Journal of Housing Economics 18 (2009) 140–149

Contents lists available at ScienceDirect

Journal of Housing Economics

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

Do house prices reflect fundamentals? Aggregate and panel data evidence

Vyacheslav Mikhed *,[1], Petr Zemčík [2]

CERGE-EI, P.O. Box 882, Politickych veznu 7, 111 21 Prague 1, Czech Republic

ARTICLE INFO

Article history:
Received 8 October 2007
Available online 3 May 2009

Keywords:
Panel data
House prices
Cointegration
Fundamentals
Bubble

ABSTRACT

We investigate whether recently high and consequently rapidly decreasing U.S. house prices have been justified by fundamental factors such as personal income, population, house rent, stock market wealth, building costs, and mortgage rate. We first conduct the standard unit root and cointegration tests with aggregate data. Nationwide analysis potentially suffers from problems of the low power of stationarity tests and the ignorance of dependence among regional house markets. Therefore, we also employ *panel data* stationarity tests which are robust to cross-sectional dependence. Contrary to previous panel studies of the U.S. housing market, we consider several, not just one, fundamental factors. Our results confirm that panel data unit root tests have greater power as compared with univariate tests. However, the overall conclusions are the same for both methodologies. The house price does not align with the fundamentals in sub-samples prior to 1996 and from 1997 to 2006. It appears that the real estate prices take long swings from their fundamental value and it can take decades before they revert to it. The most recent correction (a collapsed bubble) occurred around 2006.

1. Introduction

Prior to 2006, the possibility of a house price bubble in the U.S. housing market was an active topic of discussion in both the popular press and academic journals. This issue was of interest because a bursting bubble in a housing market could lead to a decrease in the value of household wealth. According to the 2004 Survey of Consumer Finances, primary and other residential property constituted almost 39% of the total assets in the portfolios of U.S. families (see Bucks et al., 2006). Therefore, a drop in house prices could result in a severe negative impact on consumption and GDP. Recent developments in the housing markets have confirmed that these worries had been justified.

Theoretical background for the use of various determinants of house prices can be found in Gallin (2006), Timmermann (1995), and Poterba (1984). We build on these studies, derive the housing price as a function of the underlying economic factors in both the present value and structural housing models, and explicitly illustrate the link between them. A house price bubble is then defined as a situation when a growth of the price is not supported by changes in its fundamentals (Stiglitz, 1990). There were two categories of papers which considered breaks in the relationship between house price and fundamentals. Papers in the first category argued about this issue using aggregate data. For example, McCarthy and Peach (2004) suggested that there was no bubble in the U.S. housing market and that changes in house prices reflected movements in personal income and nominal mortgage rates. Another example of this approach is Shiller (2005) or Gallin (2006) who used aggregate data on home prices, personal income, building costs, population, user costs of housing and interest rates. They showed that changes in fundamentals did not explain the rapid growth of U.S. house prices after 2000.

* Corresponding author.
E-mail addresses: vmikhed@cerge-ei.cz (V. Mikhed), petr.zemcik@cerge-ei.cz (P. Zemčík).

[1] CERGE-EI is a joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Academy of Sciences of the Czech Republic.

[2] Petr Zemčík would like to acknowledge the support from the Marie Curie International Reintegration Grant # 014864.

1051-1377/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jhe.2009.03.001

The present paper confirms the discrepancy between house prices and their determinants using similar data prior to 2006 and standard univariate unit root and cointegration tests. Adding two years of data with collapsing housing prices implies reversion to the fundamental factors. Our findings correspond to occurrence of three housing price peaks in 1979, 1989, and 2006, which have been aligned with fundamental factors' behavior only after the third price correction. Construction costs and income appear to be the driving forces of the real estate prices. Below we check whether panel data stationarity tests, which have greater power, are in line with these results.

The second stream of this literature relied on regional or micro data in order to get more insights into the behavior of the housing market. For example, Himmelberg et al. (2005) used their own calculations of owning costs of housing for 46 Metropolitan Statistical Areas (MSA) to argue that the high price-to-income and price-to-rent ratios observed in recent years were explained by shifts in real long-term interest rates and therefore there was no bubble in the U.S. housing market. Smith and Smith (2006) suggested that house prices were below their fundamental value derived from house rents where prices and rents were taken from a sample of matched single-family homes. Case and Shiller (2003) were more in favor of the existence of a speculative bubble in some regional U.S. housing markets based on the results of a survey of consumers' attitudes toward housing.

Finally, Gallin (2006) and Mikhed and Zemčík (2007) employed panel data for the U.S. MSA to analyze house prices. The former study used income and the latter rent as the only fundamental factor. Both studies employed panel data stationarity tests to find that house price dynamics could not be explained by either of the two variables. The omission of other potential demand and supply shifters on the housing market could be a reason for the lack of the relationship between the price and fundamentals at the regional level. We construct a panel with other fundamental variables to investigate this possibility. Our dataset contains series for house prices, rents, construction costs, income, population, stock index and mortgage rates. Real variables are calculated using a regional Consumer Price Index (CPI).

Individual time series in our panel are likely to be mutually correlated because close regional house markets tend to be synchronized to some extent. We confirm that cross-sectional dependence is present in our data using a test from Pesaran (2004). Then we test for unit roots in all of the involved series. Im et al. (2003) develop a panel unit root test based on an average of t-statistics for autoregressive coefficients in individual Dickey–Fuller regressions. We use an updated version of this test constructed in Pesaran (2007), which is robust to cross-sectional dependence. If house price dynamics reflects fundamentals, non-stationary house prices should be cointegrated with other variables. We implement the Pedroni (1999, 2004) statistic to test for panel data cointegration. We account for regional interdependence by bootstrapping critical values.

Using the panel data, we find that the house price series contains a unit root only prior to 2006. This finding is in contrast with the aggregate data unit root tests since these cannot reject the unit root in any sub-sample. It provides additional evidence that panel data unit root tests have greater power as compared with univariate methodology. Our results also show that house price is not cointegrated with any variables of the same order of integration. We observe the same pattern if we split the sample to the periods before and after 1996. The first sub-sample includes price peaks in 1979, and 1989 and the second sub-sample the recent rally (and fall) of the real estate prices. Therefore, there is a discrepancy between house prices and fundamentals before 2006 and the overall outcome of our panel data tests is consistent with findings using the aggregate data. The natural conclusion of our paper is that house prices swing away from fundamentals for extended periods of time. The most recent such period ended in 2006.

2. Structural model of the housing market

The present-value model may be a simple way to connect house prices to rents. Basically, this model implies that under rational expectations the price of an asset is equal to the discounted stream of expected future dividends. According to Gallin (2006), if one ignores taxes, maintenance costs, and risk premium associated with housing, the house price may be written as follows:

$$P_t = R_t + E_t\left[\frac{P_{t+1}(1-\delta)}{1+i_{t+1}}\right] \tag{1}$$

where P_t is the price of housing at time t, E_t is the expectation operator conditional on information available at date t, R_t is housing rent at time t, δ is a constant rate of depreciation, and i_{t+1} is time-variant rate of discounting.

Substituting the corresponding expressions for P_{t+1}, P_{t+2}, and so on into Eq. (1) and using the law of iterated expectations, it is possible to derive the following result:

$$\begin{aligned} P_t = E_t\Bigg[R_t &+ \frac{R_{t+1}(1-\delta)}{1+i_{t+1}} + \frac{R_{t+2}(1-\delta)^2}{(1+i_{t+1})(1+i_{t+2})} + \dots \\ &+ \frac{R_{t+k}(1-\delta)^k}{\prod_{j=1}^{k}(1+i_{t+j})} + \frac{P_{t+k+1}(1-\delta)^{k+1}}{\prod_{j=1}^{k+1}(1+i_{t+j})}\Bigg] \end{aligned} \tag{2}$$

Imposing a boundary condition

$$\lim_{k\to\infty} \frac{P_{t+k}(1-\delta)^k}{\prod_{j=1}^{k}(1+i_{t+j})} = 0 \tag{3}$$

we derive

$$P_t = E_t\left[R_t + \sum_{k=1}^{\infty} \frac{R_{t+k}(1-\delta)^k}{\prod_{j=1}^{k}(1+i_{t+j})}\right] \tag{4}$$

In the way similar to Timmermann (1995), this last equation may be transformed into

$$P_t = R_t(1+i_t)\left[\frac{1}{1+i_t} + E_t\sum_{k=1}^{\infty}\beta_t\prod_{j=1}^{k}\rho_{t+j}\beta_{t+j}\right] \tag{5}$$

where $\rho_{t+j} = (1-\delta)R_{t+k}/R_{t+k-1}$ and $\beta_{t+k} = 1/(1+i_{t+k})$. If we denote $\left[\frac{1}{1+i_t} + E_t\sum_{k=1}^{\infty}\beta_t\prod_{j=1}^{k}\rho_{t+j}\beta_{t+j}\right]$ as a_t then

$$P_t = R_t(1 + i_t)a_t. \quad (6)$$

Following Timmermann, we assume that a_t is stationary, it has first and second moments, and a_t and $R_t(1 + i_t)$ are statistically independent.[3]

A new variable S_t may be defined as follows:

$$S_t = \ln P_t - \ln R_t - \ln(1 + i_t). \quad (7)$$

This variable should be stationary because $S_t = \ln a_t$ and a_t is stationary. This implies that either $\ln P_t, \ln R_t$, and $\ln(1 + i_t)$ are stationary, or if one is not stationary, the others have to be not stationary as well and they all should be cointegrated. In other words, if the logarithm of house price is $I(1)$, then it should be cointegrated with the logarithms of rent and discount factor or there is a bubble.

One may extend this simple model in the way proposed by Gallin (2006). Similarly to this author, the inverse demand for services provided by a stock of housing can be as follows:

$$\ln R_t = \phi \ln K_t + \ln \pi_t \quad (8)$$

where K_t is a stock of housing which provides a proportional amount of services, and π_t is a vector of demand shifters. After taking logarithms from the both sides of Eq. (6) and plugging into it Eq. (8), the following result is derived:

$$\ln P_t = \phi \ln K_t + \ln \pi_t + \ln(1 + i_t) + \ln a_t \quad (9)$$

This equation implies that a consumer/investor should be indifferent between purchasing a house or renting a house. Similarly to Gallin (2006), it may be assumed that investment in the stock of housing is as follows:

$$\ln I_t = \kappa \ln P_t + \ln \varphi_t \quad (10)$$

where φ_t are housing supply shifters. The law of motion for capital in this model is

$$K_t = (1 - \delta)K_{t-1} + I_{t-1} \quad (11)$$

According to Poterba (1984), in the long run a steady state of capital growth is achieved when the housing stock is growing at a constant rate which could be zero. This implies that $K_t = (1 + \xi_k)K_{t-1}$, where ξ_k is the steady state growth rate of capital.

After this last expression is plugged into equation (11) and some rearrangements are done, the following expression can be derived

$$K_{t-1} = \frac{I_{t-1}}{\xi_k + \delta} \quad (12)$$

Shifting Eq. (12) one period forward and combining it with Eq. (10) yields

$$K_t = \frac{P_t^{\kappa} \varphi_t}{\xi_k + \delta} \quad (13)$$

which defines the steady state market equilibrium for the stock of housing.

Substituting the last equation into Eq. (9) results in

$$\ln P_t = \phi\kappa \ln P_t + \phi \ln \varphi_t - \phi \ln(\xi_k + \delta) + \ln \pi_t + \ln(1 + i_t) + \ln a_t. \quad (14)$$

Assuming that $\phi\kappa \neq 1$, Eq. (14) can be solved for $\ln P_t$ as follows:

$$\ln P_t = \frac{1}{1 - \phi\kappa}[\phi \ln \varphi_t - \phi \ln(\xi_k + \delta) + \ln \pi_t + \ln(1 + i_t) + \ln a_t]. \quad (15)$$

Once again, the theory suggests that a bubble in the price of housing may be identified if this price has a unit root, but housing demand and supply shifters are stationary or these shifters are not cointegrated with the price.

An important point of this study is the choice of supply and demand shifters. Most panel data studies of housing market bubbles concentrate on one fundamental only. This poses a problem because the no cointegration found in those studies may occur due to the ignorance of some fundamentals. In this study we are attempting to consider all the fundamentals regarded as important for a housing market and for which we could find panel data with reasonable cross-section and time dimensions. Our variables include house rent, construction costs, personal income, population, mortgage rates, and stock market wealth.

3. Data

There are two datasets used in this study. The first is the aggregate quarterly U.S. data for 1980:q2-2008:q2, and the second is annual data on 22 U.S. Metropolitan Statistical Areas for 1978–2007. The aggregate data on house prices comes from the quarterly repeat-sales price index of the Office of Federal Housing Enterprise Oversight (OFHEO).[4] The source of personal income is the Bureau of Economic Analysis. The U.S. Census Bureau provides a measure of population. The average hourly construction wage,[5] rent of primary residence, and Consumer Price Index (all urban consumers, all items) come from the Bureau of Labor Statistics. A proxy for stock market wealth is Standard and Poor's 500 stock market index. The source of an effective mortgage interest rate is the Federal Housing Finance Board. Since construction wage, rent of primary residence, CPI, mortgage rate, and the S&P 500 index are in monthly frequency, they are recalculated to the quarterly frequency by taking an arithmetic average of monthly values for a particular quarter. The measure of population is in quarterly frequency for 1980:q2-1990:q2. After that, quarterly values for population are taken from the monthly estimates.[6]

[3] Timmermann (1995) also discusses more general conditions on a_t.

[4] See Calhoun (1996) for the methodology of calculation of this price index.

[5] In the aggregate level evidence section construction wage is a proxy for building costs. This proxy may be imperfect because it does not include the price of construction materials and the amount of different hours of labor needed to finish particular construction works. The panel data section uses the building costs index, which is a better measure of the total costs of construction. Unfortunately, this index is not available in monthly or quarterly frequency for all the periods needed for the aggregate data analysis. Hence, in this analysis construction wage serves as a proxy for construction costs.

[6] Contrary to the other monthly series, the number of people at the beginning of the quarter is taken as the estimate of population (i.e. not an average).

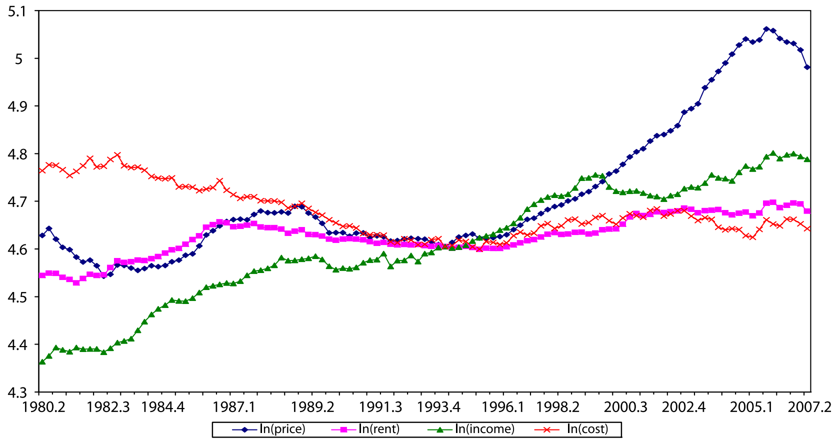

Fig. 1. House price and fundamentals, nationwide time series, 1/2.

Panel data on 22 Metropolitan Statistical Areas is annual. All the variables in this dataset and their sources are the same as in the aggregate data except for the construction wage which is substituted by a building cost index. The source of this index is the Engineering News-Record (ENR) cost indexes for 20 U.S. cities.[7] The two cost indexes published by ENR are the building cost and construction cost indexes. This study uses the building cost index because it better represents building works related to residential property, while the construction cost index is more representative for the costs of non-residential property construction. This index is available for only 16 MSA in the sample.

House rent, CPI, population, building cost index, and per capita income are in annual frequency, so they are not recalculated. The stock market index is available monthly, so annual values are obtained by taking arithmetic averages of monthly values. The house price index is in quarterly frequency, so arithmetic averages of quarterly values are used to get annual estimates. The mortgage rate is annual for 1978–2004. For 2005–2007, only quarterly values of this rate are available, and therefore averages of quarterly rates are used to obtain annual values.[8] In both nationwide and panel datasets, we use real values for the house price, rent, cost, and income. Nominal values are adjusted simply by dividing by CPI. The real mortgage rate is calculated as the difference between the nominal rate and regional inflation rate computed using CPI.

4. Aggregate data evidence

This section presents unit root and cointegration tests for house price and the set of fundamentals using aggregate national-level data in order to assess the possibility of a house price bubble. Before we report results, we illustrate the behavior of house price and various fundamentals using simple graphs. Fig. 1 depicts the house price, rent, personal income and construction costs. It may be noted from the figure that the house price follows time series patterns of income and rent in all periods except 2000–2006 when the price strongly increases, but rent and income remain constant or decrease. Construction costs seem to be weak in explaining cycles in house price since a reduction in the costs throughout the sample does not lead to a decline in the price. Fig. 2, which portrays house price, stock market wealth, mortgage rate, and population, allows for a conclusion that the fast expansion of house price at the end of the sample could be due to the declining mortgage rate. The low mortgage rate should encourage people to buy houses, increase demand, and cause house price to rise. Another possible explanation for the steep growth of house price is a crash of the stock market in 2000 and switching of many stock market investors to the housing market.[9] An increase in demand for housing generated by these investors could push house prices up. A growing population could be another force to stimulate demand for housing and bid up the price.

While the graphs provide some intuition about the possible causes of the behavior of house prices in the 2000s, they are not very useful in determining formally whether changes in fundamental factors explain movements in house price. In order to test if house prices reflect fundamentals, we use the cointegration procedure developed by Engle and Granger (1987). The cointegration regression may be assumed to have the following form:

$$y_t = \mu + \omega t + \sum_{k=1}^{K} \psi_k x_{k,t} + \theta_t \tag{16}$$

where y_t and $x_{i,t} (i = 1, \ldots, K)$ are variables with the same order of integration (usually, I(1)) which are hypothesized to be cointegrated, t is a time trend, and θ_t is an error term. According to the Engel and Granger two-stage procedure, in the first stage, it is necessary to test if all variables in

[7] ENR (2002, 2008) publish these cost series, and Grogan (2007) provides a description of these indexes.

[8] In 2005, the Federal Housing Finance Board provides data only for the last two quarters. Hence, the average of these two quarters serves as an estimate of the annual mortgage interest rate in this year.

[9] Case and Shiller (2003) also discuss this explanation.

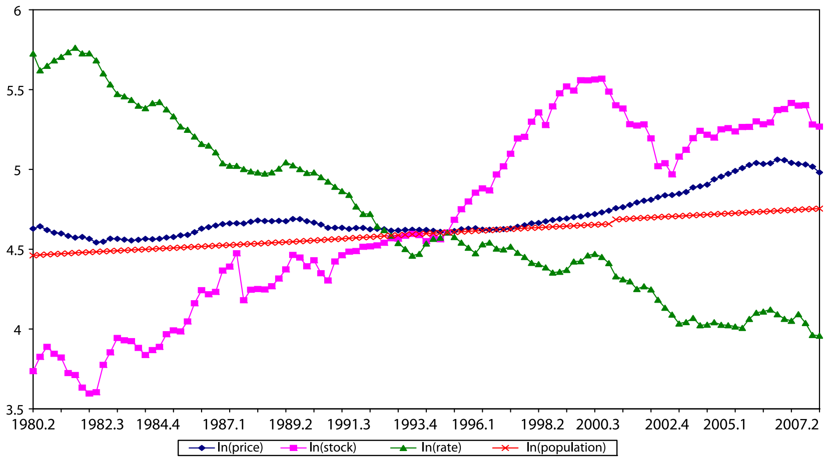

Fig. 2. House price and fundamentals, nationwide time series, 2/2.

Table 1
Aggregate data unit root tests.

	Price	Rent	Cost	Income	Pop	Rate	Stock
1980:q2-2008:q2							
Ln	−2.48	−2.70	−1.52	−2.65	−2.10	-3.63^{**}	−1.70
Growth	−0.90	-3.56^{**}	−3.04	-4.13^{***}	-4.65^{***}	-4.27^{***}	-4.58^{***}
1980:q2-2006:q4							
Ln	−0.04	−2.39	−1.23	−2.66	−1.95	−2.79	−2.10
Growth	−2.89	-3.32^{*}	−2.73	-3.73^{**}	-4.57^{***}	-3.77^{**}	-4.42^{***}
1980:q2-2005:q4							
Ln	0.47	−2.61	−1.58	−2.47	−1.83	−2.98	−1.99
Growth	−2.61	-3.30^{*}	−2.37	-3.94^{**}	-4.52^{***}	-3.83^{**}	-4.38^{***}
1997:q1-2008:q2							
Ln	−2.38	−1.75	−2.49	−2.32	−1.55	−2.34	−2.40
Growth	1.88	−2.53	−2.02	−2.69	−2.91	−1.93	−2.57
1980:q2-1996:q4							
Ln	−1.58	−1.83	-3.31^{*}	−1.76	-3.92^{**}	−2.70	-3.27^{*}
Growth	−2.50	-3.72^{**}	−2.65	−3.10	−1.41	-3.62^{**}	-4.24^{***}

Notes: (1) Included series are house price (price), rent of primary residence (*rent*), construction wage (*cost*), personal income (income), population (pop), mortgage rate (rate), and the stock market wealth (stock).
(2) The null hypothesis is that of a unit root.
(3) ADF test statistics with four lags and trend are reported for levels and first differences.
(4) Significance at the 1%, 5%, and 10% levels is denoted as ***, **, and *, respectively.

the hypothesized cointegration relation are of the same order of integration. After that, Eq. (16) is estimated using OLS and testing for the stationarity of θ_t is performed using the augmented Dickey–Fuller (ADF) test. If the error term is found to be stationary then the null hypothesis of no cointegration of y and x_i's$(i = 1, \dots, K)$ is rejected. Since a vector of cointegrating parameters (ψ's) is estimated, the usual critical values for ADF tests cannot be used in this testing, but instead critical values for cointegration tests generated by Davidson and MacKinnon (1993) should be utilized. In our case, y is the house price and x's are the per capita personal income, rent of primary residence, population, construction wage, mortgage rate, and the stock market wealth.

ADF τ-tests for levels and first differences of the variables under consideration are presented in Table 1 for five samples. The samples are given by alternating the final date to account for price decreases around 2006 and by considering a regime change around 1996. As can be seen from the table, for the full sample 1980:q2-2008:q2, the null hypothesis of a unit root fails to be rejected for house price and cost in both levels and first differences. Other series are either stationary already in levels (rate) or first differences (rent, income, population, and stock). Stationarity in house price and cost is achieved only after taking second differences which makes possible further testing for cointegration between these two variables.[10] Table 1 also shows that in 1980:q2-2006:q4 and 1980:q2-2005:q4 subsamples house price can also be cointegrated with cost because both are $I(2)$ and other variables are $I(1)$. For

[10] Unit root test statistics for second and third differences of variables are available upon request.

Table 2
Aggregate data cointegration tests.

Sample	ADF τ	Rent	Cost	Income	Pop	Rate	Stock
1980:q2-2008:q2	-4.81^{***}	–	x	–	–	–	–
1980:q2-2006:q4	-4.15^{**}	–	x	–	–	–	–
1980:q2-2005:q4	-2.18	–	x	–	–	–	–
1997:q1-2008:q2	-0.20	–	x	–	–	–	–
1980:q2-1996:q4	-2.68	–	–	x	–	–	–

Notes: (1) See Table 1 for definitions of variables.
(2) The cointegration test is ADF test with 4 lags and trend on the error term from the first stage regression of the Engle-Granger procedure. The critical values are from Davidson and MacKinnon (1993). Non-stationary variables included in the cointegrating regression are denoted x.
(3) The null hypothesis is that of no cointegration.
(4) Significance at the 1%, 5%, and 10% levels is denoted as ***, **, and *, respectively.

1997:q1-2008:q2 sample, the price, cost and rent are $I(3)$, while other variables are $I(2)$. Finally, in the last sample considered (1980:q2-1996:q4), house price and income are $I(2)$ and other variables are either $I(0)$ or $I(1)$.

Table 2 shows that the null hypothesis of no cointegration is rejected for house price and cost in the 1980:q2-2008:q2 sample. This implies that the correction of a possible house bubble may be over in 2008 because the price is in line with costs. Furthermore, the correction of house price seems to be on its way already at the end of 2006 because in 1980:q2-2006:q4 sample the price is cointegrated with the costs (however, in this case, the hypothesis of no cointegration is rejected at 5% level of significance only). Contrary to the previous two samples, the 1980:q2-2005:q4 sample shows no sign of cointegration, which may be understood as a sign of at least one house price bubble during this time period. In fact, the next to last row of Table 2 indicates that the price is not cointegrated with cost in 1997:q1-2008:q2. The cost variable can be replaced by rent in this case though they cannot be used together. This is because the structural model implies that either only rents and interest rates (see Eq. 6) or other fundamentals (see Eq. (15)) are explanatory variables. The null hypothesis of no cointegration cannot be rejected for the rent variable either. Finally, the last row of Table 2 shows that the price and income are not cointegrated in 1980:q2-1996:q4. This suggests that there could be two house price bubbles: one at the end of 1980s - the beginning of 1990s, and another in 2000–2005. However, the correction in house price after 2005 appears to remedy the influence of these two bubbles and align the price with construction cost and other fundamentals.

5. Panel data evidence

Testing for unit roots and cointegration in panel data has made rapid progress in the last fifteen years. Current tests are now robust to cross-sectional dependence and autocorrelation, allow for different autoregressive coefficient across individual units, and have favorable finite sample properties. We perform these tests using panel data on house prices and corresponding fundamental variables.

House prices tend to move together in geographically close areas, which complicates statistical testing for unit roots and cointegration in panel data. Hence, we first test how severe this problem is in our data using a general diagnostic test for cross-section dependence in panels from Pesaran (2004). The test statistic is defined as

$$CD = \sqrt{\frac{2T}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{Corr}(\hat{\varepsilon}_i, \hat{\varepsilon}_j)\right) \tag{17}$$

where $i = 1, \ldots, N$ is the the number of individual units and $t = 1, \ldots, T$ is the time dimension of the data. $\hat{\varepsilon}_i, i = 1, \ldots, N$, are $(T \times 1)$ vectors of estimated residuals from the augmented Dickey–Fuller (ADF) regression equation:

$$\Delta y_{it} = \mu_i + \omega_i t + \alpha_i y_{i,t-1} + \sum_{j=1}^{p_i}\lambda_{ij}\Delta y_{i,t-j} + \varepsilon_{it}, \tag{18}$$

where μ_i is an individual fixed effect, ω_i is an individual trend coefficient and $\alpha_i - 1$ is an autoregressive coefficient of a given series. Both α_i and the lag order p_i may vary across cross-sections. The summation term involving lagged Δy's filters out autocorrelation. CD asymptotically converges to the standardized normal distribution. We report the calculated CD statistics for all series in our sample in Table 3 ($p_i = 1$). There is a strong cross-sectional dependence for the logs. Growth rates are regionally correlated for the rent, income, mortgage rate, and the stock market.[11] Our results imply that cross-sectional dependence should be taken into account in our testing for unit roots and cointegration.

The standard ADF regression for individual series (such as Eq. 18) assumes no cross-sectional dependence. Since this assumption is clearly violated in our data, we conduct an updated version of this test proposed in Pesaran (2007). Robustness to the cross-sectional dependence in the Pesaran version of the test is achieved by adding the lagged cross-section mean and its differences to the ADF regression. The cross-sectionally augmented Dickey–Fuller regression (CADF) is then defined as

$$\Delta y_{it} = \mu_i + \omega_i t + \alpha_i y_{i,t-1} + \sum_{j=1}^{p_i}\lambda_{ij}\Delta y_{i,t-j} + \upsilon_i \bar{y}_{t-1} + \sum_{j=0}^{p_i}\varpi_{ij}\Delta\bar{y}_{i,t-j} + \epsilon_{it}, \tag{19}$$

where ϵ_{it} denotes an i.i.d. error term and $\bar{y}_t$ is the cross-section mean. The other parameters and variables are the

[11] Note that we use the nationwide stock index. The regional stock index is calculated by adjustment using regional CPI.

Table 3
Diagnostic tests for cross-section dependence in panels, sample 1978–2007, 22 MSA.

	Price	Rent	Cost	Inc	Pop	Rate	Stock
Ln	17.35^{***}	28.19^{***}	27.59^{***}	36.32^{***}	6.06^{***}	56.48^{***}	80.04^{***}
Growth	-1.20	-1.33^{*}	-0.76	-3.32^{***}	-0.28	-3.10^{***}	-3.37^{***}

Notes: (1) Included series are the housing price index (price), tenants' rent (rent), Consumer Price Index (cpi), construction costs (cost, only 16 regions), regional income (inc), the series ends in 2006), population (pop), the mortgage rate (rate, we use national average growth to calculate the rate for Honolulu in 2007), and the stock index (stock). Real values of price, rent, cost, inc, stock are calculated simply by dividing by cpi. The rate is adjusted by subtracting the inflation rate computed using cpi.
(2) ADF regression: intercept, trend, and the first lag of the dependent variable.
(3) The null hypothesis is that of no cross-sectional dependence. Pesaran (2004) CD statistic is reported.
(4) CD statistic has the standardized normal distribution. Significance at the 1%, 5%, and 10% levels is denoted as ***, **, and *, respectively.

same as in the ADF Eq. (18). The CADF equation formalizes a fairly novel idea that even individual unit root tests should account for mutual dependence among regions. Let $\tilde{t}_{i,T_i,N}(p_i)$ be the t-statistic for $\alpha_i = 0$ (a unit root) in the CADF regression. Setting $p_i = p = 1$ for all i's together with an assumption of a balanced panel, which is satisfied in our case, results in $\tilde{t}_{i,T_i,N}(p_i) = \tilde{t}_i(T, N)$. Following Pesaran (2007), we restrict the values of this statistic to the interval between -6.42 and 1.70.

We conduct the CADF test using the logarithms of house prices, rents, construction costs, income, population, and mortgage rates. Results are reported in Table 4. For all variables, majority of the series do contain unit roots. There are only five areas with stationary house prices, for example Cleveland–Arkon, OH. However, the reported *t*-statistics are mostly negative and fairly close to critical values of the CADF test, which would not have been the case for a sample ending in 2005, prior to the housing bubble collapse. Table 4 can be used to evaluate the possibility of bubble occurrence in a given region. Fundamentals in Pittsburgh, PA area all have a unit root so it does not come as a surprise that the house price index is also not stationary. Honolulu, HI provides a different story. Even though only the house price is stationary, *t*-statistics for all fundamental factors are fairly negative. Similarly, the stationarity of the area Chicago–Naperville–Joliet, IL seems to be driven by a mainly one fundamental factor, the rents.

A natural next step is to find whether a discrepancy between house prices and fundamentals exists at the national level. In other words, we would like to explore if the non-alignment of house prices and fundamentals is a local phenomenon typical for a few regions, or if this non-alignment leads to a break in the relationship between house prices and fundamentals in the U.S. housing market. Figs. 3 and 4 illustrate the relationship graphically, using a time series of cross-sectional means of the house prices and fundamentals (all variables are normalized so that the values in 1995 equal to log 100). Three peaks followed by a rapid decline can be identified in the house price ser-

Table 4
Individual unit root tests, sample 1978–2007, 22 MSA.

#	Region	ln(price)	ln(rent)	ln(cost)	ln(inc)	ln(pop)	*ln*(rate)	*ln*(stock)
1	Atlanta-Sandy Springs-Marietta, GA	-2.14	-2.82	-1.89	0.65	-4.14^{**}	-3.43	-2.99
2	Boston-Quincy, MA (MSAD)	-2.07	-4.19^{**}	-1.3	-3.03	-3.35	-2.46	-1.92
3	Chicago-Naperville-Joliet, IL (MSAD)	-6.24^{***}	-6.42^{***}	-3.16	-3.24	-2.43	-3.15	-0.39
4	Cleveland-Arkon, OH (MSAD)	-4.94^{**}	-3.27	-4.37^{**}	-1.12	-1.57	-4.81^{**}	-2.86
5	Dallas-Plano-Irving, TX (MSAD)	-3.65^{*}	-2.45	-2.32	-3.13	-5.03^{***}	-3.48	-3.29
6	Denver-Aurora, CO	-2.77	0.04	-3.77^{*}	-1.73	-1.93	-2.15	-2.13
7	Detroit-Livonia-Dearborn, MI (MSAD)	-3.18	-1.75	-2.91	-1.95	-5.57^{***}	-3.47	-3.03
8	Kansas City, MO-KS	-3.51	-1.38	-3.06	-4.11^{**}	-3.13	-3.82^{*}	-3.34
9	Los Angeles-Long Beach-Glendale, CA (MSAD)	-2.6	-2.18	-2.53	-0.31	-2.68	-2.81	-2.03
10	Minneapolis-St. Paul-Bloomington, MN-WI	-3.03	-4.56^{**}	-1.83	-3.95^{*}	-2.86	-4.19^{**}	-2.96
11	New York-White Plains-Wayne, NY-NJ (MSAD)	-2.99	-2.08	-2.04	-4.91^{**}	-2.25	-2.32	-2.78
12	Philadelphia, PA (MSAD)	-3.03	-2.03	-2.33	-2.40	-4.67^{**}	-2.24	-2.16
13	Pittsburgh, PA	-2.76	-2.45	-2.56	-2.77	-2.45	-2.94	-3.25
14	St. Louis, MO-IL	-2.84	-4.40^{**}	-3.27	-2.14	-3.51	-5.08^{***}	-3.41
15	San Francisco-San Mateo-Redwood City, CA (MSAD)	-3.01	-1.74	-3.00	-2.29	-2.34	-2.09	-2.80
16	Seattle-Bellevue-Everett, WA (MSAD)	-2.54	-5.21^{***}	-2.33	-3.59^{*}	-2.41	-3.30	-1.93
17	Milwaukee-Waukesha-West Allis, WI	-2.41	-1.77	NA	-3.26	-4.49^{**}	-5.38^{***}	-1.61
18	Houston-Sugar Land-Baytown, TX	-2.38	-3.24	NA	-1.49	-3.76^{*}	-2.73	-2.16
19	Miami-Miami Beach-Kendall, FL (MSAD)	-3.54	0.10	NA	-1.89	-1.34	-2.92	-0.48
20	Honolulu, HI	-3.72^{*}	-2.02	NA	-1.71	-0.91	-2.06	-2.27
21	Portland-Vancouver-Beaverton, OR-WA	-3.29	-2.02	NA	-2.31	-2.87	-2.6	-3.54
22	San Diego-Carlsbad-San Marcos, CA	-4.08^{**}	-2.08	NA	-0.01	-1.81	-2.09	-1.06

Notes: (1) Included series are the same as in Table 3. The cost series is not available for the last six regions (denoted NA).
(2) The cross-sectionally augmented Dickey–Fuller (CADF) regression: intercept; trend; the first lags of the difference of the dependent variable, the difference of the cross-section mean, and the cross-section mean; the difference of the cross-section mean. CADF t-statistic is reported.
(3) Critical values for the CADF t-statistic are from Pesaran (2007), Table Ic. Significance at the 1%, 5%, and 10% levels is denoted as ***, **, and *, respectively.

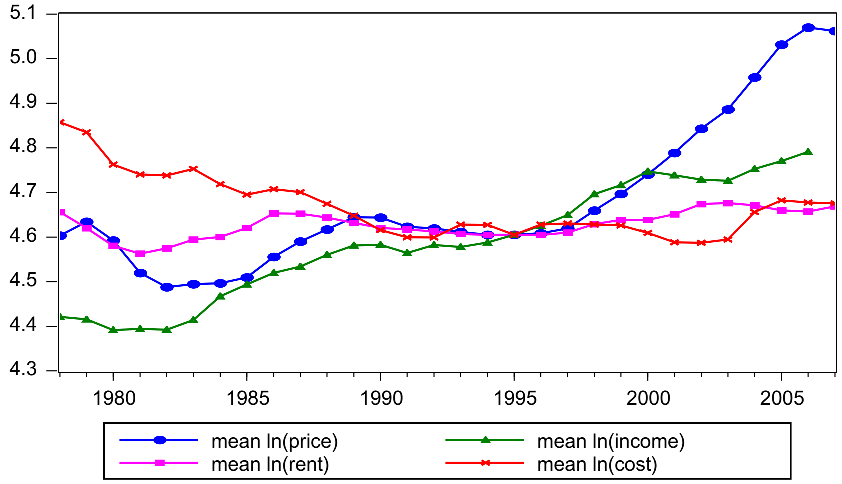

Fig. 3. House price and fundamentals, panel data, cross-sectional means, 1978–2007, 1/2.

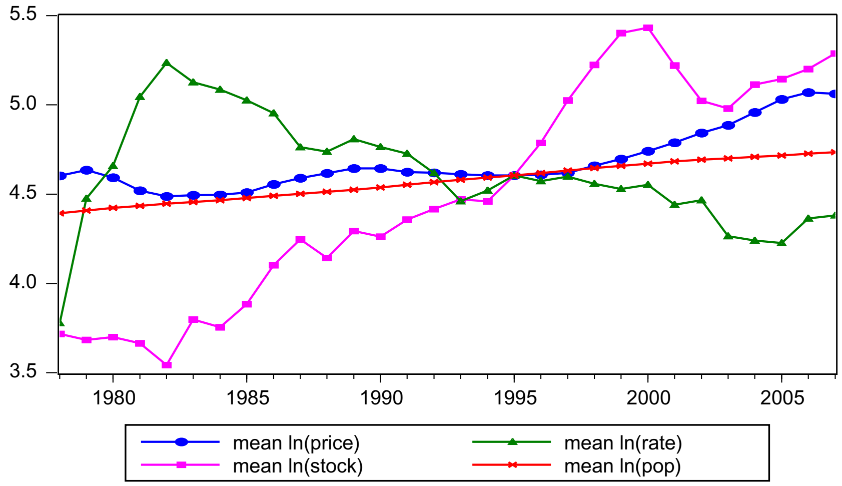

Fig. 4. House price and fundamentals, panel data, cross-sectional means, 1978–2007, 2/2.

ies: 1979, 1989, and 2006. Rents, costs, and income roughly follow a similar pattern while the mortgage rate, the stock market index, and population behave somewhat differently. These observations closely resemble patterns found in the aggregate series.

To formally assess the relationship between house prices and their determinants, we use panel data stationarity tests that combine regional test results. An intuitive and widely used panel data unit root test along these lines is developed in Im et al. (2003). This test simply averages across regions the individual t-statistics for α_i in the ADF regressions. We use an updated version of this test proposed in Pesaran (2007), which is robust to the cross-section dependence:

$$\bar{t}^{\dagger} = \frac{1}{N}\sum_{i=1}^{N}\tilde{t}_i(N,T). \tag{20}$$

The null and alternative hypotheses are, respectively,

$$H_0: \ \alpha_i = 0 \quad \text{for } i = 1,2,\ldots,N \tag{21}$$

$$H_1: \ \begin{cases} \alpha_i = 0 & \text{for } i = 1,2,\ldots,N_1 \\ \alpha_i < 0 & \text{for } i = N_1+1, N_1+2,\ldots,N. \end{cases} \tag{22}$$

Regions can be ordered as needed. H_1 states that at least one of the N series is stationary. A rejection of the null hypothesis implies that some series are stationary. Failure to reject indicates that after looking at N realizations of a given process we are not able to exclude the possibility that all series are in fact non-stationary.

We conduct the CIPS test for logs and growth rates (differences in logs) for our seven variables available as panels of data. Results are given in Table 5. The collapse of the housing prices starting in 2007 followed by dramatic events in the financial markets had been predicted by a working paper version of the present study. To analyze the collapse, we first consider three samples, all starting in 1978 and ending in 2005, 2006, and 2007, respectively. The real housing price becomes stationary in 2006, reflect-

Table 5
Panel data unit root tests.

	Price	Rent	Cost	Inc	Pop	Rate	Stock
1978–2007							
Ln	−2.76**	−2.63*	−2.67		−2.98***	−3.16***	−2.38
Growth	−3.22***	−3.34***	−4.25***	−3.48***	−3.59***	−3.42***	−3.62***
1978–2006							
Ln	−2.67*	−2.65*	−2.58	−2.30	−2.93***	−3.10***	−2.28
Growth	−3.18***	−3.25***	−4.17***	−3.48***	−3.54***	−3.33***	−3.57***
1978–2005							
Ln	−2.58	−2.64*	−2.62	−2.35	−3.05***	−2.97***	−2.24
Growth	−3.12***	−3.19***	−4.13***	−3.29***	−3.46***	−3.24***	−3.51***
1997–2007							
Ln	−1.91	−1.36	−2.56	−3.09**	−1.47*	−2.58	−2.38
Growth	−3.34*	−2.04	−3.03	−2.74*	−2.44	−3.27*	−1.84
1978–1996							
Ln	−2.29	−2.57	−2.04	−2.48	−2.72*	−2.74*	−2.09
Growth	−2.49	−2.58	−3.60***	−2.79**	−2.70*	−2.78**	−2.80**

Notes: (1) Included series are the same as in Table 3.
(2) The CIPS test is based on the individual cross-sectionally augmented Dickey–Fuller (CADF) regressions with intercept; trend; the first lags of the difference of the dependent variable, the difference of the cross-section mean, and the cross-section mean; and the difference of the cross-section mean. Critical values for the CIPS statistic are from Pesaran (2007), Table IIc. For the shortest sample 1997–2007, the critical values are not reported in Pesaran (2007) and we generate them ourselves.
(3) Significance at the 1%, 5%, and 10% levels is denoted as ***, **, and *, respectively.
(4) The null hypothesis is that of a unit root (it assumes an individual unit root process).
(5) The income series ends in 2006. To calculate an unbiased error in the CADF regression, we use the sample 1996–2006 instead of 1997–2007.
(6) For the population series in the sample 1997–2007, the t-statistic in the CADF regression is a ratio of two numbers very close to zero and hence could not be calculated. Therefore we provide the IPS statistic instead of the CIPS statistic.

ing a third through on the housing market (see Fig. 3). A stationary house price provides implicit evidence in favor of the structural model. Conventional understanding of the housing market views such a situation as an absence of a bubble. The rejection of the unit root illustrates advantages of using the panel data unit root tests - even though majority of individual time series are non-stationary, the null hypothesis of the unit root is rejected overall. A non-stationary house price prior to 2005 requires further investigation using cointegration tests to find if the price had corresponded to fundamentals. Regarding the underlying factors, rents, population, and mortgage rates are stationary, the other variables are not. The growth rates of all variables are stationary in all three samples.

In addition to altering the final date of our sample, we also account for the possibility of a regime change and calculate our tests for sub-samples 1978–1996 and 1997–2007, respectively. The former sub-sample includes the first two peaks in real estate prices (there is not enough data to analyze to first peak separately) and the latter focuses on the most recent rise. The null hypothesis of a unit root in logs in all regions is accepted for both sub-samples while the growth rate is non-stationary for the 1978–1996 period. It seems that all three price corrections are needed to make the housing prices stationary around a deterministic time trend, which is a weaker definition of stationarity. Naturally, the next step is testing for panel data cointegration of the house price with other non-stationary variables in the corresponding samples to see if the price is supported by fundamentals.

A widely used test for cointegration in panel data is constructed by Pedroni (1999, 2004) and is based on the following cointegrating regression:

$$y_{i,t} = \mu_i + \omega_i\, t + \psi_{1i}\, x_{1,i,t} + \ldots + \psi_{Mi}\, x_{M,i,t} + \zeta_{i,t} \quad \text{for } t = 1, \ldots, T,\ i = 1, \ldots, N \tag{23}$$

The slope vector ψ_i defines the cointegrating relationship between the dependent variable y (house price) and explanatory variables $x_m, m = 1, \ldots, M$ (fundamentals). Let us define γ_i as the autoregressive coefficient of the error term ζ_i. The null hypothesis of no cointegration $H_0 : \gamma_i = 1$ for all i, is tested against the alternative $H_1 : \gamma_i < 1$ forall i, where we do not assume any common value for the autoregressive coefficient. The test for cointegration is a test of stationarity of ζ_i while accounting for the fact, that ψ_i's have to be estimated. Otherwise, the CIPS test would suffice. We use the Pedroni group ADF *t*-statistic, which he has shown to have the best finite sample properties. The group *t*-statistic asymptotically converges to the standardized normal distribution under the assumption of no cross-sectional dependence. Because this assumption is violated in our case, we use bootstrapping similar to Gallin (2006) and Maddala and Wu (1999) to generate critical values.

We conduct the group-*t* Pedroni test for cointegration between the house price and explanatory variables in samples where the house price contains a unit root. Results are reported in Table 6. The chosen explanatory variables are only the ones with the same order of integration. For example, while income is not stationary from 1978 to 1996, its first difference in logs is stationary while the same does not hold true for the real estate price. We again cannot use the rent with the other determinants since one can either use the present value model (6) or the supply and demand alternative (15). However, in this case it is not relevant. In all evaluated samples, the null hypothesis

Table 6
Panel data cointegration tests.

Sample	Group t	Rent	Cost	Inc	Pop	Rate	Stock
1978–2005	−0.45	–	x	x	–	–	x
1997–2007	−3.86	–	–	–	–	x	–
1978–1996	−3.20	x	–	–	–	–	–

Notes: (1) See Table 3 for definitions of variables.
(2) The cointegration test is the group t-statistic from Pedroni (1999, 2004). The dependent variable is the housing price index. Non-stationary variables included in the cointegrating regression are denoted x. We use critical values generated by bootstrap to account for cross-sectional dependence. Significance at the 1%, 5%, and 10% levels is denoted as ***, **, and *, respectively. The null hypothesis is that of no cointegration.

of no cointegration is accepted and hence the spikes in housing prices do indeed reflect bubbles in samples 1978–2005, 1997–2007, and 1978–1996, respectively. The same conclusion applies if we include all non-stationary variables in the cointegrating relationships (even though this is not correct according to a strict definition of a cointegrating relationship). Employing nominal variables with CPI as one of the factors yields similar findings even though this approach takes one additional year to detect the collapsing bubble. Taking into account the cross-sectional dependence and using bootstrapped critical values makes a difference for the shorter samples where using the standardized normal distribution could change the results. To summarize, our results from Table 6 indicate that the house price and any combination of non-stationary fundamentals are not cointegrated. Combined with the panel data unit root tests, it appears that house prices do not reflect movements in fundamental factors, i.e., there is a house price bubble in these sub-samples.

6. Summary

This paper used a unique set of regional and aggregate data on house prices and fundamental variables in order to investigate whether the U.S. housing market experienced a bubble in recent years. The fundamental variables included real house rent, mortgage rate, personal income, building cost, stock market wealth, and population. Based on the evidence from univariate and panel unit root and cointegration tests, we conclude that there was a house price bubble in the U.S. prior to 2006. The question is whether the house price correction starting in 2006 has been sufficient, the bubble has since burst and the prices returned to fundamentals. Univariate time series tests for individual Metropolitan Statistical Areas indicate that the prices may decline even further.

References

Bucks, B., Kennikell, A., Moore, K., 2006. Recent changes in U.S. family finances: evidence from the 2001 and 2004 survey of consumer finances. Federal Reserve Bulletin 92, A1–A38.

Calhoun, C., 1996. OFHEO house price indexes: HPI technical description. Office of Federal Housing Enterprise Oversight. Available from: <http://www.fhfa.gov/webfiles/896/hpi_tech.pdf>.

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

Davidson, R., MacKinnon, J., 1993. Estimation and Inference in Econometrics. Oxford University Press, New York.

Engle, R.F., Granger, C.W., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 55 (2), 251–276.

ENR, 2008. Cost report. ENR: Engineering News-Record 260 (10), 39–42.

ENR, 2002. Cost indexes in 20 cities, 1976–2002. ENR: Engineering News-Record 248 (10), 80.

Gallin, J., 2006. The long-run relationship between house prices and income: evidence from local housing markets. Real Estate Economics 34 (3), 417–438.

Grogan, T., 2007. How to use ENR's cost indexes. ENR: Engineering News-Record 258 (11), 37.

Himmelberg, C., Mayer, C., Sinai, T., 2005. Assessing high house prices: bubbles, fundamentals, and misperceptions. Journal of Economic Perspectives 19 (4), 67–92.

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

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 61, 631–652.

McCarthy, J., Peach, R., 2004. Are home prices the next bubble? Federal Reserve Bank of New York Economic Policy Review 10 (3), 1–17.

Mikhed, V., Zemčík, P., 2007. Testing for bubbles in housing markets: a panel data approach. CERGE-EI Working Paper, 338.

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., 2004. Panel cointegration; asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. Econometric Theory 20 (3), 597–625.

Pesaran, M.H., 2004. General diagnostic tests for cross-section dependence in panels. Working Paper No. 1229, CESIFO.

Pesaran, M.H., 2007. A simple panel unit root test in the presence of cross-section dependence. Journal of Applied Econometrics 22 (2), 265–312.

Poterba, J.M., 1984. Tax subsidies to owner-occupied housing: an asset market approach. Quarterly Journal of Economics 99 (4), 729–752.

Shiller, R.J., 2005. Irrational Exuberance. Princeton University Press, Princeton.

Smith, M., Smith, G., 2006. Bubble, bubble, where's the housing bubble? Brookings Papers on Economic Activity 1, 1–50.

Stiglitz, J.E., 1990. Symposium on bubbles. Journal of Economic Perspectives 4 (2), 13–18.

Timmermann, A., 1995. Cointegration tests of present value models with a time-varying discount factor. Journal of Applied Econometrics 10 (1), 17–31.