{it}$ as proxies for the unobserved factors in (3.1). To see why such an approach could work, consider simple cross section averages of the equations in (3.4)[9]

$$\bar{\mathbf{z}}_t = \bar{\mathbf{d}} + \bar{\mathbf{C}}'\mathbf{f}_t + \bar{\boldsymbol{\upsilon}}_t, \tag{3.7}$$

where $\bar{\mathbf{z}}_t = N^{-1}\sum_{i=1}^N \mathbf{z}_{it}$, $\bar{\boldsymbol{\upsilon}}_t = N^{-1}\sum_{i=1}^N \boldsymbol{\upsilon}_{it}$, $\bar{\mathbf{d}} = N^{-1}\sum_{i=1}^N \mathbf{d}_i$, and $\bar{\mathbf{C}} = N^{-1}\sum_{i=1}^N \mathbf{C}_i$. Suppose that $\text{Rank}(\bar{\mathbf{C}}) = m \le k+1$ for all N, so that $\mathbf{f}_t = (\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\bar{\mathbf{C}}\left(\bar{\mathbf{z}}_t - \bar{\mathbf{d}} - \bar{\boldsymbol{\upsilon}}_t\right)$. Therefore, if $\bar{\mathbf{C}} \xrightarrow{p} \mathbf{C}$ and $\bar{\boldsymbol{\upsilon}}_t \to \mathbf{0}$ as $N \to \infty$,

$$\mathbf{f}_t - \left(\mathbf{C}\mathbf{C}'\right)^{-1}\mathbf{C}\left(\bar{\mathbf{z}}_t - \bar{\mathbf{d}}\right) \xrightarrow{p} \mathbf{0}, \quad \text{as } N \to \infty. \tag{3.8}$$

This suggests using $(1, \bar{\mathbf{z}}_t')'$ as observable proxies for $\mathbf{f}_t$, and is the basic insight that lies behind the common correlated effects (CCE) estimators developed in Pesaran (2006). Kapetanios et al. (2009) prove that the CCE estimators are consistent regardless of whether the common factors, $\mathbf{f}_t$, are stationary or nonstationary. It is further shown that the CCE estimation procedure in fact holds even if $\tilde{\boldsymbol{\Gamma}} = E(\tilde{\boldsymbol{\Gamma}}_i)$ turns out to be rank deficient; thus, the estimator is consistent with any fixed number of m. This contrasts with the principal component approach, which requires us to estimate the number of factors (Bai and Ng, 2002; Bai, 2003). In addition, as shown by Pesaran and Tosetti (2010), under certain regularity conditions,[10] CCE estimators are consistent even when the idiosyncratic errors, ε_{it}, are weakly cross-sectionally dependent.

Following Pesaran (2006), we focus mainly on two estimators of the mean value of $\boldsymbol{\beta}_i$, namely $E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$. First, the CCE mean group estimator (CCEMG) is a simple average of the individual CCE estimators, $\hat{\mathbf{b}}_i$ of $\boldsymbol{\beta}_i$ defined by

$$\hat{\mathbf{b}}_{\text{CCEMG}} = N^{-1}\sum_{i=1}^N \hat{\mathbf{b}}_i, \tag{3.9}$$

$$\hat{\mathbf{b}}_i = (\mathbf{X}_i'\bar{\mathbf{M}}\mathbf{X}_i)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}\mathbf{p}_i, \tag{3.10}$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})'$, $\mathbf{p}_i = (p_{i1}, p_{i2}, \ldots, p_{iT})'$, $\bar{\mathbf{M}} = \mathbf{I}_T - \bar{\mathbf{H}}\left(\bar{\mathbf{H}}'\bar{\mathbf{H}}\right)^{-1}\bar{\mathbf{H}}'$ with $\bar{\mathbf{H}} = (\boldsymbol{\tau}_T, \bar{\mathbf{Z}})$, where $\boldsymbol{\tau}_T$ is a $T \times 1$ vector of unity and $\bar{\mathbf{Z}}$ is a $T \times (k + 1)$ matrix of observations $\bar{\mathbf{z}}_t$. The (non-parametric) variance estimator for $\hat{\mathbf{b}}_{\text{CCEMG}}$ is given by

$$\widehat{\text{Var}}\left(\hat{\mathbf{b}}_{\text{CCEMG}}\right) = \frac{1}{N(N-1)}\sum_{i=1}^N \left(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{\text{CCEMG}}\right)\left(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{\text{CCEMG}}\right)'.$$

Second, when the individual slope coefficients, $\boldsymbol{\beta}_i$, are the same, efficiency gains from pooling of observations over the cross section units can be achieved. Such a pooled estimator of $\boldsymbol{\beta}$, denoted by CCEP, has been developed by Pesaran (2006) and is given by

$$\hat{\mathbf{b}}_{\text{CCEP}} = \left(\sum_{i=1}^N \mathbf{X}_i'\bar{\mathbf{M}}\mathbf{X}_i\right)^{-1}\sum_{i=1}^N \mathbf{X}_i'\bar{\mathbf{M}}\mathbf{p}_i. \tag{3.11}$$

The variance estimator for $\hat{\mathbf{b}}_{\text{CCEP}}$ is given by

$$\widehat{\text{Var}}\left(\hat{\mathbf{b}}_{\text{CCEP}}\right) = N^{-1}\hat{\boldsymbol{\Psi}}^{*-1}\hat{\mathbf{R}}^{*}\hat{\boldsymbol{\Psi}}^{*-1}, \tag{3.12}$$

8 The supply elasticity of housing units has recently been identified as an important factor behind house price movements in some US urban markets. The ease with which regulatory approval for the construction of new houses can be obtained has been identified by Glaeser and Gyourko (2005) and Glaeser et al. (2006) as a significant element in real house price increases in California, Massachusetts, New Hampshire, New Jersey and Washington, DC.

9 Pesaran (2006) considers cross section weighted averages that are more general.

10 It is sufficient that the row and column sums of the spatial weight matrix are suitably bounded in the cross section dimension. See Pesaran and Tosetti (2010).

where $\hat{\boldsymbol{\Psi}}^* = N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\bar{\mathbf{M}}\mathbf{X}_i/T$,

$$\hat{\mathbf{R}}^* = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}\mathbf{X}_i}{T}\right)\left(\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{\text{CCEMG}}\right)\left(\hat{\mathbf{b}}_i-\hat{\mathbf{b}}_{\text{CCEMG}}\right)' \times\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}\mathbf{X}_i}{T}\right). \tag{3.13}$$

In this paper we propose goodness of fit statistics based on CCEMG and CCEP estimators, denoted by $\bar{R}^2_{\text{CCEMG}}$ and $\bar{R}^2_{\text{CCEP}}$, respectively. $\bar{R}^2_{\text{CCEMG}}$ is given by

$$\bar{R}^2_{\text{CCEMG}} = 1-\bar{\bar{\sigma}}^2/\hat{\sigma}^2, \tag{3.14}$$

where $\bar{\bar{\sigma}}^2 = N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2$, with $\hat{\sigma}_i^2 = (\mathbf{p}_i-\mathbf{X}_i\hat{\mathbf{b}}_i)'\bar{\mathbf{M}}(\mathbf{p}_i-\mathbf{X}_i\hat{\mathbf{b}}_i)/(T-2k-2)$, and $\hat{\sigma}^2$ is defined by $\hat{\sigma}^2 = N^{-1}(T-1)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(p_{it}-\bar{p}_i)^2$, with $\bar{p}_i = T^{-1}\sum_{t=1}^{T}p_{it}$. Similarly, $\bar{R}^2_{\text{CCEP}}$ is defined as

$$\bar{R}^2_{\text{CCEP}} = 1-\hat{\sigma}^2_{\text{CCEP}}/\hat{\sigma}^2, \tag{3.15}$$

where $\hat{\sigma}^2_{\text{CCEP}}$ is the error variance estimator

$$\hat{\sigma}^2_{\text{CCEP}} = \sum_{i=1}^{N}\left(\mathbf{p}_i-\mathbf{X}_i\hat{\mathbf{b}}_{\text{CCEP}}\right)'\bar{\mathbf{M}}\left(\mathbf{p}_i-\mathbf{X}_i\hat{\mathbf{b}}_{\text{CCEP}}\right)/\left[N(T-k-2)-k\right].$$

3.2. A cross-section dependence test

In this paper we use a CD (cross-section dependence) test of error cross dependence, which does not require an *a priori* specification of a connection (weighting) matrix and is applicable to a variety of panel data models, including stationary and unit root dynamic heterogeneous panels with structural breaks, with short T and large N (Pesaran, 2004).[11] The CD test is based on an average of the pair-wise correlations of the OLS residuals from the individual regressions in the panel and tends to a standard normal distribution as $N\to\infty$. The CD test statistic is defined as

$$\text{CD} = \sqrt{\frac{2T}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{ij}\right)\overset{a}{\sim}N(0,1), \tag{3.16}$$

where $\hat{\rho}_{ij}$ is the sample estimate of the pair-wise correlation of the residuals. Specifically,

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt}}{\left(\sum_{t=1}^{T}\hat{u}_{it}^2\right)^{1/2}\left(\sum_{t=1}^{T}\hat{u}_{jt}^2\right)^{1/2}}, \tag{3.17}$$

where $\hat{u}_{it}$ is the OLS estimate of u_{it} defined by

$$\hat{u}_{it} = p_{it}-\hat{\alpha}_i-\hat{\boldsymbol{\beta}}_i'\mathbf{x}_{it}, \tag{3.18}$$

with $\hat{\alpha}_i$ and $\hat{\boldsymbol{\beta}}_i$ being the estimates of α_i and $\boldsymbol{\beta}_i$ computed using the OLS regression of p_{it} on an intercept and the regressors, $\mathbf{x}_{it}$, for each i, separately.

3.3. Panel unit root tests

One of the most commonly used tests for unit roots in panels is that of Im et al. (2003), called the IPS test. However, the IPS test procedure is not valid when the series are cross-sectionally dependent, and its use in the case of the house price data can lead to spurious inference. A number of panel unit root tests that allow for possible cross-sectional dependence in panels have been recently proposed in the literature.[12] Here we consider the simple test proposed by Pesaran (2007), which follows the CCE approach and filters out the cross-sectional dependence by augmenting the ADF regressions with cross section averages.[13]The panel unit root test proposed by Pesaran (known as the CIPS test) is based on cross section augmented ADF (CADF) regressions, carried out separately for each State, namely

$$\Delta\omega_{it} = a_{i0}+a_{i1}t+a_{i2}\omega_{i,t-1}+a_{i3}\bar{\omega}_{t-1}+\sum_{j=0}^{p}d_{ij}\Delta\bar{\omega}_{t-j} + \sum_{j=1}^{p}\delta_{ij}\Delta\omega_{i,t-j}+\text{v}_{it}, \tag{3.19}$$

where $\bar{\omega}_t$ denotes the cross section mean of ω_{it}. The CIPS statistic is a simple cross section average of $\tilde{t}_i$ defined by

$$\text{CIPS} = N^{-1}\sum_{i=1}^{N}\tilde{t}_i, \tag{3.20}$$

where $\tilde{t}_i$ is the OLS t-ratio of a_{i2} in the above CADF regression. The critical values for the CIPS tests are given in TablesII(a)–(c) in Pesaran (2007).

3.4. Panel cointegration tests

As noted in the introduction recently there has been a debate in the literature about whether there is cointegration between real house prices and real per capita disposable incomes. In the absence of cointegration there is no fundamentals driving real house prices so the possibility of bubbles is increased (Case and Shiller, 2003). So far the evidence is mixed.

Malpezzi (1999) uses panel data on 133 metropolitan areas in the US over 18 years from 1979 to 1996 and applies the panel unit root test of Levin et al. (2002, LLC) to house price-to-income ratios, and finds that he cannot reject the presence of a unit root in these series. But he is able to reject the null of a unit root in the residuals of the regressions of real house prices on real per capita incomes, again using the LLC panel unit root test.[14] However, the testing procedure adopted by Malpezzi suffers from two main shortcomings. The LLC's critical values are not appropriate when the panel unit root test is applied to residuals from first step regressions, and perhaps more importantly, the LLC test does not take account of possible cross-sectional dependence in the regression errors and this could bias the test results.

Capozza et al. (2002) recognize this problem and try to control for cross-sectional dependence by adding time dummies to their error correction specifications. However, as Gallin (2006) points out, local housing market shocks are likely to be correlated in ways that are not captured by simple time effects. To allow for more general error cross-sectional dependence, Gallin (2006) adopts a

[11] Frees (1995) also proposes tests based on average pair-wise sample correlations of the series across the different cross-section units. His R_{AVE} test statistic is based on Spearman rank correlations, and his C_{AVE} test statistic is based on Pearson rank correlations. The latter is closely related to the CD test also considered in Pesaran (2004).

[12] For a recent survey of the literature see Breitung and Pesaran (2008).

[13] We also considered the test proposed by Moon and Perron (2004) which is based on the t-ratio of a modified pooled OLS estimator using the de-factored panel data. However, the test is asymptotically valid only when $N/T\to 0$ as both N and T go to infinity, and this implies that T may have to be much larger than N, which does not seem appropriate for our application where $N=49$ and $T=29$.

[14] It is not clear if the panel unit root tests reported in Malpezzi are applied to the levels of price-to-income ratios or to their logarithms. See Eq. (2) and the discussions on pages 42 and 48 in Malpezzi (1999).

bootstrap version of Pedroni's 1999 residual-based cointegration test procedure and concludes that "… even these more powerful tests do not reject the hypothesis of no cointegration". However, the bootstrap approach, originally advanced in Maddala and Wu (1999), is likely to be biased. The bootstrap test statistic is not pivotal, and the bootstrap test has a finite sample error of the same order as the asymptotic test. Secondly, as Maddala and Wu (1999) show, the bootstrap procedure cannot eliminate size distortions in finite samples, particularly in cases where T is small relative to N. Also see Smith et al. (2004, p. 161–168) where they find that the bootstrap panel unit root test tends to under-reject when $N = T = 25$. They do not consider any experiments where $N > T$. Furthermore, their Monte Carlo set-up does not deal with common factor error structures, since they only consider designs where the maximum eigenvalue of the error variance–covariance matrix remains bounded in N.[15] In Gallin's application N (=95) is much larger than T (=23), and due to the presence of common factors the $N \times N$ error variance–covariance matrix is likely to be near singular. Therefore, the bootstrap panel unit root tests reported by Gallin can be subject to large size distortions.

Over the past few years a number of panel cointegration tests have been proposed in the literature that attempt to take account of error cross section dependence in their test procedures. These include the tests proposed by Bai and Kao (2006), Banerjee and Carrion-i-Silvestre (2006), Chang (2005), Gengenbach et al. (2006), Groen and Kleibergen (2003), Nelson et al. (2005), Pedroni and Vogelsang (2005) and Westerlund (2005). The tests by Groen and Kleibergen, Nelson, Ogaki and Sul, and Westerlund are applicable when N is small and T is large. For example, in their Monte Carlo experiments Groen and Kleibergen and Nelson, Ogaki, and Sul consider panels with $N \leq 8$ and $T \geq 100$. Westerlund considers panels where $N = 10$ or at most 20 and $T = 50$ or 100. The tests by Banerjee and Carrion-i-Silvestre, Chang, Pedroni and Vogelsang, in principle, can deal with panels where N is reasonably large, but they do not allow the unobserved common factors to be correlated with the observed regressors, which is an important consideration in our application. Bai and Kao allow for cross-sectional dependence using the factor approach as in (3.2), but they do not allow cross-sectional heterogeneity in the cointegrating vector. Gengenbach et al. propose a sequential procedure where in the first step unit root properties of the (extracted) common factors and idiosyncratic components are investigated, and depending on the outcomes non-cointegration of the common factors and/or the idiosyncratic components are then investigated. To deal with the joint nature of these tests, Gengenbach et al. suggest using the Bonferroni procedure, but based on Monte Carlo simulations they find the joint tests to be undersized. see Gengenbach et al. (2006, Remark 4).[16]

Following Banerjee and Carrion-i-Silvestre, Chang, Gengenbach et al., and Pedroni and Vogelsang, we adopt a two-stage procedure to assess the possibility of cointegration between the log of real house price (p_{it}) and the log of real per capita disposable income (y_{it}). But unlike these studies, in both stages we allow for unobserved common factors that could potentially be correlated with the observed regressors. In particular, using the pooled CCE estimator we first estimate the residuals, $\hat{u}_{it} = p_{it} - \hat{\beta}_{\text{CCEP}} y_{it} - \hat{\alpha}_i$. As noted earlier, the pooled estimate, $\hat{\beta}_{CCEP}$, is consistent for β under fairly general assumptions about the unobserved common factors, $\mathbf{f}_t$ (for example, irrespective of whether $\mathbf{f}_t$ is $I(0)$ or $I(1)$), and even if the slope coefficients are heterogenous. We then apply panel unit root tests to these residuals,[17] allowing for cross-sectional heterogeneity and cross-sectional dependence.[18] If the presence of a unit root in $\hat{u}_{it}$'s can be rejected, we shall conclude that the log of real house prices and the log of real per capita disposable incomes are cointegrated with the pooled cointegrating vector given by $(1, -\hat{\beta}_{CCE})'$.

4. Preliminary data analysis

We begin our empirical investigation with a preliminary analysis of spatial dependence at the USA State level, using data on the growth of real house prices and incomes. Table 1 defines the variables used. A more detailed description is provided in the Data Appendix. We use annual data on the US States, excluding Alaska and Hawaii, from 1975 to 2003. One of the features of the data in which we are interested is the extent to which real house prices are driven by fundamentals such as income, net cost of borrowing and population growth. To explore spatial interactions we calculate simple correlation coefficients between each State, within and between correlations for the Bureau of Economic Analysis (BEA) eight regions and finally the within and between correlation coefficients for three geographical regions dividing the USA into broadly the West, the Middle and the East (Table 2 shows the member States in each of these regions). The results are summarized in Tables 3 and 4.

In Table 3 we tabulate within and between correlation coefficients for the 8 BEA regions. The diagonal elements show the within region average correlation coefficients. The off-diagonal elements give the between region correlation coefficients. For many regions the within region correlation is larger than the between region correlation. But for some regions this is not so. For example, the States of the Mid-East region are more correlated on average with the States of New England than among themselves. The States of the Great Lakes are more correlated with those of the South East than they are among themselves. If we look at the correlations at the level of three geographical areas, the within correlations are always larger than the between, though the East tends to be 'closer' in some sense to the Middle than the Middle is to the West. Overall, real income growth is correlated across all States in the USA.

In Table 4 we tabulate the spatial correlations for real house prices. A similar picture to that for real incomes emerges. Within region correlations are generally larger than the between correlations, with the exception of New England and the Mid-East and the South West and the Rocky Mountains. In contrast to the results for real incomes there is a more noticeable spatial pattern. The growth of real house prices in New England is hardly correlated at all with States in the Rocky Mountains and the Far West, with the correlations on average declining with distance. This pattern is also clear when we look at the three broad geographical areas (The West, the Middle and the East).

The regional groupings also disguise some interesting correlations at the underlying State level. To save space, the State level correlation coefficients for real income growth and real house price

[15] See, for example, Chamberlain and Rothschild (1983) who show that in the case of factor models in N variables and a fixed number of factors m, the largest eigenvalue of the covariance matrix of the variables must rise with N. See also Pesaran and Tosetti (2010).

[16] For a more detailed review of this emerging literature, see, for example, Breitung and Pesaran (2008).

[17] As $H_0 : \beta = 1$ cannot be rejected, as shown later, the panel unit root test of the residual will be equivalent to testing for the presence of a unit root in $p_{it} - y_{it}$ with State specific intercepts.

[18] Abstracting from error serial correlation and deterministic components, we assume that $u_{it} = \rho_i u_{it-1} + \zeta_{it}$, $|\rho_i| \leq 1$, with $\zeta_{it} = \pi_i \psi_t + \xi_{it}$, where ψ_t and ξ_{it} are i.i.d. random variables which are independent of each other and both $I(0)$. Under the null of no cointegration, $\rho_i = 1$ for all i. See Pesaran (2007).

Table 1
List of variables and their descriptions.

$P_{it,g}$	US State general price index (1980 = 1)
$P_{it,h}$	US State house price index (1980 = 1)
PD_{it}	US State disposable income
POP_{it}	US State population
RB_t	US nominal long term interest rate, RB_t in percent
p_{it}	Natural logarithm of the US State real house price index, $p_{it} = \log(P_{it,h}/P_{it,g})$
y_{it}	Natural logarithm of the US State real per capita disposable income, $y_{it} = \log[PD_{it}/(POP_{it} \times P_{it,g})]$
r_{it}	US State real long term interest rate, $r_{it} = RB_t/100 - \ln(P_{it,g}/P_{i,t-1,g})$
g_{it}	US State population growth rate, $g_{it} = \log(POP_{it}/POP_{i,t-1})$
c_{it}	US State real cost of borrowing net of real house price appreciation/depreciation, $c_{it} = r_{it} - \Delta p_{it}$

Notes: Annual data between 1975 and 2003 ($T = 29$) for 48 States and the District of Columbia. ($N = 49$). See the Data Appendix for the data sources and a detailed description of the construction of the US State general price index.

Table 2
Regions and abbreviations.

East		Middle		West	
Regions/States	Abbrev.	Regions/States	Abbrev.	Regions/States	Abbrev.
New England region		Great lakes region		South-West region	
Connecticut	CT	Illinois	IL	Arizona	AZ
Maine	ME	Indiana	IN	New Mexico	NM
Massachusetts	MA	Michigan	MI	Oklahoma	OK
New Hampshire	NH	Ohio	OH	Texas	TX
Rhode Island	RI	Wisconsin	WI		
Vermont	VT			Rocky mountain region	
		Plains region		Colorado	CO
Mid-East region		Iowa	IA	Idaho	ID
Delaware	DE	Kansas	KS	Montana	MT
District of Columbia	DC	Minnesota	MN	Utah	UT
Maryland	MD	Missouri	MO	Wyoming	WY
New Jersey	NJ	Nebraska	NE		
New York	NY	North Dakota	ND	Far West region	
Pennsylvania	PA	South Dakota	SD	Alaska	AK
				California	CA
South-East region				Hawaii	HI
Alabama	AL			Nevada	NV
Arkansas	AR			Oregon	OR
Florida	FL			Washington	WA
Georgia	GA				
Kentucky	KY				
Louisiana	LA				
Mississippi	MS				
North Carolina	NC				
South Carolina	SC				
Tennessee	TN				
Virginia	VA				
West Virginia	WV				

Table 3
Average of correlation coefficients within and between regions first difference of log of real per capita real disposable income.

(i) Three geographical regions

	East	Middle	West
East	0.55	–	–
Middle	0.51	0.64	–
West	0.46	0.49	0.48

(ii) Eight BEA regions

	New England	Mid-East	South East	Great lakes	Plains	South West	Rocky mountain	Far West
New England	0.74	–	–	–	–	–	–	–
Mid-East	0.58	0.57	–	–	–	–	–	–
South East	0.48	0.50	0.61	–	–	–	–	–
Great lakes	0.54	0.56	0.70	0.85	–	–	–	–
Plains	0.33	0.34	0.50	0.59	0.61	–	–	–
South West	0.38	0.46	0.54	0.60	0.46	0.45	–	–
Rocky mountain	0.24	0.38	0.44	0.51	0.39	0.49	0.48	–
Far West	0.51	0.51	0.56	0.66	0.44	0.50	0.41	0.68

Notes: See Table 2 for the regions. The figures are average of sample pair-wise correlation coefficients.

growth are not included in this paper, but are available upon request. Real income growth in California is more closely correlated with many States that are geographically very distant. This reflects the common factors driving economic development in different parts of the USA, such as the growth of aerospace, information technology, etc., that stimulate growth in different States. For real house prices the average correlation coefficient between States is 0.39 compared to 0.51 for real incomes. There are also some un-

Table 4
Average of correlation coefficients within and between regions first difference of log of real house prices.

(i) Three geographical regions	East	Middle	West
East	0.48	–	–
Middle	0.42	0.65	–
West	0.19	0.45	0.50

(ii) Eight BEA regions	New England	Mid-East	South East	Great lakes	Plains	South West	Rocky mountain	Far West
New England	0.80	–	–	–	–	–	–	–
Mid-East	0.68	0.66	–	–	–	–	–	–
South East	0.40	0.32	0.52	–	–	–	–	–
Great lakes	0.40	0.35	0.57	0.81	–	–	–	–
Plains	0.27	0.20	0.53	0.62	0.61	–	–	–
South West	0.07	−0.05	0.35	0.28	0.39	0.52	–	–
Rocky mountain	−0.03	−0.11	0.40	0.52	0.53	0.57	0.70	–
Far West	0.13	0.17	0.29	0.52	0.42	0.31	0.46	0.57

Notes: See Table 2 and the notes to Table 3.

Table 5
Residual cross correlation of ADF(p) regressions.

Average cross correlation coefficients ($\bar{\hat{\rho}}$)	ADF(1)	ADF(2)	ADF(3)	ADF(4)
y_{it}	0.411	0.379	0.337	0.317
p_{it}	0.206	0.200	0.208	0.194
g_{it}	0.090	0.080	0.081	0.076
c_{it}	0.309	0.295	0.281	0.283
CD test statistics	**ADF(1)**	**ADF(2)**	**ADF(3)**	**ADF(4)**
y_{it}	68.98	63.73	56.61	53.21
p_{it}	34.62	33.55	35.00	32.52
g_{it}	15.16	13.52	13.64	12.71
c_{it}	50.84	48.54	46.22	46.55

Notes: pth-order Augmented Dickey–Fuller test statistics, ADF(p), for y_{it}, p_{it}, g_{it} and $c_{it} = r_{it} - \Delta p_{it}$ are computed for each cross section unit separately. For y_{it} and p_{it}, an intercept and a linear time trend are included in the ADF(p) regressions, but for g_{it} and c_{it} only an intercept is included. The values in 'average cross correlation coefficients' are the simple average of the pair-wise cross section correlation coefficients of the ADF(p) regression residuals. $\bar{\hat{\rho}} = [2/N(N-1)] \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}$ with $\hat{\rho}_{ij}$ being the correlation coefficient of the ADF(p) regression residuals between ith and jth cross section units. $CD = \sqrt{2T/N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}$, which tends to $N(0, 1)$ under the null hypothesis of no error cross-sectional dependence.

usual correlations at the individual State level. Real house price growth in California, for example, is more closely correlated with Washington DC and Maryland (0.86 and 0.73 respectively) than with New York (0.16) or Oregon (0.25).

Overall, there is more evidence in the raw data of a possible spatial pattern in real house prices than in real incomes, but there are also a number of between-State correlations that appear to be independent of spatial patterns.

In the next section we discuss the results of statistical tests and the estimation of the model.

5. Econometric evidence

5.1. Panel unit root tests results

The extent of cross-sectional dependence of the residuals from ADF(p) regressions of real house prices, real incomes, population growth and net cost of borrowing across the 49 States over the period 1975–2003 are summarized in Table 5. For each $p = 1, 2, 3$ and 4 we computed average sample estimates of the pair-wise correlations of the residuals, which we denote by $\bar{\hat{\rho}}$. To capture the trended nature of real incomes and real house prices, we run the ADF regressions with linear trends, but included an intercept only in the regressions for population growth and net cost of borrowing variables. The results are reasonably robust to the choice of the augmentation order, p. For real incomes and net cost of borrowing, $\bar{\hat{\rho}}$ is estimated to be around 40% and 30%, respectively, whilst for real house prices it is much lower and the estimate stands at 20%. This largely reflects the national character of changes in incomes and net cost of borrowing as compared to real house prices that are likely to be affected by State specific effects such as population growth. The results are also in line with the pair-wise correlations of the raw data discussed above and confirm the existence of a greater degree of cross State correlations in the case of real incomes as compared to real house prices.

The CD test statistics, also reported in Table 5, clearly show that the cross-correlations are statistically highly significant, and thus invalidate the use of panel unit root tests, such as the IPS test, that do not allow for error cross-sectional dependence. Therefore, in what follows we shall focus on Pesaran's *CIPS* tests.[19]

The CIPS test results, summarized in Table 6, show that for population growth and net cost of borrowing the unit root hypothesis is convincingly rejected. For p_{it} and y_{it} the unit root hypothesis cannot be rejected if the trended nature of these variables are taken into account. This conclusion seems robust to the choice of the augmentation order of the underlying CADF regressions. We proceed taking y_{it} and p_{it} as $I(1)$, and c_{it} and g_{it} as $I(0)$ variables.

5.2. The income elasticity of real house prices

To test for possible cointegration between p_{it} and y_{it}, we first estimate the following fairly general model

$$p_{it} = \alpha_i + \beta_i y_{it} + u_{it}, \quad i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T, \tag{5.1}$$

where

$$u_{it} = \sum_{\ell=1}^{m} \gamma_{i\ell} f_{\ell t} + \varepsilon_{it}. \tag{5.2}$$

19 We also conducted Moon and Perrons's t_b^*. The application of the t_b^* test requires an estimate of m, the number of common factors. We tried the various selection criteria proposed in Bai and Ng (2002), all of which require starting from an assumed maximum value of m, denoted by $m_{\max}$. But the outcomes did not prove to be satisfactory, in the sense that the choice of m often coincided with the assumed maximum number of factors, $m_{\max}$. In view of this we computed the t_b^* test statistics for various values of m in the range of 1–4. For changes in real incomes and real house prices, the t_b^* test rejects the unit root hypothesis, but for the levels of these variables the test results depend on whether linear trends are included or not. In the case of house prices the test outcomes also depend on the assumed number of factors. For population growth and the net cost of borrowing, the test results convincingly reject the unit root hypothesis. As the Moon and Perron test is valid only when T is much larger than N, we believe that the CIPS test results are more reliable for our data. Detailed results are available from the authors upon request.

Table 6
Pesaran's CIPS panel unit root test results.

With an intercept				
	CADF(1)	CADF(2)	CADF(3)	CADF(4)
Δy_{it}	−2.61[a]	−2.39[a]	−2.42[a]	−2.34[a]
Δp_{it}	−2.28[a]	−1.86	−1.76	−1.81
y_{it}	−2.52[a]	−2.44[a]	−2.39[a]	−2.49[a]
p_{it}	−2.56[a]	−2.44[a]	−2.83[a]	−2.84[a]
g_{it}	−2.76[a]	−2.29[a]	−2.20[a]	−1.97
c_{it}	−2.14[a]	−2.06[b]	−2.00	−1.95
With an intercept and a linear trend				
	CADF(1)	CADF(2)	CADF(3)	CADF(4)
y_{it}	−2.51	−2.22	−2.24	−2.09
p_{it}	−2.18	−2.02	−2.27	−2.30

Notes: The reported values are CIPS(p) statistics, which are cross section averages of Cross-sectionally Augmented Dickey–Fuller (CADF(p)) test statistics (Pesaran, 2007); see Section 3 for more details. The relevant lower 5% (10%) critical values for the CIPS statistics are −2.11 (−2.03) with an intercept case, and −2.62 (−2.54) with an intercept and a linear trend case. $c_{it} = r_{it} - \Delta p_{it}$, which is the real cost of borrowing net of real house price appreciation/depreciation.
[a] Signifies that the test is significant at the five percent level.
[b] Signifies that the test is significant at the ten percent level.

In view of discussion in Section 3, the common correlated effect (CCE) estimators are consistent regardless of $f_{\ell t}$ being stationary or non-stationary, so long as ε_{it} is stationary and m is a finite fixed number (see Pesaran, 2006; Kapetanios et al., 2009). To show the importance of allowing for the unobserved common factors in this relationship, we also provide naive estimates of β_i, $i = 1, 2, \ldots, N$ (and their mean) that do not allow for cross-sectional dependence by simply running OLS regressions of p_{it} on y_{it}. The common correlated effects (CCE) estimators are based on the cross section augmented regressions

$$p_{it} = \alpha_i + \beta_i y_{it} + d_{i0}\bar{y}_t + d_{i1}\bar{p}_t + e_{it}, \tag{5.3}$$

where $\bar{y}_t$ and $\bar{p}_t$ denote the simple cross section averages of y_{it} and p_{it} in year t. The results are reported in Table 7. The first column gives the naive mean group estimates. These suggest a small coefficient on income of only 0.30 (0.09), and considerable cross-sectional dependence.[20] The other two columns report the common correlated effects mean group (CCEMG) and the common correlated effects pooled (CCEP) estimates. The coefficient on income is now significantly larger and the residual cross-sectional dependence has been purged with the average error cross-correlation coefficient, $\bar{\hat{\rho}}$, reduced from 0.38 for the MG estimates to 0.024 and 0.003 for the CCEMG and CCEP estimates, respectively. The CCEMG and CCEP estimates of β (the mean of β_i) are 1.14 (0.20) and 1.20 (0.21), respectively, and the hypothesis that $\beta = 1$ cannot be rejected. Therefore, the long-run relation to be tested for cointegration is given by

$$\hat{u}_{it} = p_{it} - y_{it} - \hat{\alpha}_i,$$

where $\hat{\alpha}_i = T^{-1}\sum_{t=1}^{T}(p_{it} - y_{it})$.

5.3. Panel cointegration test results

The residuals $\hat{u}_{it}$ defined above can now be used to test the null of non-cointegration between p_{it} and y_{it}. Note that the CCE estimates are consistent irrespective of whether $\mathbf{f}_t$ are $I(0)$, $I(1)$ and/or cointegrated. The presence of $\mathbf{f}_t$ also requires that the panel unit root tests applied to $\hat{u}_{it}$ should allow for the cross-sectional dependence of the residuals. The extent to which these residuals are cross-sectionally dependent can be seen from the average

20 The figure in bracket is the standard error of the estimate.

Table 7
Estimation result: income elasticity of real house price: 1975–2003.

	MG	CCEMG	CCEP
$\hat{\alpha}$	3.85 (0.20)	−0.11 (0.26)	0.00 (0.24)
$\hat{\beta}$	0.30 (0.09)	1.14 (0.20)	1.20 (0.21)
Average cross correlation coefficient ($\bar{\hat{\rho}}$)	0.38	0.024	0.003
CD test statistic	71.03	4.45	0.62

Notes: Estimated model is $p_{it} = \alpha_i + \beta_i y_{it} + u_{it}$. MG stands for mean group estimates. CCEMG and CCEP signify the cross correlated effects mean group and pooled estimates, respectively. $\hat{\alpha} = N^{-1}\sum_{i=1}^{N}\hat{\alpha}_i$ for all estimates, and $\hat{\beta} = N^{-1}\sum_{i=1}^{N}\hat{\beta}_i$ for MG and CCEMG estimates. Standard errors are given in parentheses; see Section 3 for more details. The 'average cross-correlation coefficient' is computed as the simple average of the pair-wise cross section correlation coefficients of the regression residuals, namely $\bar{\hat{\rho}} = [2/N(N-1)]\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{ij}$, with $\hat{\rho}_{ij}$ being the correlation coefficient of the regression residuals of the i and j cross section units. The CD test statistic is $[TN(N-1)/2]^{1/2}\,\bar{\hat{\rho}}$, which tends to $N(0,1)$ under the null hypothesis of no error cross-sectional dependence.

cross-correlation coefficients of $\hat{u}_{it}$, within and between the eight BEA regions, which are reported in Table 8.

We computed CIPS(p) panel unit root test statistics for $p_{it} - y_{it}$, including State specific intercepts, for different augmentation and lag orders, $p = 1, 2, 3$ and 4, and obtained the results, −2.16, −2.39, −2.45, and −2.29, respectively. The 5% and 1% critical values of the CIPS statistic for the intercept case with $N = 50$ and $T = 30$ are −2.11 and −2.23, respectively. The results suggest rejection of a unit root in $p_{it} - y_{it}$ for all the augmentation orders at 5% level and rejection at 1% level in the case of the augmentation orders 2 and more.[21]

5.4. Panel error correction specifications

Having established panel cointegration between p_{it} and y_{it}, we now turn our attention to the dynamics of the adjustment of real house prices to real incomes and estimate the panel error correction model:

$$\Delta p_{it} = \alpha_i + \phi_i(p_{i,t-1} - y_{i,t-1}) + \delta_{1i}\Delta p_{i,t-1} + \delta_{2i}\Delta y_{it} + \upsilon_{it}. \tag{5.4}$$

The coefficient ϕ_i provides a measure of the speed of adjustment of house prices to a shock. The half life of a shock to p_{it} is approximately $-\ln(2)/\ln(1+\phi_i)$.

To allow for possible cross-sectional dependence in the errors, υ_{it}, we computed CCEMG and CCEP estimators of the parameters, as well as the mean group (MG) estimators that do not take account of cross-sectional dependence, as a benchmark. The former estimates as computed by OLS regressions of Δp_{it} on 1, $(p_{i,t-1} - y_{i,t-1})$, $\Delta p_{i,t-1}$, Δy_{it}, and the associated cross section averages, $(\bar{p}_{t-1} - \bar{y}_{t-1})$, $\Delta\bar{y}_t$, $\Delta\bar{p}_t$, and $\Delta\bar{p}_{t-1}$. The results are summarized in Table 9. The coefficients are all correctly signed. The CCEMG and CCEP estimators are very close and yield error correction coefficients given by −0.183(0.016) and −0.171(0.015) that are reasonably large and statistically highly significant. The average half life estimates are around 3.5 years, much smaller than the half life estimates of 6.3 years obtained using the MG estimators. But the MG estimators are likely to be biased, since the residuals from these estimates show a high degree of cross-sectional dependence. The same is not true of the CCE-type estimators.

In a number of experiments we also included the population and the cost variables, $g_{i,t-1}$ and $c_{i,t-1} = r_{i,t-1} - \Delta p_{i,t-1}$, in the

21 We also conducted Moon and Perrons's t_b^* test for $\hat{u}_{it}$, with an intercept. For the same reason given in footnote 19, t_b^* test statistics are computed for different values of $m = 1, 2, 3$, and 4. The corresponding statistics were −3.95, −2.82, −3.50, and −4.28, respectively. Thus, the Moon–Perron test also strongly reject the null of no cointegration between p_{it} and y_{it}.

Table 8
Average residual cross correlation coefficients within and between eight BEA geographical regions — $\hat{u}_{it} = p_{it} - y_{it} - \hat{\alpha}_i$.

	New England	Mid-East	South East	Great lakes	Plains	South West	Rocky mountain	Far West
New England	0.62	–	–	–	–	–	–	–
Mid-East	0.56	0.45	–	–	–	–	–	–
South East	0.18	0.29	0.95	–	–	–	–	–
Great lakes	0.04	0.21	0.75	0.82	–	–	–	–
Plains	0.10	0.24	0.92	0.80	0.93	–	–	–
South West	0.07	0.17	0.90	0.63	0.87	0.94	–	–
Rocky mountain	−0.23	−0.08	0.70	0.70	0.77	0.75	0.86	–
Far West	−0.12	−0.01	0.03	0.25	0.13	0.06	0.24	0.21

Notes: $\hat{\alpha}_i = T^{-1}\sum_{t=1}^{T}(p_{it} - y_{it})$. See also Table 2 for the abbreviations of the regions and notes to Table 3.

Table 9
Panel error correction estimates without net cost of borrowing and population growth: 1977–2003.

Δp_{it}	MG	CCEMG	CCEP
$p_{i,t-1} - y_{i,t-1}$	−0.105 (0.008)	−0.183 (0.016)	−0.171 (0.015)
$\Delta p_{i,t-1}$	0.524 (0.030)	0.449 (0.038)	0.518 (0.065)
Δy_{it}	0.500 (0.040)	0.277 (0.059)	0.227 (0.063)
Half life	6.248	3.429	3.696
$\bar{R}^2$	0.54	0.70	0.66
Average cross correlation coefficients ($\bar{\hat{\rho}}$)	0.284	−0.005	−0.016
CD test statistics	50.60	−0.84	−2.80

Notes: The State specific intercepts are estimated but not reported. MG stands for mean group estimates. CCEMG and CCEP signify the cross correlated effects mean group and pooled estimates, respectively. Standard errors are given in parentheses. The $\bar{R}^2$ for the MG and CCEP estimators are defined by (3.14) and (3.15), respectively. The half life of a shock to p_{it} is approximated by $-\ln(2)/\ln(1+\hat{\phi})$ where $\hat{\phi}$ is the pooled estimates for the coefficient on $p_{i,t-1} - y_{i,t-1}$. Also see the notes to Table 7.

error correction model (see Table 10). As predicted by the theory, we found a significant negative effect for the $c_{i,t-1}$ variable, and a positive significant effect for the population growth variable, $g_{i,t-1}$. In fact we could not reject the hypothesis that the short-term elasticity of changes in real house prices to population growth is around unity. This is in line with the supposition that State level population growth acts as a proxy for short run supply factors.

Since all the cross section variations in the real interest rates are due to the inflation variable (the long run nominal interest rate being a national variable which does not vary across the States), the inclusion of $c_{i,t-1}$ in the error correction model renders $\Delta p_{i,t-1}$ statistically insignificant. Nevertheless, judging by the average fit of the various panel regressions reported in Table 10, $c_{i,t-1}$ yields a better fit as compared to $\Delta p_{i,t-1}$, and is therefore to be preferred on theoretical grounds.

5.5. Testing for spatial autocorrelation

The previous analysis provided consistent estimates of the cointegrating relationship between real house prices and real incomes. In this section we turn to the estimation of spatial patterns based on the estimation of a spatial weighting matrix that is commonly used in the literature. We investigate the error structure (5.2), based on $\hat{u}_{it} = p_{it} - y_{it} - \hat{\alpha}_i$. Our aim is to distinguish between strong dependence which is captured by the common factors in (5.2) and the remaining dependence across the idiosyncratic components, ε_{it}, that capture weak dependence in the overall residuals, u_{it}. These idiosyncratic factors reflect forms of local dependence that are spatial in nature.[22]

To investigate possible spatial patterns in the residuals, a multi-factor decomposition of $\hat{u}_{it}$ is required. We considered the following specification

$$\hat{u}_{it} = \sum_{\ell=1}^{m} \tilde{\gamma}_{i\ell}\tilde{f}_{\ell t} + \tilde{\varepsilon}_{it}, \tag{5.5}$$

where $\tilde{f}_{\ell t}$, $\ell = 1, 2, \ldots, m$ are the common factors and $\tilde{\gamma}_{i\ell}$ are the associated factor loadings. We experimented with different values of $m = 1, 2, 3$, and estimated the factors by the principle components. The idiosyncratic components, $\tilde{\varepsilon}_{it}$, are then computed as residuals from the OLS regressions of $\hat{u}_{it}$ on the estimated factors over the period 1975–2003 for each i.[23]

To investigate the strength of spatial dependence in the idiosyncratic components, for each m we estimated the following standard spatial lag model in $\tilde{\varepsilon}_{it}$ (Cliff and Ord, 1973)

$$\tilde{\varepsilon}_{it} = \psi \sum_{j=1}^{N} s_{ij}\tilde{\varepsilon}_{jt} + \nu_{it}, \tag{5.6}$$

where ψ is a spatial autoregressive parameter, and w_{ij} is the generic element of the $N \times N$ spatial weight matrix $\mathbf{S}$, and $\nu_{it} \sim$ i.i.d. $N(0, \sigma_\nu^2)$. The log-likelihood function of this model is given by

$$L = -\left(\frac{NT}{2}\right)\ln\left(\sigma_\nu^2\right) + T\ln|\mathbf{I}_N - \psi\mathbf{S}| - \frac{1}{2\sigma_\nu^2}\sum_{t=1}^{T}\left(\tilde{\boldsymbol{\varepsilon}}_t - \psi\mathbf{S}\tilde{\boldsymbol{\varepsilon}}_t\right)'\left(\tilde{\boldsymbol{\varepsilon}}_t - \psi\mathbf{S}\tilde{\boldsymbol{\varepsilon}}_t\right),$$

where $\tilde{\boldsymbol{\varepsilon}}_t = (\tilde{\varepsilon}_{1t}, \tilde{\varepsilon}_{2t}, \ldots, \tilde{\varepsilon}_{Nt})'$, and in our application $N = 49$ and $T = 29$.[24] For $\mathbf{S}$, following the approach of Anselin, we used a contiguity criterion and assigned $s_{ij} = 1$ when State i and j share a common border or vertex, and $s_{ij} = 0$ otherwise.[25] The maximum likelihood (ML) estimates of ψ together with their standard errors given in brackets for $m = 1, 2$ and 3 are 0.653 (0.022), 0.487 (0.027) and 0.298 (0.033), respectively.[26] All the estimates are highly significant and as is to be expected, the magnitude of the spatial parameter declines with the number of factors. Nevertheless, even with 3 factors there is strong evidence that local dependence in the

[22] Pesaran and Tosetti (2010) show that most spatial models considered in the literature represent forms of weak dependence.

[23] We also tried to estimate the number of factors, m, using the information criteria (IC) proposed by Bai and Ng (2002). But, as in the case of the results reported in footnote 19, the IC procedure always ended up choosing the assumed maximum number of factors, even if $m_{\max}$ was set to 6.

[24] For computation details of maximum likelihood estimation, see Anselin et al. (2008) and references therein.

[25] The data on contiguity are obtained from Luc Anselin's web site at: http://sal.uiuc.edu/weights/index.html.

[26] We also computed generalised method of moments estimates proposed by Kelejian and Prucha (1999). These yielded very similar results to the maximum likelihood estimates. We are grateful to Elisa Tosetti for carrying out the computations of the spatial estimates.

Table 10
Panel error correction estimates with net cost of borrowing and Population growth: 1977–2003.

Δp_{it}	MG				CCEMG				CCEP			
$p_{i,t-1} - y_{i,t-1}$	−0.117 (0.013)	−0.114 (0.009)	−0.148 (0.010)	−0.138 (0.009)	−0.275 (0.021)	−0.219 (0.020)	−0.256 (0.019)	−0.215 (0.017)	−0.242 (0.021)	−0.208 (0.017)	−0.242 (0.021)	−0.195 (0.017)
$\Delta p_{i,t-1}$	0.481 (0.079)	0.444 (0.071)	–	–	−0.061 (0.098)	−0.114 (0.126)	–	–	−0.015 (0.058)	−0.134 (0.125)	–	–
Δy_{it}	0.533 (0.045)	0.544 (0.045)	0.669 (0.047)	0.679 (0.047)	0.284 (0.067)	0.272 (0.063)	0.332 (0.057)	0.309 (0.056)	0.272 (0.046)	0.258 (0.052)	0.290 (0.064)	0.272 (0.060)
$c_{i,t-1}$	0.057 (0.071)	−0.053 (0.056)	−0.375 (0.034)	−0.438 (0.027)	−0.301 (0.116)	−0.524 (0.126)	−0.269 (0.044)	−0.492 (0.036)	−0.308 (0.101)	−0.583 (0.113)	−0.304 (0.080)	−0.553 (0.054)
$\eta_{i,t-1}$	1.189 (0.331)	–	1.211 (0.331)	–	1.312 (0.532)	–	1.709 (0.532)	–	1.020 (0.318)	–	0.931 (0.381)	–
Half Life	5.57	5.73	4.33	4.67	2.16	2.80	2.34	2.86	2.50	2.97	2.50	3.12
$\bar{R}^2$	0.57	0.54	0.53	0.50	0.78	0.73	0.76	0.71	0.72	0.68	0.71	0.67
Average cross correlation coefficients ($\bar{\hat{\rho}}$)	0.286	0.293	0.346	0.339	−0.009	0.002	−0.005	−0.005	−0.017	−0.016	−0.015	−0.015
CD test statistics	50.96	52.21	61.65	60.41	−1.60	0.36	0.89	0.89	−3.03	−2.85	2.67	2.67

See the notes to Table 9. $c_{it} = r_{it} - \Delta p_{it}$, which is the real cost of borrowing net of real house price appreciation/depreciation.

form of a spatial dependence between contiguous States in the USA is present in the data.

We also checked the spatial estimates to see if they are robust to possible differences in the error variances across the States, by estimating the spatial model using standardized residuals defined by $\varepsilon^*_{it} = \tilde{\varepsilon}_{it}/s_i$, where $s_i = \sqrt{\Sigma_{t=1}^{T}\tilde{\varepsilon}^2_{it}/T}$. We obtained slightly larger estimates for ψ, namely 0.673 (0.021), 0.513 (0.027) and 0.393 (0.030), for $m = 1, 2$, and 3, respectively. These estimates confirm a highly significant and economically important spatial dependence in real house prices in the USA, even after controlling for State specific real incomes, and after allowing for a number of unobserved common factors.

5.6. Factor loading estimates across states

We have shown that the common correlated effects estimators are quite successful in taking out the cross-sectional dependence by the use of a multifactor error structure where the unobserved common factors are proxied by filtering the individual-specific regressors with cross-section aggregates. However, the sensitivity of the ith unit, in this case a State, to the factors will vary so the factor loadings differ over the cross section units. We can obtain an idea of these differential factor loadings if we regress $p_{it} - y_{it}$ on $\bar{p}_t - \bar{y}_t$, and a constant. These regressions are reminiscent of the Capital Asset Pricing regressions in finance where individual asset returns are regressed on market (or average) returns.

The results, summarized in Table 11, show an interesting pattern in the loadings on the factor $(\bar{p}_t - \bar{y}_t)$. The States are ordered by the BEA's regions. By construction, the cross section average of the estimated coefficients on $(\bar{p}_t - \bar{y}_t)$ is unity, and the cross section average of the intercepts is zero. New England and the Mid-East States all have loadings of less than one, while all of the South-East States, with the exception of Virginia, have factor loadings that are greater than one. This is also true for the States in the Plains region and the South-West region. The Far West region States all have loadings less than 1 also. But strikingly, there are a number of States that have a zero, or even a *negative* loading—Massachusetts, Rhode Island, Connecticut, New Jersey, New York, California, Oregon and Washington. In the case of Massachusetts, New York, and California, the loadings are negative, sizeable and statistically significant. These are all States that in the last 25 years have been particular beneficiaries of new technologies. These innovations interacting with restrictions on new residential buildings have resulted in real house prices in these regions deviating from the average across US States over a relatively prolonged period. This naturally raises the issue of whether these exceptional patterns are likely to be sustainable. Recent evidence on house prices and incomes can shed some light on this issue.

Table 11
Factor loading estimates.

$(p_{it} - y_{it})$	$(\bar{p}_t - \bar{y}_t)$	$(p_{it} - y_{it})$	$(\bar{p}_t - \bar{y}_t)$
Connecticut	0.35 (0.23)	Indiana	1.14* (0.05)
Maine	0.29* (0.15)	Michigan	0.54* (0.17)
Massachusetts	**−0.63*** (0.24)	Ohio	1.01* (0.09)
New Hampshire	0.81* (0.22)	Wisconsin	0.98* (0.12)
Rhode Island	**−0.11** (0.24)	Iowa	1.55* (0.11)
Vermont	0.78* (0.15)	Kansas	1.76* (0.06)
Delaware	0.32* (0.11)	Minnesota	1.20* (0.09)
District of Columbia	0.54* (0.18)	Missouri	1.37* (0.04)
Maryland	0.62* (0.10)	Nebraska	1.57* (0.10)
New Jersey	**−0.04** (0.20)	North Dakota	2.00* (0.15)
New York	**−0.39*** (0.20)	South Dakota	1.39* (0.08)
Pennsylvania	0.65* (0.13)	Arizona	1.02* (0.07)
Alabama	1.72* (0.09)	New Mexico	0.95* (0.12)
Arkansas	1.77* (0.10)	Oklahoma	2.10* (0.17)
Florida	1.44* (0.08)	Texas	2.12* (0.18)
Georgia	1.43* (0.08)	Colorado	0.80* (0.17)
Kentucky	1.21* (0.06)	Idaho	1.19* (0.11)
Louisiana	2.03* (0.15)	Montana	0.75* (0.16)
Mississippi	2.09* (0.13)	Utah	0.68* (0.19)
North Carolina	1.28* (0.05)	Wyoming	1.62* (0.18)
South Carolina	1.39* (0.06)	**California**	**−0.64*** (0.23)
Tennessee	1.53* (0.08)	Nevada	0.84* (0.11)
Virginia	0.91* (0.09)	Oregon	0.37 (0.25)
West Virginia	2.08* (0.11)	**Washington**	**−0.12** (0.17)
Illinois	0.71* (0.11)		

Notes: Reported figures are estimated slope coefficients on $(\bar{p}_t - \bar{y}_t)$ of regressions of $(p_{it} - y_{it})$ on $(\bar{p}_t - \bar{y}_t)$. Standard errors are given in parentheses. By construction, the cross section average of the estimated slope coefficients is unity, and the cross section average of the intercepts is zero (not reported). The negative slope estimates are in bold, and statistically significant slopes are denoted by $*$.

5.7. US house prices since 2003

Our analysis suggests that even if house prices deviate from the equilibrating relationship because of State-specific or common shocks, they will eventually revert. If house prices are above equilibrium they will tend to fall relative to income, and *vice versa* if they are above equilibrium. Of course, because there is heterogeneity across States, a particular State need not be in the same disequilibrium position as other States. But on average the change in the ratio of house prices to per capita incomes should be zero, consistent with a cointegrating relationship, for T sufficiently large.

The process of house price boom that had started in the USA in early 2000 accelerated during 2003–2006 which some have interpreted as a bubble. Over the period 2000–2006 the average (unweighted) rise in US house prices was 46%, as compared to a 25% rise in income per capita. However, the price increases relative to per capita incomes have been quite heterogeneous. While house prices over the period 2000 to 2006 rose by 67% in Virginia, 73%

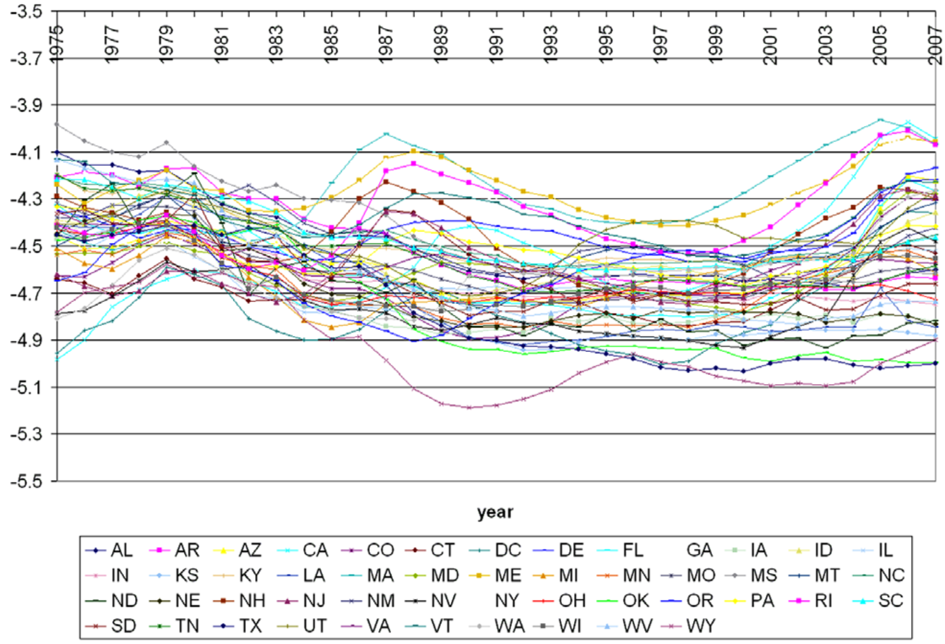

Fig. 1. Log ratio of house prices to per capita incomes over the period 1976–2007 for the 49 States of the USA. Note: figures are computed as $p_{it} - y_{it}$.

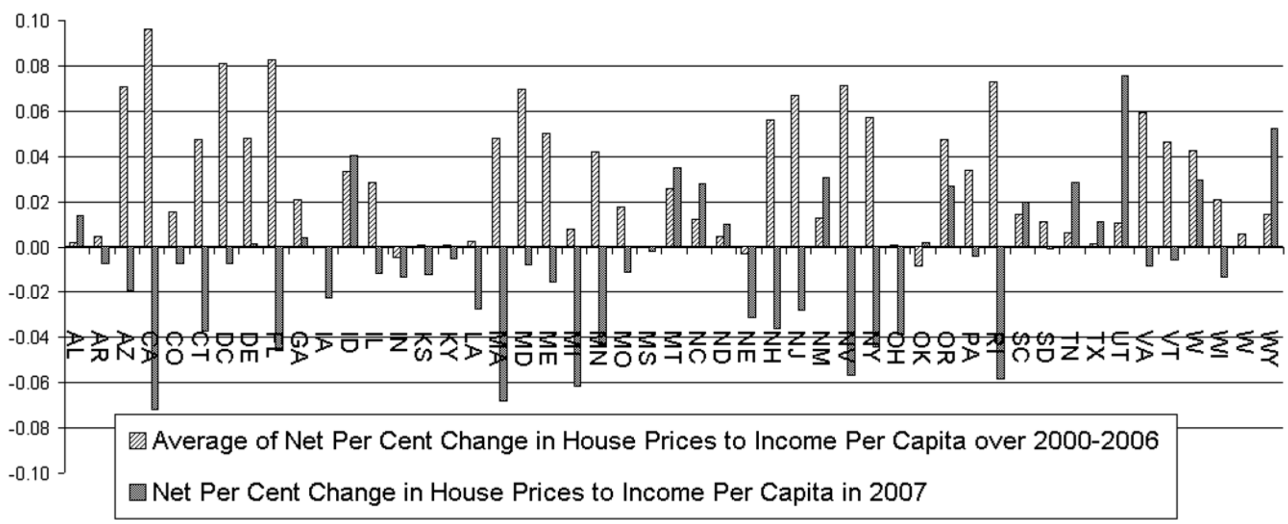

Fig. 2. Percent change in house prices to per capita incomes across the US States over 2000–2006 as compared to the corresponding ratios in 2007. Note: Figures are the average of $\Delta(p_{it} - y_{it})$ over the specified period for each state/district.

in Arizona and 92% in the District of Columbia, they rose by only 20% in Indiana and 21% in Ohio. These differences were much more pronounced that the rise in income per capita in these States (respectively 26%, 23%, 40%, 20% and 19%). Individual States can move about the average because the driving variables are behaving differently or because the initial disequilibrium is different. The extent of the heterogeneity in the disequilibrium, as measured by the time profile of the logarithm of price–income per capita over the full sample, 1976–2007, for all the 49 States is displayed in Fig. 1. It is interesting that the excess rise in house prices tends to be associated with increased dispersion in the log price–income ratios, which begins to decline with moderation of house price rises relative to incomes. This fits well with the development of house prices in 2007, where prices rose only by 4% as compared to a rise in per capita income of 5%. The range of house price changes across States was also narrowed down substantially. In fact, in the case of the five States mentioned, the price–income ratio declined by 1% in Virginia, 2% in Arizona, 0% in District of Columbia, −2% in Indiana, and 4% in Ohio.

If we calculate the average change in the log ratio of house prices to per capita income for each State over the period 2000–2006, and compare it to the average change in the ratio for 2007, it is to be expected that if a state on average is above its equilibrium before 2006 that the average change after 2006 should be negative, and vice versa otherwise. The results are plotted in Fig. 2. We find that of 49 States, 32 States have an average change in 2007 that is the opposite sign to the average for 2000–2006. Moreover, if we look at the correlation between the change in the ratio in 2007, when the house price boom began to unwind, and the average change in the ratio of house prices to per capita income over the preceding price boom period, 2000–2006, the correlation coefficient is −0.42.

6. Concluding remarks

This paper has considered the determination of real house prices in a panel made up of 49 US States over 29 years,

Table A.1
Description of the construction of State level price indices.

State/Region	Name of city/Region used	Missing data	Base of projection
District of Columbia	Washington, Washington–Baltimore	1998–2003	Merge of Washington–Baltimore
Alabama	Atlanta	–	–
Arkansas	Dallas	–	–
Arizona	San Diego	–	–
California	Los Angeles	–	–
Colorado	Denver	–	–
Connecticut	New York	–	–
Delaware	Philadelphia	–	–
Florida	Miami	1975–1977	US average
Georgia	Atlanta	–	–
Iowa	Minneapolis	–	–
Idaho	Seattle	–	–
Illinois	Chicago	–	–
Indiana	Chicago	–	–
Kansas	Kansas city	–	–
Kentucky	Kansas city	–	–
Louisiana	New Orleans	1975–1986, 1998–2003	US average
Massachusetts	Boston	–	–
Maryland	Baltimore, Washington–Baltimore	1998–2003	Merge of Washington–Baltimore
Maine	Boston	–	–
Michigan	Detroit	–	–
Minnesota	Minneapolis	–	–
Missouri	ST Louise	–	–
Mississippi	New Orleans	1975–1986, 1998–2003	US average
Montana	Seattle	–	–
North Carolina	Washington, Washington–Baltimore	1998–2003	Merge of Washington–Baltimore
North Dakota	Minneapolis	–	–
Nebraska	Kansas city	–	–
New Hampshire	Boston	–	–
New Jersey	New York	–	–
New Mexico	Denver	–	–
Nevada	San Francisco	–	–
New York	New York	–	–
Ohio	Cleveland	–	–
Oklahoma	Dallas	–	–
Oregon	Portland	–	–
Pennsylvania	Pittsburgh	–	–
Rhode Island	Boston	–	–
South Carolina	Atlanta	–	–
South Dakota	Minneapolis	–	–
Tennessee	Cincinnati	–	–
Texas	Houston	–	–
Utah	Denver	–	–
Virginia	Washington, Washington–Baltimore	1998–2003	Merge of Washington–Baltimore
Vermont	Boston	–	–
Washington	Seattle	–	–
Wisconsin	Milwaukee	–	–
West Virginia	Washington, Washington–Baltimore	1998–2003	Merge of Washington–Baltimore
Wyoming	Denver	–	–

where there is a significant spatial dimension. An error correction model with a cointegrating relationship between real house prices and real incomes is found once we take proper account of both heterogeneity and cross-sectional dependence. We do this using recently proposed estimators that use a multifactor error structure. This approach has proved useful for modelling spatial interactions that reflect both geographical proximity and unobservable common factors. We also provide estimates of spatial autocorrelation conditional on up to three common factors and find significant evidence of spatial dependence associated with contiguity.

Overall, our results support the hypothesis that real house prices have been rising in line with fundamentals (real incomes), and there seems little evidence of house price bubbles at the national level. But we also find a number of outlier States: California, New York, Massachusetts, and to a lesser extent Connecticut, Rhode Island, Oregon and Washington State, with their log house price income ratios either unrelated to the national average or even moving in the opposite direction. It is interesting that these are the States that over the past 25 years have been pioneer and major beneficiaries of technological innovations in media, entertainment, finance, and computers.

Appendix. Data Appendix

The data set are annual data 1975–2003 and cover 48 States (excluding Alaska and Hawaii), plus the District of Columbia. The US State level house price index ($P_{it,h}$) are obtained from the Office of Federal Housing Enterprise Oversight. The US State level data of disposable income (PD_{it}) and the State population (POP_{it}) are obtained from the Bureau of Economic Analysis. When only the quarterly data are available, annual simple averages of the four quarters are used.

As there is no US State level consumer price index (CPI), we constructed State level general price index, $P_{it,g}$, based on the CPIs of the cities/areas. The reasoning is summarized in Table A.1. Briefly, we choose the large cities/area of the State or next to the State which have their own CPIs, which are available from the Bureau of Labor Statistics (BLS). Note that this procedure allows multiple States to share a common price index. When the State price index have missing data, they are replaced with the US CPI average or the average of Washington–Baltimore, according to their locations.

The long-term interest rates, RB_t, which are simple annual averages of quarterly data, are taken from the Fair Model database.

References

Almeida, H., Campello, M., Liu, C., 2006. The financial accelerator: evidence from international housing markets. Review of Finance 10, 1–32.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Boston.

Anselin, L., 2001. Spatial econometrics. In: Baltagi, B. (Ed.), A Companion to Theoretical Econometrics. Blackwell, Oxford.

Anselin, L., Le Gallo, J., Jayet, H., 2008. Spatial panel econometrics. In: Matyas, L., Sevestre, P. (Eds.), The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice, 3rd ed. Springer.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–173.

Bai, J., Kao, C., 2006. On the estimation and inference of panel cointegration model with cross-sectional dependence. In: Baltagi, B. (Ed.), Panel Data Econometrics: Theoretical Contributions and Empirical Applications. In: Contributions to Economic Analysis, vol. 274. Elsevier.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 161–221.

Banerjee, A., Carrion-i-Silvestre, J.L., 2006. Cointegration in panel data with breaks and cross-section dependence. In: ECB Working Paper Series, vol. 591. European Central Bank.

Breitung, J., Pesaran, M.H., 2008. Unit roots and cointegration in panels. In: Matyas, L., Sevestre, P. (Eds.), The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice, 3rd ed. Springer.

Buckley, R., Ermisch, J., 1982. Government policy and house prices in the United Kingdom: an economic analysis. Oxford Bulletin of Economics and Statistics 44, 273–304.

Capozza, D.R., Hendershott, P.H., Mack, C., Mayer, C.J., 2002. Determinants of real house price dynamics. NBER Working Paper. 9262.

Carlino, G., DeFina, R., 1998. The differential regional effects of monetary policy. Review of Economics and Statistics 80, 9001–9921.

Case, K.E., Shiller, R., 2003. Is there a bubble in the housing market? Brookings Papers on Economic Activity 2, 299–342.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean-variance analysis in large asset market. Econometrica 51, 1305–1324.

Chang, Y., 2005. Residual based tests for cointegration in dependent panels. Mimeo. Rice University.

Cliff, A., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

Cliff, A., Ord, J.K., 1981. Spatial Processes, Models and Applications. Pion, London.

Conley, T.G., 1999. GMM estimation with cross sectional dependence. Journal of Econometrics 92, 1–45.

Conley, T.G., Topa, G., 2002. Socio-economic distance and spatial patterns in unemployment. Journal of Applied Econometrics 17, 303–327.

Feldstein, M., Green, J., Sheshinksi, E., 1978. Inflation and taxes in a growing economy with debt and equity finance. Journal of Political Economy 86, 53–70.

Fratantoni, M., Schuh, S., 2003. Monetary policy, housing investment, and heterogeneous regional markets. Journal of Money, Credit and Banking 35, 557–589.

Frees, E.W., 1995. Assessing cross-sectional correlation in panel data. Journal of Econometrics 69, 393–414.

Gallin, J., 2006. The long run relationship between house prices and income: evidence from local housing markets. Real Estate Economics 34, 417–438.

Gengenbach, C., Palm, F.C., Urbain, J.-P., 2006. Cointegration testing in panels with common factors. Oxford Bulletin of Economics and Statistics 68 (Suppl.), 638–719.

Glaeser, E.L., Gyourko, J., 2005. Urban decline and durable housing. Journal of Political Economy 113, 345–375.

Glaeser, E.L., Gyourko, J., Saks, R.E., 2006. Urban growth and housing supply. Journal of Economic Geography 6, 71–89.

Green, R.K., Malpezzi, S., 2003. A Primer on US Housing Markets and Housing Policy. In: AREUEA Monograph, Urban Institute Press, Washington.

Groen, J.J.J., Kleibergen, F., 2003. Likelihood-based cointegration analysis in panels of vector error-correction models. Journal of Business & Economic Statistics 21, 295–318.

Haining, R.P., 1978. The moving average model for spatial interaction. Transactions of the Institute of British Geographers 3, 202–225.

Haining, R.P., 2003. Spatial Data Analysis. Cambridge University Press, Cambridge.

Hendershott, P.H., Hu, S.C., 1981. Inflation and extraordinary returns on owner-occupied housing: some implications for capital allocation and productivity growth. Journal of Macroeconomics 3, 177–203.

Im, K., Pesaran, M.H., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115, 53–74.

International Monetary Fund (2004). The global house price boom. World Economic Outlook, September.

Kapetanios, G., Pesaran, M.H., Yamagata, T., 2009. Panels with nonstationary multifactor error structures. Mimeo. University of Cambridge.

Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.

Krugman, O., 1998. Space: the final frontier. Journal of Economic Perspectives 12, 161–174.

Levin, A., Lin, F., Chu, C., 2002. Unit root tests in panel data: asymptotic and finite-sample properties. Journal of Econometrics 108, 1–24.

Maddala, G.S., Wu, S., 1999. A comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and Statistics 631–652 (special issue).

Malpezzi, S., 1999. A simple error correction model of house prices. Journal of Housing Economics 8, 27–62.

McCarthy, J., Peach, R.W., 2004. Are home prices the next "bubble"? Economic Policy Review 10.

Meen, G., 2002. The time-series behavior of house prices: a transatlantic divide? Journal of Housing Economics 11, 1–23.

Moon, H.R., Perron, B., 2004. Testing for a unit root in panels with dynamic factors. Journal of Econometrics 122, 81–126.

Moran, P.A.P., 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society, Series B 37, 243–251.

Nelson, M., Ogaki, M., Sul, D., 2005. Dynamic seemingly unrelated cointegrating regressions. The Review of Economic Studies 72, 797–820.

Paelinck, J.H.P, Klaasen, L.H., 1979. Spatial Econometrics. Saxon House, Farnborough.

Pedroni, P., 1999. Critical values for cointegration tests in heterogeneous panels with multiple regressors. Oxford Bulletin of Economics and Statistics 61, 653–670.

Pedroni, P., Vogelsang, T., 2005. Robust unit root and cointegration rank tests for panels and large systems. Mimeo. Williams College.

Pesaran, M.H., 2004. General diagnostic tests for cross section dependence in panels. CESifo Working Papers. No. 1233.

Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967–1012.

Pesaran, M.H., 2007. A simple panel unit root test in the presence of cross section dependence. Journal of Applied Econometrics 22, 265–312.

Pesaran, M.H., Tosetti, E., 2010. Large panels with spatial correlations and common factors. Mimeo. University of Cambridge.

Pollakowski, H.O., Ray, R.S., 1997. Housing price diffusion patterns at different aggregation levels: an examination of housing market efficiency. Journal of Housing Research 8, 107–124.

Smith, L.V., Leybourne, S., Kim, T., Newbold, P., 2004. More powerful panel data unit root tests with an application to mean reversion in real exchange rates. Journal of Applied Econometrics 19, 147–170.

Westerlund, J., 2005. New simple tests for panel cointegration. Econometric Reviews 24, 297–316.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 41, 434–449.