Three Essays on the Application of Second Generation Cointegration Analysis in Time Series Econometrics

by

Ghislain Nono Gueye

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 5, 2017

Keywords: panel, cointegration, cross-section dependence,
inequality, growth, nonstationarity

Approved by

Dr. Hyeongwoo Kim, Chair, Associate Professor of Economics
Dr. Randy Beard, Professor of Economics
Dr. Henry Thompson, Professor of Economics
Dr. Michael Stern, Associate Professor of Economics
Dr. Gilad Sorek, Assistant Professor of Economics
Dr. Yunmi Park, Assistant Professor of Political Science

Acknowledgements

My gratitude goes first to GOD Almighty, the Creator of the universe, for His unfathomable love and His never-ending blessings in my life. I acknowledge that this work could not have been completed without His grace and His guidance. There is nothing that I can do or say, which would be enough to thank HIM. Actually, I praise GOD for the following people I acknowledge.

Two of the greatest blessings GOD has given me are my dad, Bonfils Gueye, and my mom, Pauline Gueye, whose hard work and relentless dedication towards the success of their children have paved a way for me to be where I am today. Words cannot express how grateful I am to have such wonderful parents.

One of the greatest events in my life happened right in the middle of working on this dissertation – I became the husband of the most lovely woman on earth, Esther. Her presence by my side has been a continuous source of strengh and motivation to keep moving forward even when things became hard. Life with Esther is nothing short of a bliss.

Any Ph.D. holder/student knows how important supervisors are in the progression of a dissertation. They can literally make you or break you. I was honored to be supervised by the most astonishing of all superviors, Dr. Hyeongwoo Kim. I cannot stress enough how helpful he always was. Besides being an inspiration to me, he also always kept motivating me to learn more and to become a better economist.

The kind of environment created by all professors in the Department of Economics at Auburn University is one that significantly challenges students to work hard and make progress. I would like to acknowledge the professors who have taught me at least one class. Dr. Randy Beard, Dr. Henry Thompson, Dr. Gilad Sorek, Dr. Duha Altindag, Dr. Aditi Sengupta, Dr. Alex Richter, Dr. Richard Seals. Special mention for "THE" Michael Stern, Ph.D., who is "hands down" the most epic department chair in the universe. His leadership is undoubtedly one of the reasons the Department is moving as fast as it does.

Last but in no way the least, I would like to acknowledge those who have made my experience at Auburn a wonderful one. Lising them all here would take as many pages as the dissertation itself; however, a few of them are: Dr. Cynthia Hourizene, Edgar Diomande, Dr. Omer Tadjion, Dr. Viviane Koua, Dr. Marceline Egnin, Dr. Curtis Jolly, Dr. Bijie Jia, Dr. Jon Newman and the wonderful Dr. Sara Seals. Of course, there ar

Table of Contents

List of Tables

Chapter 3

List of Figures

List of Abbreviations

ACE: Alternating Conditional Expectation

ADF: Advanced Dickey-Fuller

AIC: Akaike Information Criterion

ARDL: Autoregressive Distributive Lag

CD: Cross-Section Dependence

CPI: Consumer Price Index

CPS: Current Population Survey

CRB: Commodity Research Bureau

CSD: Cross-Section Dependence

DFGLS: Dickey-Fuller Generalized Least Squares

DGP: Data-Generating Process

DOLS: Dynamic Ordinary Least Squares

FMHPI: Freddie Mac House Price Index

FRED: Federal Reserve Economic Data

GDP: Gross Domestic Product

IMOLS: Integrated Modified Ordinary Least Squares

IPI: Industrial Production Index

IRS: Internal Revenue Service

M-TAR: Momentum-Threshold Autoregressive

OECD: Organisation for Economic Co-operation and Development

OLS: Ordinary Least Squares

PDOLS: Panel Dynamic Ordinary Least Squares

STAR: Smooth Transition Autoregressive

SVAR: Structural Vector Autoregressive

UK: United Kingdom

US: United States

USA: United States of America

VAR: Vector Autoregressive

VECM: Vector Error Correction Model

WDI: World Development Indicator

Chapter 1: Pitfalls in Testing for Cointegration between Inequality and the Real Income

Abstract

Frank (2009) constructed a comprehensive panel of state-level income inequality measures using individual tax filing data from the Internal Revenue Service. Employing an array of cointegration exercises for the data, he reported a positive long-run relationship between income inequality and the real income per capita in the US. This paper questions the validity of his findings. First, we suggest a mis-specification problem in his approach regarding the order of integration in the inequality index, which shows evidence of nonstationarity only for the post-1980 data. Second, we demonstrate that his findings are not reliable because the panel cointegration test he used requires cross-section independence, which is inappropriate for the US state-level data. Employing panel tests that allow cross-section dependence, we find no evidence of cointegration between inequality and the real income.

1 Introduction

The unprecedented rise in US income inequality since the early 1980's has been attracting the attention of researchers and policy makers over the past decades. One key question in the academic and public debates surrounding inequality regards its relation to economic growth. The current empirical literature provides mixed evidence, finding the correlation to be negative or positive, or sometimes insignificant.[1]

Early researches on this topic predominantly found a negative correlation. Many of them used modified versions of the cross-country economic growth model proposed by Barro (1991) augmented with an inequality variable. See, among others, Alesina and Perotti (1994), Alesina and Rodrik (1994), Persson and Tabellini (1994), Birdsall, Ross, and Sabot (1995), and Deininger and Squire (1998). However, Forbes (2000), later on, questioned the validity of these findings, pointing at measurement error (or omitted variable) biases in the earlier works due to the fact that inequality was measured differently in the countries studied in those cross-country analyses.

More recent studies point towards a positive relationship between income inequality and economic growth, following the significant work of Deininger and Squire (1996) who constructed an improved database of cross-country inequality measures. Using this data, Forbes (2000) reports that income inequality and growth are positively correlated, while Barro (2000) reports a positive correlation in wealthier countries and a negative one in low-income countries. On the other hand, using a cross-state panel for the US, Panizza (2002) reports that the relationship between inequality and growth is not robust, questioning the validity of previous findings. Therefore, the profession has yet failed to reach a consensus.

Recently, Frank (2009) constructed a new valuable data set for state-level income inequality measures, i.e., top percentile shares of income for the 1945-2004 period, using highly confidential data from the Internal Revenue Service (IRS).[2,3] Employing this data set for panel cointegration tests, he reported strong evidence of a positive correlation between income inequality and the real income per capita. The present paper, however, questions the validity of his findings, employing more rigorous econometric procedures for the same data set.

[1] See, name a few, Garcia-Penalosa, Caroli, and Aghion (1999) and Quadrini and Rios-Rull (2015) for a review on the theoretical literature.

[2] We use his updated data until 2011 in the present paper.

[3] Leigh (2007) find that the trends in top income shares correspond to other common measures of inequality such as the Gini coefficient. In addition, Burkhauser, Feng, Jenkins, and Larrimore (2012) find that the growth in the income share of the top income percentile substantially outpaced inequality measured by the Gini coefficient.

First, we note that income inequality and the real income are both assumed to be nonstationary in his work, which is necessary for cointegration analyses.[4] We demonstrate that the income inequality measures in most of the 49 US states follow a nonstationary stochastic process since the 1980's, whereas it is better approximated by a stationary process for the period prior to 1980.[5] This implies that Frank's (2009) conclusion of a positive relationship might not be valid because he uses cointegration tests for the entire sample period, ignoring a possible change in the stochastic process for inequality. Similar observations were also reported in Piketty and Saez (2003) and Atkinson, Piketty, and Saez (2011).

Second, the panel cointegration tests used by Frank (2009) require cross-section independence. In what follows, we show that this assumption is inappropriate for US state-level data that he has analyzed. When this assumption fails to hold, statistical inferences may suffer from severe size distortion. Applying panel cointegration tests that allow cross-section dependence, we obtain virtually no evidence of a positive correlation between inequality and the real income.

The remainder of this paper is organized as follows. Section 2 describes the data and provides preliminary discussions. In Section 3, we first describe our econometric procedures. Then, we report and discuss our empirical results. Section 4 concludes.

2 Data Descriptions and Preliminary Discussions

We employ annual observations of the state-level inequality data for 49 US states, which were compiled by Frank (2009). Using highly confidential IRS data, he constructed the top decile share of income data, that is, the percentage of total *income* held by the top 10% income earners in each state. Observations range from 1945 to 2011.[6] Also, we obtained the state-level real income per-capita data from the Federal Reserve Economic Data (FRED) for the same sample period to measure economic growth in the US. The real income per capita is then log-transformed.

We noticed a substantial degree of common tendency from each of the 49 state-level inequality measures. Similar comovements were observed from the real income variables. This observation has an important implication on our econometric test procedures, because

[4]He implemented a panel unit root test for these variables from 1945 to 2004, which fails to reject the null of nonstationarity. If the true data generating process has changed from a stationary to a nonstationary process, implementing the test for the full sample will result in invalid statistical inferences.

[5]Unit root tests are known to have low power in small samples. That is, these tests may imply nonstationarity even when the alternative hypothesis (stationarity) is correct. Since we have stationarity for inequality in small samples (pre-1980s), this greatly strengthens our argument because the test rejects the null of nonstationarity even though the test suffers from low power in small samples.

[6]We obtained the data from Frank's website at: http://www.shsu.edu/eco_mwf/inequality.html

panel cointegration tests that require cross-section independence perform poorly when the true (panel) data-generating process is given a common factor structure. One may estimate a vector of common factors via the method of the principal components to study the patterns of the cross-section dependence. It turns out that the cross-section average of the data resembles the first common factor (see Pesaran, 2007). In order to see the common dynamics of these variables, we report the cross-section averages of the inequality and income variables in Figure 1.

We noticed that the cross-section mean of the real GDP per capita is continuously trending upward since the beginning of the data in 1945, while the top decile share of income exhibits a positive trend only after 1980. The inequality variable exhibits ups and downs around 32% until around 1980. To put it differently, the real GDP per capita seems to follow a non-stationary stochastic process for the entire sample period, whereas the stochastic nature of the inequality measure might have changed from a stationary process to a nonstationary process around 1980.

We are not the first who observed such a change in inequality dynamics. Piketty and Saez (2003) also noticed a positive trend in the top 10% pre-tax income share since 1980's based on individual tax returns data. Atkinson, Piketty, and Saez (2011) summarize the literature that documents concurrent trends in other English speaking countries, but not in continental Europe or Japan.[7] Burkhauser, Feng, Jenkins, and Larrimore (2012) report remarkably similar trends from the Current Population Survey (CPS) data.[8]

Figure 1 around here

This observation casts doubt on the validity of cointegration test results in Frank (2009) for the entire sample period, since a cointegration relationship requires a set of nonstationary variables. We implement an array of econometric tests in the next section to investigate these issues.

In addition to these data, we also employ the two measures of state-level human capital data used in Frank (2009), the proportion of the population having finished high school and the percentage of those having earned college degrees, for the sample period from 1945 to

[7]Atkinson, Piketty, and Saez (2011) provide a literature review on long-run trends in the share of top-income earners for more than 20 countries.

[8]Piketty and Saez (2006) suggest that top labor compensations in the United States have risen due to executives' increased influences in setting their own salaries, which extract rents at the expense of the benefits of shareholders. Atkinson, Piketty, and Saez (2011) also note that this positive trend in inequality might be due to changes in taxation policies and politics.

2004, which were obtained from Mark Frank's website.[9] We extend our benchmark model using these series to replicate the results of Frank (2009), which confirms findings from our benchmark model. As we can see in Figure 2, the cross-section means of these human capital series exhibit an upward trend since the beginning of the data in 1945. That is, these series seem to follow a nonstationarity stochastic process.

Figure 2 around here

3 Empirical Findings

3.1 Unit Root Tests

This section implements formal econometric tests for the stochastic properties of our key variables, focusing on the state-level inequality data in the US. For this purpose, we report an array of univariate and panel unit root tests for the two sub-sample periods, the pre-1980 and the post-1980 samples. These tests are crucial for the validity of the panel cointegration tests we implement afterward.

3.1.1 Univariate Unit Root Tests

We first employ the DFGLS test proposed by Elliott, Rothenberg, and Stock (1996) for the two sub-samples: the pre-1980 (1945-1979) and the post-1980 (1980-2011) periods. The DFGLS test is known to be asymptotically more powerful than the augmented Dickey-Fuller (ADF) test. We use the year 1980 as an *ad hoc* break point based on our eye-ball inspection of the inequality graph in Figure 1. We do not attempt to estimate the structural break date, because, to the best of our knowledge, no econometric procedures are available for the cases the data generating process (DGP) changes from a stationary process to a nonstationary one in the middle of the data. However, many researchers acknowledge the late 1970's or early 1980's as the time when income inequality in the US started to grow rapidly. See, for example, Piketty and Saez (2003), Frank (2009), and Saez and Zucman (2014) for similar discussions.

The DFGLS test is based on the following regression model for each US state.

$$\Delta\tilde{y}_t = \alpha + \rho y_{t-1} + \sum_{j=1}^{p} \beta_j \Delta\tilde{y}_{t-j} + \varepsilon_t, \tag{1}$$

[9] Updated data for these series are unfortunately not available.

where $\tilde{y}_t$ is locally *demeaned* data under the local alternative of $\tilde{\alpha} = 1 + \hat{c}/T$. T is the sample size and we use $\hat{c} = -7$ as recommended by Elliott, Rothenberg, and Stock (1996). The DFGLS test statistic is defined as,

$$ADF = \frac{\hat{\rho}}{s.e.(\hat{\rho})}, \tag{2}$$

where $\hat{\rho}$ is the ordinary least squares (OLS) estimate of ρ and $s.e.(\hat{\rho})$ is the OLS standard error. We report the test results in Tables 1 and 2.[10]

In the pre-1980's sample period, the DFGLS test rejects the null of nonstationarity in the inequality series for 41 out of 49 states at the 10% significance level, which is over 83% of the total samples (see Table 1).[11] That is, we obtained very strong evidence of stationarity for the pre-1980's inequality series. On the other hand, we find no evidence of stationarity for the post 1980's inequality series as the test fails to reject the null for all 49 states even at the 10% significance level (see Table 2). Therefore, it seems that the inequality series exhibit nonstationarity only for the post-1980's data.

As to the real income series, we observe very weak evidence of stationarity in both sub-samples. The DFGLS test fails to reject the null for most states both in the pre-1980's as well as the post-1980's data.[12] We obtained similar empirical evidence in favor of nonstationarity for the human capital data used in Frank (2009) for the sample period from 1945 to 2004.[13] That is, the test implies that the real income and the human capital data obey a nonstationary stochastic process, which is consistent with the upward trend that is observed in Figures 1 and 2.

In essence, our univariate unit root test supports the nonstationarity of the inequality variable only for the post-1980's samples, while the real income and human capital data seem to follow a nonstationary process for the entire sample period. Therefore, cointegration tests for the full sample period in Frank (2009) may suffer from a mis-specification problem.

Tables 1 and 2 around here

[10] We report test results with one lag. The test with two lags yields qualitatively similar results. Results with two lags are available in the not-for-publication appendix.

[11] We also implemented the ADF test. The test rejects the null of nonstationarity in the inequality series for 32 out of 49 states, which is over 65% of the total observations. Since the DFGSL test is asymptotically more powerful than the ADF test, such weaker evidence seems to be due to low power of the ADF test. The ADF test results with 1 and 2 lags are available upon request.

[12] All test results are available from authors upon request.

[13] These test results are also available from authors upon request.

3.1.2 Panel Unit Root Test

We note that the pre-1980 and the post-1980 sub-samples include 35 and 32 annual observations, respectively. Since the univariate unit root test has low power in small samples, we investigate the possibility that weak evidence of stationarity is due to lack of power. For this purpose, we implement a panel unit root test proposed by Pesaran (2007). By adding more observations in a panel framework, we may expect greater power gains from using panel test as suggested by Taylor and Sarno (1998), for example. However, it is crucial to do a pre-test about the cross-section structure of the panel data, because panel tests that require cross-section independence suffer from severe size distortion in the presence of the cross-section dependence.

Employing the formal test proposed by Pesaran (2004), we establish the existence of cross-section dependence in our data. Consider the following test statistic,

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{i,j} \right) \xrightarrow{d} N(0,1) \tag{3}$$

where $\hat{\rho}_{i,j}$ is the pair-wise correlation coefficients from the residuals of the ADF regressions for each state.

The test results in Table 3 imply a strong presence of cross-section dependence in the panels for inequality, real income per capita, and the two measures of human capital. The test statistics reject the null of the cross-section independence at the 1% significance level for all series. Total average $\hat{\rho}$ is 0.473 and 0.524 for inequality and the real income, respectively. For the two human capital series, average $\hat{\rho}$ was much greater for the high school degree attainment data in comparison to that of the college degree attainment data, although the cross-section independence null was still strongly rejected for both human capital series. We also report average correlations of each state in Figures 3 through 6, which show high degree cross-section dependence in all variables.

Table 3, Figures 3, 4, 5, and 6 around here

Since all data including the inequality and the real income series are characterized by cross-section dependence, we employ the so-called second generation panel unit root tests, because the first generation panel unit root tests such as the ones by Im, Pesaran, and Shin (2003), Levin, Lin, and James Chu (2002), and Maddala and Wu (1999) require cross-section independence, which is clearly rejected for our state-level data in Table 3. In this paper, we implement a panel unit root test proposed by Pesaran (2007) with the following least squares regression model,

$$\Delta y_{i,t} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i \bar{y}_{t-1} + \sum_{j=0}^{p} \theta_{ij} \Delta \bar{y}_{t-j} + \sum_{j=1}^{p} \delta_{ij} \Delta y_{i,t-j} + \varepsilon_{i,t} \tag{4}$$

where $y_{i,t}$ is a variable in state $i \in \{1, 2, ..., N\}$ at time t and $\bar{y}_t$ denotes the common factor at time t, which is proxied by the cross-section mean, $N^{-1} \sum_{i=1}^{N} y_{i,t}$. Note that this is a version of the ADF regression model extended by the cross-section mean in order to control for the effect of the common factor on the panel unit root test. The panel test statistic is then computed as follows.

$$t(N, T) = N^{-1} \sum_{i=1}^{N} t_i(N, T) \tag{5}$$

where $t_i(N, T)$ is the t-statistic for β_i from the regression equation (4) for state $i \in \{1, 2, ..., N\}$.

It should be noted that the panel unit root test using this procedure requires an assumption that the common factor is *stationary*. When this assumption holds, the panel unit root test based on (5) provides meaningful inferences on the stationarity of the panel $\{y_{i,t}\}_{i=1,..,N,t=1,..,T}$. If this assumption fails, however, stationarity evidence from idiosyncratic components does not necessarily provide evidence in favor of stationarity.

Therefore, we first report the unit root test results for the common factors of the inequality and the real income data as well as the human capital data in Table 4. Table 5 provides Pesaran's (2007) panel unit root test results based on (5). Note that the ADF test rejects the null of nonstationarity only for the inequality common factor during the pre-1980 period. Combined with this, strong evidence of panel stationarity for the idiosyncratic components implies that only the inequality for the pre-1980 period obeys a stationary stochastic process. The test results imply strong evidence of nonstationarity for all other variables in both sub-sample periods.[14]

Therefore, we conclude that there is a possible mis-specification problem in Frank's (2009) approach, who uses panel cointegration tests for the state-level data for the inequality from 1945 to 2004 that includes both the pre-1980 and the post-1980 periods. Cointegration tests require nonstationarity in all variables in the cointegrating relationship. Our unit root tests imply that one may employ a panel cointegration framework only for the post-1980 sample period, because the inequality series show clear evidence of stationarity for the pre-1980 samples.

[14]High school human capital variable exhbits very strong evidence of nonstationarity as the test fails to reject the null of nonstationarity for both the common factor and the idiosyncratic components.

Tables 4 and 5 around here

3.2 Cointegration Test

In addition to the nonstationarity issue, Frank's (2009) findings may not be valid because he employed cointegration tests that require cross-section independence. In this section, we implement robust cointegration tests that incorporate cross-section dependence in the data. We demonstrate that Frank's finding of a positive relationship between inequality and the real income is not empirically supported when correct econometric procedures are used.

As explained above, it is appropriate to test for cointegration only for the post-1980 sample period, because inequality obeys a stationary stochastic process in the pre-1980 period. Nonetheless, we implement the cointegration test using the full sample to replicate the empirical findings reported by Frank (2009). Then, we compare the results with those from rigorous test procedures that allow cross-section dependence which clearly exists in the US state-level data as shown in the previous section.

We first implement our analysis for the sample period between 1945 and 2011 using inequality and the real income data, because Mark Frank's human capital data are available only until 2004. We also provide test results for the same specification used in Frank (2009) to highlight the mis-specification issues in his work.

For this purpose, we employ the error correction-based panel cointegration tests proposed by Westerlund (2007).[15] The tests allow for a large degree of heterogeneity between the cross-sectional units and can account for cross section dependence via bootstraps. The tests assume the following data-generating process,

$$\Delta y_{it} = \delta_i' \mathbf{d}_t + \alpha_i (y_{i,t-1} - \beta_i' \mathbf{x}_{i,t-1}) + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{i,t-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta \mathbf{x}_{i,t-j} + e_{it}, \tag{6}$$

where $\mathbf{d}_t$ is a vector (or scalar) of deterministic components. α_i denotes the error correction parameter with the cointegrating vector $[1 \quad -\beta_i']'$. p_i and q_i are the numbers of lags and leads, respectively. (6) can be rewritten as follows.

$$\Delta y_{it} = \delta_i' \mathbf{d}_t + \alpha_i y_{i,t-1} - \lambda_i' x_{i,t-1} + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{i,t-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta x_{i,t-j} + e_{it}, \tag{7}$$

where $\lambda_i' = -\alpha_i \beta_i'$. Note that $\alpha_i < 0$ implies that there is an error correction when deviations from the long-run equilibrium occur. If $\alpha_i = 0$, there is no cointegration because there is no adjustment toward the long-run equilibrium when shocks occur.

[15] We used the stata code following instructions from Persyn and Westerlund (2008).

Westerlund (2007) propose two types of the cointegration test with the null hypothesis $H_0 : \alpha_i = 0, \forall i$, that is, there is no cointegration for all i. Note that the test can be implemented without paying much attention on the cointegration vector β_i itself. They propose the following two tests: the group mean tests and the panel tests. The *group mean test* does not require homogeneity in α_i estimates. That is, the alternative hypothesis is $H_A : \alpha_i < 0$, for at least one i. On the other hand, his *panel test* requires homogeneity with $H_A : \alpha_i = \alpha < 0, \forall i$.

Our test results in Table 6 clearly reveal our point. When we impose a cross-section independence assumption, both the group mean test and the panel test strongly reject the null of no cointegration. However, the test that incorporates cross-section dependence via bootstraps fails to reject the null of no cointegration whichever specifications are employed.[16] To put it differently, Frank's empirical results seem to reflect size distortion caused by imposing a wrong assumption of cross-section independence in addition to the mis-specification problem which was explained in the previous section. Accounting for cross-section dependence in our cointegration tests, we find no statistically meaningful evidence for cointegration between inequality and the real income.

Table 6 around here

Next, we present further test results using the exact same model specification used in Frank (2009). That is, we added the two measures of human capital to the cointegration model for the sample period between 1945 and 2004. Results are provided in Table 7. We obtained qualitatively similar results. In 3 out of the 4 cases, the test fails to reject the null of no cointegration when cross-section dependence is allowed. The test continues to reject the null hypothesis when the panel test is implemented with an intercept. However, the panel test requires homogeneity of the error correction coefficient α, which can be restrictive. Therefore, empirical findings presented in Table 7 seem to be overall consistent with those in Table 6.[17]

Table 7 around here

[16] We implemented the same tests with different combinations of leads and lags as well as different types of kernels and bandwidths. Results are similar with each other.

[17] We also performed this test for the post-1980 samples, which yielded qualitatively similar results. The test fails to reject the null of no cointegration for 3 out of the 4 cases in both the two-variable and the four-variable models at the 5% significance level.

4 Conclusion

This paper revisits the cointegrating relationship between income inequality and economic growth using Frank's (2009) state-level inequality measures data constructed from confidential individual tax filing data from the IRS.

We question the validity of his findings that imply a positive long-run relationship between inequality and economic growth raising two issues. First, his cointegration analyses may have a mis-specification problem as to the order of integration of the data. As documented, cointegrating tests can be implemented among the integrated nonstationary variables. Via an array of univariate and panel unit root tests, we demonstrate that the nature of the stochastic process in the income inequality series has changed around 1980. More specifically, the inequality index seems to obey a stationary process during the pre-1980 sample period, while the real income data follows a non-stationary process for the entire sample period. That is, the econometric model in Frank (2009) may be mis-specified for the pre-1980 data.

Second, we note that Frank's panel cointegration tests require cross-section independence, which is strongly rejected by our test for the US state-level data. Employing rigorous panel cointegration tests that allow cross-section dependence via bootstraps, we find no such evidence of a stable long-run relationship using the same data series. We obtained the same positive cointegration results only when cross-section independence is assumed. Put it differently, the strong evidence of cointegration found in Frank (2009) is likely to be caused by size distortion by imposing a wrong assumption of cross-section independence. Using the exactly same model specification with two measures of human capital as in Frank (2009), we obtained qualitatively similar results as those from our benchmark model, which highlights our points.

References

Alesina, A., and R. Perotti (1994): "The Political Economy of Growth: A Critical Survey of the Recent Literature," *World Bank Economic Review*, 8(3), 351–71.

Alesina, A., and D. Rodrik (1994): "Distributive Politics and Economic Growth," *The Quarterly Journal of Economics*, 109(2), 465–90.

Atkinson, A. B., T. Piketty, and E. Saez (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49(1), 3–71.

Barro, R. J. (1991): "Economic Growth in a Cross Section of Countries," *The Quarterly Journal of Economics*, 106(2), 407–43.

——— (2000): " Inequality and Growth in a Panel of Countries," *Journal of Economic Growth*, 5(1), 5–32.

Birdsall, N., D. Ross, and R. Sabot (1995): "Inequality and Growth Reconsidered: Lessons from East Asia," *World Bank Economic Review*, 9(3), 477–508.

Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore (2012): "Recent Trends in Top Income Shares in the United States: Reconciling Estimates from March CPS and IRS Tax Return Data," *The Review of Economics and Statistics*, 94(2), 371–388.

Deininger, K., and L. Squire (1996): "Measuring Income Inequality: A New Database," Development Discussion Paper 537, Harvard Institute for International Development.

——— (1998): "New Ways of Looking at Old Issues: Inequality and Growth," *Journal of Development Economics*, 57(2), 259–287.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813–36.

Forbes, K. J. (2000): "A Reassessment of the Relationship between Inequality and Growth," *American Economic Review*, 90(4), 869–887.

Frank, M. W. (2009): "Inequality And Growth In The United States: Evidence From A New State-Level Panel Of Income Inequality Measures," *Economic Inquiry*, 47(1), 55–68.

Garcia-Penalosa, C., E. Caroli, and P. Aghion (1999): "Inequality and Economic Growth: The Perspective of the New Growth Theories," *Journal of Economic Literature*, 37(4), 1615–1660.

Im, K. S., M. H. Pesaran, and Y. Shin (2003): "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 115(1), 53–74.

Leigh, A. (2007): "How Closely Do Top Income Shares Track Other Measures of Inequality?," *Economic Journal*, 117(524), F619–F633.

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

Maddala, G. S., and S. Wu (1999): " A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics*, 61(0), 631–52.

Panizza, U. (2002): " Income Inequality and Economic Growth: Evidence from American Data," *Journal of Economic Growth*, 7(1), 25–41.

Persson, T., and G. Tabellini (1994): "Is Inequality Harmful for Growth?," *American Economic Review*, 84(3), 600–621.

Persyn, D., and J. Westerlund (2008): "Error-correction-based cointegration tests for panel data," *Stata Journal*, 8(2), 232–241.

Pesaran, M. H. (2004): "General Diagnostic Tests for Cross Section Dependence in Panels," IZA Discussion Papers 1240, Institute for the Study of Labor (IZA).

——— (2007): "A simple panel unit root test in the presence of cross-section dependence," *Journal of Applied Econometrics*, 22(2), 265–312.

Piketty, T., and E. Saez (2003): "Income Inequality in the United States, 19131998," *The Quarterly Journal of Economics*, 118(1), 1–41.

——— (2006): "The Evolution of Top Incomes: A Historical and International Perspective," *American Economic Review*, 96(2), 200–205.

Quadrini, V., and J.-V. Rios-Rull (2015): "Inequality In Macroeconomics," in *The Handbook of Income Distribution*, ed. by A. B. Atkinson, and B. Franois, vol. 2B, chap. 15. North-Holland Press.

Saez, E., and G. Zucman (2014): "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," CEPR Discussion Papers 10227, C.E.P.R. Discussion Papers.

Taylor, M. P., and L. Sarno (1998): "The behavior of real exchange rates during the post-Bretton Woods period," *Journal of International Economics*, 46(2), 281–312.

Westerlund, J. (2007): "Testing for Error Correction in Panel Data," *Oxford Bulletin of Economics and Statistics*, 69(6), 709–748.

Table 1. DFGLS Test for the Inequality Index: 1945 to 1979

State	DFGLS	State	DFGLS
Alabama	-2.245**	Nebraska	-2.079**
Arizona	-2.547**	Nevada	-2.663***
Arkansas	-2.265**	New Hampshire	-2.178**
California	-1.220	New Jersey	-1.810*
Colorado	-1.904*	New Mexico	-1.939*
Connecticut	-2.476**	New York	-2.330**
Delaware	-1.170	North Carolina	-1.472
District of Columbia	-1.051	North Dakota	-2.006**
Florida	-1.696*	Ohio	-2.133**
Georgia	-1.442	Oklahoma	-3.160***
Idaho	-2.406**	Oregon	-1.720*
Illinois	-2.241**	Pennsylvania	-2.148**
Indiana	-1.903*	Rhode Island	-1.765*
Iowa	-1.773*	South Carolina	-1.896*
Kansas	-1.579	South Dakota	-2.003**
Kentucky	-2.629***	Tennessee	-2.191**
Louisiana	-2.450**	Texas	-1.622*
Maine	-3.243***	Utah	-1.400
Maryland	-1.850*	Vermont	-1.875*
Massachusetts	-2.094**	Virginia	-2.035**
Michigan	-2.564**	Washington	-1.161
Minnesota	-1.757*	West Virginia	-2.290**
Mississipi	-1.798*	Wisconsin	-2.396**
Missouri	-1.757*	Wyoming	-2.245**
Montana	-1.876*		

Note: We report the DFGLS test results with an intercept. *, **, and *** denote rejections of the null hypothesis of nonstationarity in the 10%, 5%, and 1% significance level, respectively.

Table 2. DFGLS Test for the Inequality Index: 1980 to 2011

State	DFGLS	State	DFGLS
Alabama	-1.502	Nebraska	-1.127
Arizona	-1.311	Nevada	-1.131
Arkansas	-0.182	New Hampshire	-1.055
California	-0.759	New Jersey	-0.685
Colorado	-0.817	New Mexico	-0.782
Connecticut	-0.540	New York	-0.580
Delaware	-1.162	North Carolina	-0.962
District of Columbia	-0.862	North Dakota	0.305
Florida	-1.075	Ohio	-1.117
Georgia	-0.739	Oklahoma	-0.645
Idaho	-0.828	Oregon	-0.952
Illinois	-0.899	Pennsylvania	-1.287
Indiana	-1.063	Rhode Island	-1.043
Iowa	-1.263	South Carolina	-1.214
Kansas	-0.865	South Dakota	-0.692
Kentucky	-1.479	Tennessee	-1.052
Louisiana	-1.575	Texas	-0.764
Maine	-1.041	Utah	-1.192
Maryland	-0.972	Vermont	-1.134
Massachusetts	-0.785	Virginia	-1.139
Michigan	-0.639	Washington	-1.252
Minnesota	-0.974	West Virginia	-0.418
Mississipi	-1.307	Wisconsin	-0.997
Missouri	-0.990	Wyoming	-1.100
Montana	-0.734		

Note: We report the DFGLS test results with an intercept. *, **, and *** denote rejections of the null hypothesis of nonstationarity in the 10%, 5%, and 1% significance level, respectively.

Table 3. Cross-Section Dependence Test Results

	Inequality	Real Income	High School	College
CD	129.68***	143.86***	133.10***	28.48***
Average $\hat{\rho}$	0.473	0.524	0.525	0.112

Note: CD is Pesaran's (2004) cross-section dependence statistic. *** denotes a rejection of the cross-section independence at the 1% significance level. The sample period is from 1945 to 2011 for inequality and the real income, while it is from 1945 to 2004 for the human capital variables, obtained from Mark Frank's website.

Table 4. Unit Root Test Results: Common Components

	Inequality	Real Income	High School	College
1945 – 1979	-2.541*	0.488	-1.188	1.301
1980 – 2011	-2.072	-1.638	-1.986	-0.152

Note: The common components are identified by taking the cross-section means of the series. The first common factors estimated via the method of the principal components are qualitatively similar to the cross-section means. * denotes a rejection of the nonstationarity null hypothesis at the 10% significance level. The sample period is from 1945 to 2011 for inequality and the real income, while it is from 1945 to 2004 for the human capital variables, obtained from Mark Frank's website. The ADF test fails to reject the null of nonstationarity for the full sample data.

Table 5. Panel Unit Root Test Results: Idiosyncratic Components

	Inequality	Real Income	High School	College
1945 – 1979	-3.310***	-2.694***	-1.625	-2.806***
1980 – 2011	-2.473***	-2.075*	-1.930	-2.722***

Note: Test statistics are from Pesaran (2007) that controls the cross-section dependence. * and *** denote rejections of the nonstationarity null hypothesis at the 10% and 1% significance level, respectively. Critical values are obtained from Pesaran (2007).

Table 6. Panel Cointegration Test Results

	Tests with an intercept		
	Statistics	p-value	p-value with CSD
Group Mean Test	-2.005	0.048	0.548
Panel Test	-13.972	0.000	0.258

	Tests with an intercept and time trend		
	Statistics	p-value	p-value with CSD
Group Mean Test	-3.103	0.000	0.156
Panel Test	-19.177	0.000	0.438

Note: The test is implemented using inequality and the real income data from 1945 to 2011. We implement Westerlund's (2007) t-test type panel cointegration test statistics. Number of leads and lags are determined by the AIC. p-value is not sized correctly when cross-section independence fails to hold. p-value with CSD denotes p-values with cross-section dependence via 500 bootstraps. The null hypothesis is no cointegration for both tests. The group mean test does not require homogeneity and the alternative hypothesis is there is at least one cointegration. The panel test does require homogeneity and the alternative hypothesis is the common cointegration exists for all panel seris.

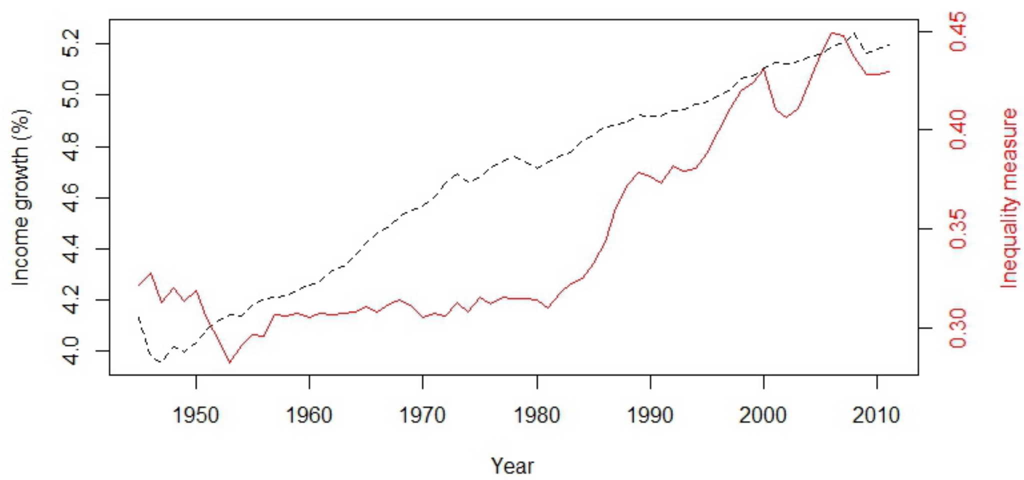
Table 7. Panel Cointegration Test Results with Frank's (2008) Model

	Tests with an intercept		
	Statistics	p-value	p-value with CSD
Group Mean Test	-2.613	0.003	0.178
Panel Test	-18.192	0.000	0.034

	Tests with an intercept and time trend		
	Statistics	p-value	p-value with CSD
Group Mean Test	-2.865	0.100	0.560
Panel Test	-19.007	0.020	0.420

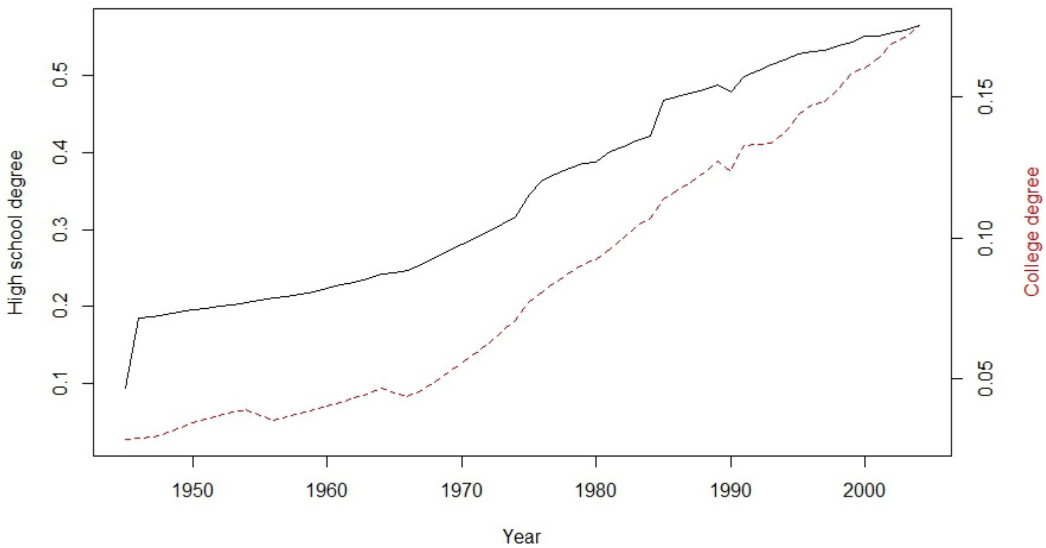
Note: The test is implemented using inequality,the real income, and the human capital data from 1945 to 2004.We implement Westerlund's (2007) t-test type panel cointegration test statistics for the empirical model used in Frank (2008). That is, we added two sets of human capital variables in the cointegration model for the same sample period (1945 - 2004) as in Frank (2008). Human capital data were obtained from Mark Frank's website. Number of leads and lags are determined by the AIC. p-value is not sized correctly when cross-section independence fails to hold. p-value with CSD denotes p-values with cross-section dependence via 500 bootstraps. The null hypothesis is no cointegration for both tests. The group mean test does not require homogeneity and the alternative hypothesis is there is at least one cointegration. The panel test does require homogeneity and the alternative hypothesis is the common cointegration exists for all panel seris.

Figure 1. Inequality (Solid) and Real Income per capita (Dashed)



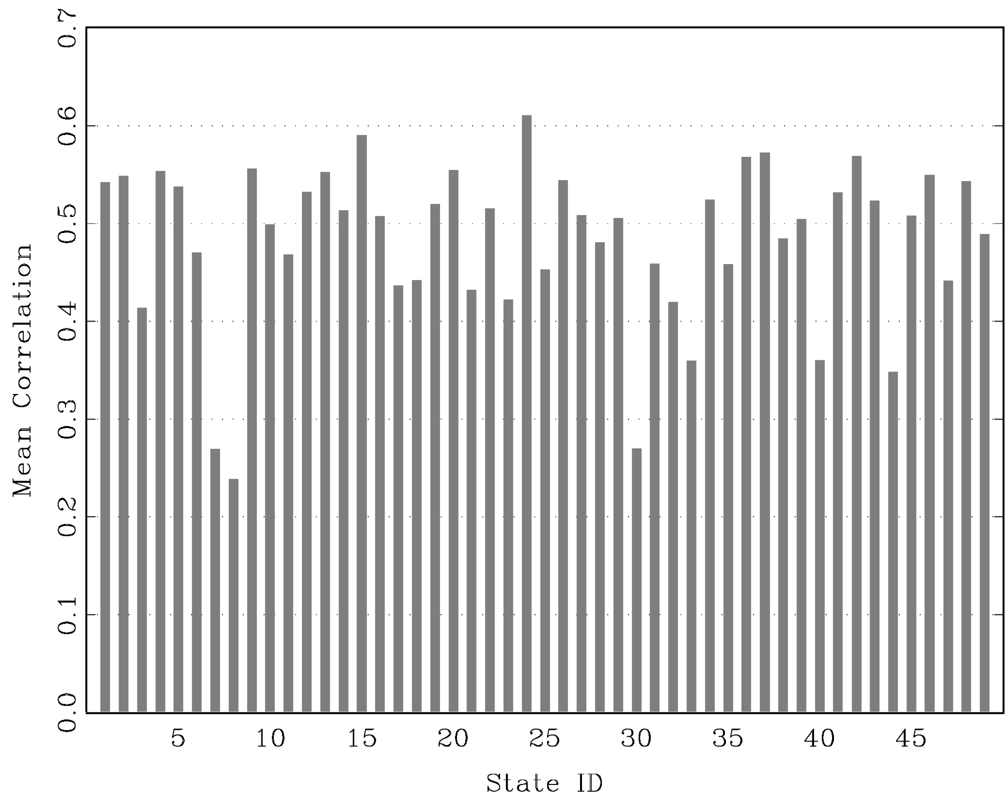
Note: Cross-section averages of the 49 state-level data are presented.

Figure 2. High School (Solid) and College (Dashed) Degree Ratios



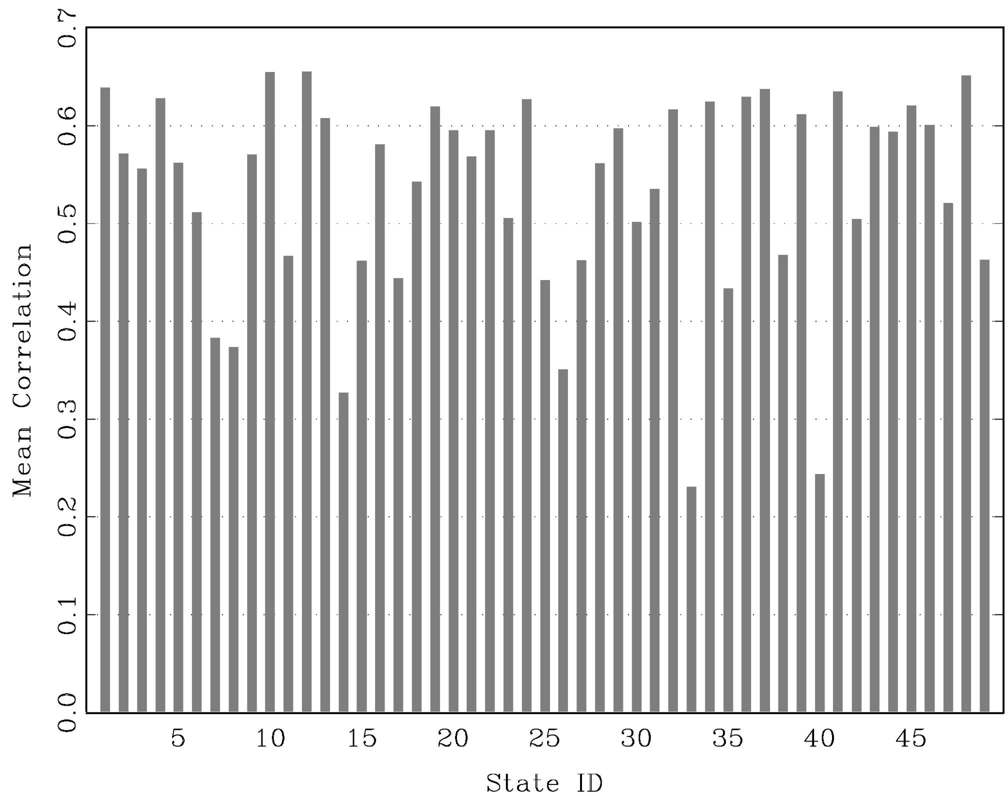
Note: Cross-section averages of the 49 state-level data are presented.

Figure 3. Mean Correlation Coefficients: Inequalty Series



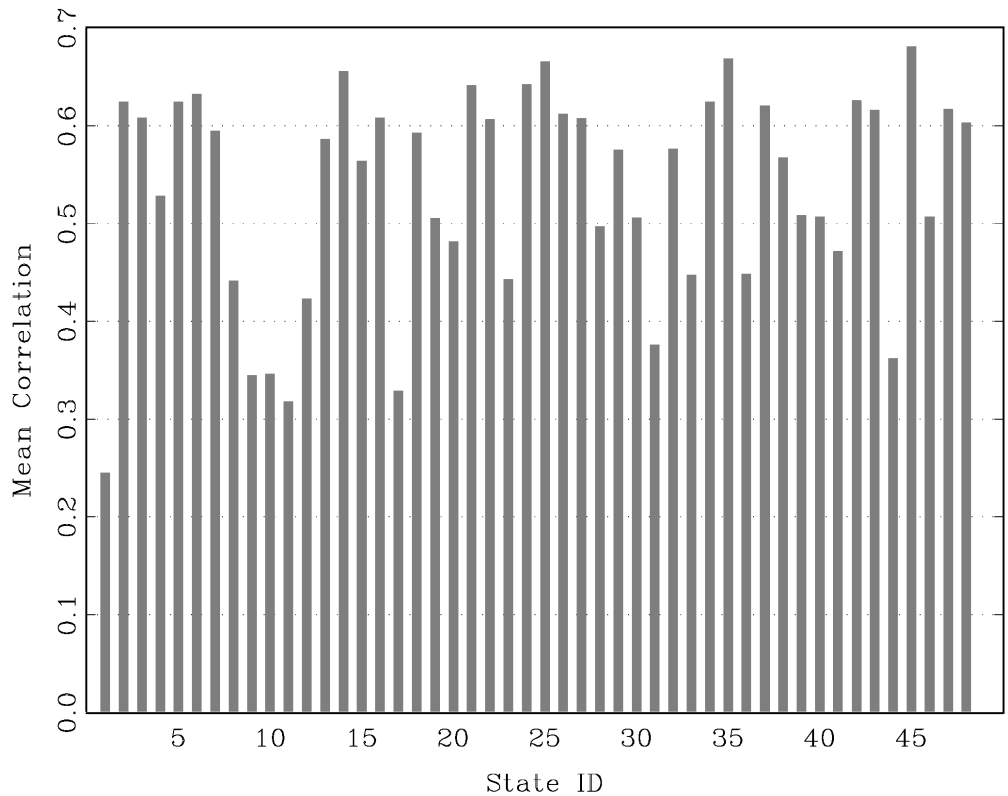
Note: We report the mean correlation coefficient of each state with respect to other 48 states.

Figure 4. Mean Correlation Coefficients: Real Income Series



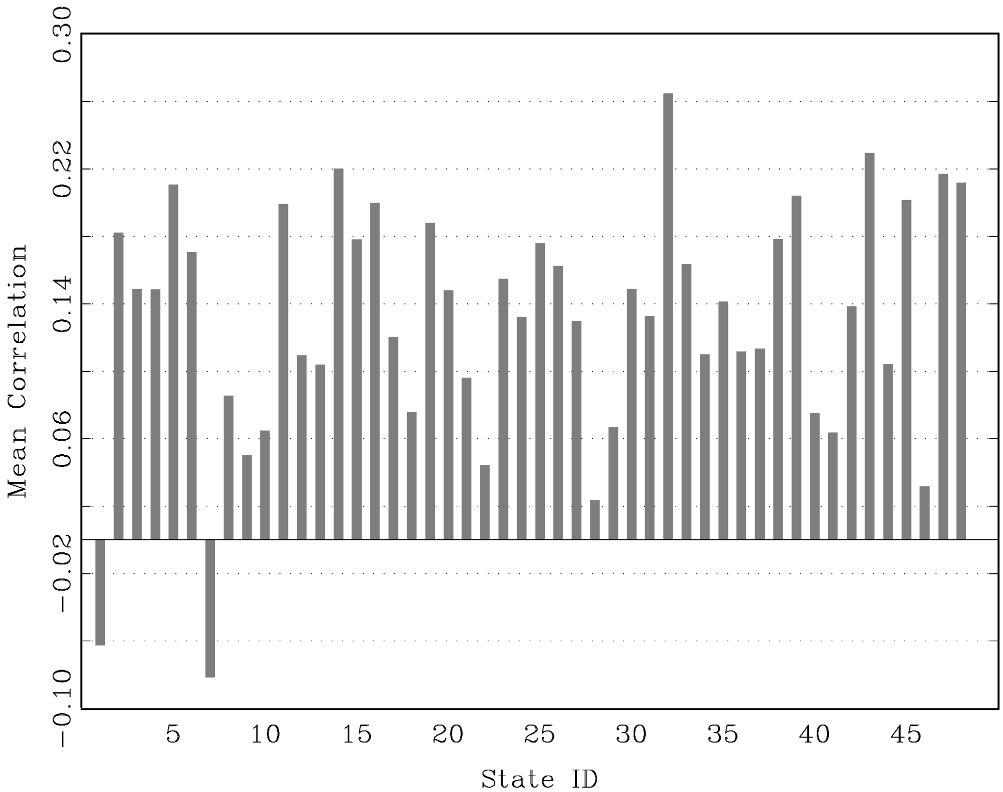
Note: We report the mean correlation coefficient of each state with respect to other 48 states.

Figure 5. Mean Correlation Coefficients: High School Ratio



Note: We report the mean correlation coefficient of each state with respect to other 48 states.

Figure 6. Mean Correlation Coefficients: College Ratio



Note: We report the mean correlation coefficient of each state with respect to other 48 states.

Chapter 2: A Second Generation Cross-Country Cointegration Analysis of Housing Prices and Fundamentals

Abstract

We study the long-run dynamics between real house prices and key macroeconomic variables namely the real GDP per capita, the real lending interest rate and the CRB raw industrial index with the use of an array of univariate and panel econometric procedures. The study is conducted on a panel of 13 OECD countries over a period of 22 years. Our empirical analysis is carried out in a four-step process: unit root testing, cointegration testing, cointegrating vector estimation and vector error correction model (VECM) estimation, which we use for an impulse-response analysis. This paper is novel for two main reasons. First, we implement two of the so-called second generation panel tests, which allow for the existence of serially correlated panels (i.e. cross-section dependence). Second, we include the CRB raw industrial index to our model as a supply-side variable, which proxies for building costs. Our results corroborate the predominant findings in the literature: housing prices have a positive long-run relationship with real GDP per capita and the commodity index and a long-run negative relationship with the real interest rate.

1 Introduction

The fluctuations in house prices across the globe over the past decades and their resulting impact on economies worldwide have increased the interest of researchers in the real estate sector. The significance of the housing market's impact on the economy was emphasized by Brocker & Hanes (2014) who made two important claims: (i) the bursting of a housing bubble has more impact on the subsequent recovery of an economy than the collapse of a stock market and (ii) the exposure of bank balance sheets to the housing market is greater than their exposure to equity and derivative markets. The most notable price swings in the international housing market in the last few decades were in the 1970s and 1980s as house prices significantly rose and fell (see Levin & Wright (1997); McCord et al. (2011)). In the 1990s, real house prices doubled in the United Kingdom and a similar increase was also seen in Spain (Niemietz, 2012). In the USA, the real estate market experienced gains in the 1980s followed by a fall in prices in the first half of the 1990s, which reversed the gains by more than half. Many experts believe that this fall in prices was a consequence of houses being over-valued in the 1980s (i.e. a bubble). Real prices started rising again over the next decade even outpacing the price increases that occurred in the 1980s. This raised the logical concern that the house market was over-valued once again (see Gallin (2006)). In the same vein the most recent crisis, the 2008 global financial crisis, is widely considered to have been caused by house price bubbles (see Duca et al. (2010), Baker (2008), Demyanyk & Van Hemert (2011)). Consequently, given how important the housing market is in driving economic cycles, much research has been carried out in order to understand the dynamics underlying the variations in house prices and their relationship to other variables. In this paper, we use a panel of 13 OECD countries over a period of 22 years to show the long-run relationship that exists between real house prices and 3 important macroeconomic variables, namely real GDP per capita, real interest rate and commodity index. Section (2) of this paper reviews the literature on house prices and their dynamics with key variables in the macroeconomy. In Section (3), we present the data used in our empirical work and provide general descriptions of the series. Section (4) presents the univariate and panel econometric procedures we use and the results we obtain. We finally conclude in Section (5).

2 Literature Review

The literature points at a number of variables as the fundamentals of housing prices – income and consumer prices being the most widely acknowledged ones. Other variables such as the mortgage rate, money supply, housing credit, stock market performance, construction costs, population, employment and other demographic factors are also recognized as important drivers.

Baffoe-Bonnie (1998) used the vector autoregressive (VAR) model to confirm the dynamic effects of employment growth, mortgage rate, anticipated and unanticipated changes in money supply and inflation on housing prices and the stock of houses sold on the national and regional levels. Using the same model, Sutton (2002) found evidence that variations in national income, interest rate and stock prices explained the fluctuations in housing prices in six advanced economies namely the US, the UK, Canada, Ireland, the Netherlands and Australia. Tsatsaronis & Zhu (2004) used the structural vector autoregressive (SVAR) model on 17 industrialized countries and determined that inflation and nominal interest rates had a strong and long-lasting impact on housing prices. They also found out that key variables in the financial sector, especially bank credit, the short-term interest rate and the term-spread, have a long-run impact on housing prices. Egert & Mihaljek (2007) use data on 8 Central and Eastern European Countries as well as 19 OECD countries in a dynamic ordinary least squares (DOLS) model to confirm that GDP per capita, real interest rate, housing credit and demographic factors drive variations in house prices. They also augment the model with transition-specific factors, especially institutional developments of housing markets. Addison-Smyth et al. (2008) analyze the Irish housing market and find that house prices are significantly influenced by interest rates and how much individuals can borrow, which is a function of their disposable income. They use a DOLS model to show that there exists a long-run relationship between house prices and the amounts borrowed by individuals. Adams & Fuss (2010) apply a panel cointegration analysis of 15 countries over 30 years and confirm the existence of a long-term relationship as well as short-term dynamics between house prices and construction costs, long-term interest rates and economic activity, which they define as a set of macroeconomic variables including real money supply, real GDP, real consumption, real industrial production and employment.

More recently, some researchers found that the linear framework, assumed by the majority of the literature on house prices and fundamentals, is not appropriate due to the restrictive nature of the assumption. Kim & Bhattacharya (2009) analyze house prices in the US and the four regions [1] over 36 years using a smooth transition autoregressive (STAR) model. They find that house prices in the US in 3 out of the 4 regions exhibit nonlinear properties. Zhou (2010) apply the ACE algorithm, a nonparametric method, and confirm the existence of nonlinear cointegration between house prices and fundamentals in many US cities. Tsai et al. (2012) studied the relationship between the housing market and the stock markets in

[1] The USA is subdivided into four main regions according to the Census Bureau: the West, Midwest, South and Northeast.

the US. Using a momentum-threshold autoregressive (M-TAR) model, they show evidence of the existence of asymmetric wealth effects in the markets. Katrakilidis & Trachanas (2012) apply the asymmetric autoregressive distributive lag (ARDL) cointegration methodology to monthly data on the Greek housing market spanning over a period of 13 years and their results indicate the presence of asymmetric long-run and short-run effects from the consumer price index (CPI) and the industrial production index (IPI) towards house prices.

3 Data Description

The data used in the empirical work is a panel spanning from 1992 to 2013 (i.e. 22 years) and across 13 OECD countries. The countries considered are: Australia, Denmark, Finland, France, Ireland, Italy, Japan, New Zealand, Norway, Spain, Switzerland, the United Kingdom and the United States. The real house price index was obtained from the databank of the Federal Reserve Bank of Dallas (FRED, 2016). The real values for the house price index were computed by dividing the nominal values by the personal consumption expenditure deflator (see Mack & MartÃnez-GarcÃa (2011)). The real lending interest rate is available in the World Development Indicators (WDI) database. The nominal lending rate is adjusted for inflation using the GDP deflator. The real GDP per capita data were collected from the World Bank Databank. Finally, the commodity index data were obtained from the Commodity Research Bureau (CRB) database [2]. Note that we applied the log-transformation to the real GDP per capita, real house prices and commodity index series.

Real house prices have fairly similar patterns across countries throughout the sample period except for Japan. While house prices have an increasing trend in most countries, the trend in Japan is a decreasing one. Japan is known to be an international outlier as it differentiates itself from other countries in many areas including the housing market (see Bardhan & Kroll (2013)). Other countries that have a slightly distinct pattern are Italy, with an oscillating series with long swings, and Switzerland, with a U-shaped series. The cross-section means plot of real house prices in Figure (1), however, is similar to the pattern exhibited by most countries. Real interest rates series have a similar downward trend across all countries (see Figure (6)) and so does the cross-section means plot in Figure (3). There is also a uniformity in the real GDP per capita series (see Figure (7)) with an overall positive trend, which the cross-section means plot also reflects (see Figure (2)). The commodity index falls up till the period right after 2000, then picks up an increasing trend till the end of the sample period. There is, however, a structural break around 2008, which points out the recession (see Figure (4)).

[2] The commodity index data denotes the CRB raw industrial index.

Expectations of the long-run relationships between our variables can be formulated by simply eye-balling the series. These informal expectations are corroborated by the literature. We expect:

- the real GDP per capita to have a positive long-run relationship with real house prices (see Sutton (2002), Egert & Mihaljek (2007), Adams & Fuss (2010))
- the real interest rate to have a negative long-run relationship with real house prices (see Sutton (2002), Egert & Mihaljek (2007), Adams & Fuss (2010))
- the commodity index to have the positive long-run relationship with real house prices (Adams & Fuss (2010))

4 Empirical Findings

We use an array of econometric procedures on the international data described in Section (3). We implement univariate and panel econometric procedures for the tests and estimations we carry out. Panel procedures are generally known to have more power than their univariate counterparts.

4.1 Unit root tests

4.1.1 Dickey-Fuller Generalized Least Squares (DF-GLS) (Elliott et al. (1996))

One of the requirements for two or more series to be cointegrated is for them to be individually integrated of order n (with $n \geq 1$). In other words, they need to be nonstationary. We test the null of nonstationarity in our series using the DF-GLS test. The test is known to be asymptotically more powerful than the more popular augmented Dickey-Fuller (ADF) test. It is based on the following regression model for each country in our sample:

$$\Delta \tilde{y}_t = \alpha + \rho y_{t-1} + \sum_{j=1}^{p} \beta_j \Delta \tilde{y}_{t-j} + \varepsilon_t, \tag{1}$$

where $\tilde{y}_t$ is the locally demeaned data under the local alternative of $\tilde{\alpha} = 1 + \hat{c}/T$. T is the sample size, p is the number of lags. We use $\hat{c} = 7$ as recommended by Elliott et al. (1996). The DF-GLS statistic is defined as:

$$DFGLS = \frac{\hat{\rho}}{se(\hat{\rho})},$$

where $\hat{\rho}$ is the ordinary least squares (OLS) estimate of ρ in equation (1) and $se(\hat{\rho})$ is the standard error of the coefficient estimate. The DF-GLS results are reported in Tables (1), (2), (3) and (4).

The test is first performed on the cross-section means of our variables and then on each series for each country. The test results on the variables' cross-section means fail to reject the null of nonstationarity for all 4 variables (see Table (1)). It is important to note that the commodity index series is an international variable and its values are, therefore, the same for every country at each time period. Consequently, the actual series' values and the cross-section means are equal.

The results at the individual country level are also strongly in favor of the null of nonstationarity. The test shows evidence against stationarity in:

- 9 out of 13 countries (i.e. 69.23%) for the real house price index series (see Table (2))
- 12 out of 13 countries (i.e. 92.31%) for the real GDP per capita series (see Table (3))
- 13 out of 13 countries (i.e. 100%) for the real interest rate series (see Table (4))

In a nutshell, the unit root test on cross-section means, which shows evidence for nonstationarity in our time series, is strongly corroborated by the test results at the individual country level. We can, therefore, postulate that the prerequisite for the existence of a cointegrating relationship between the variables is satisfied.

(Table 1 around here)

(Table 2 around here)

(Table 3 around here)

(Table 4 around here)

4.1.2 Panel unit root test (Pesaran (2007))

There are 22 time series observations in each country sample. Such small samples are known to decrease the statistical power of univariate unit root tests such as the DF-GLS. Therefore, we investigate the possibility that the weak evidence of stationarity could be a consequence of lack of power. For this reason, we consider a panel unit root test (Pesaran (2007)), which benefits from greater power gains due to the fact that panel frameworks have more observations (see Sarno & Taylor (1998)). It is important to note that the panel unit root test we implement presupposes the existence of cross-section dependence in the panel. Therefore, we conduct a pretest on the cross-section structure of the panels using Pesaran (2004)'s cross-section dependence test, which is as follows:

$$CD = \sqrt{\frac{2T}{N(N-1)}}\Big(\sum_{i=1}^{N-1}\sum_{j=1}^{N}\hat{\rho}_{i,j}\Big) \xrightarrow{d} N(0,1)$$

where N is the number of cross-sections in the panel, T is the number of time periods and $\hat{\rho_{i,j}}$ is the pair-wise correlation coefficients from the residuals of the ADF regressions for each country. The results in Table (5) suggest a strong presence of cross-section dependence in the panels for real house prices, real GDP per capita and real interest rate. The null of cross-section independence is rejected at the 1% significance level for all variables. The total average correlation coefficients are 0.219, 0.527 and 0.603 for real house prices, real interest rates and real GDP per capita, respectively. We also report the average correlations for each country in Figures (8), (9) and (10), which show high degrees of cross-section dependence.

The aforementioned results make Pesaran (2007)'s test an appropriate unit root test for our study. The test is based on the following regression equation:

$$\Delta y_{i,t} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i \bar{y}_{t-1} + \sum_{j=0}^{p}\theta_{ij}\Delta\bar{y}_{t-j} + \sum_{j=1}^{p}\delta_{ij}\Delta y_{i,t-j} + \varepsilon_{i,t} \tag{2}$$

where $y_{i,t}$ is a variable in state $i \in \{1, 2, \dots, N\}$ at time t, and $\bar{y}_t$ denotes the common factor at time t, which is proxied by the cross-section mean: $N^{-1}\sum_{i=1}^{N} y_{i,t}$. This model is essentially an extension of the ADF regression model, which specifically controls for the effect of the common factor on the panel unit root test. The test statistic is then obtained with the following:

$$t(N,T) = N^{-1}\sum_{i=1}^{N} t_i(N,T)$$

where $t_i(N,T)$ is the t-statistic for the coefficient β_i from Equation (2) for state $i \in \{1, 2, \dots, N\}$.

It should be pointed out that Pesaran (2007)'s panel unit root test assumes that the common factor is stationary. In case this requirement is met, the results in Table (6) provide meaningful inferences on the stationarity of the panel $\{y_{i,t}\}_{i=1,\dots,N,t=1,\dots,T}$. However, if the assumption does not hold, the rejection of the null of nonstationarity from the idiosyncratic components does not necessarily provide evidence for stationarity.

The unit root test results for the common factors are reported in Table (1) [3] and the results from Pesaran (2007)'s panel unit root test are in Table (6). The test rejects the null of nonstationarity only for the real house price index series at the 1% significance level. However, since the null of nonstationarity fails to be rejected for all common factors, we cannot deductively say that there is evidence for stationarity of the panel. Our panel unit root test results are, therefore, in favor of the nonstationarity of the panels, which is the prerequisite for the existence of a cointegrating relationship between two or more series.

4.2 Cointegration tests

4.2.1 Johansen cointegration test (Johansen (1988))

Despite the existence of a number of univariate cointegration tests, namely the tests proposed by Robert F. Engle (1987) and James H. Stock (1988), we implement Johansen (1988)'s procedure for its desirable properties of which the treament of all time series as endogenous variables (see Gonzalo (1994)). Let's consider the following vector autoregressive process of first order (VAR(1)):

$$X_t = A_t X_{t-1} + \varepsilon_t, \tag{3}$$

where X_t is a vector of n variables which are individually integrated of order 1. ε_t is a vector containing white noise disturbances. Equation (3) can therefore be rewritten as:

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t, \tag{4}$$

where $\Pi = A_t - I$, A_t is an $n \times n$ matrix of parameters, I is an n-dimensional identity matrix and ε_t is an $n \times 1$ vector.

If the rank of the matrix Π is 0, then all the elements of the matrix are 0 and A_t is an identity matrix. In such a case, X_t is a VAR(1) process represented by:

$$\Delta X_t = \varepsilon_t, \tag{5}$$

which means that each sequence in X_t is difference stationary. Since each variable in X_t is nonstationary, we can then conclude that no linear combination of these variables is stationary and there exists no cointegrating relationship.

In case the matrix Π is full rank, then the long-run equilibrium solution is given by the following n independent equations, which represent n independent restrictions:

[3] We pointed out earlier that the common factors are proxied by the cross-section means.

$$
\begin{aligned}
\Pi_{11}X_{1t} + \Pi_{12}X_{2t} + \cdots + \Pi_{1n}X_{nt} &= 0 \\
\Pi_{21}X_{1t} + \Pi_{22}X_{2t} + \cdots + \Pi_{2n}X_{nt} &= 0 \\
&\vdots \\
\Pi_{n1}X_{1t} + \Pi_{n2}X_{2t} + \cdots + \Pi_{nn}X_{nt} &= 0
\end{aligned}
$$

All the n variables in the system face n long-run constraints. In such a case, each sequence in X_t must be stationary with the long-run values given by the system.

More generally, the number of cointegrating vectors is equal to the rank of Π denoted as r. It is known that the rank of a matrix is equal the number of characteristic roots significantly different from 0. Consequently, the main goal of Johansen's cointegration test is to determine the number of significant characteristic roots. This is achieved by the following test statistics:

$$
\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} ln(1 - \hat{\lambda_i}),
$$

$$
\lambda_{max}(r, r+1) = -T\, ln(1 - \hat{\lambda_{r+1}}),
$$

where $\hat{\lambda_i}$ represents the estimated values of the characteristics roots computed from the estimated matrix Π, and T is the number of observations. The null hypothesis of the first statistic is that there is no distinct cointegrating vector or the number of cointegrating vectors is less than r. The null hypothesis is tested against a more general hypothesis. The second statistic, on the other hand, tests the null hypothesis that the number of cointegrating relationships is equal to r, against the alternative $r + 1$.

The trace and eigen statistics are reported in Table (7). The null hypothesis, which states that there exists no cointegrating relationship between the sequences, is significantly rejected for every country in our sample using both statistics. Our univariate results therefore show evidence of cointegration between the variables. Consequently, the variables are related by a stable long-run relationship and any deviation from the long-run equilibrium will trigger an error-correction mechanism for the equilibrium to be restored.

4.2.2 Panel cointegration test (Westerlund (2007))

The error-correction based panel cointegration tests we employ were proposed by Westerlund (2007). What makes these tests appropriate for this study is that they can account for cross-section dependence in the panels via bootstraps. We already established the fact that our series are cross-sectionally dependent in Section (4.1.2). The two tests, which are the group mean test and the panel test, start with the following regression equation:

$$\Delta y_{it} = \delta_i^T d_t + \alpha_i(y_{i,t-1} - \beta_i^T x_{i,t-1}) + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{i,t-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta x_{i,t-j} + e_{it}, \tag{6}$$

where d_t is a vector (or scalar) of deterministic components, α_i is the error-correction coefficient associated with the cointegrating vector $(1 - \beta_i^T)^T$. p_i and q_i are, respectively, the number of lags and the number of leads chosen for each cross-section. We can rewrite Equation (6) as follows:

$$\Delta y_{it} = \delta_i^T d_t + \alpha_i y_{i,t-1} - \lambda_i^T x_{i,t-1} + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{i,t-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta x_{i,t-j} + e_{it}, \tag{7}$$

where $\lambda_i^T = -\alpha_i \beta_i^T$. If $\alpha_i < 0$, then any deviations from the long-run equilibrium will be followed by an error-correction. On the other hand, $\alpha_i = 0$ implies that there is no reversion to the long-run equilibrium when deviations occur (i.e. there is no cointegration). The tests are implemented under the null hypothesis of no cointegration for all cross-sections.

The group mean test does not require homogeneity in α_i estimates. This means that the alternate hypothesis is $H_A : \alpha_i < 0$ for at least one of the cross-sections i. The panel test, however, is more restrictive in the fact that it requires homogeneity and the alternate hypothesis is the same for all cross-sections ($H_A : \alpha_i = \alpha < 0$, $\forall i$).

Table (8) reports the results of the test. The p-values are obtained under the assumption of cross-section independence and the p-values with cross-section dependence (i.e. robust p-values) are obtained via 500 bootstraps. The p-values with CSD are therefore our statistics of interest given the serially correlated nature of our panels.

The group mean test results do not find enough evidence to reject the null of no cointegration (p-value with CSD = 0.662). However, the panel test rejects the null at the 10% significance level (p-value with CSD = 0.068). Westerlund (2007) points out the fact that the panel test has the highest power of the two tests because it is based on the pooled least squares estimator of α_i, which is efficient under the homogeneity assumption. The test with the highest statistical power, therefore, substantiates the univariate results and thus shows evidence for an existing stable long-run relationship between real house prices and the other macroeconomic variables considered in our study.

4.3 Cointegrating vector estimation

4.3.1 Dynamic Ordinary Least Squares (DOLS) (Saikkonen (1991) and Stock & Watson (1993))

We define y_t to be a vector of difference stationary random variables such that: $y_t = \begin{bmatrix} p_t & g_t & r_t & c_t \end{bmatrix}$, where p_t is the real house price index, g_t represents the real GDP per capita, r_t is the real lending interest rate, c_t is the commodity index. The subscript t represents the time period. The cointegration test results we obtained in Section (4.2) provide evidence that there exists a non-zero vector of real numbers $\gamma = \begin{bmatrix} 1 & -\beta^T \end{bmatrix}$ such that $\gamma^T y_t$ is stationary, where $\beta = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \end{bmatrix}^T$. In other words, y_t is cointegrated with a cointegrating vector γ. Then, according to Phillips (1991), the triangular representation of the cointegrated vector process is:

$$p_t = \alpha + \beta_1 g_t + \beta_2 r_t + \beta_3 c_t + \varepsilon_t \tag{8}$$

$$\Delta x_t = \delta + u_t \tag{9}$$

where $\Delta x_t = \begin{bmatrix} \Delta g_t & \Delta r_t & \Delta c_t \end{bmatrix}^T$ is a 3×1 vector of differenced series, α is a vector of constants, ε_t is zero-mean stationary for β and u_t is a 3×1 zero-mean stationary vector. The cointegrating vector γ is assumed to eliminate all trends (i.e. deterministic and stochastic trends), which explains the non-inclusion of a time trend in the cointegrating regression (8) [4].

Our goal is to estimate the cointegrating vector $\hat{\beta}$ from equation (8). The ordinary least squares estimator $\hat{\beta_{LS}}$ is super-consistent as it converges to the true value β at the rate of T (the sample size) even in the presence of serial correlation (i.e when x_t and ε_t are correlated). However, it is asymptotically biased and inefficient and its asymptotic distribution is non-normal [5]. As a consequence, the standard errors produced by the least squares estimator cannot be reliably used for statistical inference. A more adequate estimation technique for the cointegrating vector in this case scenario is the Dynamic Ordinary Least Squares (DOLS). Employing DOLS consists in regressing p_t on x_t as well as leads and lags of Δx_t. Newey-West standard errors (Newey & West (1994)) are then used. We report the DOLS cointegrating vector estimates for each country in Table (10) [6]. The standard errors reported in the table

[4] Ogaki & Park (1998)show that this is the case when the deterministic cointegrating restriction is satisfied. When the stochastic trend only is eliminated by the cointegrating vector, $\gamma^T y_t$ is trend stationary.

[5] See Stock (1987) and Phillips (1991) for details.

[6] We check the robustness of our results by using two other equally appropriate estimation techniques,

are based on Andrews & Monahan (1992)'s pre-whitening method.

All the coefficient estimates for all variables are significant at least at the 10% significance level with the only exception of the coefficient for real GDP per capita in Japan. These results are strong evidence for the existence of a long-run relationship between real house prices and each of the other variables in our model. Also, the signs of the coefficients are predominantly in line with the expectations we formulated in Section (3) except for the real interest rate coefficients:

- the real GDP per capita coefficient is positive for 11 out of 13 countries (i.e. 84.61%)
- the commodity index coefficient is positive for 10 out of 13 countries (i.e. 76.92%)
- the real interest rate coefficient is negative only for 5 out of 13 countries (i.e. 38.46%)

(Table 10 around here)

The relative small size of our sample (i.e. 22 observations) could be the reason why the coefficients on the interest rate poorly reflect our expectations. The small nature of our sample and the model we estimate may fail to fully capture the idiosyncracies of each country. Inder (1993) and Stock and Watson (1993) show that DOLS results may perform poorly on small samples.

4.3.2 Panel Dynamic Ordinary Least Squares (PDOLS) (Mark & Sul (2003))

The panel DOLS extends the univariate DOLS to a panel framework. Kao and Chiang (2000) discuss the estimation technique's features in the presence of fixed-effects in the cointegrating regression and Mark & Sul (2003) use their approach as a starting point. They assume that the cointegrating vector is homogenous accross cross-section units (long-run dynamics), but individual heterogeneity is allowed through distinct short-run dynamics, individual-specific fixed effecs and individual-specific time trends. In addition, they allow for a limited degree of cross-sectional dependence through the inclusion of time-specific effects. In this paper, the only heterogeneity we account for is the heterogeneity through country-specific fixed effects.

The benefits of using the panel DOLS over its univariate counterpart are manifold. Inder (1993) and Stock & Watson (1993) show that the statistical properties of the univariate DOLS can be quite poor when used on relatively small sample sizes. Also, the existence of heterogeneity in the short-run dynamics across cross-section units can yield significant

namely: the Fully Modified Ordinary Least Squares (FM-OLS) (Phillips & Hansen (1990)) and the Integrated Modified Ordinary Least Squares (IM-OLS) (Vogelsang & Wagner (2014)). The results obtained are very similar to those obtained by the DOLS and are available upon request.

disparities in the univariate DOLS estimates of the cointegrating vector. Therefore, using a panel leverages cross-sectional and time-series information and can yield much more precise point estimates.

The estimated coefficients and the standard errors are reported in Table (9). They are as follows (standard errors are in parentheses):

- The coefficient for real GDP is positive: 1.950 (0.343)
- The coefficient for real interest rate is negative: -0.001 (0.026)
- The coefficient for commodity index is positive: 0.009 (0.182)

The positive coefficients of the real GDP per capita and commodity index variables depict a positive long-run relationship with real house prices as also shown by the univariate results. In the panel framework, the coefficient on real interest rates is negative as expected, which shows a negative long-run relationship with real house prices. These panel estimates provide an empirical support for our expectations of the long-run dynamics of real house prices as they relate to the macroeconomy.

(Table 9 around here)

4.4 Vector Error Correction Model

4.4.1 Univariate VECM

Given the cointegrating vector β estimated by DOLS, we construct a vector error correction model (VECM) in order to analyze the short-run and long-run dynamics between housing prices, real GDP per capita, real lending interest rates and commodity prices. Abstracting from deterministic components,

$$\Delta y_t = \alpha + \rho\gamma^T y_{t-1} + \sum_{j=1}^{k} \theta_j \Delta y_{t-j} + C\varepsilon_t \tag{10}$$

where α is a vector of constants, $\rho = \begin{bmatrix} \rho_1 & \rho_2 & \rho_3 & \rho_4 \end{bmatrix}^T$ is a 4×1 vector of speed of convergence coefficients, C is a matrix that defines the contemporaneous structural relationship among the four variables, and $\varepsilon_t = \begin{bmatrix} \varepsilon_{p,t} & \varepsilon_{g,t} & \varepsilon_{r,t} & \varepsilon_{c,t} \end{bmatrix}^T$ is a vector of mutually orthogonal structural shocks. That is, $E\varepsilon_t\varepsilon_t^T = I$, where I is 4×4 identity matrix, and $Eu_tu_t^T = EC\varepsilon_t\varepsilon_t^TC^T = CC^T = \sum$, where $\sum$ is the variance-covariance matrix. The matrix of contemporaneous structural relationship C is identified with the Cholesky decomposition of the variance-covariance matrix using the following ordering: $\begin{bmatrix} g_t & r_t & c_t & p_t \end{bmatrix}$ [7].

In order to study the effects of $\varepsilon_{g,t}$, $\varepsilon_{r,t}$, $\varepsilon_{c,t}$ on real house prices in the short-run and long-run, we implement an impulse-response function based on the VECM system described in equation (10). We, therefore, rewrite equation (10) into the following state-space representation:

$$z_t = Fz_{t-1} + \zeta_t, \tag{11}$$

$$z_t = \begin{bmatrix} y_t & y_{t-1} & ... & y_{t-k} \end{bmatrix},$$

$$F = \begin{bmatrix} \vartheta_1 & \vartheta_2 & \dots & \vartheta_{k+1} \\ & & & \vdots \\ & I_4 & & \\ & & & 0 \end{bmatrix},$$

$$\zeta_t = \begin{bmatrix} Ce_t & 0 & \dots & 0 \end{bmatrix}^T,$$

$$\vartheta_1 = I_4 + \rho\gamma^T + \theta_1,$$

$$\vartheta_j = \theta_{j+1} - \theta_j,\ j = 2, \dots, k\ ,$$

$$\vartheta_{k+1} = -\theta_j,$$

I_4 is a 4×4 identity matrix. Based on the above, we obtain the n^{th} period impulse-response function as follows:

$$(S^T F^n S)C, \tag{12}$$

$S = \begin{bmatrix} I_4 & 0 & \dots & 0 \end{bmatrix}^T$ is a selection matrix of dimensions $2(k \times 1) \times 4$. Note that γ is obtained using the DOLS β estimates (i.e. $\hat{\beta}_{DOLS}$). The response of real house prices to shocks in real GDP per capita, real interest rate and commodity prices are estimated with

[7] For the purpose of our research, it is important for p_t to be the last series in the vector. Even though this specific ordering is the one we believe reflects the transmission mechanisms in the actual economy more adequately , we obtained the impulse-response functions for all permutations of the ordering keeping p_t at the same position (i.e. 6 distinct orderings in total). The results are qualitatively similar. These impulse-response functions are available upon request.

the use of Expression (12). The resulting impulse-response functions are reported in Figures (11), (12) and (13) along with their respective confidence bands at the 95% confidence level. We estimate the impulse-response functions for 20 periods ahead and we report the responses of p_t to shocks in g_t, r_t and c_t:

- Responses of house prices to shocks in real GDP per capita (see Figure (11))

  All impulse-response functions show that a shock in real GDP per capita leads to a positive response in real house prices with the exception of Japan, which is the international outlier. This explains why, unlike every other country in our sample, the impulse-response function for Japan is the only one, which does not confirm our expectations on the long-run relationship between real house prices and real GDP per capita. Even though most of the response-functions have the lower bound of their confidence bands below 0, the distribution of responses is skewed towards positive responses. Italy and the UK, on the other hand, have significant long-run responses with their confidence bands entirely in the first quadrant.

- Responses of real house prices to shocks in real interest rate (see Table (12))

  9 out of 13 countries (i.e. 69.23%) show negative house price responses to shocks in real interest rates. Of the 4 countries with different response functions, one is Japan (i.e. the international outlier), 2 countries, New Zealand and Spain, show evenly distributed responses across the negative and the positive spectrum. Australia is the only country besides Japan showing positive long-run responses.

- Responses of real house prices to shocks in the commodity index (see Table (13))

  10 out of 13 countries (i.e. 76.92%) exhibit positive responses in real house prices due to shocks in commodity index. Once again, Japan's response function is different from the other countries. The other 2 countries are Switzerland and the UK, who show negative long-run responses of house prices to shocks in the commodity index.

Overall, the impulse-response functions obtained from the VECM are strongly substantiative of our expectations despite the relative small size of our samples.

(Table 11 around here)

(Table 12 around here)

(Table 13 around here)

4.4.2 Panel VECM

We extend the Vector Error Correction Model discussed and used in Section (4.4.1) to a panel framework. We achieve this by implementing the econometric procedure on pooled and demeaned data [8]. Consequently, the first step of the panel VECM procedure, which is to estimate the cointegrating vector, is carried out without an intercept:

$$\Delta y_t = \rho\gamma^T y_{t-1} + \sum_{j=1}^{k} \theta_j \Delta y_{t-j} + C\varepsilon_t \tag{13}$$

The impulse-response functions obtained are shown in Figures (14), (15) and (17).

- Figure (14) shows that the distribution of the responses of real house prices to a shock in the real GDP per capita is entirely in the first quadrant. We can, therefore, infer a positive long-run relationship between the variables.
- Despite a short-lived positive initial response, real house prices maintain a negative long-run response to shocks in the real interest rate (see Figure 15). The distribution of responses is predominantly negative even though the upper bound of the confidence band being in the positive side.
- Real house prices respond positively to shocks in the commodity index. A small portion of the confidence band lies in the negative side; however, overall responses are positively skewed.

The response functions estimated from the panel VECM corroborate the predominant results obtained by the univariate counterpart and provide empirical evidence for our expectations of house price dynamics.

5 Conclusion

This article is a cross-country study of the cointegrating relationship between real house prices and each of the real GDP per capita, real interest rate and commodity prices time series. We begin by formulating expectations of the direction of the long-run relationships by using a simple eye-balling technique. Eye-balling the series leads to the conclusion that real house prices have a positive long-run relationship with real GDP per capita and the commodity index, and a negative long-run relationship with real interest rate. We then

[8] After pooling the data from all the countries in our sample, we substract each country's mean from the country's values at each time period.

implement an array of econometric procedures on international data in order to find empirical evidence for our theoretical postulations. We use are both univariate and panel procedures.

Since cointegration is only possible among nonstationary series, we implement the DF-GLS test on each cross-section and find evidence for the nonstationarity of our series. Pesaran (2007)'s panel unit root test is also implemented and the results are also in favor of the nonstationarity of the panels.

Having established the nonstationarity of our variables, we then estimate the cointegrating vectors with the dynamic OLS (Saikkonen,1991; Stock and Watson,1993) and find very significant coefficient estimates for each country. The signs of the DOLS coefficients substantiate our theoretical expectations with the exception of the real interest rate coefficients. In order to avoid statistical issues related to small sample size, we also implement the panel DOLS and obtain coefficients in line with our expectations.

Finally, we implement a Vector Error Correction Model and estimate the response functions of real house prices to shocks in real GDP per capita, real interest rate and commodity index. The response functions obtained predominantly support our expectations. Japan, which is known as an outlier in the housing market, always exhibits responses which are contrary to what theory predicts. In order the leverage the cross-section and time series informations of our panel, we also implement a panel VECM. The response functions derived from the panel model are also strongly substantiative of our theoretical assertions.

References

Adams, Z. & Fuss, R. (2010), 'Macroeconomic determinants of international housing markets', *Journal of Housing Economics* **19**(1), 38–50.

Addison-Smyth, D., McQuinn, K. & O' Reilly, G. (2008), 'Estimating the Structural Demand for Irish Housing', (1/RT/08).

Andrews, D. W. & Monahan, J. C. (1992), 'An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator', *Econometrica: Journal of the Econometric Society* pp. 953–966.

Baffoe-Bonnie, J. (1998), 'The dynamic impact of macroeconomic aggregates on housing prices and stock of houses: A national and regional analysis', *The Journal of Real Estate Finance and Economics* **17**(2), 179–197.

Baker, D. (2008), 'The housing bubble and the financial crisis', *Real-world economics review* **46**(20), 73–81.

Bardhan, Ashok, R. H. E. & Kroll, C. A. (2013), 'Global housing markets: Crises, policies, and institutions', *Journal of Social Policy* **42**(03), 646–647.

Brocker, M. & Hanes, C. (2014), 'The 1920s american real estate boom and the downturn of the great depression: Evidence from city cross-sections', pp. 161–201.

Demyanyk, Y. & Van Hemert, O. (2011), 'Understanding the subprime mortgage crisis', *Review of Financial Studies* **24**(6), 1848–1880.

Duca, J. V., Muellbauer, J. & Murphy, A. (2010), 'Housing markets and the financial crisis of 2007-2009: Lessons for the future', (0049).

Egert, B. & Mihaljek, D. (2007), 'Determinants of house prices in central and eastern europe', (2152).

Elliott, G., Rothenberg, T. J. & Stock, J. H. (1996), 'Efficient tests for an autoregressive unit root', *Econometrica* **64**(4), 813–836.

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

Gonzalo, J. (1994), 'Five alternative methods of estimating long-run equilibrium relationships', *Journal of Econometrics* **60**(1-2), 203–233.

Inder, B. (1993), 'Estimating long-run relationships in economics: a comparison of different approaches', *Journal of econometrics* **57**(1), 53–68.

James H. Stock, M. W. W. (1988), 'Variable trends in economic time series', *The Journal of Economic Perspectives* **2**(3), 147–174.

Johansen, S. (1988), 'Statistical analysis of cointegration vectors', *Journal of Economic Dynamics and Control* **12**(2), 231 – 254.

Katrakilidis, C. & Trachanas, E. (2012), 'What drives housing price dynamics in greece: New evidence from asymmetric ardl cointegration', *Economic Modelling* **29**(4), 1064–1069.

Kim, S.-W. & Bhattacharya, R. (2009), 'Regional housing prices in the usa: an empirical investigation of nonlinearity', *The Journal of Real Estate Finance and Economics* **38**(4), 443–460.

Levin, E. J. & Wright, R. E. (1997), 'The impact of speculation on house prices in the united kingdom', *Economic Modelling* **14**(4), 567 – 585.

Mack, A. & MartÃnez-GarcÃa, E. (2011), 'A cross-country quarterly database of real house prices: a methodological note', *Globalization and Monetary Policy Institute Working Paper* **99**.

Mark, N. C. & Sul, D. (2003), 'Cointegration vector estimation by panel dols and long-run money demand', *Oxford Bulletin of Economics and Statistics* **65**(5), 655–680.

McCord, M., McGreal, S., Berry, J., Haran, M. & Davis, P. (2011), 'The implications of mortgage finance on housing market affordability', *International Journal of Housing Markets and Analysis* **4**(4), 394–417.

Newey, W. K. & West, K. D. (1994), 'Automatic lag selection in covariance matrix estimation', *The Review of Economic Studies* **61**(4), 631–653.

Ogaki, M. & Park, J. (1998), 'A cointegration approach to estimating preference parameters', *Journal of Econometrics* **82**(1), 107–134.

Pesaran, H. M. (2004), 'General diagnostic tests for cross section dependence in panels', (0435).

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.

Phillips, P. C. (1991), 'Optimal inference in cointegrated systems', *Econometrica: Journal of the Econometric Society* pp. 283–306.

Phillips, P. C. & Hansen, B. E. (1990), 'Statistical inference in instrumental variables regression with i (1) processes', *The Review of Economic Studies* **57**(1), 99–125.

Robert F. Engle, C. W. J. G. (1987), 'Co-integration and error correction: Representation, estimation, and testing', *Econometrica* **55**(2), 251–276.

Saikkonen, P. (1991), 'Asymptotically efficient estimation of cointegration regressions', *Econometric Theory* **7**(01), 1–21.

Sarno, L. & Taylor, M. P. (1998), 'Real exchange rates under the recent float: unequivocal evidence of mean reversion', *Economics Letters* **60**(2), 131–137.

Stock, J. H. (1987), 'Asymptotic properties of least squares estimators of cointegrating vectors', *Econometrica: Journal of the Econometric Society* pp. 1035–1056.

Stock, J. & Watson, M. (1993), 'A simple estimator of cointegrating vectors in higher-order integrated systems', *Econometrica* **61**(4), 783–820.

Sutton, G. D. (2002), ‘Explaining changes in house prices1’, *BIS quarterly review* p. 47.

Tsai, I.-C., Lee, C.-F. & Chiang, M.-C. (2012), ‘The asymmetric wealth effect in the us housing and stock markets: evidence from the threshold cointegration model’, *The Journal of Real Estate Finance and Economics* **45**(4), 1005–1020.

Tsatsaronis, K. & Zhu, H. (2004), ‘What drives housing price dynamics: cross-country evidence’, *BIS Quarterly Review* .

Vogelsang, T. J. & Wagner, M. (2014), ‘Integrated modified ols estimation and fixed-b inference for cointegrating regressions’, *Journal of Econometrics* **178**(2), 741–760.

Westerlund, J. (2007), ‘Testing for error correction in panel data’, *Oxford Bulletin of Economics and Statistics* **69**(6), 709–748.

Zhou, J. (2010), ‘Testing for cointegration between house prices and economic fundamentals’, *Real Estate Economics* **38**(4), 599–632.

Appendix

Variable	DFGLS	P-value
Real house price index	-1.271	0.22
Real GDP per capita	-1.155	0.26
Real interest rate	-0.044	0.97
Commodity index	-0.893	0.38

Table 1: DFGLS results for cross-section means

Note: We report the DFGLS test results with an intercept. *, **, and *** denote rejections of the null hypothesis of nonstationarity in the 10%, 5%, and 1% significance level, respectively.

Country	DFGLS	P-value
Australia	-0.341	0.74
Denmark	-1.200	0.25
Finland	-1.257	0.22
France	-1.541	0.14
Ireland	-1.945 *	0.07
Italy	-4.373 ***	0.00
Japan	0.119	0.91
New Zealand	-0.513	0.61
Norway	-0.442	0.66
Spain	-2.511 **	0.02
Switzerland	-0.777	0.45
UK	-0.887	0.39
USA	-1.868 *	0.08

Table 2: DFGLS results for real house price index

Note: We report the DFGLS test results with an intercept. *, **, and *** denote rejections of the null hypothesis of nonstationarity in the 10%, 5%, and 1% significance level, respectively.

Country	DFGLS	P-value
Australia	-0.379	0.71
Denmark	-1.566	0.13
Finland	-1.294	0.21
France	-1.187	0.25
Ireland	-1.688	0.11
Italy	-1.647	0.12
Japan	-0.580	0.57
New Zealand	-0.006	1.00
Norway	-1.279	0.22
Spain	-1.924 *	0.07
Switzerland	-0.646	0.53
UK	-0.751	0.46
USA	-0.734	0.47

Table 3: DFGLS results for real GDP per capita

Note: We report the DFGLS test results with an intercept. *, **, and *** denote rejections of the null hypothesis of nonstationarity in the 10%, 5%, and 1% significance level, respectively.

Country	DFGLS	P-value
Australia	-0.244	0.81
Denmark	0.086	0.93
Finland	0.214	0.83
France	0.159	0.88
Ireland	-1.568	0.13
Italy	-0.939	0.36
Japan	-0.463	0.65
New Zealand	0.411	0.69
Norway	0.268	0.79
Spain	-1.056	0.30
Switzerland	0.320	0.75
UK	0.088	0.93
USA	0.289	0.78

Table 4: DFGLS results for the real interest rate

Note: We report the DFGLS test results with an intercept. *, **, and *** denote rejections of the null hypothesis of nonstationarity in the 10%, 5%, and 1% significance level, respectively.

Variable	CD	Average $\hat{\rho}$	P-value
Real house price index	9.077 ***	0.219	0.00
Real interest rate	21.846 ***	0.527	0.00
Real GDP per capita	24.985 ***	0.603	0.00

Table 5: Pesaran cross-section dependence test

Note: We report the DFGLS test results with an intercept. *, **, and *** denote rejections of the null hypothesis of nonstationarity in the 10%, 5%, and 1% significance level, respectively.

Variable	Statistc
Real house price index	-2.333 ***
Real interest rate	-1.260
Real GDP per capita	-1.248

Table 6: Pesaran panel unit root test

Note: We report the DFGLS test results with an intercept. *, **, and *** denote rejections of the null hypothesis of nonstationarity in the 10%, 5%, and 1% significance level, respectively.

Country	Statistic (trace)	Statistics (eigen)
Australia	10.22 **	10.22 **
Denmark	32.75 *	32.42 **
Finland	12.08 **	12.08 **
France	8.41 *	8.41 *
Ireland	21.89 **	17.25 **
Italy	9.99 **	9.99 **
Japan	35.05 **	21.60 *
New Zealand	33.94 *	20.09 *
Norway	33.11 *	32.35 **
Spain	12.11 **	12.11 **
Switzerland	20.12 **	13.87 *
UK	9.96 **	9.96 **
USA	9.79 **	9.79 **

Table 7: Johansen cointegration test

Note: We report the Johansen test statistic with an intercept. *, **, and *** denote rejections of the null hypothesis in the 10%, 5%, and 1% significance level, respectively.

	Statistics	P-value	P-value with CSD
Group mean test	-1.460	0.999	0.662
Panel test	-6.551	0.638	0.068 *

Table 8: Panel cointegration test

Note: We implement Westerlund (2007) t-test type statistics with 1 lag. The number of leads is determined by the AIC. The p-value is not sized correctly when cross-section independence fails to hold. The p-value with CSD denotes p-values with cross-section dependence via 500 bootstraps.

Variable	PDOLS	Standard error
Real GDP Per Capita	1.950	0.343
Real Interest Rate	-0.001	0.026
Commodity index	0.009	0.182

Table 9: Panel Dynamic OLS results

Country	Variable	DOLS	Standard error	P-value
Australia	Real GDP per capita	3.250 ***	0.333	0.00
	Real interest rate	0.069 **	0.025	0.01
	Commodity index	-0.107 *	0.061	0.10
Denmark	Real GDP per capita	4.837 ***	0.099	0.00
	Real interest rate	0.084 ***	0.005	0.00
	Commodity index	0.321 ***	0.016	0.00
Finland	Real GDP per capita	1.485 ***	0.062	0.00
	Real interest rate	-0.020 ***	0.004	0.00
	Commodity index	0.186 ***	0.013	0.00
France	Real GDP per capita	5.115 ***	0.608	0.00
	Real interest rate	0.122 ***	0.032	0.00
	Commodity index	0.532 ***	0.045	0.00
Ireland	Real GDP per capita	1.113 ***	0.040	0.00
	Real interest rate	-0.115 ***	0.006	0.00
	Commodity index	0.236 ***	0.013	0.00
Italy	Real GDP per capita	4.513 ***	0.396	0.00
	Real interest rate	0.069 ***	0.008	0.00
	Commodity index	0.250 ***	0.033	0.00
Japan	Real GDP per capita	-0.884	0.616	0.17
	Real interest rate	0.072 ***	0.010	0.00
	Commodity index	-0.275 ***	0.065	0.00
New Zealand	Real GDP per capita	3.351 ***	0.508	0.00
	Real interest rate	0.178 **	0.060	0.01
	Commodity index	0.401 ***	0.085	0.00
Norway	Real GDP per capita	1.195 **	0.397	0.01
	Real interest rate	-0.076 **	0.032	0.03
	Commodity index	0.237 **	0.078	0.01
Spain	Real GDP per capita	4.686 ***	1.226	0.00
	Real interest rate	0.117 **	0.047	0.02
	Commodity index	0.313 **	0.100	0.01
Switzerland	Real GDP per capita	-1.356 ***	0.416	0.00
	Real interest rate	-0.098 ***	0.024	0.00
	Commodity index	0.299 ***	0.044	0.00
UK	Real GDP per capita	2.849 ***	0.086	0.00
	Real interest rate	-0.031 ***	0.006	0.00
	Commodity index	-0.073 ***	0.020	0.00
USA	Real GDP per capita	2.732 ***	0.161	0.00
	Real interest rate	0.123 ***	0.014	0.00
	Commodity index	0.092 **	0.037	0.02

Table 10: Dynamic OLS results

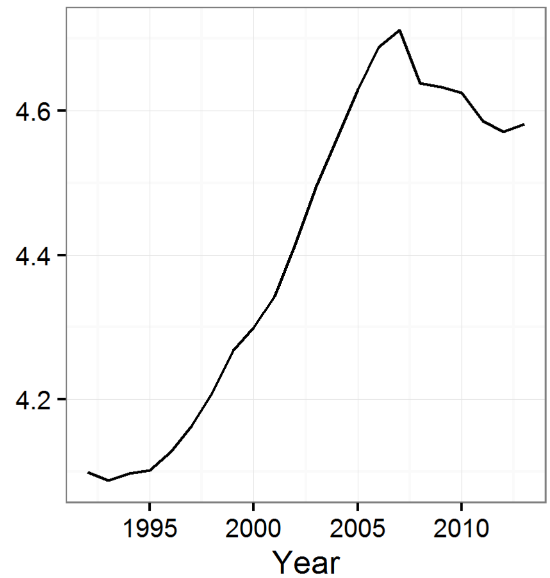


Figure 1: Log of real house price index (cross-section means)

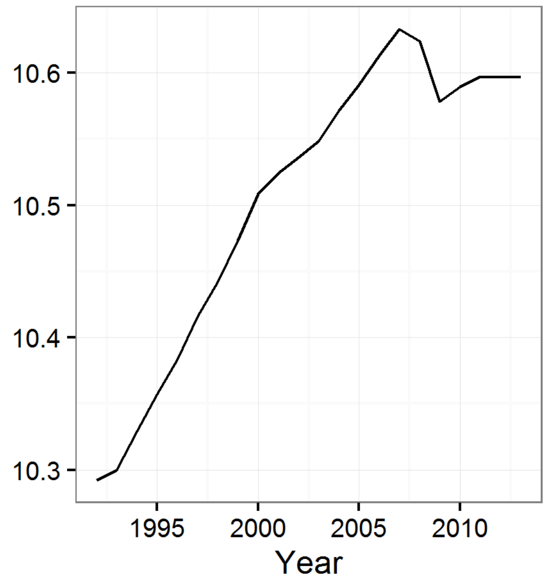


Figure 2: Log of real GDP per capita (cross-section means)

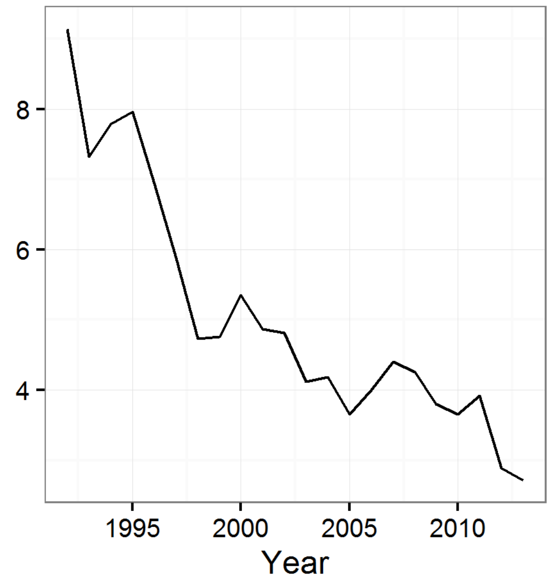


Figure 3: Real interest rate (cross-section means)

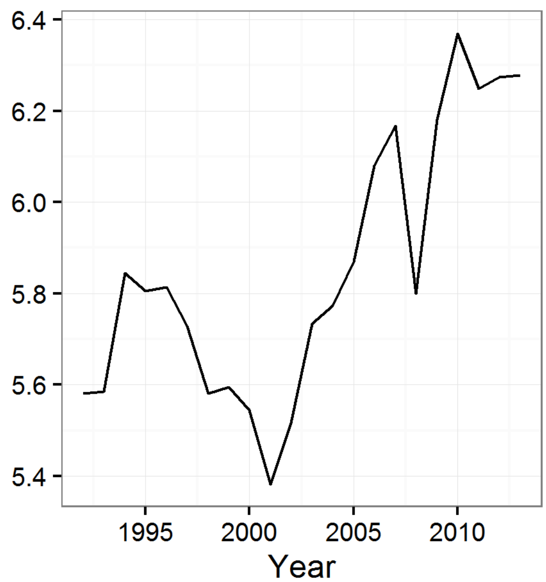


Figure 4: Log of commodity index

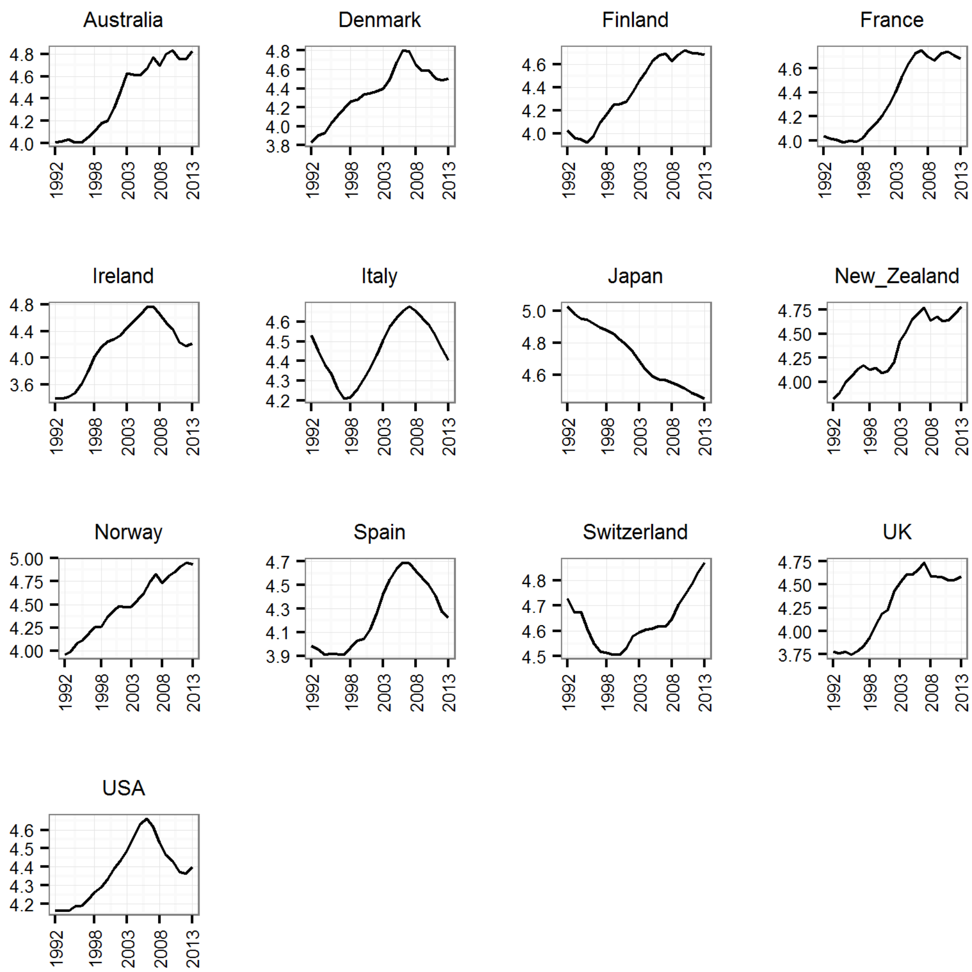


Figure 5: Log of real house price index

Figure 6: Real interest rate

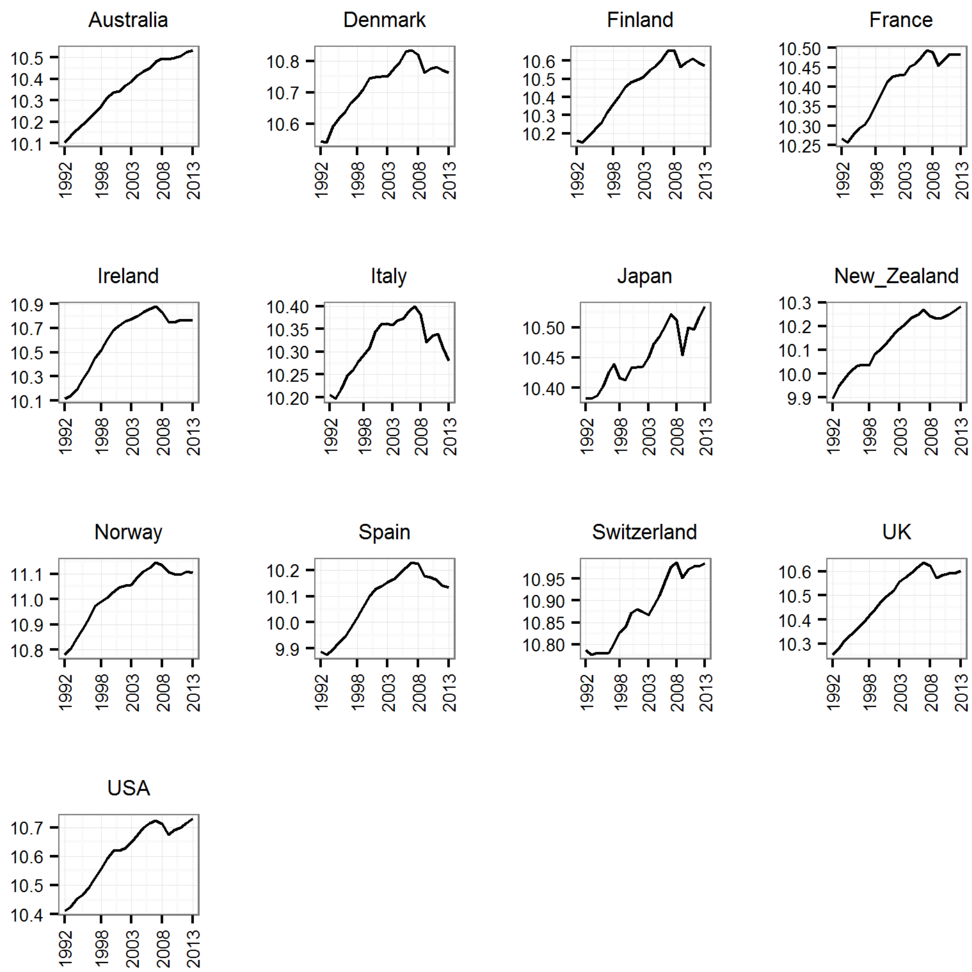


Figure 7: Log of real GDP per capita

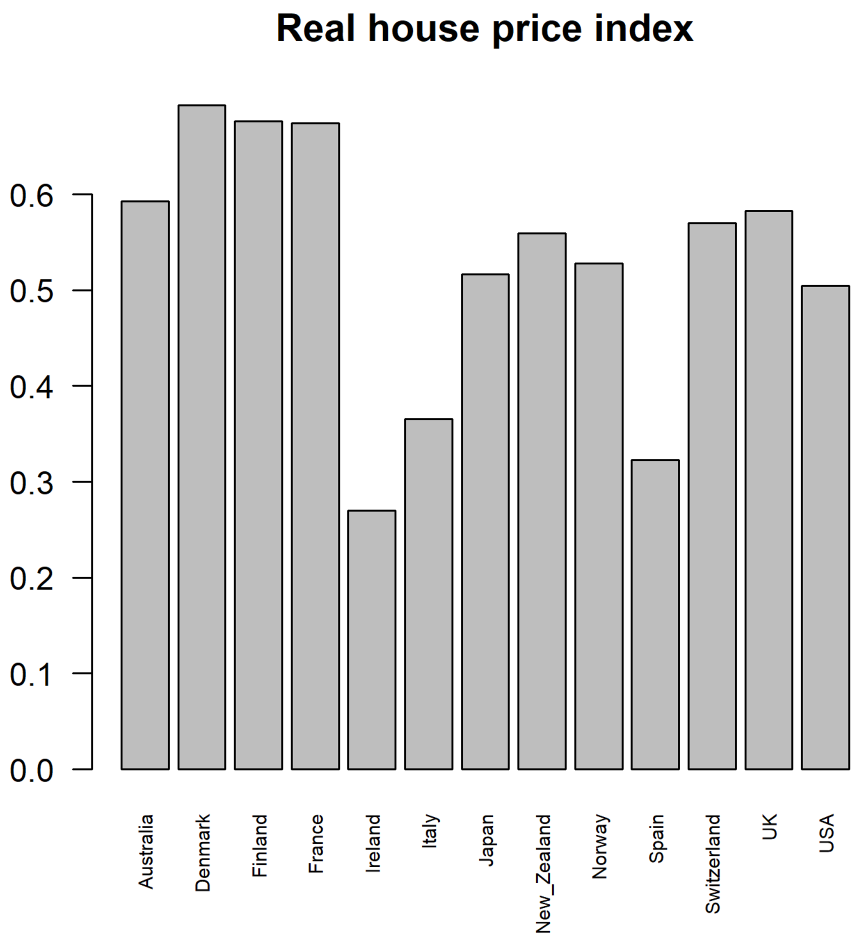


Figure 8: Mean correlation coefficients: real house price index

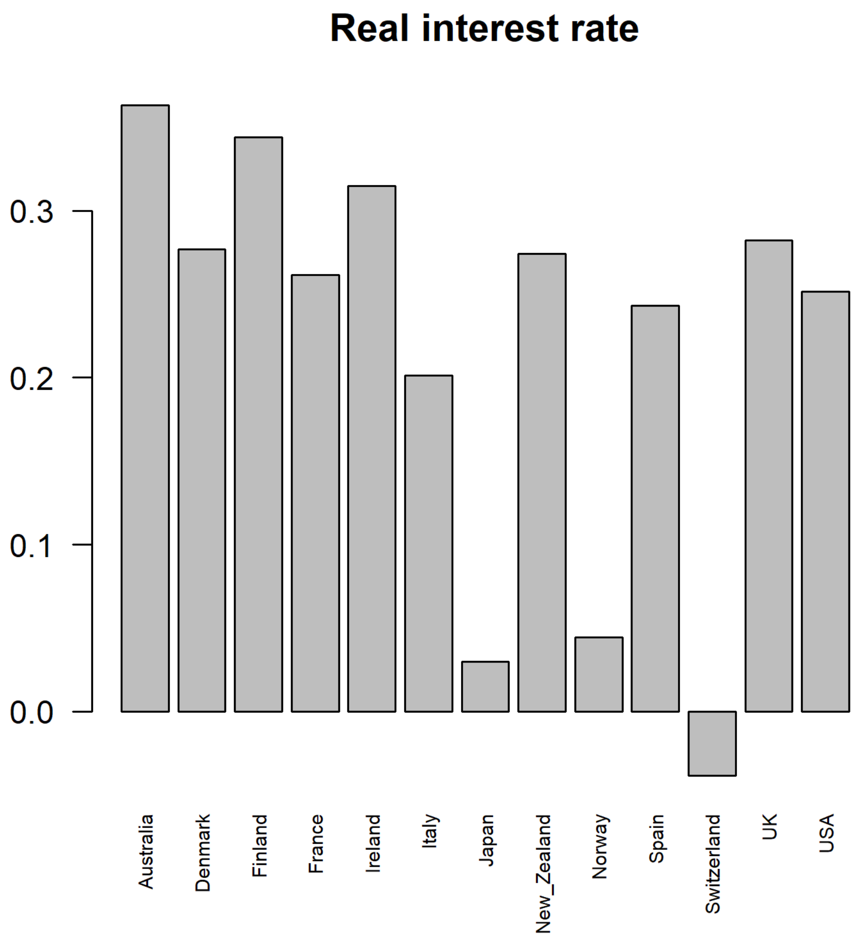


Figure 9: Mean correlation coefficients: real interest rate

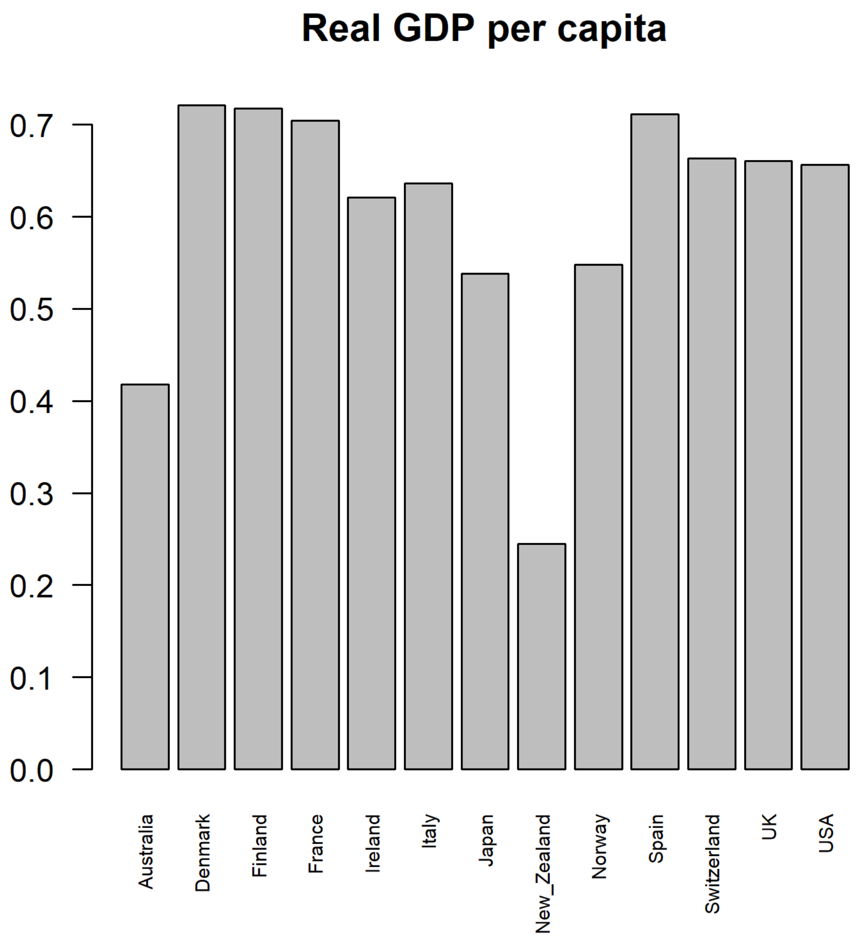


Figure 10: Mean correlation coefficients: real GDP per capita

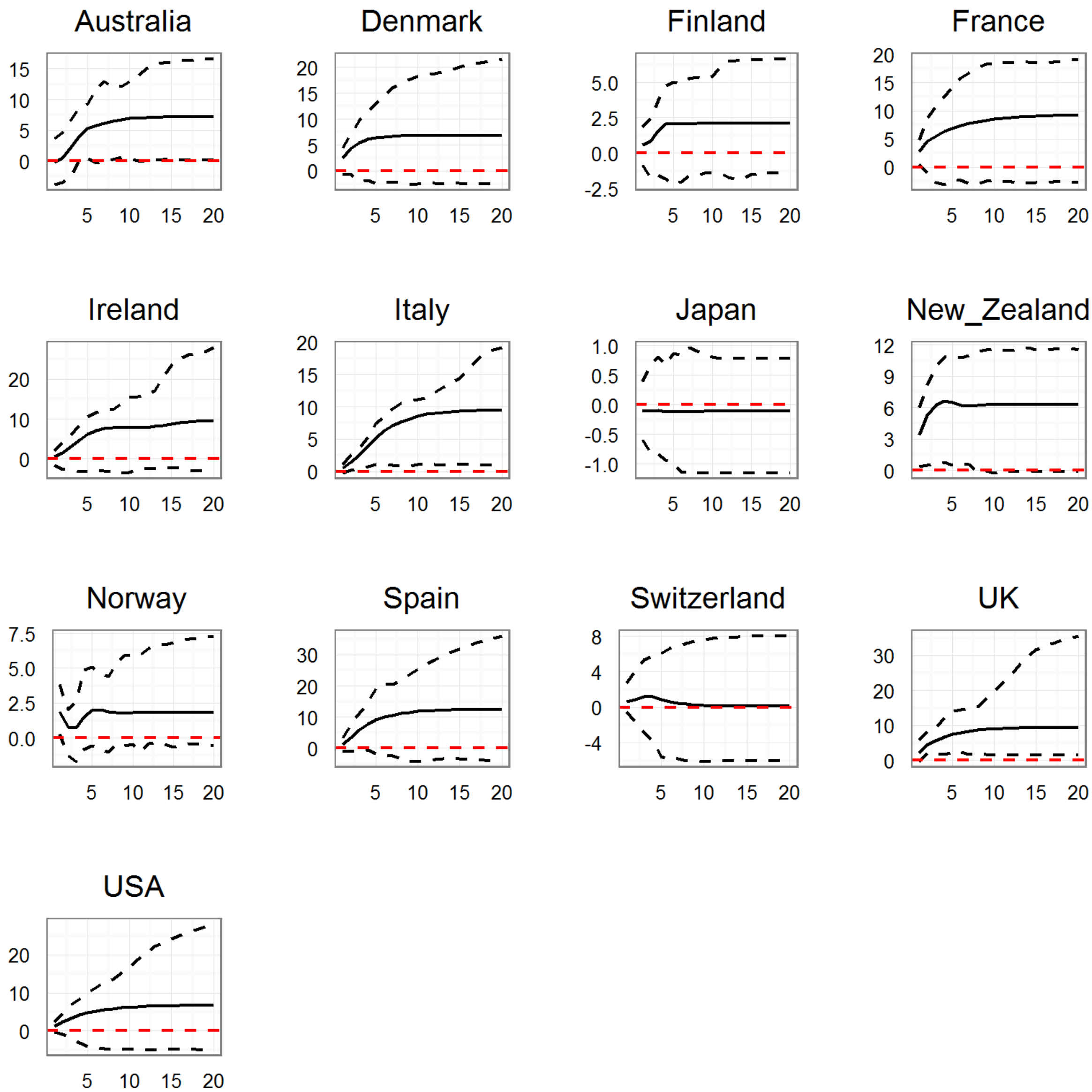


Figure 11: Impulse-response function: impulse (log of real GDP per capita)/response (log of real house price index)

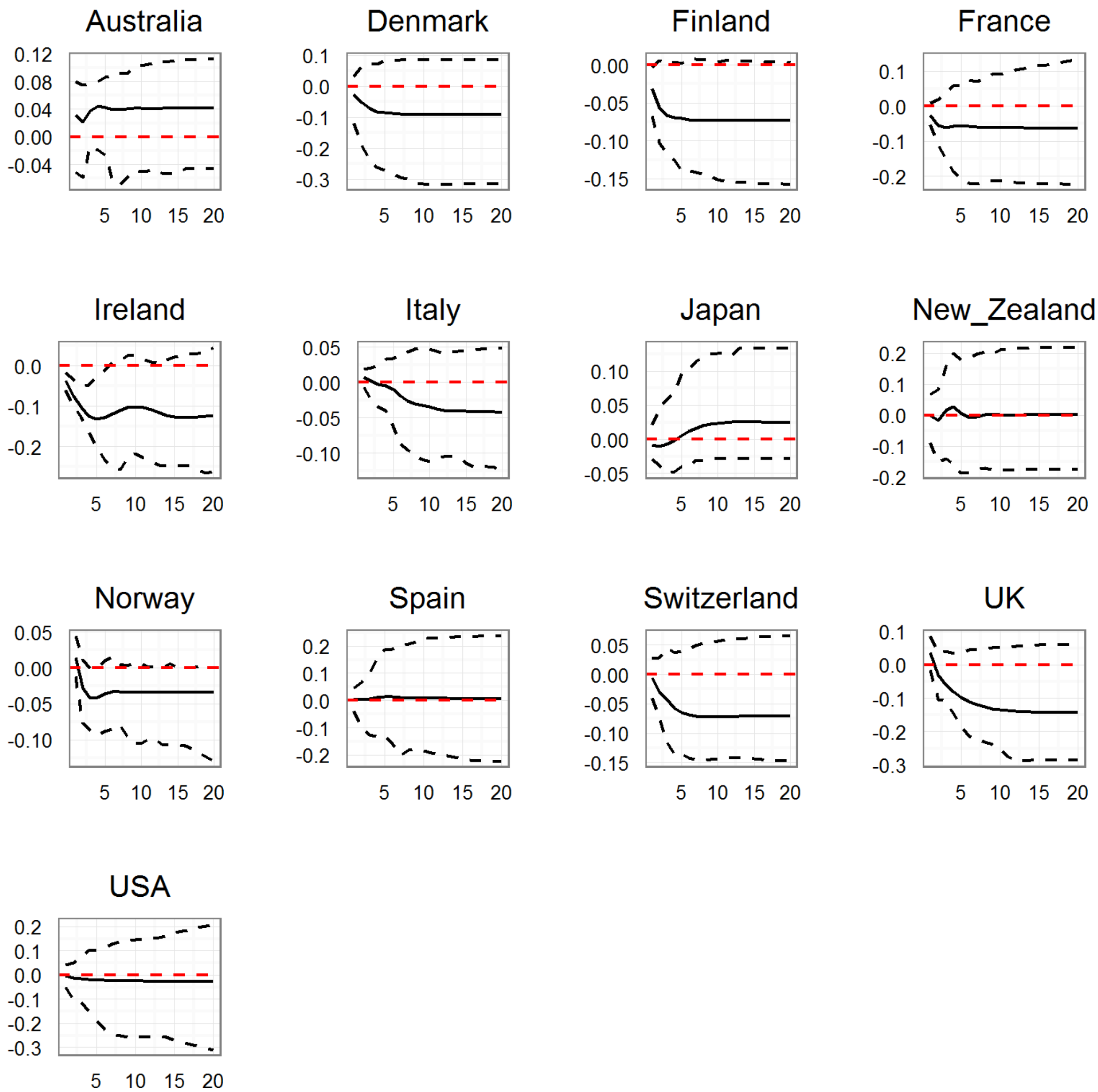


Figure 12: Impulse-response function: impulse (real interest rate)/response (log of real house price index)

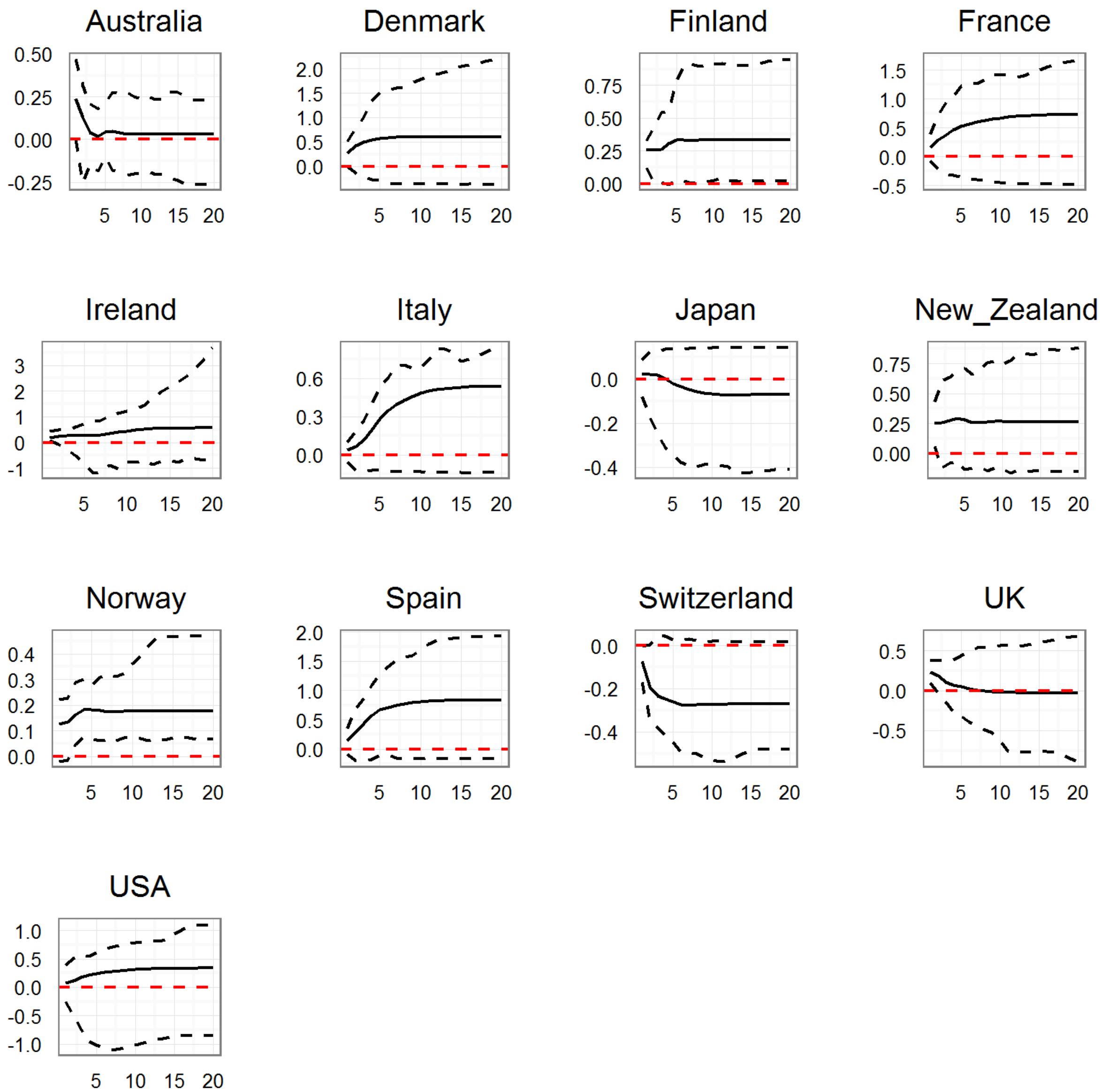


Figure 13: Impulse-response function: (log of commodity index)/response (log of real house price index)

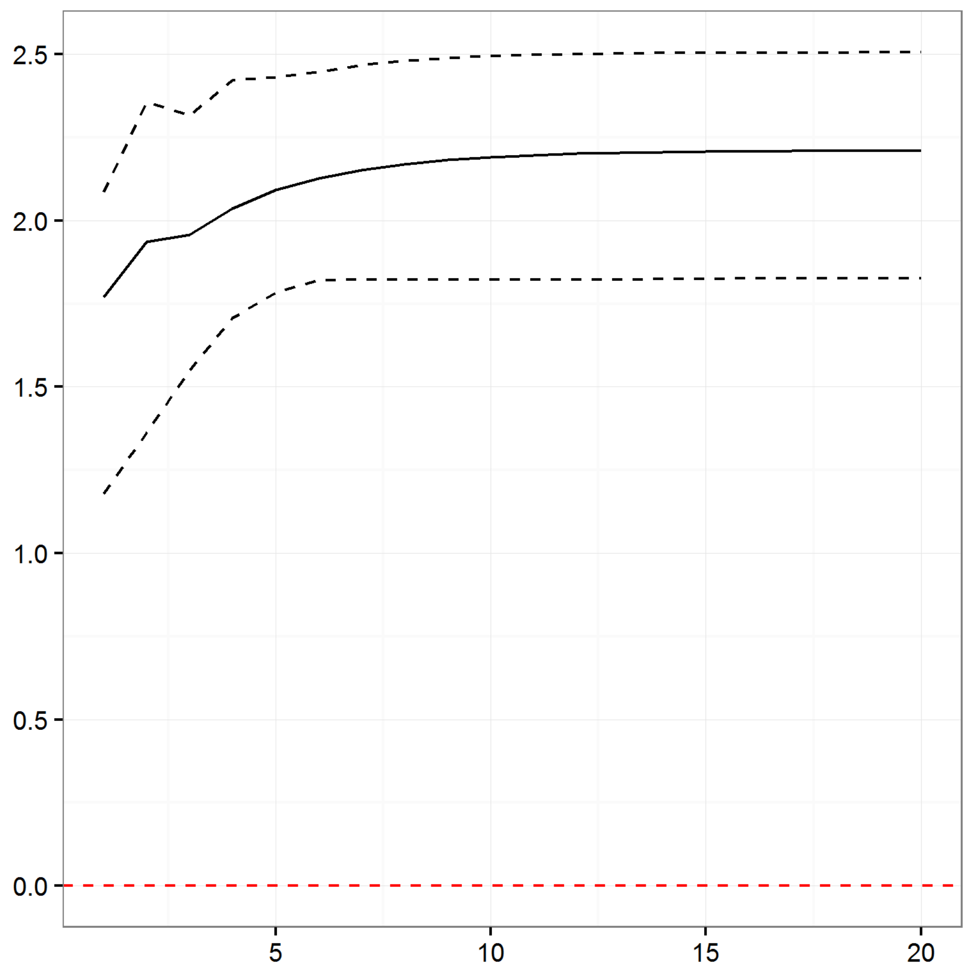


Figure 14: Impulse-response function: impulse (log of real GDP per capita)/response (log of real house price index)

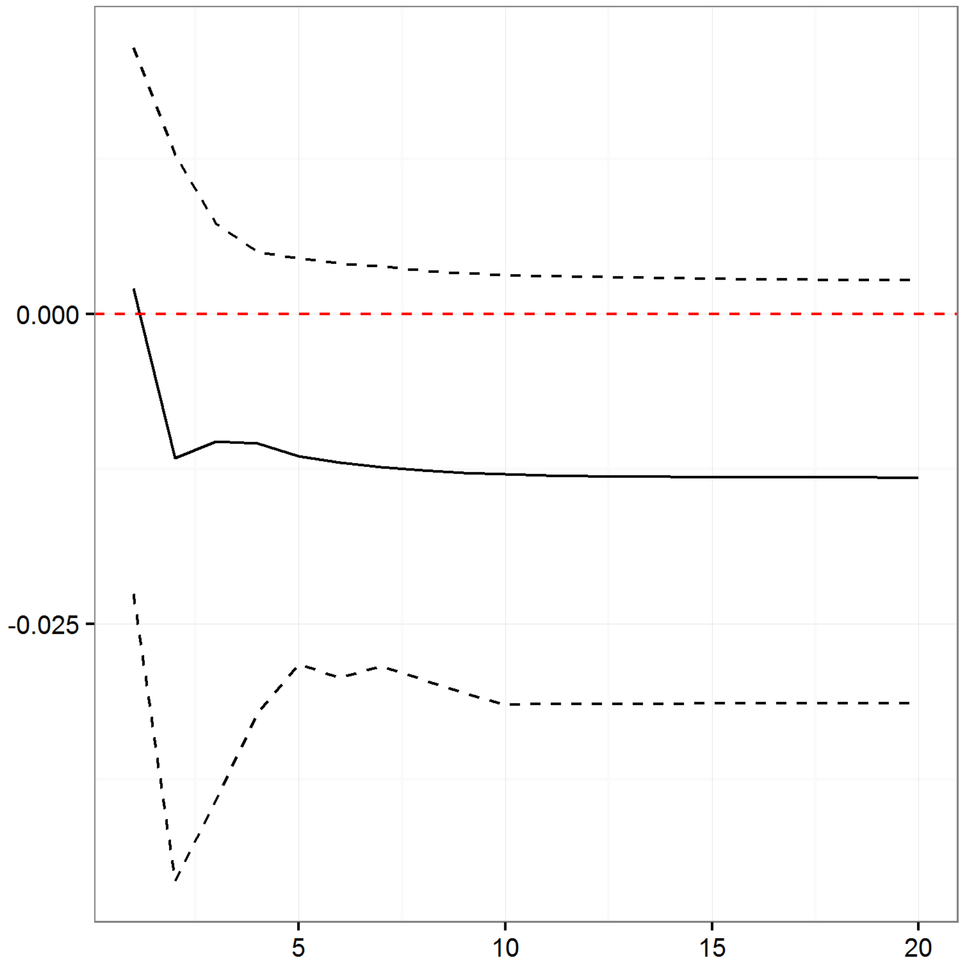


Figure 15: Impulse-response function: impulse (real interest rate)/response (log of real house price index)

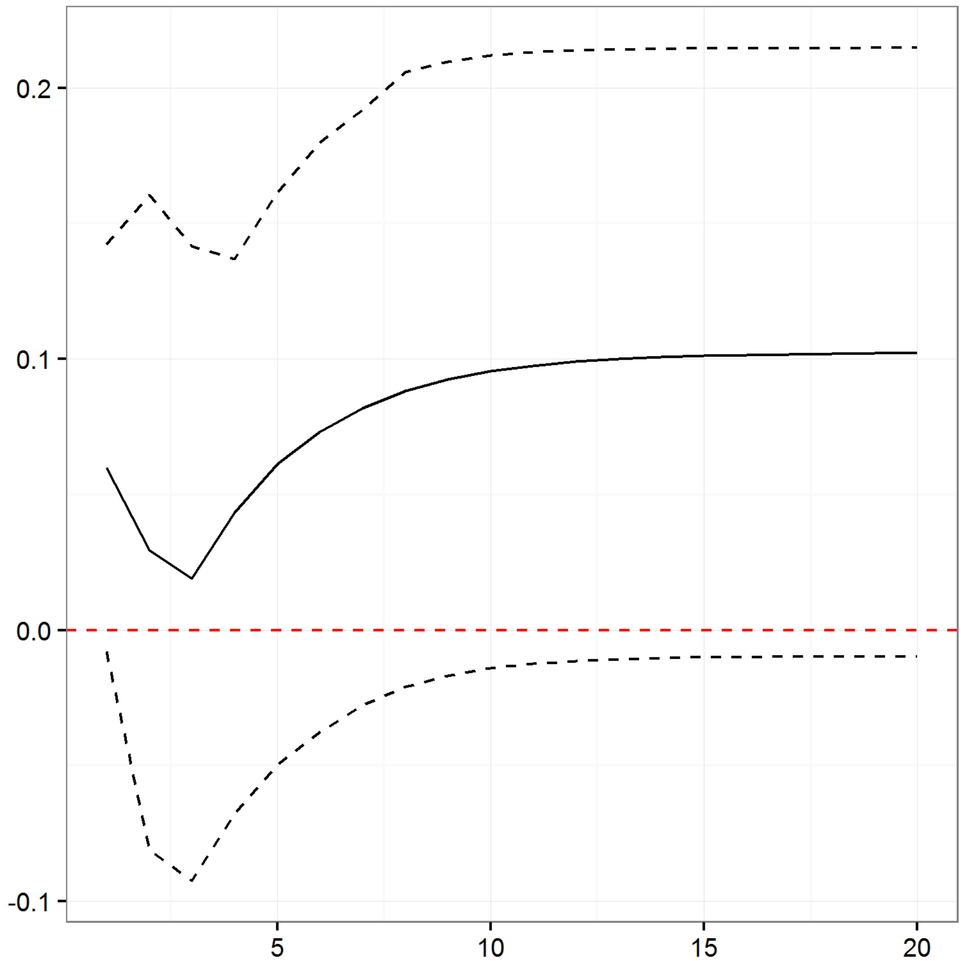


Figure 16: Impulse response function: impulse (commodity index)/response (real house price index)

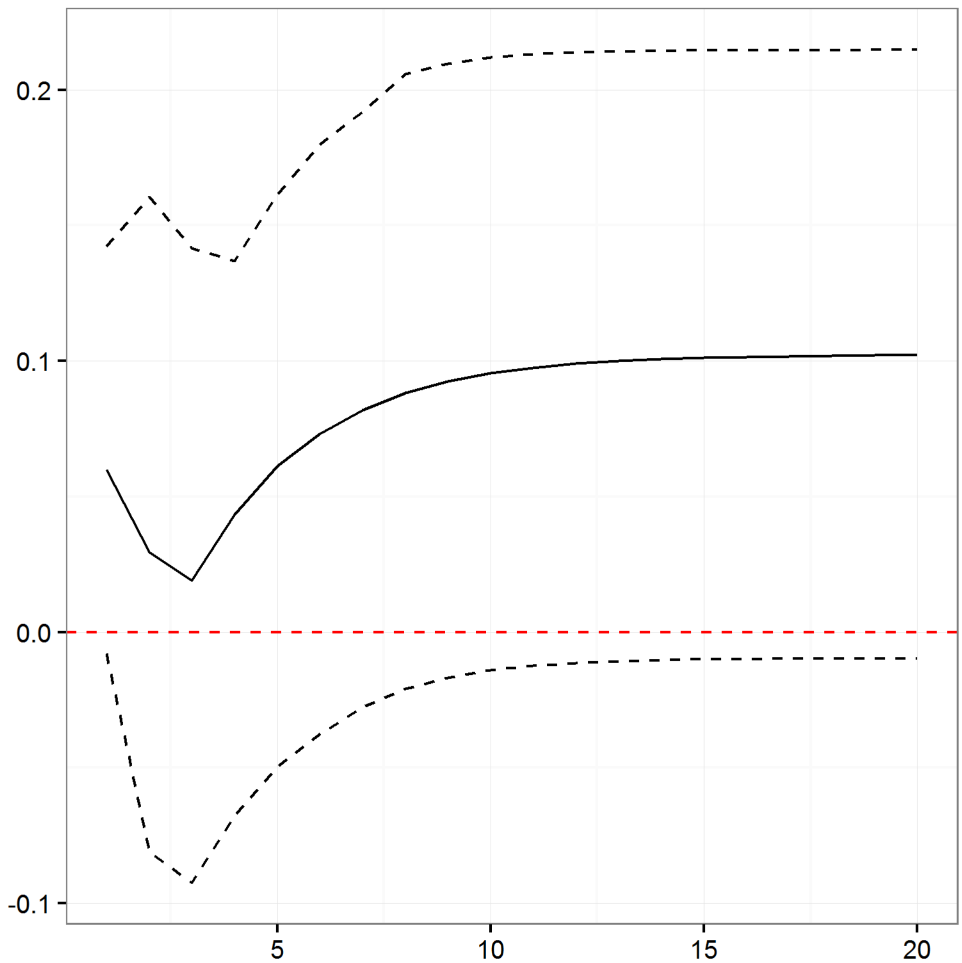


Figure 17: Impulse response function: impulse (commodity index)/response (real house price index)

Chapter 3: A Second Generation Cross-State Cointegration Analysis of Housing Prices and Key Fundamentals

Abstract

The cointegrating relationship between housing prices, real income per capita, interest rates and real building costs is studied with state-of-the-art econometric procedures in this paper. The panel spans over a time period of 37 years and across 48 contiguous states in the USA. Despite an abundant literature on housing prices dynamics with the macroeconomy, the existence of cross-section dependence (i.e. serial correlation) in the panels is seldom accounted for in the empirical analyses, which is known to be problematic in econometric tests and estimation techniques. In this paper, two of the so-called second generation econometric procedures are used, namely Westerlund (2007)'s error-correction based cointegration analysis and Pesaran (2007)'s panel unit root test. An array of univariate and panel models are used for unit root testing, cointegrating testing, cointegrating vector estimation and vector error correction modeling. Strong evidence is found for the existence of cointegration between real housing prices and the macroeconomy; however, the results obtained do not support the hypothesized negative long-run relationship with real interst rates.

1 Introduction

Major economic downturns have historically been preceded by the burst of bubbles in asset prices. The American economy corroborates this observation with 3 major recessions in the last century. The earliest one is the Great Depression, which occurred in 1929. It immediately followed a crash in the American stock market. In the early 1990s, the so-called dotcom bubble began and increased throughout the decade. It burst in the early 2000s and led the US economy into a recession. The most recent one is the Great Recession of 2008; it is widely agreed that its main cause was the burst of a bubble in the housing market, which developed in the course of the 2000s. It is, therefore, important to seek to understand the existing long-run relationships between asset prices and the underlying fundamentals. A deviation from the long-run relationship is an indicator of the creation of a bubble in which case appropriate measures can be taken to avoid the negative economic outcomes of a burst. Long-run relationships between time series can be investigated with the use of cointegration analysis.

The housing market in particular has a very significant impact on the economy. Brocker & Hanes (2014) claim that the impact of the burst of a housing bubble has more impact on the economy than a stock market crash. They also postulate that banks' balance sheets are more vulnerable to the housing market than they are to the equity and derivatives markets. Therefore, the study of the co-movements between housing prices and significant macroeconomic variables is important.

Housing prices in the USA have known several major variations in the course of the recent decades. The market experienced significant gains in the 1970s and then began to fall sharply at the end of the decade up till the early 1980s. Housing prices began to rise again in the first half of the 1980s, thereby reversing the losses by about half; however, the market experienced a downturn again shortly after. Many experts attributed the fall in prices to the overvaluation of housing. The subsequent boom in the American real estate, which began in the first half of the 1990s, was one of the most significant booms experienced in the USA in the century. It lasted for a decade and it outpaced all the recent previous gains in the market by more than double. In light of the oscillatory pattern in housing prices in the course of the previous decades, these sharp gains raised the concern that a bubble was growing in the market (see Gallin (2006)). These concerns were justified by the Great Recession that the American economy experienced in the late 2000s when the housing bubble burst (see Duca et al. (2010), Baker (2008), Demyanyk & Van Hemert (2011)).

The next section of the paper is a review of the research work that has been done on the dynamics between the housing market and other macroeconomic variables. In Section (3), a thorough description of the data used in the empirical analysis is provided. Section (4) is devoted to the empirical analysis of the data in order to study the underlying dynamics between real house prices and the macroeconomy in the USA. A summary of the empirical process and the main results is provided in the conclusion section (Section (5)).

2 Literature review

The literature points at a number of variables as the fundamentals of housing prices – income and consumer prices being the most widely acknowledged ones. Other variables such as the mortgage rate, money supply, housing credit, stock market performance, construction costs, population, employment and other demographic factors are also recognized as important drivers.

Baffoe-Bonnie (1998) used the vector autoregressive (VAR) model to confirm the dynamic effects of employment growth, mortgage rate, anticipated and unanticipated changes in money supply and inflation on housing prices and the stock of houses sold on the national and regional levels. Using the same model, Sutton (2002) found evidence that variations in national income, interest rate and stock prices explained the fluctuations in housing prices in six advanced economies namely the US, the UK, Canada, Ireland, the Netherlands and Australia. Tsatsaronis & Zhu (2004) used the structural vector autoregressive (SVAR) model on 17 industrialized countries and determined that inflation and nominal interest rates had a strong and long-lasting impact on housing prices. They also found out that key variables in the financial sector, especially bank credit, the short-term interest rate and the term-spread, have a long-run impact on housing prices. Egert & Mihaljek (2007) use data on 8 Central and Eastern European Countries as well as 19 OECD countries in a dynamic ordinary least squares (DOLS) model to confirm that GDP per capita, real interest rate, housing credit and demographic factors drive variations in house prices. They also augment the model with transition-specific factors, especially institutional developments of housing markets. Addison-Smyth et al. (2008) analyze the Irish housing market and find that house prices are significantly influenced by interest rates and how much individuals can borrow, which is a function of their disposable income. They use a DOLS model to show that there exists a long-run relationship between house prices and the amounts borrowed by individuals. Adams & Fuss (2010) apply a panel cointegration analysis of 15 countries over 30 years and confirm the existence of a long-term relationship as well as short-term dynamics between house prices and construction costs, long-term interest rates and economic activity, which they define as a set of macroeconomic variables including real money supply, real GDP, real consumption, real industrial production and employment.

More recently, some researchers found that the linear framework, assumed by the majority of the literature on house prices and fundamentals, is not appropriate due to the restrictive nature of the assumption. Kim & Bhattacharya (2009) analyze house prices in the US and the four regions [1] over 36 years using a smooth transition autoregressive (STAR) model. They find that house prices in the US in 3 out of the 4 regions exhibit nonlinear properties. Zhou (2010) apply the ACE algorithm, a nonparametric method, and confirm the existence of nonlinear cointegration between house prices and fundamentals in many US cities. Tsai et al. (2012) studied the relationship between the housing market and the stock markets in the US. Using a momentum-threshold autoregressive (M-TAR) model, they show evidence of the existence of asymmetric wealth effects in the markets. Katrakilidis & Trachanas (2012) apply the asymmetric autoregressive distributive lag (ARDL) cointegration methodology to monthly data on the Greek housing market spanning over a period of 13 years and their results indicate the presence of asymmetric long-run and short-run effects from the consumer price index (CPI) and the industrial production index (IPI) towards house prices.

3 Data description

The models estimated in this paper involve 4 major macroeconomic variables: the real house price index, the real GDP per capita, the long interest rate and the real building cost index. The data collected on the variables are log-transformed except for the long interest rate. The panels have 48 cross-sections, which represent 48 contiguous US states, and spans over the period between 1975 and 2011 (37 years). The real house price index and the real GDP per capita are state-level data; however, the long interest rate and the real building cost index data are national data.

The house price index data is the Freddie Mac House Price Index (FMHPI). The values collected are nominal values and they are transformed into real values by inflation adjustment. The values are adjusted for inflation by dividing them by the USA consumer price index (CPI). The state-level real GDP data are from the US Bureau of Economic Analysis and the state population values are from the US Census Bureau. The real GDP is divided by the population for each state in order to obtain the real GDP per capita data used in the paper. The long interest rate and the real building cost index are obtained from Shiller (2015).

The exhibited oscillatory pattern in the national real house price index is an indication of housing bubbles growing and bursting (see Figure 1). The same pattern is also generally

[1] The USA is subdivided into four main regions according to the Census Bureau: the West, Midwest, South and Northeast.

experienced at the state level even though the price swings are more accentuated in some states than in others (see Figure 6). The US real GDP per capita has an overal increasing trend in the sample period (see Figure 2). However, there are 2 notable declines in real income in the late 1970s and in the late 2000s. These periods correspond to the 2 sharpest falls in housing prices in our sample period. The trend in real income per capita in individual states is overall similar to the national trend (see Figure 7). The long interest rate has an overall decreasing trend with high volatility throughout the sample period (see Figure 4). Building costs decreased sharply until the early 1990s when they began to stagnate for about a decade. In the first half of the 2000s, they began to rise again until the end of the sample period (see Figure 3). Section (5) concludes the paper by reviewing the main results.

(Figure 1 **around here)**

(Figure 2 **around here)**

(Figure 3 **around here)**

(Figure 4 **around here)**

(Figure 6 **around here)**

(Figure 7 **around here)**

4 Empirical Analysis

The primary tool for the study of the co-movements between real house prices and the macroeconomy is cointegration analysis. Two or more series are cointegrated if they are individually nonstationary and if they share a stable long-run relationship. Therefore, the first step of the empirical analysis is the test for unit roots. Cointegration tests are then carried out. Furthermore, in order to understand the existing dynamics between housing prices and the macroeconomy with more depth, the cointegrating vectors are estimated in the third step. Finally, a vector error correction model (VECM) is estimated in order to analyze the responses of the housing market to macreconomic shocks. Each step in the empirical analysis is carried out at the univariate level (i.e. in each individual state) and in a panel framework. Panel models are known to be generally statistically more powerful than univariate models because they leverage both cross-section and time series information.

4.1 Unit root tests

4.1.1 Augmented Dickey Fuller (ADF) test

The null of nonstationarity is tested with the Augmented Dickey-Fuller (ADF) test. The procedure is based on the following regression equation:

$$\Delta y_t = \alpha + \rho y_{t-1} + \sum_{j=1}^{p} \beta_j \Delta y_{t-j} + \varepsilon_t, \tag{1}$$

where y_t is the series being tested for stationarity and ε_t is a stochastic error. The equation is estimated with the least squares estimation technique and the ADF test statistic is then computed as:

$$ADF = \frac{\hat{\rho}}{se(\hat{\rho})},$$

where $\hat{\rho}$is the OLS estimate of ρ in Equation (1) and $se(\hat{\rho})$ is the standard error of the estimate. As previously stated, the real house price index and the real GDP per capita as state-level variables. The real building cost index and the long interest rate are national data. The test is, therefore, implemented on the data for each state as well as on cross-section means, which are used as a proxy for national dat. The long interest rate and building cost data are values pertaining to the USA as a whole. Consequently, their values are the same for each cross-section (i.e. state) in the panels. All unit root test results are reported in Table (1), Table (2) and Table (3).

At the national level, the test fails to reject the null of nonstationarity for the real income per capita, the long interest rate and the real building cost index variables. On the other hand, there is evidence for the rejection of the null for the real house price index series (see Table (1)). It can be argued; however, that the test result for the real house price index is not very strong. First, the sample size is only limited to 37 observations per state. Consequently, the cross-section means have a relatively small sample size, which is known to affect a negative effect on the power of the ADF. Second, by eyeballing the series in Figure (1), we can see that the values do not oscillate around a mean; long swings can be observed instead, which does not reflect the behavior of a series with a stationary process. Finally, similar data on real house prices at the national level are provided by Shiller (2015). The data spans over a longer period (1930-2011) and a strong positive trend can be observed in the series (see Figure(5)). These 3 arguments are corroborated by the unit root tests for each cross-section. At the state level, no evidence is found to reject the null hypothesis in:

- 30 states out of 48 (62.5%) for the real house price index series

- 47 states out of 48 (97.9%) for the real income per capita series

The results of the univariate unit root tests at the national level and at the state level are strongly pointing at a nonstationary stochastic process in the variables.

(Table 1 around here)

(Table 2 around here)

(Table 3 around here)

4.1.2 Pesaran (2007) panel unit root test

The relatively small number of time series observations for each state in our dataset is a potential issue with for the ADF test because the test is generally known to have low statistical power in small samples. Thus, the panel unit root test proposed by Pesaran (2007) is a more appropriate test. Its adequacy in such cases comes from its use of all observations in a panel rather than the observations of a single cross-section. In our specific case, the panel test makes use of 1778 observations (i.e. 48 states over 37 years) instead of 37 observations used by the ADF test. Sarno & Taylor (1998) discuss the statistical benefits of panel tests.

The Pesaran (2007) panel unit root test is a so-called second generation test because its underlying assumption is the existence of cross-section dependence in the panels tested (i.e. the cross sections are serially correlated). Therefore, the cross section structure of the panels is analyzed prior to implementing the test. Pesaran (2004) proposes a cross-section dependence test which is computed by:

$$CD = \sqrt{\frac{2T}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=1}^{N}\hat{\rho}_{i,j}\right) \xrightarrow{d} N(0,1),$$

where N is the number of cross-sections in the panel, T is the number of time periods and $\hat{\rho_{i,j}}$ is the pair-wise correlation coefficients from the residuals of the ADF regressions for each state. The test finds strong evidence for the presence of cross-section dependence in the panels as the null of nonstationarity is rejected at the 1% significance level (see 4). The dependence of each state to all other states can be visualized in Figure (8) and Figure (9). The average correlation coefficient for the real house price index is about 0.4 and it is about 0.72 for the real income per capita. This means that, on average, 40% of housing prices in US states are dependent on housing prices in other states. In the same way, about 72% of the real income per capita in states is correlated with the real income in other states.

The above test results provide evidence that the requirement of cross-section dependence by Pesaran (2007)'s panel unit root test is met. The test is essentially an extension of the ADF test to a panel framework. The ADF regression is augmented with components to control for the common factor in the panel:

$$\Delta y_{i,t} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i \bar{y}_{t-1} + \sum_{j=0}^{p} \theta_{ij} \Delta \bar{y}_{t-j} + \sum_{j=1}^{p} \delta_{ij} \Delta y_{i,t-j} + \varepsilon_{i,t}, \tag{2}$$

where $y_{i,t}$ is the observation from variable y in state i at time t and $\bar{y}_t$ is the cross-section mean of series y at time t, which proxies for the common factor of the panel[2]. Equation (2) is estimated with the least squares estimator for each state and the panel test statistic is computed as the mean of the t-statistics of coefficient β:

$$t = N^{-1} \sum_{i=1}^{N} t_i,$$

$$t_i = \frac{\beta_i}{se(\beta_i)},$$

where t_i is the t-statistic for the coefficient β_i from Equation (2) for state i and $se(\beta_i)$ is the standard error of coefficient β_i.

As aforementioned, Equation (2) contains terms that control for the common components of the panel (proxied by cross-section means) and the null hypothesis of nonstationarity is then tested on the idiosyncratic components of the panel. The common components are, therefore, required to be stationary for the test statistic to be meaningful. In case the common components are nonstationary, the test statistic cannot be meaningfully used for inferential purposes. Table (1) reports the ADF unit root tests on cross-section means and the test fails to reject the null of nonstationarity only for the real house price index. However, the weakness of the real house price index result was argued in Section (4.1.1). Consequently, the very significant test statistic for the real house price index reported in Table (5) cannot be a viable basis to infer a stationary process for the series.

Overall, the two unit root tests carried out do not provide enough evidence to reject the hypothesis that the 4 variables considered are stationary.

(Table (4) around here)

(Table (5) around here)

[2] $\bar{y}_t = N^{-1} \sum_{i=1}^{N} y_{i,t}$

4.2 Cointegration tests

4.2.1 Johansen cointegration test (Johansen (1988))

The test proposed by Johansen (1988) is implemented to test for cointegration between the series at the state level. The test has desirable properties such as the treament of all time series as endogenous variables (see Gonzalo (1994)). The following vector autoregressive process of order 1 is condered:

$$X_t = A_t X_{t-1} + \varepsilon_t, \tag{3}$$

where X_t is a vector of n $I(1)$ variables, A_t is an $n \times n$ coefficient matrix andε_t is an $n \times 1$ vector of white noise error terms. Equation (3) can be rewritten as:

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t,$$

where $\Pi = A_t - I$, and I is an n-dimensional identity matrix.

If Π is a matrix of rank 0, then all elements in the matrix are zeros and A_t is an identity matrix. In this case, X_t can be expressed as:

$$\Delta X_t = \varepsilon_t, \tag{4}$$

The implication of Equation (4) is that X_t is difference stationary since ε_t is a vector of white noise terms. Given that each variable in X_t follows a nonstationary process, it follows that no linear combination of the variables may lead to a stationary series. Therefore, no cointegrating relationship may exist between the variables.

If Π is full rank, then the long-run equilibrium solution is given by the following n independent equations, which represent n independent restrictions:

$$\begin{aligned}
\Pi_{11}X_{1t} + \Pi_{12}X_{2t} + \cdots + \Pi_{1n}X_{nt} &= 0 \\
\Pi_{21}X_{1t} + \Pi_{22}X_{2t} + \cdots + \Pi_{2n}X_{nt} &= 0 \\
&\vdots \\
\Pi_{n1}X_{1t} + \Pi_{n2}X_{2t} + \cdots + \Pi_{nn}X_{nt} &= 0
\end{aligned}$$

All the n variables in the system face n long-run constraints. In such a case, each series in X_t must be stationary with the long-run values given by the system.

More generally, the number of cointegrating vectors is equal to the rank of Π denoted as r.

It is known that the rank of a matrix is equal the number of characteristic roots significantly different from 0. Consequently, the main goal of Johansen's cointegration test is to determine the number of significant characteristic roots. This is achieved by the following test statistics:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} ln(1 - \hat{\lambda}_i),$$

$$\lambda_{max}(r, r+1) = -T\, ln(1 - \hat{\lambda_{r+1}}),$$

where $\hat{\lambda}_i$ represents the estimated values of the characteristics roots computed from the estimated matrix Π, and T is the number of observations. The null hypothesis of the first statistic is that there is no distinct cointegrating vector or the number of cointegrating vectors is less than r. The null hypothesis is tested against a more general hypothesis. The second statistic tests the null hypothesis that the number of cointegrating relationships is equal to r, against the alternative $r + 1$.

The eigen and trace statistics are reported in Table (6). The eigen statistic provides evidence for the existence of cointegrating relationships between the variables in 31 states out of 48 (64.58%) and the trace statistic provides evidence of cointegration in 48 states out 48 (100%). There is, therefore, univariate evidence for a long-run equilibrium relationship between real house prices, real income per capita, the long interest rate and the building costs.

4.2.2 Panel cointegration test (Westerlund (2007))

The error-correction based panel cointegration tests implemented here are proposed by Westerlund (2007). The tests are also second generation tests because they account for cross-section dependence in the tested panels. They do so by the use of bootstraps. The tests are the group mean test and the panel test. They are both based on the following regression equation:

$$\Delta y_{it} = \delta_i^T d_t + \alpha_i(y_{i,t-1} - \beta_i^T x_{i,t-1}) + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{i,t-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta x_{i,t-j} + e_{it}, \tag{5}$$

where d_t is a vector (or scalar) of deterministic components, α_i is the error-correction coefficient associated with the cointegrating vector $(1 - \beta_i^T)^T$. p_i and q_i are, respectively, the number of lags and the number of leads chosen for each cross-section. Equation (5) can be rewritten as:

$$\Delta y_{it} = \delta_i^T d_t + \alpha_i y_{i,t-1} - \lambda_i^T x_{i,t-1} + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{i,t-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta x_{i,t-j} + e_{it}, \tag{6}$$

where $\lambda_i^T = -\alpha_i \beta_i^T$. If $\alpha_i < 0$, then any short-run deviation from the long-run equilibrium will be followed by an error-correction. On the other hand, $\alpha_i = 0$ implies that there is no cointegration. The tests are implemented under the null hypothesis of no cointegration for all cross-sections.

The group mean test does not require homogeneity in α_i estimates. This means that the alternate hypothesis is $H_A : \alpha_i < 0$ for at least one of the cross-sections. The panel test, however, is more restrictive because the null hypothesis is that the estimated α_i is less than 0 for all cross sections in the panels ($H_A : \alpha_i = \alpha < 0$, $\forall i$).

The test statistics are reported in Table (7). The p-values computed without bootstraps and the p-values computed with bootstraps are reported. The p-values of interest for this study are the latter ones because they account for the presence of cross-section dependence in the panels. The cross-section structure of the panels was tested in Section (4.1.2) and evidence was found that the panels are strongly serially correlated. The test results provide strong evidence for cointegration between real house prices and the 3 macroeconomic variables considered: real income per capita, real building costs and the long interest rate. The group mean test is significant at the 1% significance level. Even though the more restrictive test, the panel test, has a slightly higher p-value, it is still significant at the 1% significant level as well.

Section (4.2.1) and Section (4.2.2) provide univariate evidence and panel evidence of an existing cointegration relationship between the variables in consideration. Therefore, the following sections will study the relationship dynamics more deeply. First, the cointegrating vector between the series isestimated. Then, a vector error correction model is estimated and its results are used for an impulse-response analysis.

(Table (7) around here)

4.3 Cointegrating vector estimation

4.3.1 Integrated Modified OLS (Vogelsang & Wagner (2014))

In our study, we consider the vector of difference stationary series: $y_t = \begin{bmatrix} p_t & g_t & r_t & c_t \end{bmatrix}$, where p_t is the real house price index, g_t is the real GDP per capita, r_t is the long interest rate and c_t is the real building cost index. The time period is represented by subscript t.

The results from the previous section provide evidence for the existence of a vector of real numbers $\gamma = \begin{bmatrix} 1 & -\beta^T \end{bmatrix}$ such that $\gamma^T y_t$ is stationary, where $\beta = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \end{bmatrix}^T$. In other words, even though y_t is a vector of nonstationary variables, a linear combination of the variables yields a series, which is stationary. The vector γ is called the cointegrating vector of y_t. Phillips (1991) represents the cointegrated process in the following way:

$$p_t = \alpha + \beta_1 g_t + \beta_2 r_t + \beta_3 c_t + \varepsilon_t \tag{7}$$

$$\Delta x_t = \delta + u_t \tag{8}$$

where $\Delta x_t = \begin{bmatrix} \Delta g_t & \Delta r_t & \Delta c_t \end{bmatrix}^T$ is a 3×1 vector of first differenced series, α is a vector of constants, ε_t is zero-mean stationary for β and u_t is a 3×1 zero-mean stationary vector. The cointegrating vector γ is assumed to eliminate all trends (i.e. deterministic and stochastic trends). This is why a deterministic trend was not included in Equation (7) [3].

We seek to estimate the vector β. Given that the series are cointegrated, estimating β with the least squares estimator will provide estimates that converge to the true value at the rate of T (the sample size) even when x_t is serially correlated with ε_t. However, the estimate is asymptotically biased and inefficient and its asymptotic distribution is non-normal [4]. Consequently, the standard errors produced by the least squares estimator cannot be reliably used for statistical inference. A suitable estimator among others in such a case is the integrated modified OLS (IMOLS) estimator proposed by Vogelsang & Wagner (2014). The results of the cointegrating vector estimation are reported in Table (8), Table (9), Table (10) and Table (11):

- 43 of the 48 coefficient estimates (89.58%) for the real income per capita are positive.
- Only 20 of the 48 estimated coefficients (41.67%) for the long interest rate are negative.
- 45 of the 48 coefficient estimates (93.75%) for the real building cost index are positive.

The IMOLS estimates predominantly corroborate the theoretical expectations about the long-run relationship between real house prices and the macroeconomic variables considered. The estimator provides evidence for a positive cointegrating relationship with real income per capita and the real building cost index. As for the long interest rate, the states for which the test results support the expectations formulated are only about 42% of the sampled states. The hypothesis that real house prices have a negative long-run relationship with interest rates is, therefore, not supported by the data and the model used.

[3] Ogaki & Park (1998)show that this is the case when the deterministic cointegrating restriction is satisfied. When the stochastic trend only is eliminated by the cointegrating vector, $\gamma^T y_t$ is trend stationary.

[4] See Stock (1987) and Phillips (1991) for details.

(Table (8) around here)

(Table (9) around here)

(Table (10) around here)

(Table (11) around here)

4.3.2 Panel dynamic OLS

The dynamic OLS (DOLS) is an estimation technique, which was proposed by Saikkonen (1991) and Stock & Watson (1993). It is similar to the IMOLS described above because it is useful for the estimation of cointegrating vectors. Mark & Sul (2003) extend the procedure to a panel framework, which has many benefits over the univariate estimator. Inder (1993) and Stock & Watson (1993) point out the poor statistical properties of the DOLS in small samples. In the panel DOLS, Mark & Sul (2003) assume that the cross sections are individually heterogenous in their short run dynamics especially due to individual-specific fixed effects as well as their individual-specific time trends. In the long-run, however, the cointegrating vector is assumed to be homogenous across cross-sections.

The panel DOLS results are reported in Table (12) and they corroborate the univariate results (standard errors are in parenthesis):$\beta_1 = 1.0151(0.1150)$, $\beta_2 = 0.0060(0.0085)$ and $\beta_3 = 0.6687(0.1665)$. The estimates point to a positive cointegrating relationship between real house prices and the 3 macroeconomic variables considered: real building costs, interest rates and real income per capita. Therefore, there is panel evidence to support the theoretical expectations on real income per capita and the real building costs. However, the panel results do not provide support for the formulated expectations about the cointegrating relationship between real house prices and interest rates.

4.4 Vector error correction model

4.4.1 Univariate VECM

The Vector Error Correction Model (VECM) is used to study the responses of housing prices to shocks in real income per capita, interest rates and real building costs. The model makes use of the cointegrating vector β estimated with IMOLS. The following equation is considered:

$$\Delta y_t = \alpha + \rho\gamma^T y_{t-1} + \sum_{j=1}^{k} \theta_j \Delta y_{t-j} + C\varepsilon_t \tag{9}$$

where α is a vector of constants, $\rho = \left[\begin{array}{cccc} \rho_1 & \rho_2 & \rho_3 & \rho_4 \end{array}\right]^T$ is a 4×1 vector of speed of convergence coefficients, C is a matrix that defines the contemporaneous structural relationship among the four variables, and $\varepsilon_t = \left[\begin{array}{cccc} \varepsilon_{p,t} & \varepsilon_{g,t} & \varepsilon_{r,t} & \varepsilon_{c,t} \end{array}\right]^T$ is a vector of mutually orthogonal structural shocks. Orthogonality in the structural shocks implies $E\varepsilon_t\varepsilon_t^T = I$, where I is 4×4 identity matrix.$Eu_tu_t^T = EC\varepsilon_t\varepsilon_t^T C^T = CC^T = \sum$, where $\sum$ is the variance-covariance matrix. The matrix of contemporaneous structural relationship C is identified with the Cholesky decomposition of the variance-covariance matrix using the following ordering of y_t: $\left[\begin{array}{cccc} g_t & r_t & c_t & p_t \end{array}\right]^T$. The ordering of y_t is relevant for the model. For instance, the chosen order implies that a structural shock in interest rates will induce a contemporaneous response in real building costs and real housing prices. However, real income per capita will not be contemporaneously affected by the shock.

The effects of the structural shocks $\varepsilon_{g,t}$, $\varepsilon_{r,t}$, $\varepsilon_{c,t}$ on real house prices are studied with an impulse-response analysis based on the vector error correction system described in Equation (9). We, therefore, rewrite the equation into the following state-space representation:

$$z_t = Fz_{t-1} + \zeta_t, \tag{10}$$

$$z_t = \left[\begin{array}{cccc} y_t & y_{t-1} & \dots & y_{t-k} \end{array}\right],$$

$$F = \left[\begin{array}{cccc} \vartheta_1 & \vartheta_2 & \dots & \vartheta_{k+1} \\ & & & \vdots \\ & I_4 & & \\ & & & 0 \end{array}\right],$$

$$\zeta_t = \left[\begin{array}{cccc} Ce_t & 0 & \dots & 0 \end{array}\right]^T,$$

$$\vartheta_1 = I_4 + \rho\gamma^T + \theta_1,$$

$$\vartheta_j = \theta_{j+1} - \theta_j,\ j = 2, \dots, k\ ,$$

$$\vartheta_{k+1} = -\theta_j,$$

I_4 is a 4×4 identity matrix. The n^{th} period impulse-response point estimate is, therefore, obtained by:

$$(S^T F^n S)C, \tag{11}$$

where $S = \begin{bmatrix} I_4 & 0 & \dots & 0 \end{bmatrix}^T$ is a selection matrix of dimensions $2(k+1) \times 4$. Note that γ is obtained using the estimates of β as computed by the IMOLS estimator.

The impulse-response functions are estimated with confidence bands at 95% significance level with 20 periods ahead. The reported functions are the responses of the real house price index to orthogonal shocks in the real income per capita (see Figure (10)), in the long interest rate (see Figure (11)) and in the real building cost index (see Figure (12)). The impulse-response functions generally corroborate the formulated expectations about the long-run equilibrium relationships. Real house prices generally respond positively to shocks in the real income per capita. Despite that the IMOLS coefficients for the long interest rate are mostly positive, the response functions show a negative response of the housing prices overall. The 95% confidence band in most states has a distribution, which is weakly negative. In other words, a larger portion of the confidence band lies below zero. Finally, overall the housing price index responds positively to shocks in the real building cost index. However, a very small number of states show opposite results. An example of such states is Ohio.

(Figure (10) around here)

(Figure (11) around here)

(Figure (12) around here)

4.4.2 Panel VECM

The panel VECM is an extension of the univarate VECM to a panel framework. First, the panel data is pooled and demeaned. The demeaning process consists in substracting each country's arithmetic mean from the country's observations. The rest of the procedure is similar to the univariate VECM described in Section (4.4.1) with the exception that the intercept is removed from the equation. This is due to the demeaning process:

$$\Delta y_t = \rho \gamma^T y_{t-1} + \sum_{j=1}^{k} \theta_j \Delta y_{t-j} + C\varepsilon_t$$

The impulse-response functions of interest are the responses of housing prices to a shock in the real income per capita (see Figure (13)), the long interest rate (see Figure (14)) and building costs (see Figure (15)). The panel results corroborate the univariate results and the expectations about the cointegrating relationships for the the real income per capita and the real building cost index variables. As for the long interest rate, the initial response of the house price index is negative. However, the negative response is very short-lived because it becomes a significantly positive response after about 3 time periods and throughout the rest of the time periods. There is, therefore, no panel evidence supporting a negative cointegrating relationship between housing prices and interest rates.

(Figure (13) around here)

(Figure (14) around here)

(Figure (15) around here)

5 Conclusion

The literature studying the relationship between housing prices and the macroeconomy is plentiful. Many techniques have been used to study the dynamics between the housing market and other variables. In this paper, econometric procedures known as second generation models are employed in the analysis. The macreconomic variables considered are the real income per capita, the long interest rate and the real building cost index. Expectations about the relationships between the variables are formulated with the usage of a basic demand and supply model and these expectations are generally corroborated by the literature. The empirical analysis consists in four main steps each involving univariate models and panel models. The first step involves unit root tests which show strong evidence that the variables in the model obey a non stationary process. The panel unit test used (Pesaran (2007)) required the testing of the structure of the panels for cross-section dependence. The test results showed significant serial correlation between the panel cross-sections. In the second step, strong evidence for cointegration between the variables considered was found at both univariate and panel levels. The purpose of the third step was to estimate the cointegrating vector between the series. The sign of the coefficients in the cointegrating vector is important because it indicates the nature of the long-run relationship between the variables. The univariate and panel results strongly support the demand/supply predictions as well as the literature when it comes to the real income per capita and real building cost index series. As for the long rate series, the cointegrating coefficients were predominantly positive at the state level and so was the cointegrating coefficient of the panel model, which is contrary to what theory predicts as well as the predominant results in the literature. The impulse-response functions in step 4, however, showed mostly negative responses of the real housing price index to shocks in the long interest rate in the univariate analysis. On the other hand, the panel technique showed a positive response of housing prices to a shock in interest rates. The expectations and the literature were supported by the response functions produced for shocks in the real income per capita and the building cost index. The responses of housing prices were estimated bo be predominantly positive in a 20-year time period.

Overall, this paper shows evidence for cointegration between real housing prices and the real income per capita, the long interest rate and the real building cost index in the case of the US economy.

References

Adams, Z. & Fuss, R. (2010), 'Macroeconomic determinants of international housing markets', *Journal of Housing Economics* **19**(1), 38–50.

Addison-Smyth, D., McQuinn, K. & O' Reilly, G. (2008), 'Estimating the Structural Demand for Irish Housing', (1/RT/08).

Baffoe-Bonnie, J. (1998), 'The dynamic impact of macroeconomic aggregates on housing prices and stock of houses: A national and regional analysis', *The Journal of Real Estate Finance and Economics* **17**(2), 179–197.

Baker, D. (2008), 'The housing bubble and the financial crisis', *Real-world economics review* **46**(20), 73–81.

Brocker, M. & Hanes, C. (2014), 'The 1920s american real estate boom and the downturn of the great depression: Evidence from city cross-sections', pp. 161–201.

Demyanyk, Y. & Van Hemert, O. (2011), 'Understanding the subprime mortgage crisis', *Review of Financial Studies* **24**(6), 1848–1880.

Duca, J. V., Muellbauer, J. & Murphy, A. (2010), 'Housing markets and the financial crisis of 2007-2009: Lessons for the future', (0049).

Egert, B. & Mihaljek, D. (2007), 'Determinants of house prices in central and eastern europe', (2152).

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

Gonzalo, J. (1994), 'Five alternative methods of estimating long-run equilibrium relationships', *Journal of Econometrics* **60**(1-2), 203–233.

Inder, B. (1993), 'Estimating long-run relationships in economics: a comparison of different approaches', *Journal of econometrics* **57**(1), 53–68.

Johansen, S. (1988), 'Statistical analysis of cointegration vectors', *Journal of Economic Dynamics and Control* **12**(2), 231 – 254.

Katrakilidis, C. & Trachanas, E. (2012), 'What drives housing price dynamics in greece: New evidence from asymmetric ardl cointegration', *Economic Modelling* **29**(4), 1064–1069.

Kim, S.-W. & Bhattacharya, R. (2009), 'Regional housing prices in the usa: an empirical investigation of nonlinearity', *The Journal of Real Estate Finance and Economics* **38**(4), 443–460.

Mark, N. C. & Sul, D. (2003), 'Cointegration vector estimation by panel dols and long-run money demand', *Oxford Bulletin of Economics and Statistics* **65**(5), 655–680.

Ogaki, M. & Park, J. (1998), 'A cointegration approach to estimating preference parameters', *Journal of Econometrics* **82**(1), 107–134.

Pesaran, H. M. (2004), 'General diagnostic tests for cross section dependence in panels', (0435).

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.

Phillips, P. C. (1991), 'Optimal inference in cointegrated systems', *Econometrica: Journal of the Econometric Society* pp. 283–306.

Saikkonen, P. (1991), 'Asymptotically efficient estimation of cointegration regressions', *Econometric Theory* **7**(01), 1–21.

Sarno, L. & Taylor, M. P. (1998), 'Real exchange rates under the recent float: unequivocal evidence of mean reversion', *Economics Letters* **60**(2), 131–137.

Shiller, R. J. (2015), *Irrational exuberance*, Princeton university press.

Stock, J. H. (1987), 'Asymptotic properties of least squares estimators of cointegrating vectors', *Econometrica: Journal of the Econometric Society* pp. 1035–1056.

Stock, J. & Watson, M. (1993), ‘A simple estimator of cointegrating vectors in higher-order integrated systems’, *Econometrica* **61**(4), 783–820.

Sutton, G. D. (2002), ‘Explaining changes in house prices1’, *BIS quarterly review* p. 47.

Tsai, I.-C., Lee, C.-F. & Chiang, M.-C. (2012), ‘The asymmetric wealth effect in the us housing and stock markets: evidence from the threshold cointegration model’, *The Journal of Real Estate Finance and Economics* **45**(4), 1005–1020.

Tsatsaronis, K. & Zhu, H. (2004), ‘What drives housing price dynamics: cross-country evidence’, *BIS Quarterly Review* .

Vogelsang, T. J. & Wagner, M. (2014), ‘Integrated modified ols estimation and fixed-b inference for cointegrating regressions’, *Journal of Econometrics* **178**(2), 741–760.

Westerlund, J. (2007), ‘Testing for error correction in panel data’, *Oxford Bulletin of Economics and Statistics* **69**(6), 709–748.

Zhou, J. (2010), ‘Testing for cointegration between house prices and economic fundamentals’, *Real Estate Economics* **38**(4), 599–632.

Appendix

Variable	ADF	P-value
Real house price index	-3.2567	0.09 *
Real income per capita	-2.2248	0.49
Long interest rate	-2.9306	0.21
Real building cost index	-0.9901	0.93

Table 1: ADF test statistics (cross-section means)

State	ADF	P-value	State	ADF	P-value
Alabama	-2.7632	0.28	Montana	-2.2633	0.47
Arkansas	-2.4415	0.40	North Carolina	-2.6446	0.32
Arizona	-3.4546	0.06 *	North Dakota	-1.6395	0.71
California	-3.7901	0.03 **	Nebraska	-2.1957	0.50
Colorado	-2.2309	0.48	New Hampshire	-4.8976	0.01 **
Connecticut	-3.8733	0.03 **	New Jersey	-3.8739	0.03 **
District of Columbia	-3.6177	0.05 *	New Mexico	-3.3708	0.08 **
Delaware	-3.1968	0.11	Nevada	-3.9814	0.02 **
Florida	-3.4968	0.06 *	New York	-4.3325	0.01 **
Georgia	-2.4430	0.40	Ohio	-2.0325	0.56
Iowa	-3.3010	0.09 *	Oklahoma	-2.0626	0.55
Idaho	-3.3938	0.07 *	Oregon	-2.5595	0.36
Illinois	-2.6871	0.31	Pennsylvania	-4.4360	0.01 **
Indiana	-2.5785	0.35	Rhode Island	-4.2002	0.01 **
Kansas	-2.0559	0.55	South Carolina	-1.9314	0.60
Kentucky	-2.2753	0.47	South Dakota	-2.3114	0.45
Louisiana	-2.0035	0.57	Tennessee	-2.2967	0.46
Massachusetts	-4.1791	0.01 **	Texas	-2.1752	0.50
Maryland	-4.6079	0.01 **	Utah	-3.1818	0.11
Maine	-2.6468	0.32	Virginia	-3.7197	0.04 **
Michigan	-1.2349	0.87	Vermont	-3.0123	0.18
Minnesota	-2.7133	0.30	Washington	-3.8568	0.03 **
Missouri	-2.7322	0.29	Wisconsin	-1.6998	0.69
Mississippi	-1.8265	0.64	West Virginia	-1.8018	0.65

Table 2: ADF test statistics: real house price index (log)

State	ADF	P-value	State	ADF	P-value
Alabama	-1.4962	0.77	Montana	-1.9494	0.59
Arkansas	-3.3872	0.07 *	North Carolina	-0.7632	0.96
Arizona	-2.2085	0.49	North Dakota	-2.1671	0.51
California	-2.3365	0.44	Nebraska	-3.2364	0.10
Colorado	-1.4423	0.79	New Hampshire	-1.4566	0.79
Connecticut	-1.6222	0.72	New Jersey	-1.4668	0.78
District of Columbia	-2.7826	0.27	New Mexico	-2.6365	0.33
Delaware	-1.2104	0.88	Nevada	-1.8155	0.65
Florida	-1.3914	0.81	New York	-2.2630	0.47
Georgia	-0.7175	0.96	Ohio	-1.8773	0.62
Iowa	-2.5910	0.34	Oklahoma	-1.7966	0.65
Idaho	-2.0267	0.56	Oregon	-1.8857	0.62
Illinois	-1.9802	0.58	Pennsylvania	-2.3298	0.44
Indiana	-1.5699	0.74	Rhode Island	-1.8637	0.63
Kansas	-2.8834	0.23	South Carolina	-0.8017	0.95
Kentucky	-1.9077	0.61	South Dakota	-2.7573	0.28
Louisiana	-2.2373	0.48	Tennessee	-0.9131	0.94
Massachusetts	-1.8546	0.63	Texas	-2.3337	0.44
Maryland	-2.0898	0.54	Utah	-2.6375	0.32
Maine	-1.7847	0.66	Virginia	-1.7959	0.65
Michigan	-2.0501	0.55	Vermont	-2.0360	0.56
Minnesota	-1.6688	0.70	Washington	-2.4383	0.40
Missouri	-2.0208	0.57	Wisconsin	-1.6310	0.72
Mississippi	-2.2389	0.48	West Virginia	-2.5505	0.36

Table 3: ADF test statistics: Real income per capita (log)

Variable	CD	Average $\hat{\rho}$	P-value
Real house price index	80.5698 ***	0.3944	0.00
Real income per capita	147.3117 ***	0.7211	0.00

Table 4: Cross-section dependence test statistics

Variable	Statistic
Real house price index	-2.5638 ***
Real income per capita	-1.7632

Table 5: Panel unit root test statistics

State	Eigen	Trace	State	Eigen	Trace
Alabama	25.12	33.48 *	Montana	23.98 **	39.33 **
Arizona	27.95 *	59.27 **	Nebraska	22.63 **	35.91 **
Arkansas	27.66 *	32.66 *	Nevada	25.74 *	56.92 **
California	25.71 *	55.78 **	New Hampshire	24.44	32.83 *
Colorado	21.57 *	35.61 **	New Jersey	23.94	53.78 **
Connecticut	21.65	53.05 *	New Mexico	29.16 **	32.51 *
Delaware	24.38	54.19 **	New York	23.27	51.9 *
District of Columbia	25.6 *	55.73 **	North Carolina	22.9	34.75 *
Florida	22.67	51.75 *	North Dakota	23.56 **	37.43 **
Georgia	23.41	33.82 *	Ohio	26.23 *	32.01 *
Idaho	27.57 *	59.07 **	Oklahoma	27.46 *	32.29 *
Illinois	27.44 *	58.93 **	Oregon	20.81 *	35.1 **
Indiana	26.89 *	33.61 *	Pennsylvania	25.23	55.79 **
Iowa	20.33 *	33.55 *	Rhode Island	23.49	54.18 **
Kansas	20.78 *	34.39 *	South Carolina	24.13	34.83 *
Kentucky	27.56 *	32.51 *	South Dakota	22.5 **	35.71 **
Louisiana	27.23 *	57.09 **	Tennessee	24.06	55.11 **
Maine	23.81	53.45 **	Texas	25.94 *	57.54 **
Maryland	25.59 *	55.5 **	Utah	14.34 *	36.6 **
Massachusetts	23.68	54.52 **	Vermont	24.41	56.3 **
Michigan	25.4	33.02 *	Virginia	29.02 **	33.03 *
Minnesota	19.84 *	34.62 *	Washington	26.13 *	33.88 *
Mississippi	22.1 **	36.43 **	West Virginia	25.83 *	57.74 **
Missouri	30.09 **	61.63 ***	Wisconsin	26.03 *	34.38 *

Table 6: Johansen cointegration test statistics

Test	Statistic	P-value	P-value with CSD
Group mean test	-5.540 ***	0.000	0.006
Panel test	-8.321 ***	0.000	0.000

Table 7: Westerlund (2007) panel cointegration test

State	Variable	Statistic	Standard error
Alabama	Real income per capita	-0.0298	0.2076
	Long interest rate	-0.0284	0.0125
	Real building cost index	0.895	0.2879
Arkansas	Real income per capita	0.1193	0.203
	Long interest rate	-0.0122	0.0121
	Real building cost index	1.2711	0.2553
Arizona	Real income per capita	1.7407	0.2512
	Long interest rate	0.0143	0.0147
	Real building cost index	1.8059	0.3063
California	Real income per capita	2.6514	0.6845
	Long interest rate	0.0651	0.0372
	Real building cost index	0.1279	0.7299
Colorado	Real income per capita	1.6107	0.2827
	Long interest rate	0.0103	0.0184
	Real building cost index	1.6453	0.4451
Connecticut	Real income per capita	1.2549	0.4904
	Long interest rate	0.0438	0.0356
	Real building cost index	1.2571	0.8006
District of Columbia	Real income per capita	1.7424	0.327
	Long interest rate	0.0056	0.0358
	Real building cost index	2.4063	0.6405
Delaware	Real income per capita	2.2594	0.5257
	Long interest rate	0.0586	0.0286
	Real building cost index	0.9943	0.5692
Florida	Real income per capita	1.8461	0.3098
	Long interest rate	0.0062	0.0205
	Real building cost index	2.5041	0.4299
Georgia	Real income per capita	0.9336	0.5042
	Long interest rate	0.0259	0.0326
	Real building cost index	0.9668	0.7239
Iowa	Real income per capita	-0.0098	0.3921
	Long interest rate	-0.0347	0.0215
	Real building cost index	0.8995	0.3538
Idaho	Real income per capita	1.1076	0.2631
	Long interest rate	0.0219	0.0127
	Real building cost index	0.9647	0.3079

Table 8: IMOLS estimates (1 of 4)

State	Variable	Statistic	Standard error
Illinois	Real income per capita	0.5705	0.2091
	Long interest rate	-0.0172	0.0109
	Real building cost index	0.254	0.203
Indiana	Real income per capita	-0.2106	0.3013
	Long interest rate	-0.022	0.0128
	Real building cost index	-0.0174	0.2663
Kansas	Real income per capita	0.2678	0.1667
	Long interest rate	-0.0149	0.0096
	Real building cost index	1.3392	0.1766
Kentucky	Real income per capita	0.1162	0.2839
	Long interest rate	-0.0177	0.0145
	Real building cost index	0.4643	0.323
Louisiana	Real income per capita	0.7022	0.2
	Long interest rate	0.0152	0.0128
	Real building cost index	1.8546	0.3207
Massachusetts	Real income per capita	2.1814	0.508
	Long interest rate	0.0913	0.0473
	Real building cost index	0.7709	0.8169
Maryland	Real income per capita	1.9102	0.4232
	Long interest rate	0.0601	0.0345
	Real building cost index	1.158	0.5547
Maine	Real income per capita	1.5412	0.4919
	Long interest rate	0.0538	0.0353
	Real building cost index	1.1152	0.6731
Michigan	Real income per capita	2.1007	0.6422
	Long interest rate	0.0492	0.0234
	Real building cost index	-0.0402	0.508
Minnesota	Real income per capita	1.4214	0.1843
	Long interest rate	0.0282	0.0122
	Real building cost index	1.4049	0.2419
Missouri	Real income per capita	0.5922	0.1578
	Long interest rate	0.0007	0.0088
	Real building cost index	0.9393	0.1678
Mississippi	Real income per capita	0.1874	0.2039
	Long interest rate	-0.0055	0.0126
	Real building cost index	1.2832	0.2972

Table 9: IMOLS estimates (2 of 4)

State	Variable	Statistic	Standard error
Montana	Real income per capita	1.1132	0.2394
	Long interest rate	-0.0143	0.0135
	Real building cost index	1.043	0.2978
North Carolina	Real income per capita	0.5162	0.0967
	Long interest rate	-0.0045	0.0063
	Real building cost index	0.7321	0.1342
North Dakota	Real income per capita	0.309	0.1053
	Long interest rate	-0.0134	0.0074
	Real building cost index	1.7895	0.1344
Nebraska	Real income per capita	0.242	0.1437
	Long interest rate	-0.0089	0.0092
	Real building cost index	0.6338	0.1613
New Hampshire	Real income per capita	1.8604	0.271
	Long interest rate	0.0791	0.0212
	Real building cost index	1.8333	0.4617
New Jersey	Real income per capita	2.1193	0.6014
	Long interest rate	0.0624	0.0418
	Real building cost index	1.604	0.8533
New Mexico	Real income per capita	0.8176	0.1836
	Long interest rate	0.0106	0.0109
	Real building cost index	0.9143	0.2279
Nevada	Real income per capita	3.0153	0.8326
	Long interest rate	0.0592	0.0305
	Real building cost index	2.2483	0.7342
New York	Real income per capita	1.7593	0.5639
	Long interest rate	0.0251	0.0425
	Real building cost index	0.9632	0.6806
Ohio	Real income per capita	-0.3000	0.3081
	Long interest rate	-0.0228	0.0132
	Real building cost index	-0.3585	0.2717
Oklahoma	Real income per capita	0.505	0.1947
	Long interest rate	0.0152	0.0111
	Real building cost index	2.0281	0.263
Oregon	Real income per capita	1.9233	0.626
	Long interest rate	0.0089	0.029
	Real building cost index	0.9337	0.5435

Table 10: IMOLS estimates (3 of 4)

State	Variable	Statistic	Standard error
Pennsylvania	Real income per capita	0.6534	0.4272
	Long interest rate	-0.0063	0.0232
	Real building cost index	0.6213	0.4667
Rhode Island	Real income per capita	2.057	0.3802
	Long interest rate	0.0599	0.029
	Real building cost index	0.8119	0.5206
South Carolina	Real income per capita	0.6151	0.1016
	Long interest rate	0.0018	0.0064
	Real building cost index	1.0271	0.1358
South Dakota	Real income per capita	0.3282	0.1588
	Long interest rate	-0.017	0.0119
	Real building cost index	1.074	0.2095
Tennessee	Real income per capita	0.4055	0.2442
	Long interest rate	-0.0037	0.0138
	Real building cost index	0.9294	0.3386
Texas	Real income per capita	0.8388	0.1635
	Long interest rate	0.0415	0.0095
	Real building cost index	1.539	0.2057
Utah	Real income per capita	1.254	0.2448
	Long interest rate	-0.0182	0.0145
	Real building cost index	1.5407	0.2741
Virginia	Real income per capita	1.4644	0.361
	Long interest rate	0.0328	0.0299
	Real building cost index	1.3994	0.5465
Vermont	Real income per capita	1.1131	0.3262
	Long interest rate	0.0328	0.0252
	Real building cost index	0.7072	0.4794
Washington	Real income per capita	1.2238	0.3638
	Long interest rate	-0.0128	0.0205
	Real building cost index	0.2905	0.4238
Wisconsin	Real income per capita	0.6996	0.4061
	Long interest rate	-0.0194	0.0201
	Real building cost index	0.7063	0.4222
West Virginia	Real income per capita	-0.2824	0.263
	Long interest rate	-0.0285	0.0137
	Real building cost index	1.2778	0.2733

Table 11: IMOLS estimates (4 of 4)

Variable	Statistic	Standard error
Real income per capita	1.0151	0.1150
Real building cost index	0.6687	0.1665
Long interest rate	0.0060	0.0085

Table 12: Panel dynamic ordinary least squares (PDOLS) estimates

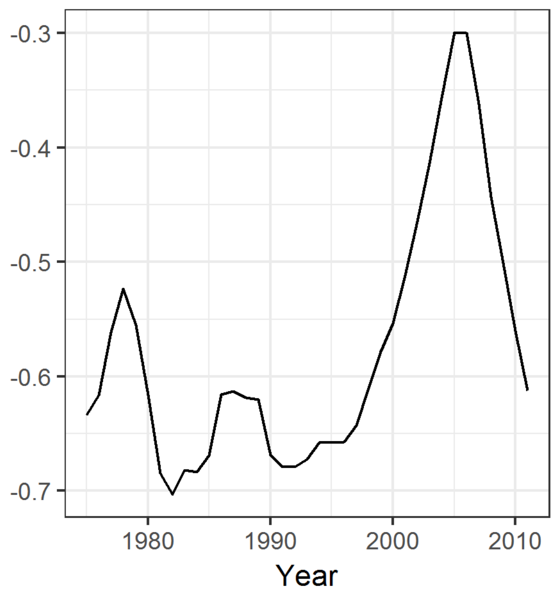


Figure 1: Real house price index (log) (cross-section means)

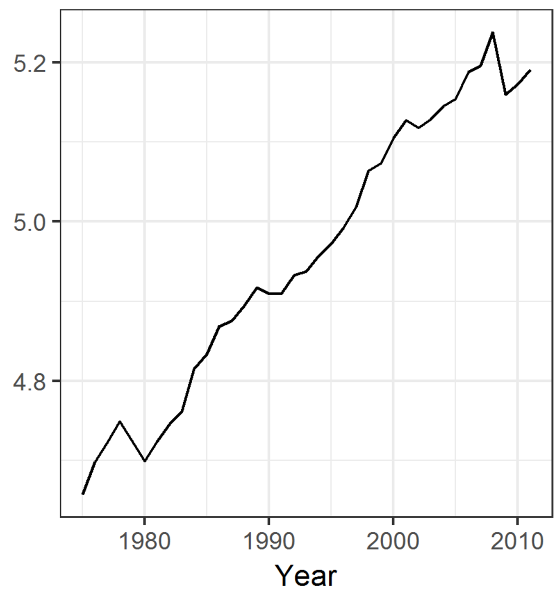


Figure 2: Real income per capita (log) (cross-section means)

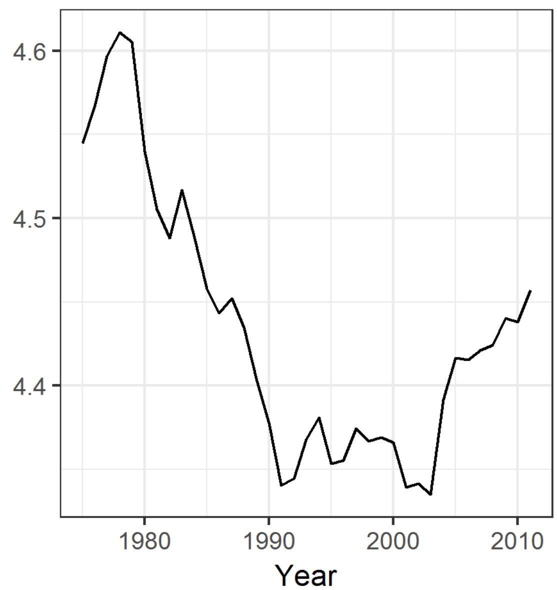


Figure 3: Real building cost index (cross-section means) (log)

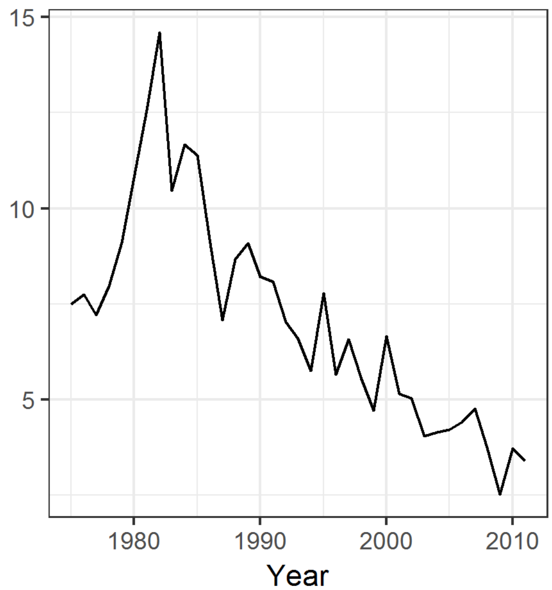


Figure 4: Long interest rate

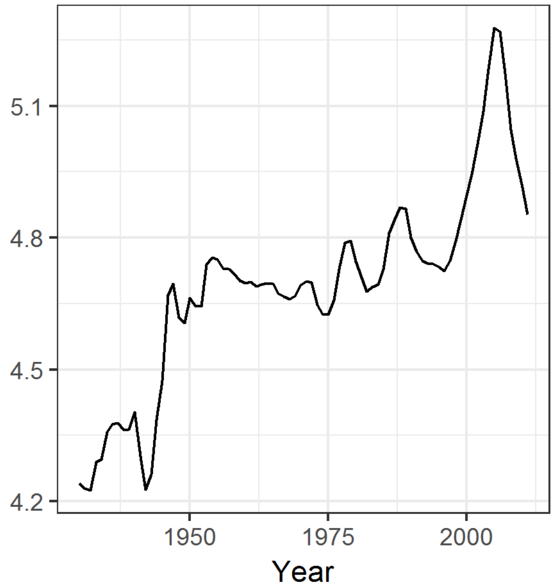


Figure 5: Real house price index (Shiller, 2009) (log)

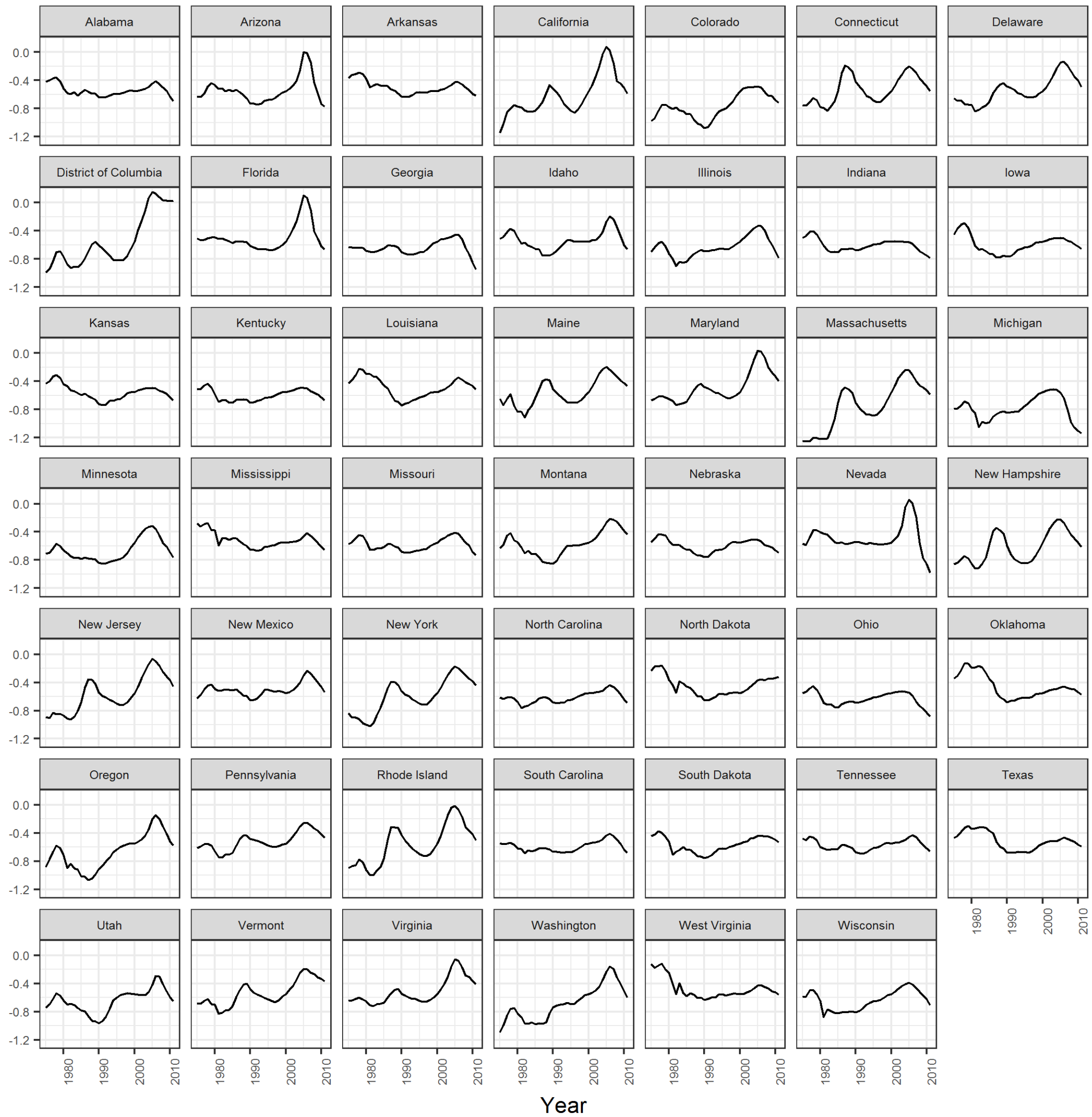


Figure 6: Real house price index (log) (by state)

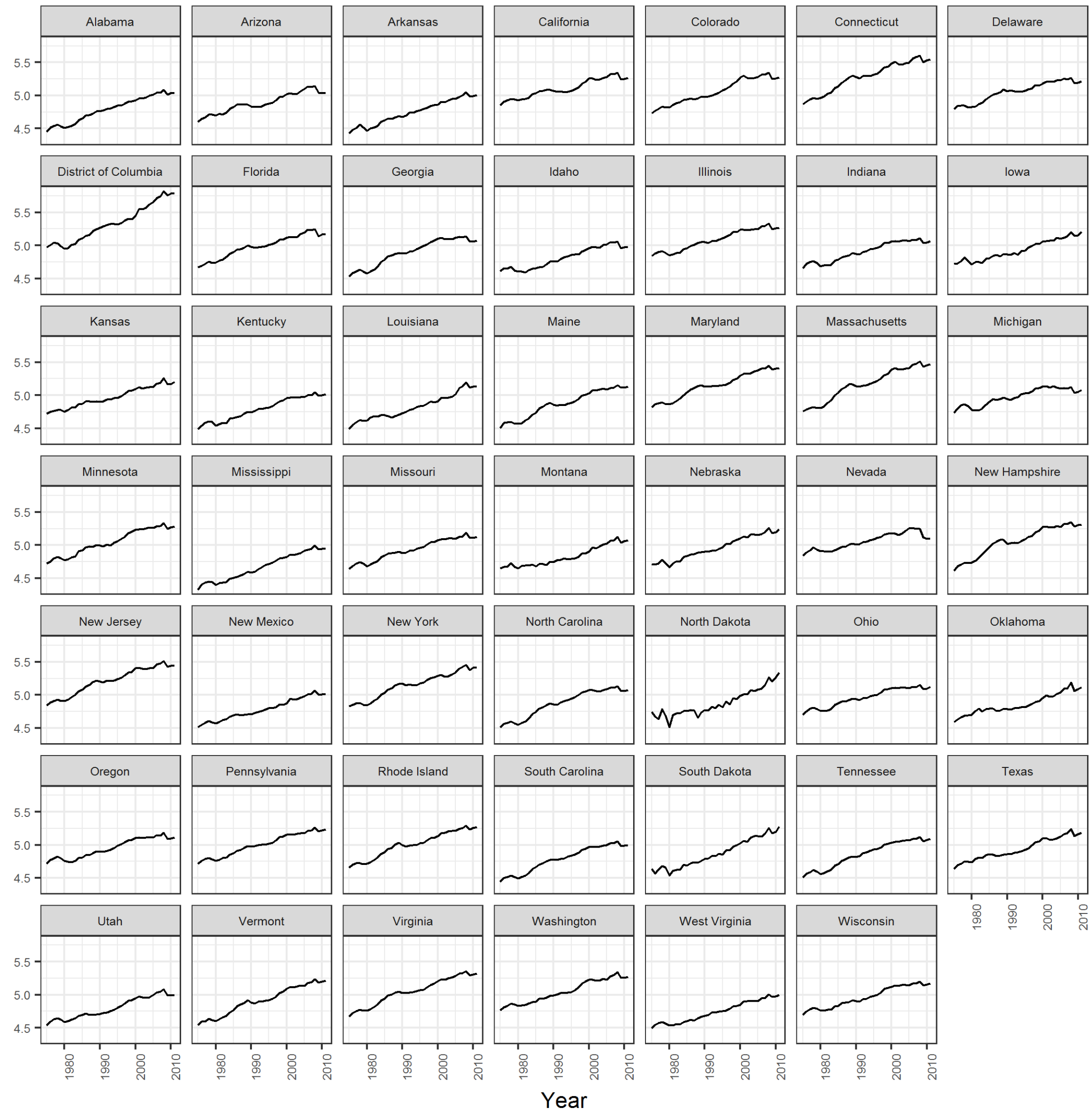


Figure 7: Real income per capita (log) (by state)

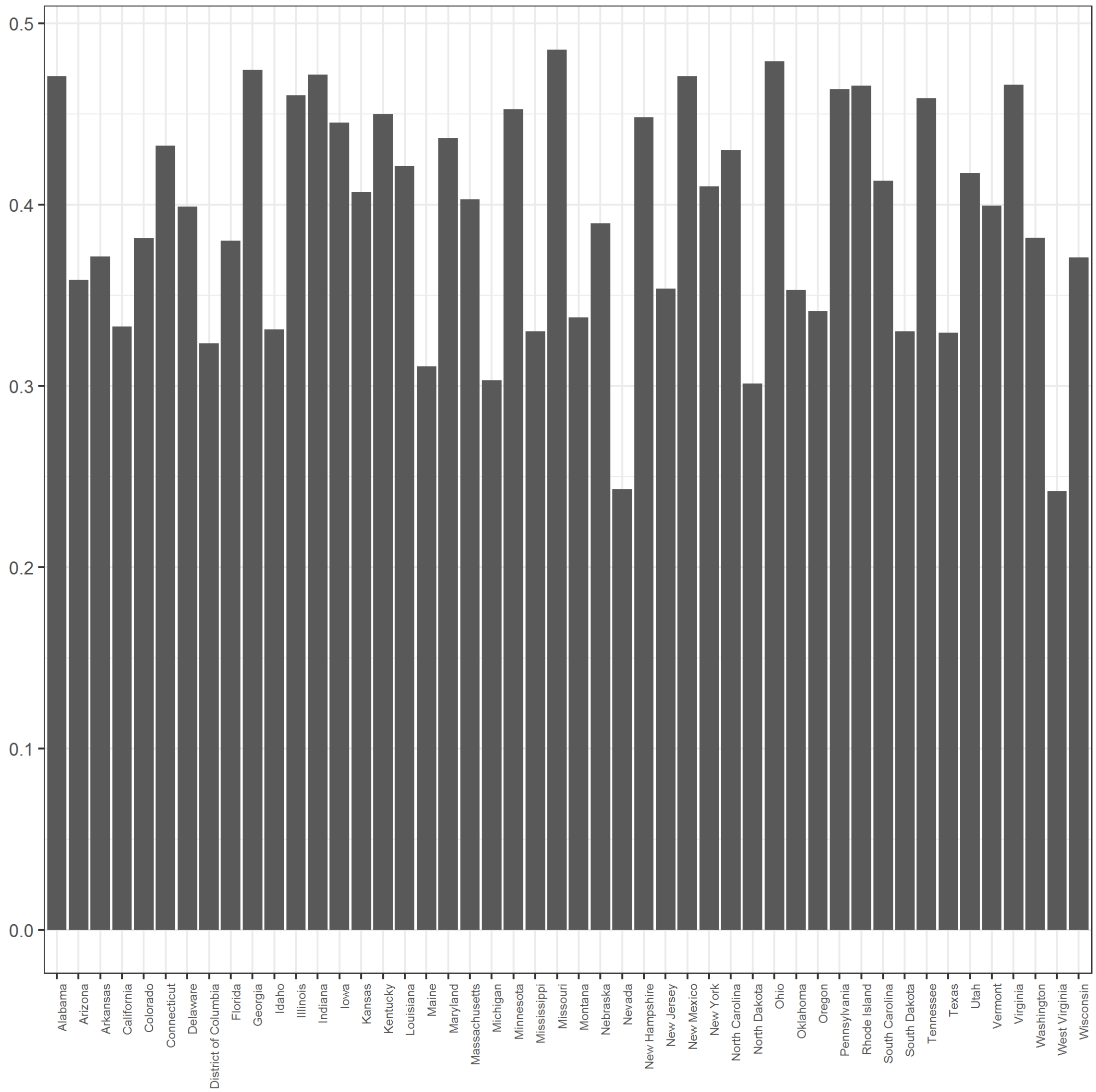


Figure 8: Correlation coefficients: real house price index (by state)

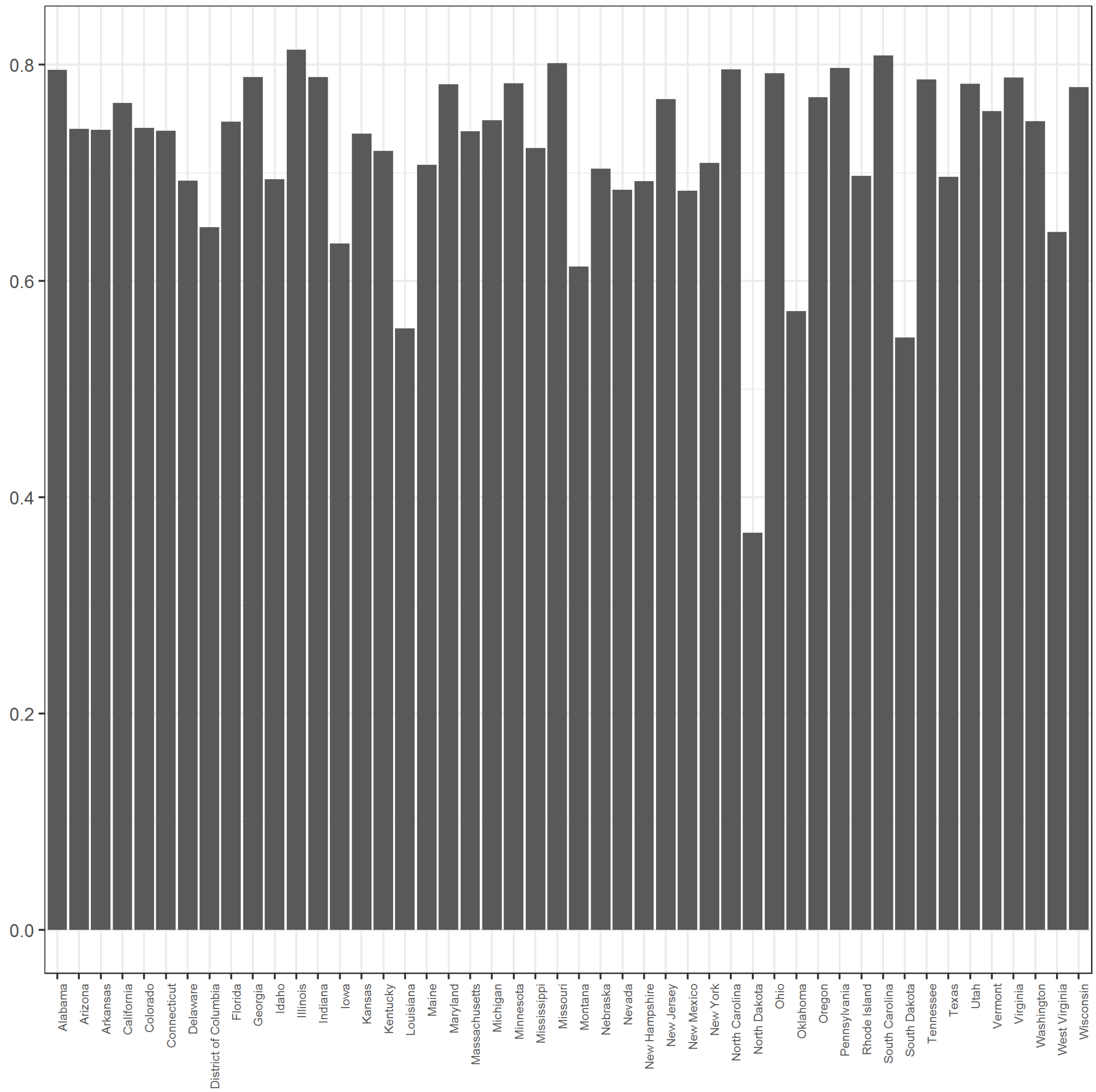


Figure 9: Correlation coefficients: real GDP per capita (by state)

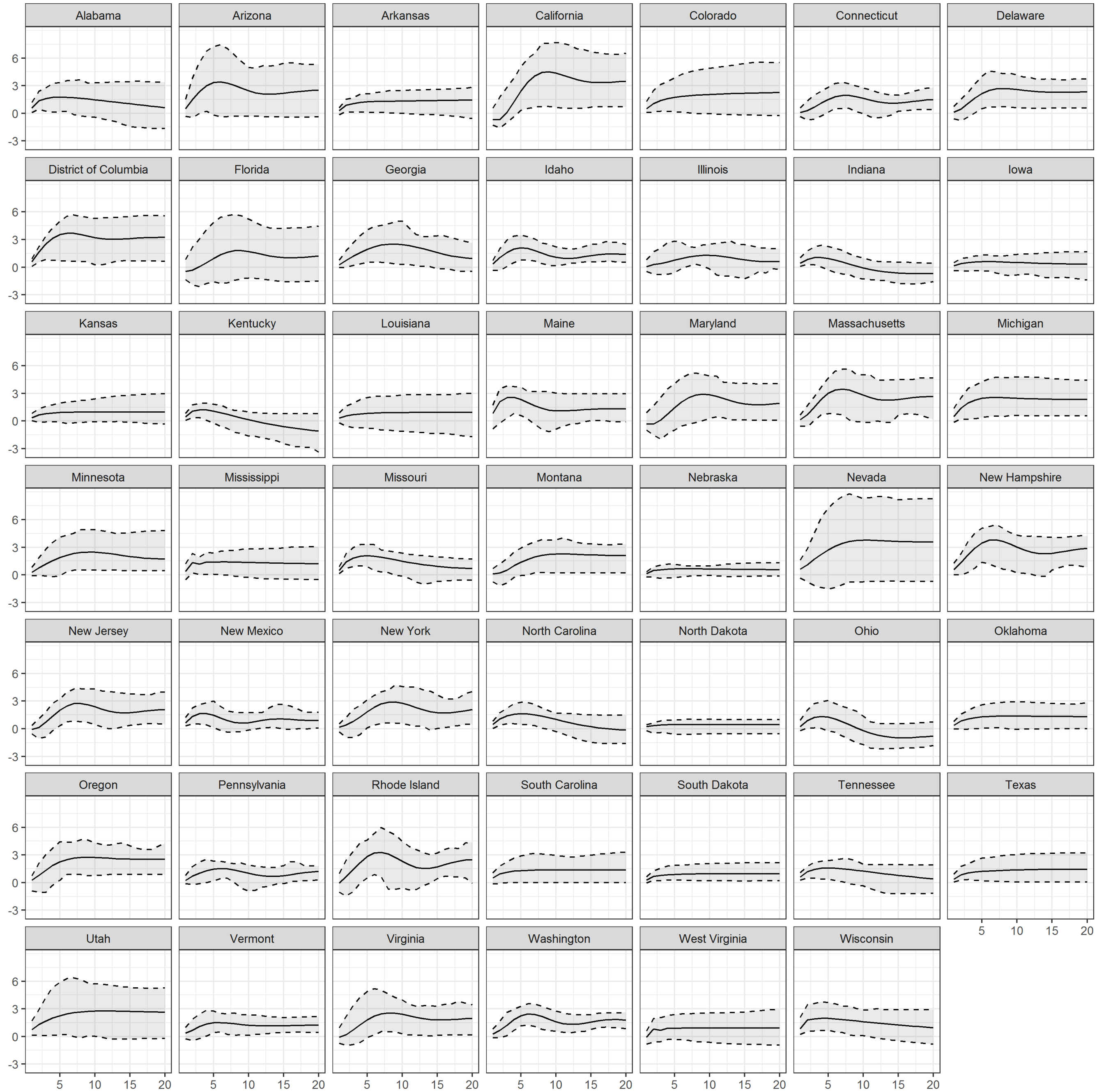


Figure 10: Responses of the house price index to a shock in real income per capita

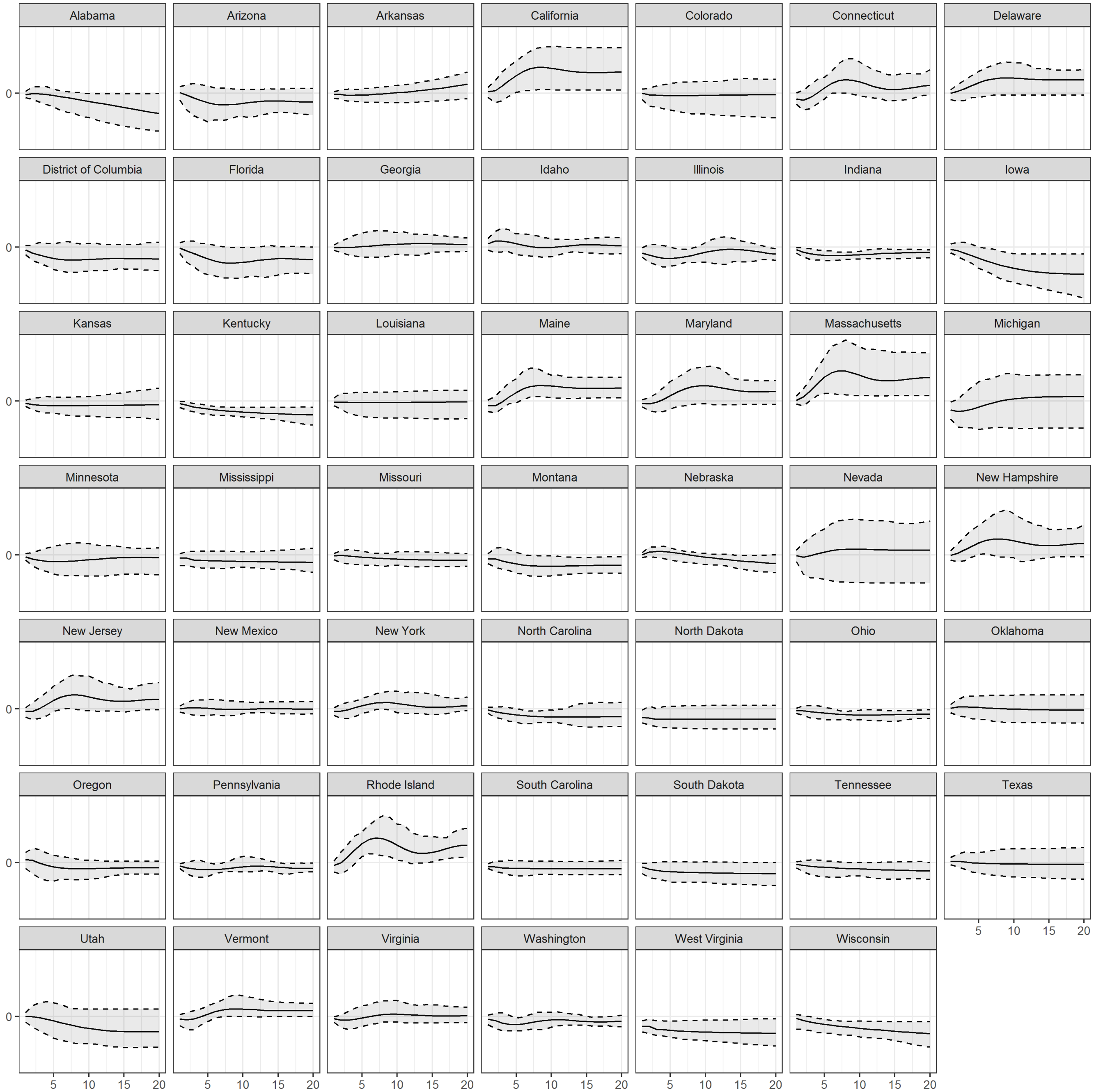


Figure 11: Responses of the real house price index to a shock in the long interest rate

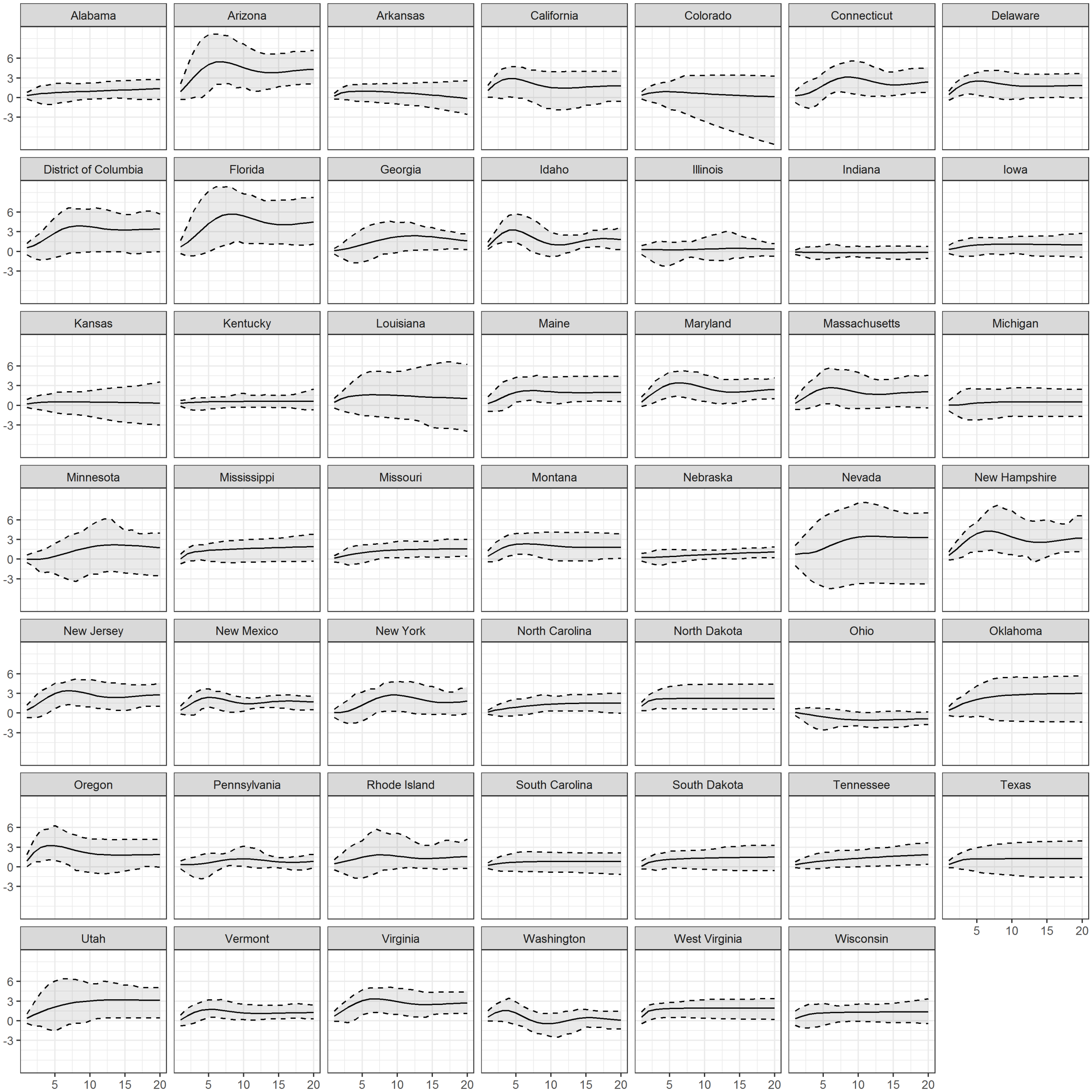


Figure 12: Responses of the real house price index to a shock in real building costs

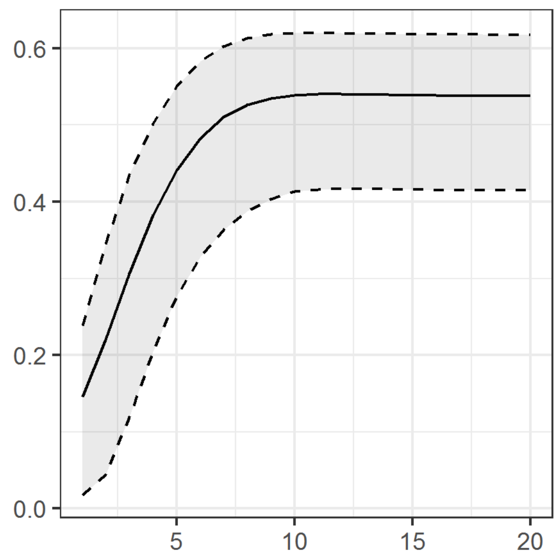


Figure 13: (Panel) responses of the real house price index to a shock in the real income per capita

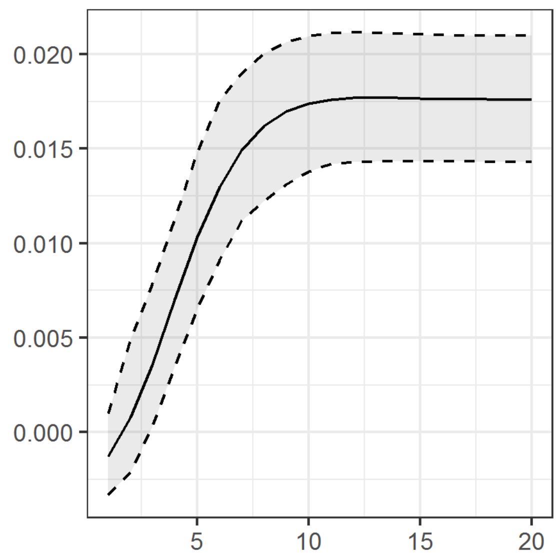


Figure 14: (Panel) responses of the real house price index to a shock in the long rate

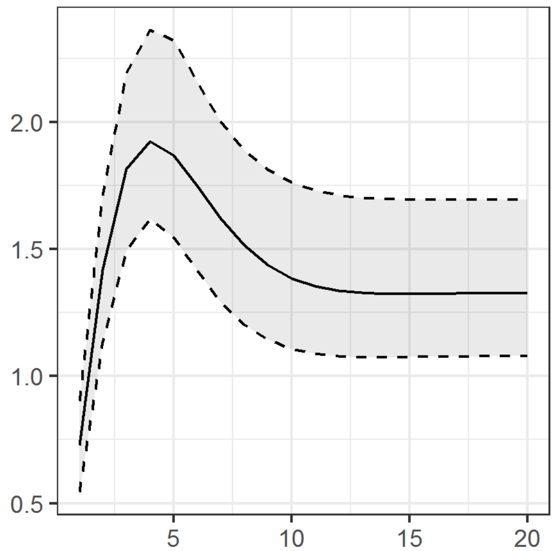


Figure 15: (Panel) responses of the real house price index to a shock in the real building cost index