

A Spatio-Temporal Model of House Prices in the US*

Sean Holly M. Hashem Pesaran
Takashi Yamagata

Faculty of Economics and Centre for International Macroeconomics and Finance (CIMF),
University of Cambridge, Cambridge, CB3 9DD, UK

19 September 2006

Abstract

In this paper we model the dynamic adjustment of real house prices using data at the level of US States. We consider interactions between housing markets by examining the extent to which real house prices at the State level are driven by fundamentals such as real income, as well as by common shocks, and determine the speed of adjustment of 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 house prices and incomes and to identify a small role for real interest rates. Using this model we 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.

JEL Classification: C21, C23

Keywords: House Price, Cross Sectional Dependence, Spatial Dependence

*We would like to thank Ron Smith and Elisa Tosetti for useful comments. Financial support from the ESRC under Grant No. RES-000-23-0135 is gratefully acknowledged.

1 Introduction

The long standing interest of geographers and regional scientists in spatial issues has spelt over into economics and into the development of spatial econometrics (Paelinck and Klaasen, 1979, Anselin, 1988, Krugman 1998) with a particular emphasis placed on interactions in space (spatial autocorrelation) and spatial structures (spatial heterogeneity). At the same time there are many economic studies based on panels of economic data at the city, state, regional and country level that have an implicit spatial structure, but which effectively ignore possible spatial interactions and interdependencies. This may not matter if the spatial interactions are captured by common observed factors which are themselves included in the panel model. However, in practice there may be spatial interactions that are not adequately captured in this way.

In the literature on spatial econometrics the degree of cross section dependence is calibrated by means of a weighting matrix. For example the (i, j) elements of a weighting matrix, w_{ij} , could take a value of 1 if the i^{th} and j^{th} areas/regions/countries are contiguous and zero otherwise¹. Of course there are many other forms of contiguity that draw on the metaphor of a chessboard with the type of connectiveness reflecting the scope for movements by the rook, the bishop and the queen. Weights can also be based on distance, squared distance or the number of nearest neighbours. Often, however, in economic applications space may not be the appropriate metric (Conley and Dupor, 2003, and Conley and Topa, 2002). In some instances trade flows might be relevant, whilst in the case of inter-industry dependencies input-output matrices might provide the appropriate spatial metric. Alternatively, there may be dependencies between geographical areas that reflect cultural similarity, and migration or commuting relationships (Eff, 2004).

Much of these forms of interaction are unobservable, or difficult to measure. We need a method to test for possible hidden interactions. Recently tests have been proposed (Pesaran, 2004) for spatial dependence based on the average of pair-wise correlation coefficients of the OLS residuals from the individual regressions in the panel. Where spatial dependence is detected (perhaps due to an unobservable common factor or factors) a widely employed way of taking account of this in modelling is to use a fixed, non-stochastic spatial weights matrix. Another way would be to use (for a sufficiently large number of cross section observations) the common correlated effects estimator of Pesaran (2006a). This is a new approach to estimation and inference in panel data models with a multifactor error structure where the unobserved common factors are (possibly) correlated with exogenously given individual-specific regressors, and the factor loadings differ over the cross section units. The basic idea behind the proposed estimation procedure is to filter the individual-specific regressors by means of (weighted) cross-section aggregates such that asymptotically as the cross-section dimension (N) tends to infinity the differential effects of unobserved common factors are eliminated. The estimation procedure has the advantage that it can be computed by 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. Two different but related problems are addressed: one that concerns the coefficients of the individual-specific regressors, and the other that focuses on the mean of the individual coefficients assumed random. In both cases appropriate estimators, referred to as common correlated effects (CCE) estimators, are proposed and their asymptotic distribution as N tends to infinity, with T (the time-series dimension) fixed or as N and T tend to infinity (jointly) are de-

¹See Moran (1948), Cliff and Ord (1973, 1981), Anselin (1988, 2001) and Haining (2003).

rived under different regularity conditions. One important feature of the proposed CCE mean group (CCEMG) estimator is its invariance to the (unknown but fixed) number of unobserved common factors as N and T tend to infinity (jointly). In this paper we apply this methodology to an analysis of house prices at the State level in the USA.

Recently there has been considerable interest in the behaviour of house prices not only in the US but also internationally (IMF, 2004). The majority of OECD countries have experienced a considerable rise in real house prices in the last decade. Because of the role housing wealth plays in household behaviour, fears have been expressed that there is a ‘bubble’ in house prices, with prices moving well away from their fundamental drivers (Case and Shiller, 2003, McCarthy and Peach, 2004). Changes in real house prices relative to household income can also have important consequences for household budgets, with implications for the design of social policy, and possibly, the behaviour of the macroeconomy. Maclennan, Muellbauer, and Stephens (1998), for example, argue that there are huge differences in housing and financial market institutions across Europe and that this has profound effects on the way in which output and inflation in the different countries respond to changes in short-term interest rates, as well as to external shocks to asset-markets. One important aspect of the interaction between the housing market and the macroeconomy arises from the link to the labour market. For example, Bover, Maullbauer and Murphy (1989) argue that differences in the level of house prices between regions lowers labour mobility. See also Meen (2002).

House prices at the regional level also exhibit much more volatility both over time and across regions. Pollakowski and Ray (1997) examine the spatial diffusion process of the change in the log of US regional house prices, using vector autoregressive (VAR) models. Their results suggest that at the national level (dividing US into nine regions) evidence confirms the significance of some (non spatial) diffusion patterns, but at the metropolitan area level there is evidence of contiguous spatial diffusion.

Recently Cameron *et al.* (2006) have investigated the evolution of house prices across nine UK regions between 1972 and 2003. Their model of real house prices is an error correction panel data model with regional effects and time effects, derived from a system of inverted housing demand equations, including additional terms to take into account spatial correlation, as well as supply side effects, credit conditions, etc. Specifically, a seemingly unrelated regression (SUR) estimation approach is adopted to control for contemporaneous correlation, and lagged dependent variables of contiguous regions and Greater London are also included to control for the spatial diffusion process. This estimation strategy, apart from its rather complex structure, can not be applied when the number of regions is relatively large, as is the case in the present study.

Housing is typically a largely immovable asset. However, as an asset it is reasonable to assume that at the margin the price of two identical houses located spatially will differ only by a (fixed) factor which reflects general aspects of physical location. In this paper the fundamental driver of real house prices is real income. However, there is considerable differences among US States in both the level and rates of growth of real incomes.² This heterogeneity should in turn be reflected in real house prices. We examine this possibility in the context of a panel error correction model (Malpezzi, 1999, Capozza *et al.*, 2002, and Gallin, 2003). The importance of heterogeneity in spatially distributed housing markets has been highlighted recently 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

²For a recent review of the US housing market see Green and Malpezzi (2003).

effects on particular US regions (Carlino and DeFina, 1998).

The plan of the paper is 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. A standard model of the demand for housing services includes permanent income, the real price of housing services and a set of other influences affecting changes in household formation such as demographic shifts. In equilibrium the real price of houses, P_h/P_g , is equal to the real price of household services from home ownership, s , divided by the user cost of housing, c :

$$P_h/P_g = s/c,$$

where P_g is a general price index. Assume that alternative assets are taxed at the rate τ . c is then equal to the expected real, after-tax rate of return on other assets with a similar degree of risk:

$$c = (r + \pi)/(1 - \tau) - \pi^e, \quad (2.1)$$

where r is the risk-equivalent real interest rate on alternative assets and π^e is the expected rate of price inflation. Following the approach of Feldstein *et al.* (1978), Hendershott and Hu (1981) and Buckley and Ermisch (1982), assume that the alternative asset is some aggregate capital which can be financed by the issue of equity or the sale of bonds. The bonds are of an equivalent degree of risk to house ownership. Equity is riskier, so there is a market determined risk premium, ρ , on the holding of equity. In equilibrium the risk adjusted return on equity, ε , is equal to the return on bonds:

$$(1 - \tau)\varepsilon - \tau\pi - \rho = (r + \pi)/(1 - \tau) - \pi^e. \quad (2.2)$$

The return to equity is expressed as the dividend payout per unit of equity.

Another way of deriving the user cost of housing is to use the full intertemporal model of consumption in which in equilibrium the marginal rate of substitution between housing services and the flow of utility from consumption is:

$$u_h/u_c = (P_h/P_g) \{(r + \pi)/(1 - \tau) - \pi^e - \Delta^e(P_h/P_g)\}, \quad (2.3)$$

where $\Delta^e(P_h/P_g)$ denotes the expected appreciation in the real price of houses and u_h and u_c are the marginal utilities of housing services and consumption, respectively. The price of houses that satisfies the market for housing services and the asset market arbitrage condition is then:

$$P_h/P_g = s/\{(r + \pi)/(1 - \tau) - \pi^e - \Delta^e(P_h/P_g)\}. \quad (2.4)$$

The empirical model that can be derived from this form of analysis employs the device of proxying the unobservable real rental price of the flow of housing services, s , by the determinants of the demand for housing services, such as real disposable income.

3 The Econometric Model and Tests

The long-run relation compatible with the theory can be written most conveniently in the following log-linear form:

$$p_{it} = \alpha_i + \beta_i' \mathbf{x}_{it} + u_{it}, i = 1, 2, \dots, N; t = 1, 2, \dots, T, \quad (3.1)$$

where $p_{it} = \log(P_{it,h}/P_{it,g})$ is the logarithm of real price of housing in the i^{th} State, $\mathbf{x}_{it} = (y_{it}, rl_{it})'$, y_{it} is real personal disposable income in the i^{th} State, and rl_{it} is the logarithm of real interest rate. The price dynamics and their adjustments to the long-run equilibrium across States are captured in the error terms u_{it} . The common unobserved factors as well as the spatial effects will also be modelled through the error terms. In particular, we shall assume that u_{it} has the following multi-factor structure

$$u_{it} = \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \quad (3.2)$$

in which \mathbf{f}_t is an $m \times 1$ vector of unobserved common effects, and ε_{it} are the individual-specific (idiosyncratic) errors assumed to be distributed independently of \mathbf{x}_{it} and \mathbf{f}_t . However, we allow ε_{it} s to be weakly dependent across i . This, for example, allows the idiosyncratic errors to follow the Spatial Autoregressive (SAR), or the Spatial Moving Average (SMA) processes introduced by Whittle (1954), Cliff and Ord (1973, 1981), and Haining (1978).

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. Some of the supply factors that are particularly difficult to measure accurately can be captured through the unobserved common components of u_{it} .³ The specification also allows for the possible effects of the transmission of monetary policy onto house prices at the aggregate level. We are also able to test for cointegration between real house prices and real disposable income, whilst allowing for a high degree of dependence across States.

3.1 The Common Correlated Effects (CCE) Estimator

We use the Common Correlated Effects (CCE) type estimator, which asymptotically eliminates the cross dependence (Pesaran 2006a). To illustrate, suppose the $(k \times 1)$ vector \mathbf{x}_{it} is generated as

$$\mathbf{x}_{it} = \mathbf{a}_i + \mathbf{\Gamma}_i' \mathbf{f}_t + \mathbf{v}_{it}, \quad (3.3)$$

where \mathbf{a}_i is a $k \times 1$ vector of individual effects, $\mathbf{\Gamma}_i$ is a $m \times k$ 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) and (3.3) we now have

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

³The 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 Gyourno (2005) and Glaeser, Gyourko and Saks (2005) as a significant element in real house price increases in California, Massachusetts, New Hampshire, New Jersey and Washington, DC.

where

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

$$\mathbf{d}_i = \begin{pmatrix} 1 & \boldsymbol{\beta}'_i \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \begin{pmatrix} \alpha_i \\ \mathbf{a}_i \end{pmatrix}, \quad \mathbf{C}_i = \begin{pmatrix} \gamma_i & \boldsymbol{\Gamma}_i \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \quad (3.6)$$

\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

$$\tilde{\boldsymbol{\Gamma}}_i = \begin{pmatrix} \gamma_i & \boldsymbol{\Gamma}_i \end{pmatrix}. \quad (3.7)$$

Pesaran (2006a) 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)⁴

$$\bar{\mathbf{z}}_t = \bar{\mathbf{d}} + \bar{\mathbf{C}}' \mathbf{f}_t + \bar{\mathbf{v}}_t, \quad (3.8)$$

where

$$\bar{\mathbf{z}}_t = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_{it}, \quad \bar{\mathbf{v}}_t = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_{it},$$

and

$$\bar{\mathbf{d}} = \frac{1}{N} \sum_{i=1}^N \mathbf{d}_i, \quad \bar{\mathbf{C}} = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_i. \quad (3.9)$$

Suppose that

$$\text{Rank}(\bar{\mathbf{C}}) = m \leq k+1, \text{ for all } N. \quad (3.10)$$

Then, we have

$$\mathbf{f}_t = \left(\bar{\mathbf{C}} \bar{\mathbf{C}}' \right)^{-1} \bar{\mathbf{C}} (\bar{\mathbf{z}}_t - \bar{\mathbf{d}} - \bar{\mathbf{v}}_t). \quad (3.11)$$

But since

$$\bar{\mathbf{v}}_t \xrightarrow{q.m.} \mathbf{0}, \text{ as } N \rightarrow \infty, \text{ for each } t, \quad (3.12)$$

and

$$\bar{\mathbf{C}} \xrightarrow{p} \mathbf{C} = \tilde{\boldsymbol{\Gamma}} \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{I}_k \end{pmatrix}, \text{ as } N \rightarrow \infty, \quad (3.13)$$

where

$$\tilde{\boldsymbol{\Gamma}} = (E(\gamma_i), E(\boldsymbol{\Gamma}_i)) = (\boldsymbol{\gamma}, \boldsymbol{\Gamma}). \quad (3.14)$$

It follows, assuming that $\text{Rank}(\tilde{\boldsymbol{\Gamma}}) = m$, that

$$\mathbf{f}_t - (\mathbf{C} \mathbf{C}')^{-1} \mathbf{C} (\bar{\mathbf{z}}_t - \bar{\mathbf{d}}) \xrightarrow{p} \mathbf{0}, \text{ as } N \rightarrow \infty.$$

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 (2006a). Kapetanios *et al.* (2006) 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 $\tilde{\boldsymbol{\Gamma}}$ turns out to be rank deficient, thus the estimator is consistent with any fixed number of m . This contrasts to the principal

⁴Pesaran (2006a) considers cross section weighted averages that are more general. However, we used the weights N^{-1} , which is asymptotically ($N \rightarrow \infty$) equivalent to other weights.

component approach, which requires us to estimate the number of factors (Bai and Ng, 2002, and Bai, 2003).

The CCEMG estimator is a simple average of the individual CCE estimators, $\hat{\mathbf{b}}_i$ of β_i defined by

$$\hat{\mathbf{b}}_{MG} = N^{-1} \sum_{i=1}^N \hat{\mathbf{b}}_i, \quad (3.15)$$

$$\hat{\mathbf{b}}_i = (\mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i)^{-1} \mathbf{X}'_i \bar{\mathbf{M}} \mathbf{p}_i, \quad (3.16)$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})'$, $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{iT})'$, and $\bar{\mathbf{M}}$ is defined by

$$\bar{\mathbf{M}} = \mathbf{I}_T - \bar{\mathbf{H}} (\bar{\mathbf{H}}' \bar{\mathbf{H}})^{-1} \bar{\mathbf{H}}', \quad (3.17)$$

and $\bar{\mathbf{H}} = (\boldsymbol{\tau}_T, \bar{\mathbf{Z}})$, $\boldsymbol{\tau}_T$ is a $(T \times 1)$ vector of unity, $\bar{\mathbf{Z}}$ is a $T \times (k+1)$ matrix of observations $\bar{\mathbf{z}}_t$. The (non-parametric) variance estimator for $\hat{\mathbf{b}}_{MG}$ is given by

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

Efficiency gains from pooling of observations over the cross section units can be achieved when the individual slope coefficients, β_i , are the same. Such a pooled estimator of β , denoted by CCEP, has been developed by Pesaran (2006a) and is given by

$$\hat{\mathbf{b}}_P = \left(\sum_{i=1}^N \mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \bar{\mathbf{M}} \mathbf{p}_i. \quad (3.18)$$

The variance estimator for $\hat{\mathbf{b}}_P$ is given by

$$\widehat{Var}(\hat{\mathbf{b}}_P) = N^{-1} \hat{\Psi}^{*-1} \hat{\mathbf{R}}^* \hat{\Psi}^{*-1}, \quad (3.19)$$

where

$$\hat{\Psi}^* = N^{-1} \sum_{i=1}^N \left(\frac{\mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i}{T} \right), \quad (3.20)$$

$$\hat{\mathbf{R}}^* = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{\mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i}{T} \right) (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG}) (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{MG})' \left(\frac{\mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i}{T} \right). \quad (3.21)$$

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). 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 $N \rightarrow \infty$. The CD test statistic is defined as

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij} \right) \stackrel{a}{\sim} N(0, 1), \quad (3.22)$$

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

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^T \hat{u}_{it} \hat{u}_{jt}}{\left(\sum_{t=1}^T \hat{u}_{it}^2\right)^{1/2} \left(\sum_{t=1}^T \hat{u}_{jt}^2\right)^{1/2}}, \quad (3.23)$$

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

$$\hat{u}_{it} = p_{it} - \hat{\alpha}_i - \hat{\beta}'_i \mathbf{x}_{it}, \quad (3.24)$$

with $\hat{\alpha}_i$ and $\hat{\beta}_i$ being the estimates of α_i and β_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, Pesaran and Shin (2003), called the IPS test. Consider a p^{th} order Augmented Dickey-Fuller (ADF) regression for the i^{th} cross section unit:

$$\Delta\omega_{it} = a_i + \lambda_i t + b_i \omega_{i,t-1} + \sum_{j=1}^p \delta_{ij} \Delta\omega_{i,t-j} + v_{it}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T. \quad (3.25)$$

The unit root hypothesis of interest can be expressed as

$$H_0 : b_i = 0 \text{ for all } i,$$

against the possibly heterogeneous stationary alternatives,

$$H_1 : b_i < 0, \quad i = 1, 2, \dots, N_1, \quad b_i = 0, \quad i = N_1 + 1, N_1 + 2, \dots, N.$$

It is assumed that N_1/N is non-zero and tends to a fixed constant δ such that $0 < \delta \leq 1$ as $N \rightarrow \infty$, which is necessary for the consistency of the panel unit root test. The individual p^{th} order ADF statistic, $ADF(p)$, is obtained as the OLS t-ratio of b_i , \hat{t}_i , and the IPS statistic is defined as⁵

$$IPS = \frac{\sqrt{N} \{t\text{-bar} - E[\hat{t}_i | b_i = 0, p]\}}{\sqrt{Var[\hat{t}_i | b_i = 0, p]}}, \quad (3.26)$$

where

$$t\text{-bar} = N^{-1} \sum_{i=1}^N \hat{t}_i. \quad (3.27)$$

The values of $E[\hat{t}_i | b_i = 0, p]$ and $Var[\hat{t}_i | b_i = 0, p]$ with various combinations of T and p are tabulated in Im, Pesaran and Shin (2003). Under the null hypothesis, the distribution of the IPS statistic is well approximated by the standard normal distribution for sufficiently large N and T .

However, the IPS test procedure is not valid when the errors, v_{it} , are dependent across i , 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 section dependence in panels

⁵The dependence of the statistics on N and/or T are suppressed for ease of notation.

have been recently proposed in the literature.⁶ Here we consider two of these that are particularly relevant to our application. One is 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. For de-factoring, they make use of principal components estimator of m common factors, where the number of factors m is estimated using the model selection criteria proposed in Bai and Ng (2002). In their paper two test statistics, t_a^* and t_b^* , are proposed. Here only the latter is considered, which was reported to have better finite sample performance in Moon and Perron (2004). Following Moon and Perron (2004) the long-run variances are estimated using Andrews and Monahan (1992) estimator based on a quadratic spectral kernel and prewhitening. Under the null, the statistic t_b^* tends to a standard normal variate as both N and T go to infinity so long as $N/T \rightarrow 0$.

The second panel unit root test considered in our application is the test proposed by Pesaran (2006b), which follows the CCE approach and filters out the cross section dependence by augmenting the ADF regressions with cross section averages. The cross section augmented ADF (CADF) regressions, carried out separately for each State, are given by

$$\Delta\omega_{it} = a_i + \lambda_i t + b_i \omega_{i,t-1} + c_i \bar{\omega}_{t-1} + \sum_{j=0}^p d_{ij} \Delta \bar{\omega}_{t-j} + \sum_{j=1}^p \delta_{ij} \Delta \omega_{i,t-j} + v_{it}, \quad (3.28)$$

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

$$CIPS = N^{-1} \sum_{i=1}^N \tilde{t}_i, \quad (3.29)$$

where \tilde{t}_i is the OLS t-ratio of b_i in the above CADF regression. Pesaran (2006b) also considers a truncated version of \tilde{t}_i so that these statistics have finite moments even for relatively small N and/or T . The truncated version of the CIPS statistic is defined as $CIPS^* = N^{-1} \sum_{i=1}^N \tilde{t}_i^*$, where \tilde{t}_i^* are truncated CADF statistics such that

$$\begin{cases} \tilde{t}_i^* = \tilde{t}_i, & \text{if } -K_1 < \tilde{t}_i < K_2, \\ \tilde{t}_i^* = -K_1, & \text{if } \tilde{t}_i \leq -K_1, \\ \tilde{t}_i^* = K_2, & \text{if } \tilde{t}_i \geq K_2, \end{cases}$$

K_1 and K_2 are positive constants that are sufficiently large so that $\Pr(-K_1 < \tilde{t}_i < K_2)$ is sufficiently large. The choice of the value of K_1 and K_2 depends on the specification of CADF's. For models with an intercept only, $K_1 = 6.19$ and $K_2 = 2.61$, and for models with an intercept and a linear trend, $K_1 = 6.42$ and $K_2 = 1.70$, are used. The critical values for the CIPS tests are given in Tables 2a-2c in Pesaran (2006b).⁷

3.4 Cointegration between Real House Prices and Real Incomes

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.

⁶For a recent survey of the literature see Breitung and Pesaran (2006).

⁷As Pesaran (2006b) notes, the CADF(p) approach is valid when there is single common factor.

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, Lin and Chu (2002, LLC) to house price-to-income ratios, and finds that he can not 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.⁸ 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 section dependence of house prices and this could seriously bias the test results.

Capozza, *et al.* (2002) recognize this problem and try to control for cross section dependence by adding time dummies to their error correction specifications. However, as Gallin (2003) 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 section dependence, Gallin (2003) adopts a 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 seriously 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 N is small relative to T . Also see Smith, Leybourne, Kim and Newbold (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 it ensures that the maximum eigenvalue of the error variance-covariance matrix remains bounded in N , by design.⁹ 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. 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 Groen and Kleibergen (2003), Nelson, Ogaki, and Sul (2005), Westerlund (2005), Pedroni and Vogelsang, (2005), Chang (2005), and Bai and Kao (2006). The tests by Groen and Kleibergen, Nelson, Ogaki, and Sul, and Westerlund are applicable when N is small and T 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 Pedroni and Vogelsang and Chang, in principle, can deal with panels where N is reasonably large, but their models do not allow for unobserved common factors, \mathbf{f}_t , that could be correlated with the observed regressors, \mathbf{x}_{it} , which is an important consideration in our application. When \mathbf{f}_t and

⁸It 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 equation (2) and the discussions on pages 42 and 48 in Malpezzi (1999).

⁹See, 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 .

\mathbf{x}_{it} are correlated the first stage residuals used in the tests by Chang and Pedroni and Vogelsang will be inconsistent which invalidates their residual-based testing procedures. Bai and Kao model cross section dependence using the factor approach as in (3.2), but assume that the innovations to the factors and the observed regressors are independently distributed. (see their Assumption 2). However, as Bai and Kao note in their Remark 1, it is possible to relax their Assumption 2 by allowing the innovations in \mathbf{x}_{it} to be correlated with \mathbf{f}_t . For a more detailed review of this emerging literature see, for example, Breitung and Pesaran (2006).

We follow Chang, Pedroni and Vogelsang, and Bai and Kao and 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 be potentially correlated with the observed regressors. Using the pooled CCE estimator we first estimate the residuals, $\hat{u}_{it} = p_{it} - \hat{\beta}_{CCE}y_{it} - \hat{\alpha}_i$. As noted earlier, the pooled estimate, $\hat{\beta}_{CCE}$ is consistent for β , under fairly general assumptions about the unobserved common factors, \mathbf{f}_t . We then apply panel unit root tests to these residuals. 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 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 US 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 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. 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. The results are shown 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 coefficient. 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 Mideast 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 3 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 the USA.

In Tables 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 Middle 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 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 unusual 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.

5 Econometric Evidence

In this section we turn to the test and estimation results.

5.1 Panel Unit Root Tests Results

The extent of cross section dependence of the residuals from $ADF(p)$ regressions of real house prices, real incomes and interest rates across the 49 States over the period 1975 to 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 $\bar{\rho}$. 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 real interest rates. The results are reasonably robust to the choice of the augmentation order, p . For real incomes and real interest rates, $\bar{\rho}$ is estimated to be around 40% and 50%, 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 interest rates as compared to real house prices that are likely to be affected by State specific effects as well. 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 the IPS test, that do not allow for error cross section dependence. Therefore, in what follows we shall focus on the Moon-Perron's t_b^* and Pesaran's *CIPS* tests.

The results for Moon and Perrons's t_b^* test are summarized in Table 6. 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 in Table 6 we present the t_b^* test

results for various values of m in the range of 1 to 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. Only for real interest rates do the test results convincingly reject the unit root hypothesis.

The CIPS test results, summarized in Table 7, show a similar outcome for the real interest rates. But for p_{it} and y_{it} the unit root hypothesis can not 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. As the Moon and Perron test is valid only when T is much larger than N , we believe the CIPS test results are more reliable for our data. We proceed taking y_{it} and p_{it} as $I(1)$, and rl_{it} as $I(0)$ variables.

5.2 The Income Elasticity of Real House Prices

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

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

where

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

In view of discussion in Section 3, the common correlated effects (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, 2006a, Kapetanios *et al.*, 2006). To show the importance of allowing for the unobserved common factors in this relationship we also provide naive estimates of β_i , $i = 1, 2, \dots, N$ (and their mean) that do not allow for cross section 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

$$p_{it} = \alpha_i + \beta_i y_{it} + g_{i0} \bar{y}_t + g_{i1} \bar{p}_t + e_{it}, \quad (5.3)$$

where \bar{y}_t and \bar{p}_t denote the cross section averages of y_{it} and p_{it} in year t . The results are reported in Table 8. The first column gives the naive mean group estimates. These suggest a small coefficient on income of only 0.30, and considerable cross sectional dependence. 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 correlation coefficient, $\bar{\rho}$, reduced from 0.38 for the MG estimates to 0.024 and 0.003 respectively for the CCEMG and CCEP estimates.

The CCEMG and CCEP estimates and their standard errors also support the hypothesis of a unit elasticity between changes in house prices and real incomes. The t-ratios of both β_{CCE} estimates in Table 8 are less than unity for the null hypothesis of interest. 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} can now be used to test for 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 section dependence of the residuals. The extent to which these residuals are cross-sectionally dependent can be seen from the average cross correlation coefficients of \hat{u}_{it} , within and between the eight BEA regions, which are reported in Table 9.

We computed CIPS(p) panel unit root test statistics for $p_{it}-y_{it}$, including an intercept, 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 and panels 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.

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:

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

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 section dependence in the errors, v_{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 section dependence. The former estimates as computed by OLS regressions of Δp_{it} on $1, (p_{i,t-1} - y_{i,t-1}), \Delta p_{i,t-1}, \Delta y_{it}$, 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 10. The coefficients are all correctly signed. The CCEMG and CCEP estimators are very close and yields error correction coefficients (-0.183 and -0.171) 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.

To check for the effect of real interest rates on real house prices we also estimated the above ECM with $rl_{i,t-1}$ included. (see Table 11).¹⁰ We found a small negative effect from real interest rates on real house prices, but the effect was not statistically significant at the conventional levels. This could partly be due to the fact that $rl_{i,t-1}$ is defined in terms of a common nominal rate of interest and the cross section variations of $rl_{i,t-1}$ is limited to the differences of the lagged inflation rates across the States.¹¹ A cross country analysis of house prices might be needed for a more reliable estimate of the effects of real

¹⁰Since the cross section variations of $rl_{i,t-1}$ is due solely to the cross section variations of $\Delta p_{i,t-1}$, cross section averages of $rl_{i,t-1}$ are not included in the panel regressions.

¹¹We also tried contemporaneous values of rl_{it} in the panel regressions and found even weaker results.

interest rates on real house prices, as such a study is likely to yield a much higher degree of cross section variations in real interest rates.

5.5 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 i^{th} 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 the cointegrating relationship $(p_{it} - y_{it})$ on $(\bar{p}_t - \bar{y}_t)$ and a constant. These regressions are reminiscent of the Capital Asset Pricing Models (CAPM) in finance where individual asset returns are regressed on market (or average) returns.

The results, summarized in Table 12, 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 loadings greater than one on the common factor. 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, Rhode Island, 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 building, have resulted in real house prices deviating from the average across US States.

5.6 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 $\hat{u}_{it} = p_{it} - y_{it} - \hat{\alpha}_i$. We want to distinguish between a common dependence which is captured by the common factors in (5.2) and idiosyncratic components in ε_{it} . These idiosyncratic factors reflect forms of local dependence that are spatial in nature. Initially we applied all six information criteria (IC) for the estimation of the number of factors proposed by Bai and Ng (2002), with the maximum number of factors varied between 1 to 8. Similar to the results in Table 7, all the criteria suggested the same maximum number of factors. In view of this, we analysed \hat{u}_{it} , instead, by principal components for $m = 1, 2$ and 3 (following Bai (2003, p.140), for example) so that

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

where $\tilde{f}_{\ell t}$, $\ell = 1, 2, \dots, m$ are the extracted factors and $\tilde{\gamma}_{i\ell}$ are the associated factor loadings. 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 to 2003 for each i .

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)

$$\tilde{\varepsilon}_{it} = \psi \sum_{j=1}^N w_{ij} \tilde{\varepsilon}_{jt} + \nu_{it}, \quad (5.6)$$

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

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

where $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_{1t}, \tilde{\varepsilon}_{2t}, \dots, \tilde{\varepsilon}_{Nt})'$, and in our application $N = 49$ and $T = 29$.¹² For \mathbf{W} , following the approach of Anselin, we used a contiguity criterion and assigned $w_{ij} = 1$ when State i and j share a common border or vertex, and $w_{ij} = 0$ otherwise.¹³ The 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.¹⁴ All the estimates are highly significant and as is to be expected, the magnitude of the spatial parameter declines markedly with the number of factors. Nevertheless, even with 3 factors there is strong evidence that local dependence in the form of a spatial dependence between contiguous states in the US 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 standard residuals defined by $\varepsilon_{it}^* = \tilde{\varepsilon}_{it}/s_i$, where $s_i = \sqrt{\sum_{t=1}^T \tilde{\varepsilon}_{it}^2/T}$. 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 US, even after controlling for State specific real incomes, and allowing for a number of unobserved common factors.

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, 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 3 common factors and find significant evidence of spatial dependence associated with contiguity.

¹²For computation details of maximum likelihood estimation, see Anselin *et al.* (2006) and references therein.

¹³The data on contiguity are obtained from Luc Anselin's web site at: <http://sal.uiuc.edu/weights/index.html>.

¹⁴We 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.

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.

References

- Andrews, D. and Monahan, C., (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60, 953-966.
- Anselin, L., (1988). *Spatial Econometrics: Methods and Models*, Boston: Kluwer Academic Publishers.
- Anselin, L., (2001). Spatial econometrics. In Baltagi, B. (Ed.), *A Companion to Theoretical Econometrics*, Oxford: Blackwell.
- Anselin, L., Le Gallo, J. and Jayet, H., (2006). Spatial panel econometrics. In Matyas, L. and Sevestre, P. (Eds.), *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*, (3rd Edition), Dordrecht, Kluwer, forthcoming.
- Bai, J., (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71, 135-173.
- Bai, J. and 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 (Contributions to Economic Analysis, Volume 274)*. Elsevier.
- Bai, J. and Ng, S., (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70, 161-221.
- Bover, O., Muellbauer, J. and Murphy, A., (1989). Housing, wages and the UK labour market. *Oxford Bulletin of Economics and Statistics*, 51, 97-136.
- Breitung, J. and Pesaran, M.H., (2006). Unit roots and cointegration in panels. Forthcoming in Matyas, L. and Sevestre, P. (Eds.), *The Econometrics of Panel Data* (Third Edition), Kluwer Academic Publishers.
- Buckley, R. and Ermisch, J., (1982). Government policy and house prices in the United Kingdom: An economic analysis. *Oxford Bulletin of Economics and Statistics*, 44, 273-304.
- Cameron, G., Muellbauer, J. and Murphy, A., (2006). Was there a British house price bubble?: Evidence from a regional panel. Mimeo, University of Oxford.
- Capozza, D.R., Hendershott, P.H., Mack, C. and Mayer, C.J., (2002). Determinants of real house price dynamics. *NBER Working Paper*, 9262.
- Carlino, G. and DeFina, R., (1998). The differential regional effects of monetary policy. *Review of Economics and Statistics*, 80, 9001-9921.
- Case, K.E. and 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. and Ord, J.K., (1973). *Spatial Autocorrelation*, London: Pion.

Cliff, A. and Ord, J.K., (1981). *Spatial Processes, Models and Applications*, London: Pion.

Conley, T.G. and Dupor, B., (2003). A spatial analysis of sectoral complementarity. *Journal of Political Economy*, 111, 311-352.

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

Eff, E.A., (2004). Spatial, cultural and ecological autocorrelation in U.S. regional data. Working Paper, Middle Tennessee State University.

Feldstein, M., Green, J. and Sheshinski, E., (1978). Inflation and taxes in a growing economy with debt and equity finance. *Journal of Political Economy*, 86, 53-70.

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

Gallin, J., (2003). The long run relationship between house prices and income: evidence from local housing markets. Federal Reserve Board.

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

Glaeser, E.L., Gyourko, J. and Saks, R.E., (2005). Urban growth and housing supply. Harvard Institute of Economic Research, Discussion Paper Number 2062.

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

Groen, J.J.J. and 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: Cambridge University Press.

- Hendershott, P.H. and 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, H. and 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. and Yamagata, T., (2006). Panels with nonstationary multifactor error structures. *Cambridge Working Papers in Economics* 0651, University of Cambridge.
- Kelejian, H.H. and Prucha, I.R., (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40, 509-533.
- Krugman, O., (1998). Space: the final frontier. *Journal of Economic Perspectives*, 12, 161-174.
- Levin, A., Lin, F. and Chu, C., (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, 108, 1-24.
- Maclennan, D, Muellbauer, J. and Stephens, M., (1998). Asymmetries in housing and financial market institutions and EMU. *The Oxford Review of Economic Policy*, 14, 54-80.
- Maddala, G.S. and Wu, S., (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics, Special Issue*, 631-652.
- Malpezzi, S., (1999). A simple error correction model of house prices. *Journal of Housing Economics*, 8, 27-62.
- McCarthy, J. and Peach, R.W., (2004). Are home prices the next “bubble”? *Economic Policy Review*, 10, Federal Reserve Board of New York.
- Meen, G., (2002). The time-series behavior of house prices: A transatlantic divide? *Journal of Housing Economics*, 11, 1-23.
- Moon, H.R. and 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, 243-51.
- Paelinck, J.H.P and Klaasen, L.H., (1979). *Spatial Econometrics*, Saxon House.
- Nelson, M., Ogaki, M. and Sul, D., (2005). Dynamic seemingly unrelated cointegrating regressions. *The Review of Economic Studies*, 72, 797-820.

- 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. and 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. *CEifo Working Papers*, No.1233.
- Pesaran, M.H., (2006a). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74, 967-1012.
- Pesaran, M.H., (2006b). A simple panel unit root test in the presence of cross section dependence. mimeo, University of Cambridge.
- Pollakowski, H.O. and 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. and 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.

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
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, $y_{it} = \log[PD_{it}/(POP_{it} \times P_{it,g})]$
rl_{it}	US State real long term interest rate, $rl_{it} = RB_t - \ln(P_{it,g}/P_{it-1,g}) \times 100$,

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	NENG	Great Lakes Region	GLAK	Southwest Region	SWST
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	RKMT
		Plains Region	PLNS	Colorado	CO
Mideast Region	MEST	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	FWST
Pennsylvania	PA	South Dakota	SD	Alaska	AK
				California	CA
Southeast Region	SEST			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								
	NENG	MEST	SEST	GLAK	PLNS	SWST	RKMT	FWST
NENG	0.74	–	–	–	–	–	–	–
MEST	0.58	0.57	–	–	–	–	–	–
SEST	0.48	0.50	0.61	–	–	–	–	–
GLAK	0.54	0.56	0.70	0.85	–	–	–	–
PLNS	0.33	0.34	0.50	0.59	0.61	–	–	–
SWST	0.38	0.46	0.54	0.60	0.46	0.45	–	–
RKMT	0.24	0.38	0.44	0.51	0.39	0.49	0.48	–
FWST	0.51	0.51	0.56	0.66	0.44	0.50	0.41	0.68

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

**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								
	NENG	MEST	SEST	GLAK	PLNS	SWST	RKMT	FWST
NENG	0.80	–	–	–	–	–	–	–
MEST	0.68	0.66	–	–	–	–	–	–
SEST	0.40	0.32	0.52	–	–	–	–	–
GLAK	0.40	0.35	0.57	0.81	–	–	–	–
PLNS	0.27	0.20	0.53	0.62	0.61	–	–	–
SWST	0.07	–0.05	0.35	0.28	0.39	0.52	–	–
RKMT	–0.03	–0.11	0.40	0.52	0.53	0.57	0.70	–
FWST	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 ($\hat{\rho}$)				
	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
rl_{it}	0.509	0.509	0.469	0.485

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
rl_{it}	85.46	85.50	78.71	81.44

Notes: p^{th} -order Augmented Dickey-Fuller test statistics, $ADF(p)$, for y_{it} , p_{it} and rl_{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 rl_{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. $\bar{\rho} = [2/N(N-1)] \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij}$ with $\hat{\rho}_{ij}$ being the correlation coefficient of the $ADF(p)$ regression residuals between i^{th} and j^{th} 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 section dependence.

Table 6: Moon and Perron t_b^* Panel Unit Root Test Results

m	1	2	3	4
With an intercept only				
Δy_{it}	-21.90*	-24.27*	-29.67*	-27.85*
Δp_{it}	-14.51*	-15.34*	-15.03*	-14.13*
y_{it}	6.85	3.29	4.57	3.73
p_{it}	-1.06	-0.15	0.73	0.03
rl_{it}	-10.22*	-9.39*	-9.56*	-8.39*
With a Linear Trend				
y_{it}	-2.49*	-1.92*	-5.39*	-8.35*
p_{it}	-0.48	-2.17*	-4.28*	-5.00*

Notes: The t_b^* test is the Moon and Perron (2004) panel unit root test statistic. The t_b^* statistic is computed for a given number of factors, $m = 1, 2, 3, 4$. The long-run variances are estimated using the Andrews and Monahan (1992) estimator, using a quadratic spectral kernel and prewhitening. Under the null, the t_b^* statistic tends to a standard normal distribution as both N and T go to infinity such that $N/T \rightarrow 0$. The 5% critical value (one-sided) is -1.645. The superscript “*” signifies the test is significant at the five per cent level.

Table 7: Pesaran’s CIPS Panel Unit Root Test Results

With an Intercept				
	CADF(1)	CADF(2)	CADF(3)	CADF(4)
Δy_{it}	-2.61*	-2.39*	-2.42*	-2.34*
Δp_{it}	-2.28*	-1.86	-1.76	-1.81
y_{it}	-2.52*	-2.44*	-2.39*	-2.49*
p_{it}	-2.56*	-2.44*	-2.83*	-2.84*
rl_{it}	-3.53*	-2.69*	-2.05	-1.87
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 2006b); see Section 3 for more details. The relevant lower 5 per cent critical values for the CIPS statistics are -2.11 with an intercept case, and -2.62 with an intercept and a linear trend case. The superscript “*” signifies the test is significant at the five per cent level.

Table 8: Estimation Result: Income Elasticity of Real House Price

	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 ($\bar{\rho}$)	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 parenthesis; 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 $\bar{\rho} = [2/N(N-1)] \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij}$, 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 $[TN(N-1)/2]^{1/2} \bar{\rho}$, which tends to $N(0,1)$ under the null hypothesis of no error cross section dependence.

Table 9: Average Residual Cross Correlation Coefficients Within and Between Eight BEA Geographical Regions - $\hat{u}_{it} = p_{it} - y_{it} - \hat{\alpha}_i$,

	NENG	MEST	SEST	GLAK	PLNS	SWST	RKMT	FWST
NENG	0.62	–	–	–	–	–	–	–
MEST	0.56	0.45	–	–	–	–	–	–
SEST	0.18	0.29	0.95	–	–	–	–	–
GLAK	0.04	0.21	0.75	0.82	–	–	–	–
PLNS	0.10	0.24	0.92	0.80	0.93	–	–	–
SWST	0.07	0.17	0.90	0.63	0.87	0.94	–	–
RKMT	-0.23	-0.08	0.70	0.70	0.77	0.75	0.86	–
FWST	-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 10: Panel Error Correction Estimates
without Long-Term Interest Rate**

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
Average R^2	0.54	0.70	0.66
Average Cross Correlation Coefficients ($\bar{\rho}$)	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 parenthesis. The Average R^2 is computed as $1 - \sum_{i=1}^N \hat{\sigma}_i^2 / \sum_{i=1}^N \hat{\sigma}_{\Delta p_{it}}^2$, where $\hat{\sigma}_i^2$ are estimated error variances and $\hat{\sigma}_{\Delta p_{it}}^2$ are sample variances of Δp_{it} , for each State i . 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 8.

**Table 11: Panel Error Correction Estimates
with Long-Term Interest Rate**

Δp_{it}	MG	CCEMG	CCEP
$p_{i,t-1} - y_{i,t-1}$	-0.114 (0.009)	-0.212 (0.016)	-0.174 (0.017)
$\Delta p_{i,t-1}$	0.494 (0.031)	0.390 (0.041)	0.518 (0.069)
Δy_{it}	0.536 (0.046)	0.244 (0.063)	0.236 (0.057)
$rl_{i,t-1}$	-0.0004 (0.0006)	-0.0014 (0.0008)	-0.0012 (0.0007)
Half Life	5.727	3.072	3.626
Average R^2	0.54	0.72	0.66
Average Cross Correlation Coefficients ($\bar{\rho}$)	0.291	0.005	-0.011
CD Test Statistics	51.94	0.87	-1.93

See the notes to Table 10. Since the cross section variations of $rl_{i,t-1}$ is due solely to the cross section variations of $\Delta p_{i,t-1}$ (already included in the model), cross section averages of $rl_{i,t-1}$ were not included in the regressions for the CCE type estimates.

Table 12: Factor Loading Estimates

$(p_{it} - y_{it})$	$(\bar{p}_t - \bar{y}_t)$	Constant
Connecticut	0.35 (0.23)	1.42
Maine	0.29* (0.15)	1.90
Massachusetts	-0.63* (0.24)	3.99
New Hampshire	0.81* (0.22)	0.51
Rhode Island	-0.11 (0.24)	2.76
Vermont	0.78* (0.15)	0.67
Delaware	0.32* (0.11)	1.63
District of Columbia	0.54* (0.18)	0.82
Maryland	0.62* (0.10)	0.81
New Jersey	-0.04 (0.20)	2.37
New York	-0.39* (0.20)	3.35
Pennsylvania	0.65* (0.13)	0.88
Alabama	1.72* (0.09)	-1.54
Arkansas	1.77* (0.10)	-1.70
Florida	1.44* (0.08)	-1.09
Georgia	1.43* (0.08)	-0.89
Kentucky	1.21* (0.06)	-0.35
Louisiana	2.03* (0.15)	-2.45
Mississippi	2.09* (0.13)	-2.33
North Carolina	1.28* (0.05)	-0.50
South Carolina	1.39* (0.06)	-0.73
Tennessee	1.53* (0.08)	-1.14
Virginia	0.91* (0.09)	0.22
West Virginia	2.08* (0.11)	-2.47
Illinois	0.71* (0.11)	0.61
Indiana	1.14* (0.05)	-0.33
Michigan	0.54* (0.17)	1.01
Ohio	1.01* (0.09)	-0.08
Wisconsin	0.98* (0.12)	0.01
Iowa	1.55* (0.11)	-1.38
Kansas	1.76* (0.06)	-1.90
Minnesota	1.20* (0.09)	-0.59
Missouri	1.37* (0.04)	-0.90
Nebraska	1.57* (0.10)	-1.41
North Dakota	2.00* (0.15)	-2.44
South Dakota	1.39* (0.08)	-0.90
Arizona	1.02* (0.07)	-0.08
New Mexico	0.95* (0.12)	0.23
Oklahoma	2.10* (0.17)	-2.70
Texas	2.12* (0.18)	-2.77
Colorado	0.80* (0.17)	0.34
Idaho	1.19* (0.11)	-0.39
Montana	0.75* (0.16)	0.63
Utah	0.68* (0.19)	0.85
Wyoming	1.62* (0.18)	-1.73
California	-0.64* (0.23)	3.70
Nevada	0.84* (0.11)	0.16
Oregon	0.37 (0.25)	1.40
Washington	-0.12 (0.17)	2.51

Notes: Standard errors are given in parentheses. 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. The negative slope estimates are in bold, and statistically significant slopes are denoted by *.

A 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 A1. 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 rate, RB_t , which are simple annual averages of quarterly data, are taken from the Fair Model database.

Table A1: 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	–	–