Three Essays on House Prices: Stationarity, Dynamics, and Expectations

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Jing Zhang, M.A.

Graduate Program in Economics

The Ohio State University

2014

Dissertation Committee:

Donald Haurin, Co-Advisor
Robert de Jong, Co-Advisor
Stephen Cosslett
Lucia Dunn
Abstract

My research focuses on house prices. I study whether house prices are stationary or have a unit root. I then develop and estimate a model to explain the stationarity property of house prices as well as the observed short-run momentum and long-run reversion of house price changes. In the third essay, I examine the unbiasedness and efficiency of forecasters’ expectations of house price changes.

The first essay “Are Real House Prices Stationary? Evidence from New Panel and Univariate Data” questions the common treatment in the housing literature that the logarithm of real house price is a unit root process. Those papers study the cointegration relationship of the logarithm of real house price and economic fundamental variables such as the logarithm of income, and apply the error correction specification for modeling and forecasting real house prices. My study argues that the logarithm of real house price is not a unit root process. Instead, the evidence from a 120-year national dataset and metro area level and state level panel data sets point towards trend stationarity with structural breaks. I also find that the apparent reason that the most cited papers in the literature do not reject a unit root is that they do not include the most recent house price data, which includes the post-2006 house price bust. One result of this conclusion is that the
validity of analyses of house prices based on cointegration and error correction models is questioned.

The second essay “House Price Dynamics” develops and estimates a model that explains the “stationarity-puzzle” discovered in the first essay. The puzzle is that real house prices are trend stationary while real income, one of the most important factors identified in the literature as a determinant of real house prices, contains a unit root. Therefore, the challenge for a coherent house price model is to combine the different stationarity properties of house prices and income. Another phenomenon that has not been well explained in the literature is the short-run positive serial correlation (positive one-year autoregressive coefficient) and long-run mean reversion (negative five-year autoregressive coefficient) of house price changes. This essay presents a dynamic spatial equilibrium model to explain the above phenomena. I solve the model under both rational expectations and adaptive expectations. Separating Metropolitan Statistical Areas into a coastal group and an inland group, I estimate the model’s parameters. Next, house price series are simulated and the autoregressive coefficients of one-year and five-year changes as well as the unit root test statistic are calculated. The model with adaptive expectations fits the empirical features better than the one that assumes rational expectations: it generates a positive one-year autoregressive coefficient, a negative five-year autoregressive coefficient, and the unit root test rejects a unit root.

The third essay “House Price Expectations: Unbiasedness and Efficiency of Forecasters” is the first paper to rigorously test the rationality of expectations of house price changes using survey data. House price expectations are a critical
determinant of housing demand and supply; however, the level of understanding how house price expectations are formed is quite limited. Using a panel dataset of economic forecasters surveyed from 2007 through 2012, I test for unbiasedness and efficiency by implementing the econometric methodology in Davies and Lahiri (1995). Their method implements these tests for a “three-dimensional” panel dataset. I find that, after controlling for aggregate shocks, 9 out of the 47 forecasters have statistically significant biases, and their biases are all negative, indicating that they persistently predicted too high of a change in house prices. The hypothesis of efficiency cannot be rejected. When the year 2012 is excluded, the unbiasedness test shows that 25 out of the 47 forecasters systematically over predicted house price changes during the housing bust. Again, the hypothesis of efficiency cannot be rejected.
Dedicated to my beloved parents.
Acknowledgments

I would like to express my deepest gratitude and respect to my two advisors, Donald Haurin and Robert de Jong, for their intellectual inspiration, trustworthy advice, tremendous kindness, and unbelievable patience in guiding me through the writing of this thesis.

Don was an amazing real estate and urban economist. He always provided insightful advice that helped me obtain a much deeper understanding of housing and urban economics. Don was also very generous with his time and was always available to discuss my research, which greatly improved my papers. I also benefited a lot from his instinct of the important issues in housing, and this inspired me on my research.

Robert was a great econometrician who lightened the darkness when I encountered a time series problem in writing my essays. Discussing ideas with him made my logic become much clearer. I would not be able to have this dissertation without his intelligence and kindness. I was also deeply infected by his enthusiasm about research, and this made the doctoral study a joyful journey.

I am also most grateful to my committee members, Stephen Cosslett and Lucia Dunn, for their constructive guidance, generous support and inspired suggestions on my work. Their shrewd insight and valuable suggestions have greatly helped to improve my research.
I would also like to express my warmest thanks to faculty, staff, and graduate students in the Economics Department for their assistance, support and friendship during the course of my doctoral studies.

Finally, and the most importantly, I devote my deepest gratitude to my parents. They have always been the foundation of my life, and their love and support carried me throughout the years of my study.
Vita

2006 ................................. B.A. Economics, Wuhan University, China
2006 ................................. B.S. Mathematics, Wuhan University, China
2008 ................................. M.A. Economics, Wuhan University, China
2009 ................................. M.A. Economics, The Ohio State University
2009-present ......................... Graduate Teaching Associate, The Ohio State University

Fields of Study

Major Field: Economics
# Table of Contents

Abstract ................................................................. ii  
Dedication ................................................................. v  
Acknowledgements ..................................................... vi  
Vita ................................................................. viii  
List of Tables ........................................................... xii  
List of Figures ........................................................... xiv  
Chapter 1: Are Real House Prices Stationary? Evidence from New Panel and Univariate Data ........................................................... 1  
  1.1 Introduction ..................................................... 1  
  1.2 Literature Review .............................................. 3  
  1.3 Evidence from Shiller’s 120-Year National Real Home Price Index ...... 8  
    1.3.1 Data ......................................................... 8  
    1.3.2 Methodology: Univariate Unit Root Tests ......................... 9  
    1.3.3 Empirical Results ........................................... 11  
  1.4 Evidence from MSA and State Level Panel Data ....................... 13  
    1.4.1 Data ......................................................... 13  
    1.4.2 Methodology: Panel Unit Root Test .............................. 13  
    1.4.3 Empirical Results .......................................... 15  
    1.4.4 Panel Unit Root Tests for Samples Studied in the Literature ...... 17
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4.5 Implications of the Results</td>
<td>20</td>
</tr>
<tr>
<td>1.5 Conclusions</td>
<td>21</td>
</tr>
<tr>
<td>1.6 Figures and Tables</td>
<td>24</td>
</tr>
<tr>
<td>Chapter 2: House Price Dynamics</td>
<td>38</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>38</td>
</tr>
<tr>
<td>2.2 Related Literature</td>
<td>44</td>
</tr>
<tr>
<td>2.3 Model</td>
<td>50</td>
</tr>
<tr>
<td>2.3.1 Setup</td>
<td>50</td>
</tr>
<tr>
<td>2.3.2 Trend-adjusted Equations</td>
<td>56</td>
</tr>
<tr>
<td>2.3.3 Rational Expectations</td>
<td>56</td>
</tr>
<tr>
<td>2.3.4 Adaptive Expectations</td>
<td>58</td>
</tr>
<tr>
<td>2.3.5 Impulse Response</td>
<td>60</td>
</tr>
<tr>
<td>2.3.6 Implications of the Impulse Response Analyses</td>
<td>62</td>
</tr>
<tr>
<td>2.4 Estimation</td>
<td>63</td>
</tr>
<tr>
<td>2.4.1 Data</td>
<td>63</td>
</tr>
<tr>
<td>2.4.2 Methodology</td>
<td>64</td>
</tr>
<tr>
<td>2.4.3 Moment Conditions under Rational Expectations</td>
<td>66</td>
</tr>
<tr>
<td>2.4.4 Moment Conditions under Adaptive Expectations</td>
<td>67</td>
</tr>
<tr>
<td>2.4.5 Estimation Results</td>
<td>68</td>
</tr>
<tr>
<td>2.5 Simulation</td>
<td>69</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
<td>71</td>
</tr>
<tr>
<td>2.7 Figures and Tables</td>
<td>73</td>
</tr>
</tbody>
</table>

Chapter 3: House Price Expectations: Unbiasedness and Efficiency of Forecasters

83
3.1 Introduction ................................................................................. 83
3.2 Data ......................................................................................... 85
3.3 Econometrics Methodology ......................................................... 86
    3.3.1 Forecast Error Covariance Matrix ........................................ 86
    3.3.2 Test for Unbiasedness .......................................................... 90
    3.3.3 Test for Efficiency ............................................................... 90
3.4 Empirical Results .......................................................................... 92
    3.4.1 Unbiasedness ...................................................................... 93
    3.4.2 Efficiency ........................................................................... 94
    3.4.3 Excluding Year 2012 ............................................................ 96
3.5 Conclusions ................................................................................. 97
3.6 Figures and Tables ........................................................................ 98

Bibliography ....................................................................................... 107

Appendices ......................................................................................... 113
A. Deterministic trends and trend-adjusted equations ......................... 113
B. Blanchard and Kahn (1980) .......................................................... 115
C. Proof of Proposition 1 ................................................................. 117
D. Proof of Proposition 2 and derivation of the stability condition ........ 120
E. Moment conditions for rational expectations ................................. 125
F. Moment conditions for adaptive expectations ............................... 128
G. House Price Expectation Surveys ............................................... 130
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Lee and Strazicich (2003) minimum LM endogenous two breaks unit root test for Shiller's 120-year log real home price index</td>
<td>27</td>
</tr>
<tr>
<td>1.2</td>
<td>Carrion-I-Silvestre, Kim, and Perron (2009) three breaks unit root tests for Shiller's 120-year log real home price index</td>
<td>28</td>
</tr>
<tr>
<td>1.3</td>
<td>ADF and KPSS unit root test for Shiller's 120-year log real home price index</td>
<td>29</td>
</tr>
<tr>
<td>1.4</td>
<td>Summary of unit root test results for log of Shiller's real home price index</td>
<td>30</td>
</tr>
<tr>
<td>1.5</td>
<td>Pesaran (2004) cross-section dependence test (CD test) for log real house prices, 1975-2011</td>
<td>31</td>
</tr>
<tr>
<td>1.6</td>
<td>Pesaran (2007) CIPS panel unit root test for log real house prices</td>
<td>32</td>
</tr>
<tr>
<td>1.7</td>
<td>Panel unit root test for log real house prices of the MSAs in Malpezzi (1999)</td>
<td>33</td>
</tr>
<tr>
<td>1.8</td>
<td>Panel unit root test for log real house prices of the MSAs in Gallin (2006)</td>
<td>34</td>
</tr>
<tr>
<td>1.9</td>
<td>Panel unit root test for log real house prices of the MSAs in Mikhed and Zemcik (2009)</td>
<td>35</td>
</tr>
<tr>
<td>1.10</td>
<td>Panel unit root test for log real house prices of the states in Holly, Pesaran, and Yamagata (2010)</td>
<td>36</td>
</tr>
<tr>
<td>1.11</td>
<td>Pesaran (2007) CIPS panel unit root test for log real per capita income</td>
<td>37</td>
</tr>
</tbody>
</table>
2.1 CIPS panel unit root test, 1975-2010 ........................................ 77
2.2 Summary statistics .......................................................... 78
2.3 Income parameters .......................................................... 79
2.4 Rational expectations ....................................................... 80
2.5 Adaptive expectations ....................................................... 81
2.6 Simulation: autocorrelation of house price changes over 1-year and
               5-year and unit root test for house prices ............................. 82
3.1 Test for Unbiasedness, 2007-2012 ........................................ 99
3.2 Test for Unbiasedness, 2007-2012 II ..................................... 100
3.3 Test for Efficiency, 2007-2012 ........................................... 101
3.4 Joint Test for Efficiency, 2007-2012 ..................................... 102
3.5 Test for Unbiasedness, 2007-2011 ........................................ 103
3.6 Test for Unbiasedness, 2007-2011 II ..................................... 104
3.7 Test for Efficiency, 2007-2011 ............................................ 105
3.8 Joint Test for Efficiency, 2007-2011 ..................................... 106
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Log of Shiller's Real Home Price Index, 1890-2011</td>
<td>24</td>
</tr>
<tr>
<td>1.2</td>
<td>Average log real house price index, 1975-2011</td>
<td>25</td>
</tr>
<tr>
<td>1.3</td>
<td>Average log real per capita personal income, 1975-2009</td>
<td>26</td>
</tr>
<tr>
<td>2.1</td>
<td>U.S. log real per capita income and log real house price index, 1975-2010</td>
<td>73</td>
</tr>
<tr>
<td>2.2</td>
<td>Log of real house prices, 1975-2010</td>
<td>74</td>
</tr>
<tr>
<td>2.3</td>
<td>Impulse response to an income shock under rational expectations</td>
<td>75</td>
</tr>
<tr>
<td>2.4</td>
<td>Impulse response to an income shock under adaptive expectations</td>
<td>76</td>
</tr>
<tr>
<td>3.1</td>
<td>Box-and-Whisker plots for each target year</td>
<td>98</td>
</tr>
</tbody>
</table>
Chapter 1: Are Real House Prices Stationary? Evidence from New Panel and Univariate Data

1.1 Introduction

Understanding real house price dynamics is an important issue because the housing market is an important economic sector. Many papers in the housing literature apply cointegration analysis and error correction specifications for modeling real house prices (Abraham and Hendershott (1996), Maplezzi (1999), Capozza, Hendershott, and Mack (2004), Gallin (2006), Mikhed and Zemcik (2009), and Holly, Pesaran, and Yamagata (2010)). For cointegration and error correction analysis to be valid, real house price should be a unit root process.

In a regression equation the stationarity of the variables ensures that hypothesis tests are valid. If a series is not stationary but trend-stationary, it can be transformed into a stationary series by subtracting the deterministic trend. If a nonstationary series has a unit root, a stationary series can be generated by differencing. Moreover, in a context with multiple series with unit roots, if they have a long run equilibrium relationship (or more formally, if there exists a linear

\footnote{Following conventions in the literature, all series under consideration in this paper are in logarithms, and we take statements such as “unit root in real house prices” to mean “unit root in the log of real house prices.”}
combination among them that is stationary), they are said to be cointegrated. In the cointegration case, variables adjust to discrepancies from the long-run relationship, and hence an error correction specification is appropriate to capture the impact of the deviation from the long-run equilibrium on the short-run dynamics. Therefore, a prerequisite for applying the error correction model is the existence of a cointegrating relationship, and a prerequisite for the cointegration analysis is that the variables contain unit roots.

Determining whether there is a unit root in real house prices also sheds light on the appropriateness of theoretical urban models that explain real house prices. If real income has a unit root and real house price is trend stationary as our results suggest, then the models such as the one by Capozza and Helsley (1989, 1990) that suggest an equilibrium relationship between real house price and real income are puzzling.

Our paper studies a fundamental and important question: do real house prices contain a unit root? Surprisingly, the literature has not taken a careful look at this question. To answer the question of whether real house prices have a unit root, we apply unit root tests to national data, to state level panel data, and to MSA (Metropolitan Statistical Area) panel data. We first study a national real home price index over 120 years constructed by Robert Shiller. This is a long time series that is of particular interest because the literature typically uses data sets that begin around 1975. A central feature of the data is that there appear to be structural breaks, and therefore we apply unit root tests that explicitly allow for such breaks. Whether there is a national housing market has been debated and thus we next
apply the Pesaran (2007) panel unit root test to a data set of 48 states and Wash-
ington, DC and to one that contains 363 MSAs for the 1975-2011 period. Because
areas with an inelastic supply of housing usually have more severe house price
cycles, we conduct separate tests in set of MSAs where supply is most inelastic
and most elastic.

To understand why our results differ from results found in the literature, we
restrict our data to the areas and years in the samples used in the most cited pa-
ers that apply panel unit root tests and cointegration tests to real house prices.
We find that these paper's results regarding whether real house price has a unit
root change when the sample period is extended to the most recent data.

The remainder of this paper is organized as follows. The next section dis-
cusses the related literature. The next section provides unit root tests results for
Shiller's 120-year home price index. Then we report the panel unit root test re-
sults for the state and MSA panels as well as for supply elastic and inelastic sub-
groups. Next, we return to the literature and determine what happens to the
results in the literature when the time period of analysis is extended to include
recent data. Last, we discuss the implications of our results for theoretical house
price models.

1.2 Literature Review

Error correction models are widely used in the housing literature to model
house price dynamics. Examples include Abraham and Hendershott (1996), Maplezzi
(1999), and Capozza, Hendershott, and Mack (2004). Abraham and Hendershott
(1996) use annual data for 30 metropolitan areas over the period 1977-1992 to
investigate the determinants of real house price appreciation. The explanatory variables consist of three parts: the change in fundamental price, the lagged real house price appreciation with its coefficient called the “bubble builder,” and the deviation of house price from its equilibrium level in the previous period with its coefficient called the “bubble burster.” They find a positive bubble builder coefficient and a negative bubble burster coefficient. Moreover, the absolute values of these coefficients are higher for the coastal city group than for the inland city group.

Capozza, Hendershott, and Mack (2004) empirically study which variables explain the significant difference in the geographic patterns of the bubble builder and the bubble burster coefficients found in Abraham and Hendershott’s (1996) error correction specification. Using a panel dataset of 62 MSAs from 1979-1995, they find that higher real income growth, higher population growth, and a higher level of real construction costs increase the bubble builder coefficient, while higher real income growth, a larger population, and a lower level of real construction costs increase the absolute value of the bubble burster coefficient.

Abraham and Hendershott (1996) and Capozza, Hendershott, and Mack (2004) do not provide formal unit root and cointegration tests results to justify the existence of a long-run equilibrium relationship among house prices and fundamental variables, which should be a prerequisite for the validity of error correction models.

Malpezzi (1999) uses a dataset which includes 133 MSAs and covers 18 years from 1979 through 1996 and states that short run real house price changes are well modeled by an error correction formulation. The panel unit root test of
Levin, Lin, and Chu (2002, LLC test) is applied to real house price changes, the house-price-to-income ratio, and the residuals of the regression of real house prices on real per capita incomes. The first two are the dependent variables in Malpezzi’s error correction model. A unit root is rejected for price changes, but cannot be rejected for the price-to-income ratio. Moreover, a unit root is rejected for the residuals of the regression of real house prices on real per capita incomes, and hence Malpezzi concludes that real house prices and real incomes are cointegrated. This cointegration test procedure suffers from several shortcomings. First, before applying the cointegration test, Malpezzi does not examine if real house prices and income have a unit root, respectively. Second, critical values of the LLC panel unit root test have not been shown to work for residuals from the first stage regression, so the claim that house prices and incomes are cointegrated based on the LLC critical values could be misleading. Third, the LLC test does not allow cross-sectional dependence in the regression errors, hence the test result may be biased.

The 2000s’ housing boom, which reached its peak in 2006, raises the question of whether real house prices are supported by fundamentals; that is, whether real house prices and the fundamental economic variables such as income have a long run equilibrium relationship. The housing literature formalizes this argument by discussing the cointegration relationship among real house prices and the fundamental variables. Thus far, the results are mixed. Some papers such as Gallin (2006) and Mikhed and Zemcik (2009) apply cointegration tests and claim that there is no long run equilibrium relationship, which cast doubts on the validity of applying error correction models to real house prices. Other papers argue
that there is a cointegration relationship, such as Holly, Pesaran, and Yamagata (2010).

Gallin (2006) tests for the existence of a long run relationship among house prices and economic fundamental variables by applying cointegration tests to both national level data and city level panel data. The augmented Engle-Granger cointegration test is applied to national level house prices, per capita income, population, construction wage, user cost of housing, and the Standard and Poor's 500 stock index. No cointegration relationship is found. He also applies panel cointegration tests to city level house prices, per capita income, and population, for 95 MSAs over 23 years from 1978 to 2000. The panel cointegration tests he uses are Pedroni (1999) and Maddala and Wu (1999). He also applies a bootstrapped version of the tests to take into account cross-sectional dependence. The null hypothesis of no cointegration cannot be rejected, neither by the original tests nor the bootstrapped version. To test for a unit root, Gallin applies the ADF unit root test to the national level real house prices and a unit root is not rejected. But he does not provide panel unit root test result for the city level panel data.

Mikhed and Zemcik (2009) also examines if house price and fundamental factors are cointegrated. The innovation in their paper is that they include more fundamental variables to avoid the possibility that the omission of potential demand and supply shifters cause the lack of cointegration relationships. The fundamentals included are house rent, a building cost index, per capita income, population, mortgage rate, and the Standard and Poor's 500 stock index. Their sample includes 22 MSAs over 1978-2007, and they examine several different time periods (1978-2007, 1978-2006, 1978-2005, 1997-2007, 1978-1996). They apply the
CIPS panel unit root tests to real house prices and the fundamental variables for all periods, setting the time lag in the CIPS test to one year. For real house prices, a unit root is rejected at the 5% level for 1978-2007 and rejected at the 10% level for 1978-2006, but cannot be rejected for other periods. The authors interpret this as a correction of the house price bubble around 2006. They further investigate the cointegration relationship of house prices and fundamentals for periods prior to 2005 when a unit root cannot be rejected in house prices. They apply the Pedroni (1999, 2004) panel cointegration test and bootstrap the critical values for possible cross-sectional dependence. No evidence of cointegration relationships is found in any of the cases. Hence they claim that real house price dynamics are not explained by fundamentals, and this is evidence of housing bubbles in those subsamples.

Holly, Pesaran, and Yamagata (2010) study the determination of real house prices using a panel of state level data (48 states, excluding Alaska and Hawaii and they include the District of Columbia) over 29 years from 1975-2003. Unlike Gallin (2006) and Mikhed and Zemcik (2009), Holly et al. find that real house prices and real income are cointegrated. The innovation of Holly et al. is that they apply the common correlated effects (CCE) estimators of Pesaran (2006) to study a panel of real house prices. They first use an asset non-arbitrage model to show that the ratio of real house price and real income should be stationary, hence the log of real house price and the log of real income should be cointegrated with a cointegration vector of (1, -1). Then they empirically find such a cointegration relationship and estimate a panel error correction model for the dynamics of adjustment of real house prices to real incomes.
Most of the above papers use time periods that end before 2006 and thus they do not include the recent housing bust period in their sample. An exception is Mikhed and Zemcik (2009), which reports unit root tests for periods ending after 2006, and a unit root is rejected at the 5% level over 1978-2007 for real house prices. The authors then apply cointegration tests to the subperiods ending before 2006 where a unit root is not rejected. This is an ad hoc solution that does not address the issue of whether real house prices have a unit root or are stationary.

1.3 Evidence from Shiller’s 120-Year National Real Home Price Index

1.3.1 Data

The periods examined in the literature usually start at or after 1975 because reliable house price data are available since then, both for national and regional house prices. A problem is that a 35 year period may be too short to draw reliable conclusions about the time series properties of house prices. Another problem is that if the period of the recent substantial house price run up and decline is omitted, then only 23 years remain to be studied.

For the above reasons, we study the U.S. real home price index constructed by Shiller (2005).\[^2\] It is a national level index that begins in 1890. Shiller constructed this index “by linking together various available series that were designed to provide estimates of the price of a standard, unchanging, house…” He also created an index for the period 1934-1953. Before 1953 there are only annual data, and therefore we use the logarithm of the annual data for this analysis. Figure 1 shows

the path of this index. Real house prices appear to be relatively stable except for
the period during the two world wars and the Great Depression, and a sharp spike
during the 2000s.

1.3.2 Methodology: Univariate Unit Root Tests

We apply the Lee and Strazicich (2003) (L-S) minimum LM endogenous two
breaks unit root test, the multiple breaks unit root tests in Carrion-I-Silvestre, Kim, and Perron (2009) (CKP), the Augmented Dickey-Fuller (ADF) unit root test, and the Kwiatkowski-Phillips- Schmidt-Shin (KPSS) test to the 120 year time se-
ries of house prices.

Because two breaks clearly appear in Figure 1, with one break around World
War I (1914-1918) and the second around World War II (1939-1945), we apply the
Lee-Strazicich minimum LM endogenous two structural breaks unit root test. The L-S test estimates the points of structural break by searching for the two
break points where the unit root t-statistic is minimized, and tests the null of a
unit root against the alternative of trend stationarity. It allows for breaks under
both the null and alternative hypotheses. If breaks are not included in the null,
as in Lumsdaine and Papell (1997), then a rejection of the null may imply a unit
root process with breaks and not necessarily a trend stationary series. Model C in
the Lee and Strazicich paper is adopted here, which is the most general case that
allows for changes in both level and slope.

We apply unit root test statistic in Lee and Strazicich (2003), which is obtained
from the following regression:

\[ \Delta y_t = \delta' Z_t + \phi \tilde{S}_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta \tilde{S}_{t-i} + u_t, \]  

(1.1)
where \( \tilde{S}_t = y_t - \tilde{\Psi}_x - Z_t \delta, \ t = 2, \ldots, T; \tilde{\Psi}_x = y_1 - Z_1 \delta; \delta \) are the regression coefficients from a regression of \( \Delta y_t \) on \( \Delta Z_t; Z_t = [1, t, D_{1t}, D_{2t}, DT_{1t}, DT_{2t}]' \), where \( D_{jt} = 1 \) and \( DT_{jt} = t - T_{Bj} \) for \( t \geq T_{Bj} + 1 \) and 0 otherwise, \( j = 1, 2 \), and \( T_{Bj} \) is the time of a break point. The lagged terms \( \Delta \tilde{S}_{t-i} \) are included to correct for serial correlation, and the number of lags \( k \) is selected following the general-to-specific procedure described in their paper. The LM test statistic is given by

\[
\tilde{\tau} = t \text{ statistic testing the null hypothesis } \phi = 0.
\] (1.2)

The two breaks \( (\lambda_j = T_{Bj}/T, \ j = 1, 2) \) are determined endogenously by a grid search over the time span \([0.1T, 0.9T]\) (to avoid end-point issues).

Because of the potential existence of three breaks during 1890-2011 due to the two world wars and the 2000s’ housing bubble, we also apply Carrion-i-Silvestre, Kim, and Perrons (2009) unit root tests that allow for multiple structural breaks. They extend the unit root tests of Elliott, Rothenberg, and Stock (1996) and Ng and Perron (2001) to allow for multiple breaks (we use their tests with three breaks), and provide five test statistics \((P_G^{GLS}, MZ^{GLS}, MSB^{GLS}, MZ_I^{GLS}, MP_T^{GLS})\) to test the null of a unit root against the alternative of trend stationarity. The tests considered in Carrion-I-Silvestre et al. (2009) estimate the break points, allow for breaks under both the null and the alternative hypotheses, and allow for breaks in both the level and the slope.

We also apply the ADF and KPSS tests because they are popular in the literature that tests the stationarity of a series. The ADF test tests the null of a unit root against the alternative of trend stationarity, while the KPSS test tests the null of trend stationarity against the alternative of a unit root.
1.3.3 Empirical Results

Table 1.1 reports the results of the Lee and Strazicich (2003) two breaks test. A unit root cannot be rejected for the period 1890-2011. Break points are found in 1916 and 1949. These values closely correspond to the visual inspection of Figure 1.1. However, it is possible that a third break occurs during the most recent housing bubble, hence biasing the test towards non-rejection of a unit root. If the period of the recent housing bubble is excluded, a unit root is rejected at the 1% level over 1890-1998. The estimated break points are 1916 and 1947 for this time span. We also apply this two-break test to the 1920-2011 period to avoid the structural break at World War I, thus allowing for the break at World War II and a possible break during the recent housing bubble. A unit root again is rejected at the 1% level. The estimated break points are 1946 and 2000. Therefore, a unit root is rejected for both 1890-1998 and 1920-2011.

Table 1.2 reports the tests statistics for three breaks in Carrion-i-Silvestre, Kim, and Perron (2009), for different numbers of lags. The estimated break points are 1917, 1944, and 1998, which are similar to the above results. The rejection of a unit root is quite robust to the number of lags. None of the statistics can reject a unit root when the number of lags equals zero; three of them reject a unit root at the 10% significance level when one lag is included; all of them reject a unit root at the 5% or 1% level when more than one lag is included. The selected number of lags using the Bayesian Information Criterion (BIC) is two. Therefore, the three

3The selected lag length is zero using the Modified Akaike Information Criterion (MAIC) by Ng and Perron (2001) and Perron and Qu (2007). Because real house prices and real house price changes are documented in the literature as having strong serial correlation, we think the number of lags selected based on BIC is more reliable.
breaks tests provide strong evidence of trend stationarity with three breaks over 1890-2011.

Results for the ADF and the KPSS tests are reported in Table 1.3. For the period 1890-2011, the ADF test cannot reject a unit root and the KPSS test rejects stationarity, which is not surprising because ignoring the possibility of structural breaks will bias these test statistics towards the unit root hypothesis. To avoid the effects of dramatic changes due to the world wars and the Great Depression, we restrict the test period to 1950-2011, and the unit root hypothesis is rejected at the 5% level by the ADF test, suggesting that real house price is trend stationary during this 60-year period. Moreover, to examine the possible effect of the recent bubble on the unit root test results, we examine the period 1950-2006. The ADF test cannot reject a unit root and the KPSS test rejects trend stationarity at the 5% level. When we look at the series that ends at 1998 (excluding the recent bubble), the ADF test rejects a unit root at the 1% level and KPSS test cannot reject trend stationarity, implying strong evidence of trend stationarity over 1950-1998. Therefore, including or excluding the entire recent bubble-bust cycle results in trend stationarity, but including only the boom period of this bubble will result in a unit root conclusion. To further emphasize the potential problem in the literature when only a subperiod is studied, we apply the ADF test and the KPSS test to periods starting at 1975 for both annual data and quarterly data. A unit root is weakly rejected at the 10% level for 1975-2011, but cannot be rejected for

As a robustness check, we also apply unit root tests to the period 1950-1989 and 1950-1982, which respectively ends at the peak and the start of the housing boom prior to the 2000s bubble, and find evidence of trend stationarity for both periods: a unit root is rejected by the ADF test at the 5% level and the 1% level respectively. Therefore, of the examined five periods since 1950, 1950-2006 is the only one for which we cannot reject a unit root.
1975-2006 and 1975-2002. This result coincides with the findings in the literature which generally claim that real house prices have a unit root.

Table 1.4 presents a summary of the findings of the ADF test, the L-S two breaks test and the CKP three breaks test for different time periods. Our preferred method is the CKP three breaks test for the full sample period, 1890-2011, resulting in a conclusion that house prices are trend stationary with breaks.

1.4 Evidence from MSA and State Level Panel Data

1.4.1 Data

We consider two panel data sets: the first includes house prices for 363 MSAs in the U.S. and the second reports data for 48 states and Washington, D.C. Both data sets cover the 1975-2011 period. We use the Freddie Mac House Price Index (FMHPI), which is a monthly nominal repeated-sales index estimated using data on house price transactions (including refinanced) on one-family detached and townhome properties whose mortgage has been purchased by Freddie Mac or Fannie Mae.\(^5\) We deflate the nominal house price index by the CPI-U for each month and then average over a year to obtain annual data and then take logarithms.\(^6\)

1.4.2 Methodology: Panel Unit Root Test

In the literature, early specifications of panel unit root tests did not allow for cross-sectional dependence (Levin, Lin and Chu 2002; Im, Pesaran and Shin

\(^5\)The FMHPI series can be found at \url{http://www.freddiemac.com/finance/fmhpi/}

\(^6\)The CPI-U series can be found at \url{ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai}.
2003). More recently, panel unit root tests allow for cross-sectional dependence (Moon and Perron 2004; Bai and Ng 2004; Pesaran 2007).

We apply the Pesaran CIPS test (2007). There are several reasons for this choice. First, the CIPS test allows for cross-sectional dependence, which is an important consideration for house prices because house prices in different areas are very likely to be affected by common effects such as changes in interest rates or technology. Second, it allows for cross-sectional heterogeneity in the intercept, trend, and autoregressive coefficients. The third reason is its favorable size and power for large N and small T, which is the case for our sample. In particular, Moon and Perron (2004) and Bai and Ng (2004) both assume that \( N/T \to 0 \) as \( N \) and \( T \to \infty \) when deriving the asymptotic properties of the test, while Pesaran (2007) assumes \( N/T \to k \) (where \( k \) is a finite positive constant) as \( N \) and \( T \to \infty \).

To justify the use of Pesaran’s (2007) CIPS test, we need to establish the presence of cross-sectional dependence. We use the CD (cross-section dependence) test proposed in Pesaran (2004) for this purpose. This test statistic asymptotically converges in distribution to a standard normal distribution as \( T \) and \( N \to \infty \) in any order under the null of no cross-sectional dependence, and is defined as

\[
CD = \sqrt{\frac{2T}{N(N-1)}} \left( \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right), \\
(1.3)
\]

where \( \hat{\rho}_{ij} = Corr(\hat{\epsilon}_i, \hat{\epsilon}_j) \) and \( \hat{\epsilon}_i \) is the estimated residual from the ADF(\(p\)) regression that estimates the model

\[
\Delta y_{it} = a_{i0} + a_{i1}t + a_{i2}y_{i,t-1} + \sum_{j=1}^{p} \delta_{ij} \Delta y_{i,t-j} + \epsilon_{it}, \\
(1.4)
\]

Pesaran (2007)’s CIPS panel unit root test is based on a cross-sectionally augmented ADF (CADF) regression, which filters out the cross-sectional dependence
by augmenting the ADF regressions with the lagged cross-section mean and the
lagged first differences of the cross-sectional mean. The CADF regression esti-
mates the model

\[ \Delta y_{it} = a_{i0} + a_{i1} t + a_{i2} y_{i,t-1} + a_{i3} \bar{y}_{t-1} + \sum_{j=0}^{p} d_{ij} \Delta \bar{y}_{t-j} + \sum_{j=1}^{p} \delta_{ij} \Delta y_{i,t-j} + v_{it}. \]  

(1.5)

Let \( \tilde{t}_i \) denote the t-ratio for \( a_{i2} \) in the above regression. Then the CIPS statistic is

\[ CIPS = N^{-1} \sum_{i=1}^{N} \tilde{t}_i. \]  

(1.6)

The null and alternative hypotheses for the CIPS test are

\[ H_0 : a_{i2} = 0 \text{ for } i = 1, 2, \ldots, N, \]  

(1.7)

and

\[ H_1 : \begin{cases} a_{i2} = 0 & \text{for } i = 1, 2, \ldots, N_1 \\ a_{i2} < 0 & \text{for } i = N_1 + 1, N_1 + 2, \ldots, N. \end{cases} \]  

(1.8)

1.4.3 Empirical Results

To confirm the presence of cross-sectional dependence, we carry out CD tests
using 1, 2 and 3 lags and using both the MSA and state data sets. We also consider
subsamples of the top 20 and 50 most supply inelastic and elastic MSAs. These
MSAs are selected based on the supply elasticity estimated in Saiz (2010), who
provides housing supply elasticity measures for 269 metro areas. After matching
his data with the MSA sample used here, supply elasticity measures for 254 MSAs
are available.

The cross-sectional dependence is confirmed by the CD test statistics in Table
1.5, which are statistically highly significant. This implies that the panel unit root
tests that do not allow for cross-sectional dependence are inappropriate.
Table 1.6 reports the CIPS statistics with an intercept and a linear time trend included for varying augmented orders and for the state and MSA sample and the MSA subsamples. We examine the periods 1975-2011, 1975-2006, 1975-2002, and 1975-1998 to determine if and how the recent housing bubble affects unit root test results. For the 363 MSAs group, the period 1975-2011 strongly rejects a unit root at the 1% level, but a unit root cannot be rejected for all other periods except 1975-1998 which rejects a unit root at the 10% level when the augmented order is one. This result indicates that we can reject a unit root if we examine the period that includes both the boom and bust years of the recent bubble, but a unit root cannot be rejected if we exclude the bust. For the state sample, the periods 1975-2011 and 1975-2006 both reject a unit root, but 1975-2002 and 1975-1998 cannot reject.

The CIPS test statistics for the supply elastic and inelastic MSA groups also are reported in Table 1.6. The recent literature has identified that housing supply elasticity plays an important role in house price dynamics. For example, Glaeser, Gyourko, and Saiz (2008) find, both theoretically and empirically, that places with more elastic housing supply have smaller house price increases and shorter house price bubbles. Also, we can see from Figure 1.2 that the average of the log real house price index for the top 20 supply inelastic MSAs is much more volatile and has much bigger and prolonged cycles than that for the top 20 supply elastic MSAs.

Table 1.6 shows that the top 50 supply inelastic MSA group strongly rejects a unit root over 1975-2011 with the test statistics much higher in absolute value than that for the 363 MSA group, while all other periods cannot reject a unit root.
The top 20 supply inelastic MSA group reveals even stronger evidence for trend stationarity - a unit root is very strongly rejected for all time periods when a one-year lag is selected. In contrast, the supply elastic groups reveal little evidence of trend stationarity. The top 50 supply elastic MSA group cannot reject a unit root at the 5% level for all periods when we select a one-year lag, and the top 20 supply elastic group cannot reject a unit root even at the 10% level. Therefore, it appears that the test statistics are impacted greatly by the house supply elasticity of the areas under consideration.

1.4.4 Panel Unit Root Tests for Samples Studied in the Literature

Our panel unit root test results indicate trend stationarity if the data extends through 2011. This result differs from many studies in the literature, which find that real house prices have a unit root process. This raises the question as to the causes of the difference in results. Possible explanations are (1) the different time periods covered, (2) the different areas included, and (3) the different unit root tests that are applied. To answer this question, we apply the CIPS panel unit root test.

7 Pesaran, Smith, and Yamagata (2013) propose a CIPSM test which is an extension of the CIPS test. The CIPS test allows for only one common factor affecting the cross-sectional dependence, while the CIPSM test allows for multiple common factors. In the CIPS test, an additional variable is included when there are two common factors, and the lagged cross-sectional mean and the lagged first differences of the cross-sectional mean of that additional variable are used to filter out the cross-sectional dependence created by the second common factor. Similarly, in the case of three common factors, two additional variables are included in the CIPSM test. We carry out the CIPSM test with lag order being one year, and consider three cases of the additional variables included: income, population, and income plus population. (By income we mean log real per capita income, and population means log population.) For the 363 MSAs group over 1975-2011, a unit root is rejected at the 5% level when income is the additional variable included, but it cannot be rejected if the additional variable(s) included is(are) population or income plus population. For the top 20 supply inelastic MSAs group and the to 20 supply elastic MSAs group, results are very similar to the one common factor CIPS test: the former group strongly rejects a unit root in all the cases of the additional variables included, and the latter group has little evidence of rejecting a unit root.
test to the area samples studied in the most cited papers that examine panel unit root tests and cointegration tests for U.S. real house prices.

Malpezzi (1999)

We apply the CIPS panel unit root test to a subsample that matches the MSAs and years in Malpezzi's paper. Of the 133 MSAs studied by Malpezzi, 124 of them are available in our dataset. The first row of Table 1.7 presents the CIPS statistics. A unit root cannot be rejected over the years studied in Malpezzi's sample 1979-1996, and this result is robust to the varying augmented lags, and thus we replicate his result. The second row of Table 1.7 presents the CIPS statistics for the same group of MSAs over 1975-2011, and a unit root is rejected at the 1% level, regardless of the number of the augmented lags. This dramatic change of the test results is clearly a result of the different time periods available in his and our samples.

Gallin (2006)

Of the 95 MSAs in Gallin's sample, 83 of them can be matched with our data and we apply the CIPS test to this group of MSAs. The first row of Table 1.8 reports the CIPS statistics over 1978-2000, the years examined in Gallin's sample. A unit root is not rejected. But if the data are extended to cover the 1975-2011 period, a unit root is rejected as reported in the second row. The results are robust to changing the number of lags. Therefore, the 1975-2011 period data contain new information that may invalidate the cointegration analysis in Gallin's paper.

Mikhed and Zemcik (2009)
We apply the CIPS test to the 22 MSAs in Mikhed and Zemcik’s paper that our in our sample. The first and second rows of Table 1.9 report the CIPS statistics over 1978-2005 and 1978-2007, their sample periods. Our results are very similar to theirs when one augmented lag is selected (which is the case in their paper). Specifically, a unit root is rejected at the 5% level for 1978-2007, but cannot be rejected for 1978-2005. The third row of Table 1.9 reports the test statistics for the entire period 1975-2011 and a unit root is rejected at the 1% level when the lag is set to one year. The non-rejection of a unit root for periods before 2005 could be an indication of less test power caused by the limited number of time periods. As before, the application of cointegration methodology in Mikhed and Zemcik's paper may well have been inappropriate.

Holly, Pesaran, and Yamagata (2010)

Table 1.10 reports the CIPS statistics using our sample for 48 states and Washington, D.C. The house price data in Holly et al. are the house price index from the Office of Federal Housing Enterprise Oversight and are deflated by a state level consumer price index, while we use the Freddie Mac house price index deflated by the national consumer price index. The first row shows the test statistics for the time period 1975-2003 studied in Holly et al., and a unit root cannot be rejected. This agrees with their unit root test result. However, the second row shows that, if the data are extended to also cover the 2004-2011 period, a unit root is strongly rejected. Again, this change in unit root test results indicates that the cointegration analysis and the error correction model discussed in Holly et al.'s paper may well have been inappropriate.
For each of these studies, we find that a unit root cannot be rejected if we consider samples similar to those in the original study and thus we confirm their results. However, if these samples are extended to 1975-2011, a unit root is always rejected. Therefore, the reason for our disagreement with the literature is the difference in the time span covered in the analysis. In particular, stopping an analysis midway through the 1998-2011 housing cycle yields different results than including the complete cycle.

1.4.5 Implications of the Results

Both the panel unit root tests and the univariate unit root tests provide supportive evidence that real house price is a trend stationary series with structural breaks, rather than a unit root process. This result presents a challenge to urban models that aim to find the determinants of real house prices. As mentioned in the introduction, house price models usually identify real income as one of the most important factors determining real house prices. But our study suggests some income-house-price “puzzles.” Real house prices and real income are very likely two processes with different stationarity properties. Real house prices reject a unit root based on findings in our paper, while there is strong evidence that real income is a unit root process. Table 1.11 reports the CIPS panel unit root test statistics when applied to log of real per capita personal income, for the all-MSA group, top-20-supply-inelastic-MSA group, and top-20-supply-elastic-MSA group, and covering different time periods: 1975-2009, 1975-2006, and 1975-1998. None of the test statistics are significant. How a unit root real

---

8 The per capita personal income is annual data from Bureau of Economic Analysis, and we convert this nominal data to real terms by deflating it by CPI-U.
income process determines a trend stationary real house price is a challenge for urban models.

In addition to the stationarity puzzle, there is also an issue related to house price trends. Figure 1.1 indicates that there is hardly any trend for the national level real home price index, but Figure 1.3 shows that there is an obvious upward trend for real income. A possible explanation is to include housing supply because if housing supply is elastic, house price trends will follow trends in construction costs. As documented in Gyourko and Saiz (2006), national level real construction costs trended down over 1980-2003. Therefore, in supply elastic MSAs, one would expect real house prices to decline even when real income is increasing. In contrast, if housing supply is inelastic, then real house prices will rise as income increases. This explanation is supported by Figure 1.2. Real house prices have an upward trend for the top 20 supply inelastic MSAs, and a downward trend for the top 20 supply elastic MSAs. In contrast, Figure 1.3 shows that, for real income, there are obvious upward trends for all the three groups, and the three paths are almost parallel.

1.5 Conclusions

This paper examines a fundamental and important question regarding the time series properties of real house prices; specifically, do real house prices contain a unit root. This stationarity property is of vital importance for analyzing univariate and panel time series data, modeling and forecasting the dynamics of
the series, and conducting tests such as cointegration. Choosing an inappropriate model due to misunderstanding of whether a unit root exists could invalidate the usual statistical tests, such as t-tests and F-tests.

We first examine Shiller’s national level real home price index over a 120-year period. We apply the multiple breaks unit root tests in Carrion-i-Silvestre, Kim, and Perron (2009) (CKP), the Augmented Dickey-Fuller (ADF) unit root test, and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The CKP three breaks test rejects a unit root for 1890-2011. When the ADF and the KPSS tests are applied to periods beginning at 1950, whether the recent bubble is included or excluded affects the results. Selecting a period that includes the recent house price boom, but not the bust, results in the ADF test being unable to reject a unit root. Moreover, if the period typically used in the literature (starting in 1975) is examined, the ADF test generally cannot reject a unit root at the 5% level. However, these results are reversed if the complete price cycle is included.

Arguably, the housing market is not a single national market, but is composed on state or MSA level markets. Thus, we also conduct the Pesaran (2007) CIPS panel unit root test using samples of states and MSAS. We also conduct this test in supply elastic and inelastic MSAs given that house price cycles differ between them. We find that, first, real house prices are trend stationary if both the boom and bust periods of the recent price bubble are included, both for the 363 MSA group and the state group. But not including the bust period of this bubble results in non-rejection of a unit root for the 363 MSA group. For the state group, a unit root in real house prices can be rejected if the period ends at 2006, but a unit root cannot be rejected if the period ends at 2002. Second, housing supply elasticity
affects unit root test results; specifically, there is stronger evidence of rejecting a unit root in supply inelastic MSAs than elastic MSAs.

We then apply the CIPS panel unit root test to the samples studied in the literature. When our data is restricted to the same areas and time periods, we verify these studies’ results. In nearly all cases the result is that a unit root cannot be rejected. But if the dataset is extended to 2011, a unit root is always rejected at a highly significantly level. Thus, the difference in unit root test results between the literature and this paper occur because of the smaller number of time periods studied in the literature.

The unit root tests results in this study have important implications, both empirically and theoretically. Empirically, if real house prices are indeed trend stationary, then the cointegration analysis and the error correction models widely used in modeling and forecasting house price dynamics would be invalid. Theoretically, a model should be able to explain the following issues. First, the model should explain the relationships between real income, which has a unit root, and real house prices, which does not have a unit root. Second, it should explain that the evidence in favor of trend stationarity in real house prices for supply inelastic cities is stronger than for supply elastic cities. Third, although areas with inelastic and elastic housing supply have similar trends in real income, the former group has an upward trend and the latter group has a negative trend in real house prices.
1.6 Figures and Tables

![Graph of Shiller's Real Home Price Index, 1890-2011](image)

**Figure 1.1:** Log of Shiller's Real Home Price Index, 1890-2011

Note: The construction of Shiller's home price index is described in Shiller (2005).
Figure 1.2: Average log real house price index, 1975-2011

Note: House prices are based on the Freddie Mac house price index. Housing supply elasticity measure is provided in Saiz (2010).
Figure 1.3: Average log real per capita personal income, 1975-2009

Note: Personal income is from the Bureau of Economic Analysis. Housing supply elasticity measure is provided in Saiz (2010).
Table 1.1: Lee and Strazicich (2003) minimum LM endogenous two breaks unit root test for Shiller’s 120-year log real home price index

<table>
<thead>
<tr>
<th>Break points</th>
<th>Test stat.</th>
<th>Lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>1890-2011</td>
<td>1916, 1949</td>
<td>-5.02</td>
</tr>
<tr>
<td>1890-1998</td>
<td>1916, 1947</td>
<td>-6.54***</td>
</tr>
<tr>
<td>1920-2011</td>
<td>1946, 2000</td>
<td>-6.53***</td>
</tr>
</tbody>
</table>

Notes: Model C which allows for changes in both level and trend is adopted here. The number of lags ($k$) is selected following the general-to-specific procedure described in their paper. Critical values of the unit root statistic, which depend upon the location of the breaks, are tabulated in Lee and Strazicich (2003 Table 2). For $T = 100$, the 1%, 5%, 10% critical values are -6.16, -5.59, -5.27 for $\lambda_1 = 0.2$, $\lambda_2 = 0.4$, are -6.41, -5.74, -5.32 for $\lambda_1 = 0.2$, $\lambda_2 = 0.6$, are -6.32, -5.73, -5.32 for $\lambda_1 = 0.6$, $\lambda_2 = 0.8$, and are -6.33, -5.71, -5.33 for $\lambda_1 = 0.2$, $\lambda_2 = 0.8$. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.
Table 1.2: Carrion-I-Silvestre, Kim, and Perron (2009) three breaks unit root tests for Shiller’s 120-year log real home price index

<table>
<thead>
<tr>
<th>Number of lags</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{T}^{GLS}$</td>
<td>12.34</td>
<td>9.58</td>
<td>6.19**</td>
<td>6.14**</td>
<td>4.55***</td>
<td>1.63***</td>
</tr>
<tr>
<td>$MZ_{a}^{GLS}$</td>
<td>-23.77</td>
<td>-30.89*</td>
<td>-48.37***</td>
<td>-48.72***</td>
<td>-66.11***</td>
<td>-186.44***</td>
</tr>
<tr>
<td>$MSB_{T}^{GLS}$</td>
<td>0.14</td>
<td>0.13*</td>
<td>0.10***</td>
<td>0.10***</td>
<td>0.09***</td>
<td>0.05***</td>
</tr>
<tr>
<td>$MZ_{t}^{GLS}$</td>
<td>-3.38</td>
<td>-3.87*</td>
<td>-4.87***</td>
<td>-4.89***</td>
<td>-5.71***</td>
<td>-9.63***</td>
</tr>
<tr>
<td>$MP_{T}^{GLS}$</td>
<td>11.53</td>
<td>8.95</td>
<td>5.78***</td>
<td>5.74***</td>
<td>4.25***</td>
<td>1.52***</td>
</tr>
</tbody>
</table>

Notes: The estimated break points are 1917, 1944 and 1998. Breaks are allowed in both level and trend. The test statistics for different number of lags in CKP(2009) are reported. The 1%, 5% and 10% critical values are 5.78, 7.59 and 8.80 for $P_{T}^{GLS}$, -43.82, -34.31 and -30.01 for $MZ_{a}^{GLS}$, 0.11, 0.12 and 0.13 for $MSB_{T}^{GLS}$, -4.68, -4.14 and -3.87 for $MZ_{t}^{GLS}$, and 5.78, 7.59 and 8.80 for $MP_{T}^{GLS}$. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF t-stat.</td>
<td>-2.46</td>
<td>-3.82**</td>
<td>-1.07</td>
<td>-4.93***</td>
</tr>
<tr>
<td>1%, 5% and 10% critical values</td>
<td>(-4.04, -3.45, -3.15)</td>
<td>(-4.12, -3.49, -3.17)</td>
<td>(-4.14, -3.50, -3.18)</td>
<td>(-4.17, -3.51, -3.18)</td>
</tr>
<tr>
<td>KPSS LM-stat.</td>
<td>0.18**</td>
<td>0.14*</td>
<td>0.17**</td>
<td>0.07</td>
</tr>
<tr>
<td>1%, 5% and 10% critical values</td>
<td>(0.22, 0.15, 0.12)</td>
<td>(0.22, 0.15, 0.12)</td>
<td>(0.22, 0.15, 0.12)</td>
<td>(0.22, 0.15, 0.12)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>annually</td>
<td>quarterly</td>
<td>annually</td>
</tr>
<tr>
<td>ADF t-stat.</td>
<td>-3.48*</td>
<td>-3.35*</td>
</tr>
<tr>
<td>1%, 5% and 10% critical values</td>
<td>(-4.24, -3.54, -3.20)</td>
<td>(-4.03, -3.44, -3.15)</td>
</tr>
<tr>
<td>KPSS LM-stat.</td>
<td>0.09</td>
<td>0.12*</td>
</tr>
<tr>
<td>1%, 5% and 10% critical values</td>
<td>(0.22, 0.15, 0.12)</td>
<td>(0.22, 0.15, 0.12)</td>
</tr>
</tbody>
</table>

Notes: An intercept and a linear time trend are included. The null hypothesis for the Augmented Dickey-Fuller test is that the series has a unit root. The null hypothesis for the Kwiatkowski-Phillips-Schmidt-Shin test is that the series is trend stationary. The number of lags included in the ADF test are based on the Schwarz Info Criterion. For 1950-2006, the SIC selected lag length is 2, and SIC selected lag lengths for all other annual periods are 1. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.
Table 1.4: Summary of unit root test results for log of Shiller's real home price index

<table>
<thead>
<tr>
<th>Period</th>
<th>ADF</th>
<th>L-S two breaks</th>
<th>CKP three breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1890-2011</td>
<td>unit root</td>
<td>unit root</td>
<td>trend stationary</td>
</tr>
<tr>
<td>1890-1998</td>
<td>–</td>
<td>trend stationary</td>
<td>–</td>
</tr>
<tr>
<td>1920-2011</td>
<td>–</td>
<td>trend stationary</td>
<td>–</td>
</tr>
<tr>
<td>1950-2011</td>
<td>trend stationary</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1950-2006</td>
<td>unit root</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1950-1998</td>
<td>trend stationary</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Table 1.5: Pesaran (2004) cross-section dependence test (CD test) for log real house prices, 1975-2011

<table>
<thead>
<tr>
<th></th>
<th>ADF(1)</th>
<th>ADF(2)</th>
<th>ADF(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>all 363 MSAs</td>
<td>403.44</td>
<td>405.37</td>
<td>358.66</td>
</tr>
<tr>
<td>top 50 supply inelastic MSAs</td>
<td>85.90</td>
<td>85.49</td>
<td>80.71</td>
</tr>
<tr>
<td>top 20 supply inelastic MSAs</td>
<td>38.21</td>
<td>38.12</td>
<td>36.95</td>
</tr>
<tr>
<td>top 50 supply elastic MSAs</td>
<td>69.16</td>
<td>66.70</td>
<td>57.39</td>
</tr>
<tr>
<td>top 20 supply elastic MSAs</td>
<td>26.53</td>
<td>26.69</td>
<td>24.84</td>
</tr>
<tr>
<td>48 states + D.C.</td>
<td>94.96</td>
<td>91.92</td>
<td>81.55</td>
</tr>
</tbody>
</table>

Notes: An intercept and a trend are included. The CD statistic tends to $N(0, 1)$ under the null hypothesis of no error cross-sectional dependence. The 2-sided 5% critical value is 1.96.
Table 1.6: Pesaran (2007) CIPS panel unit root test for log real house prices

<table>
<thead>
<tr>
<th></th>
<th>CADF(1)</th>
<th>CADF(2)</th>
<th>CADF(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>all 363 MSAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975-2002</td>
<td>-2.410</td>
<td>-2.222</td>
<td>-2.278</td>
</tr>
<tr>
<td>top 50 supply inelastic MSAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975-2006</td>
<td>-2.507</td>
<td>-2.103</td>
<td>-2.178</td>
</tr>
<tr>
<td>1975-2002</td>
<td>-2.415</td>
<td>-1.926</td>
<td>-1.741</td>
</tr>
<tr>
<td>top 20 supply inelastic MSAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975-2006</td>
<td>-2.796**</td>
<td>-1.816</td>
<td>-1.947</td>
</tr>
<tr>
<td>1975-2002</td>
<td>-3.028***</td>
<td>-1.768</td>
<td>-1.610</td>
</tr>
<tr>
<td>1975-1998</td>
<td>-3.319***</td>
<td>-2.103</td>
<td>-1.789</td>
</tr>
<tr>
<td>top 50 supply elastic MSAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975-2002</td>
<td>-2.561*</td>
<td>-2.437</td>
<td>-2.528</td>
</tr>
<tr>
<td>top 20 supply elastic MSAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975-2002</td>
<td>-2.567</td>
<td>-2.476</td>
<td>-2.183</td>
</tr>
<tr>
<td>1975-1998</td>
<td>-2.296</td>
<td>-2.174</td>
<td>-1.995</td>
</tr>
<tr>
<td>48 states + D.C.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975-2011</td>
<td>-3.093***</td>
<td>-2.970***</td>
<td>-2.995***</td>
</tr>
<tr>
<td>1975-2002</td>
<td>-2.312</td>
<td>-2.387</td>
<td>-2.608*</td>
</tr>
<tr>
<td>1975-1998</td>
<td>-2.051</td>
<td>-2.079</td>
<td>-2.386</td>
</tr>
</tbody>
</table>

Notes: An intercept and a trend are included. Reported are truncated CIPS statistics. 1%, 5%, 10% critical values for $N = 363$, $T = 30$ and $T = 20$ are -2.609, -2.532, -2.485 and -2.627, -2.534, -2.482. For $N = 50$, $T = 30$ and $T = 20$, they are -2.73, -2.61, -2.54 and -2.76, -2.62, -2.54. For $N = 20$, $T = 30$ and $T = 20$, they are -2.88, -2.72, -2.63 and -2.92, -2.73, -2.63. Critical values for $N = 50$ and $N = 20$ are given in Table II(c) in Pesaran (2007). Critical values for $N = 363$ are not reported in Pesaran's paper and we generate them by ourselves. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.
Table 1.7: Panel unit root test for log real house prices of the MSAs in Malpezzi (1999)

<table>
<thead>
<tr>
<th></th>
<th>CADF(1)</th>
<th>CADF(2)</th>
<th>CADF(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979-1996</td>
<td>-1.831</td>
<td>-1.894</td>
<td>-1.459</td>
</tr>
<tr>
<td>1975-2011</td>
<td>-2.855***</td>
<td>-2.802***</td>
<td>-3.001***</td>
</tr>
</tbody>
</table>

Notes: Pesaran (2007) truncated CIPS statistics are reported. An intercept and a trend are included. 1%, 5%, 10% critical values for $N = 100$, $T = 30$ and $T = 20$ are -2.66, -2.56, -2.51 and -2.70, -2.57, -2.51. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.
Table 1.8: Panel unit root test for log real house prices of the MSAs in Gallin (2006)

<table>
<thead>
<tr>
<th></th>
<th>CADF(1)</th>
<th>CADF(2)</th>
<th>CADF(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978-2000</td>
<td>-2.244</td>
<td>-2.090</td>
<td>-2.206</td>
</tr>
<tr>
<td>1975-2011</td>
<td>2.921***</td>
<td>-2.720***</td>
<td>-2.941***</td>
</tr>
</tbody>
</table>

Notes: Pesaran (2007) truncated CIPS statistics are reported. An intercept and a trend are included. 1%, 5%, 10% critical values for $N = 70$, $T = 30$ and $T = 20$ are -2.69, -2.58, -2.52 and -2.72, -2.59, -2.53. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.
Table 1.9: Panel unit root test for log real house prices of the MSAs in Mikhed and Zemcik (2009)

<table>
<thead>
<tr>
<th>Period</th>
<th>CADF(1)</th>
<th>CADF(2)</th>
<th>CADF(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978-2005</td>
<td>-2.567</td>
<td>-2.142</td>
<td>-1.866</td>
</tr>
</tbody>
</table>

Notes: Pesaran (2007) truncated CIPS statistics are reported. An intercept and a trend are included. 1%, 5%, 10% critical values for $N = 20$, $T = 30$ and $T = 20$ are -2.88, -2.72, -2.63 and -2.92, -2.73, -2.63. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.
Table 1.10: Panel unit root test for log real house prices of the states in Holly, Pesaran, and Yamagata (2010)

<table>
<thead>
<tr>
<th></th>
<th>CADF(1)</th>
<th>CADF(2)</th>
<th>CADF(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975-2003</td>
<td>-2.358</td>
<td>-2.455</td>
<td>-2.614*</td>
</tr>
<tr>
<td>1975-2011</td>
<td>-3.093***</td>
<td>-2.970***</td>
<td>-2.995***</td>
</tr>
</tbody>
</table>

Notes: Pesaran (2007) truncated CIPS statistics are reported. An intercept and a trend are included. 1%, 5%, 10% critical values for $N = 50$, $T = 30$ are -2.73, -2.61, -2.54. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.
Table 1.11: Pesaran (2007) CIPS panel unit root test for log real per capita income

<table>
<thead>
<tr>
<th></th>
<th>CADF(1)</th>
<th>CADF(2)</th>
<th>CADF(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>all 363 MSAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975-2009</td>
<td>-1.740</td>
<td>-1.657</td>
<td>-1.705</td>
</tr>
<tr>
<td>1975-2006</td>
<td>-1.788</td>
<td>-1.708</td>
<td>-1.704</td>
</tr>
<tr>
<td>1975-1998</td>
<td>-1.754</td>
<td>-1.539</td>
<td>-1.372</td>
</tr>
<tr>
<td>top 20 supply inelastic MSAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975-2009</td>
<td>-1.785</td>
<td>-1.786</td>
<td>-1.585</td>
</tr>
<tr>
<td>1975-2006</td>
<td>-1.589</td>
<td>-1.712</td>
<td>-1.491</td>
</tr>
<tr>
<td>1975-1998</td>
<td>-1.512</td>
<td>-1.478</td>
<td>-1.308</td>
</tr>
<tr>
<td>top 20 supply elastic MSAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975-2009</td>
<td>-1.897</td>
<td>-1.671</td>
<td>-1.720</td>
</tr>
<tr>
<td>1975-2006</td>
<td>-2.144</td>
<td>-2.068</td>
<td>-2.057</td>
</tr>
<tr>
<td>1975-1998</td>
<td>-2.042</td>
<td>-1.967</td>
<td>-1.385</td>
</tr>
</tbody>
</table>

Notes: An intercept and a trend are included. Reported are truncated CIPS statistics. 1%, 5%, 10% critical values for \( N = 363, T = 30 \) and \( T = 20 \) are -2.609, -2.532, -2.485 and -2.627, -2.534, -2.482. For \( N = 20, T = 30 \) and \( T = 20 \), they are -2.88, -2.72, -2.63 and -2.92, -2.73, -2.63. Critical values for \( N = 20 \) are given in Table II(c) in Pesaran (2007). Critical values for \( N = 363 \) is not reported in Pesarans paper and we generate them by ourselves.
Chapter 2: House Price Dynamics

2.1 Introduction

The housing market is a large economic sector and house value forms a key component of household assets. In 2010, the value of the primary residence was $20.7 trillion, representing around 30% of households’ total assets, and 62% of the median homeowner's total assets. Therefore, house price movements have a large impact on households’ wealth. Also, the performance of the housing market is closely linked to the overall economy. The recent housing bubble triggered a financial crisis and a recession. In light of these facts, understanding house price dynamics is an important issue.

One topic related to house price movements that gained a lot of attention is the relationship between house prices and income. Comparisons between house prices and income are often regarded as an indication of whether the housing market is overvalued or undervalued. Figure 2.1 presents the log real house prices.

9The numbers are from a National Association of Home Builders study based on Survey of Consumer Finance 2010:
http://www.nahb.org/generic.aspx?sectionID=734&genericContentID=215073&channelID=311#Footnote4
and log real per capita income at the U.S. national level over 1975-2010.10 To improve visual clarity, we scale the vertical axis for house prices so that the two series cross each other in 1975. The figure shows that house prices and income behave quite differently. House prices are much more volatile than income. For example, during the recent housing boom, U.S. real house prices rose by 56% over the period 1998-2006, while U.S. real income only rose by 12% during the same period of time. This rapid appreciation in real house prices relative to real income raised the publics’ concern that house prices were much higher than the “fundamental value” and would tend to revert to a level that is justified by fundamentals such as income.

Another widely discussed issue is the inter-temporal evolution of house prices: the short-run momentum and long-run reversion of house price changes. Figure 2.2 plots the log real house prices for a coastal metropolitan statistical area (MSA) (San Francisco, CA) and an inland MSA (Columbus, OH). Two features are noticeable from the figure. First, house prices are predictable over short periods of time. There are apparent house price cycles. House prices keep rising for several years and then revert back in the following years. Second, the cycles are much bigger in the coastal city than the inland city: San Francisco experienced three sharp cycles during 1975-2010, while the house price series for Columbus is smoother. These two MSAs are fairly typical of coastal and inland MSAs.

10In Figure 2.1, house price is measured by the Freddie Mac House Price Index (FMHPI), and income is measured by the Bureau of Economic Analysis (BEA) per capita income. We deflate them by CPI.
The goal of this paper is to explain several house price phenomena that bring challenges to housing models. The first phenomenon is about a time series property of house prices: are house prices stationary or do they have a unit root\textsuperscript{11} This time series property has important implications for studying the relationship between house prices and income and modeling house prices. The housing literature indicates income as one of the most important factors that determine house prices. For example, Case and Shiller (2003) report that “income alone explains patterns of home price changes since 1985 in all but eight states.” Also, the idea of a stable long-run relationship between house prices and income, as in the discussion for Figure 2.1, is studied in the literature by cointegration analyses. If house prices and income are cointegrated, they can drift away from each other temporarily, but they will tend to adjust back to their long-run relationship. Many papers test for a cointegrating relationship between house prices and income (for example, Gallin (2006), Holly, Pesaran, and Yamagata (2010)). Note that for cointegration analyses to be valid, house prices and income should both be unit root processes, which is the common treatment in the housing literature. However, Zhang, de Jong, and Haurin (2013) find that, when examining extended time periods that include the recent full housing cycle, there is strong evidence that house prices are trend stationary while income has a unit root\textsuperscript{12} Therefore,

\textsuperscript{11}When we state “house price” or “income” in the paper, we refer to real house price and real income.

\textsuperscript{12}They examine log of real house price and log of real income. However, we also get similar results using real house price and real income: real house price rejects a unit root and real income does not reject a unit root.
the relationship between house prices and income needs to be carefully reconsidered, and a challenge for a coherent house price model is to combine the different stationarity properties of house prices and income.

Another phenomenon that has not been well explained in the literature is the short-run positive serial correlation and long-run mean reversion of house price changes, as well as their geographic differences, as depicted in Figure 2.2. This momentum and reversion closely relate to house price cycles. Two questions are raised: what generates the house price fluctuations, and why do coastal cities have much bigger cycles than inland cities. In the literature, empirical studies document that there is short-run inertia in house price changes (Case and Shiller (1989)). If house prices rise this year, they have the tendency to continue this appreciation in the next year. The literature also documents that as house prices rise over a relatively long period, they will tend to revert. For example, Glaeser et al. (2012) report that an increase of house prices in the past five years is associated with a decrease of house prices over the next five-year period. Another way to describe the mean reversion is that when house prices deviate from their “fundamental value” they tend to adjust back, as in error correction models (Abraham and Hendershott (1996), Malpezzi (1999), Capozza, Hendershott, and Mack (2004), etc.). Momentum and reversion generate house price cycles. An equally important issue is the geographic differences of house price cycles. The magnitudes of momentum and reversion are both larger in coastal areas than inland areas (Abraham and Hendershott (1996), Glaeser et al. (2012)). Therefore, understanding the serial correlation and mean reversion of house price changes is
useful for explaining how cycles happen, and why house price dynamics differ in different areas.

This paper modifies the dynamic spatial equilibrium model in Glaeser et al. (2012) to explain the above house price issues. Specifically, we study the stationarity property and the autocorrelation of one-year and five-year house price changes. In Glaeser et al.’s model, households’ indirect utility of living in a city depends on income, amenities, and housing cost. Hence, differences in house prices across cities reflect the differences in income and amenities across cities. They assume migration among cities is costless and they have a spatial equilibrium condition requiring households to equalize their utility across cities in each period. On the supply side, they assume that the current value of the expected house prices in the next period equals the marginal cost of building a house, and this housing supply condition determines the amount of construction each period and therefore the housing stock in a city. Therefore, house prices are endogenously determined and reflect the local heterogeneity of demand side and supply side factors. They solve and estimate the model under rational expectations.

The main differences between this paper and Glaeser et al.’s are the assumptions about income and expectations. Glaeser et al. assume that income can be expressed as an exogenous productivity variable minus a coefficient times the population. We assume that income is an exogenous unit root process based on results from unit root tests. Our assumption about the exogeneity of income can be interpreted as following Capozza and Helsley (1990)’s urban model assumption. They assume a small open urban area which is affected by external demand shocks through the labor market, and thus household income in that small open
city is an exogenous process. The main reason for our assumption of the income process is to study the stationarity puzzle: how a unit root income process is related to a trend stationary house price process. Glaeser et al. do not consider this issue in their model.

The second difference from Glaeser et al. is that they only consider rational expectations, while we solve and estimate the model under both rational and adaptive expectations. Their model under rational expectations failed to explain the short-run momentum of house price changes. Moreover, there are claims, both in the media and in the literature, that people in the housing market are irrational and backward-looking (Case and Shiller (2003), Shiller (2005), Glaeser, Gyourko, and Saiz (2008)). Therefore, we also consider the case where people adaptively adjust their expectations.

Our sample consists of 356 MSAs over 1980-2010. We separate the sample into a coastal group and an inland group and estimate the model parameters separately for each group. The estimation results reveal that the coastal group has more persistence in income changes, more inelastically supplied housing, and that the population increase has a smaller effect on amenities in the coastal cities. Based on estimation of the parameters, we simulate the house price series and compare how the predictions from the simulation match the data. The model with adaptive expectations fits the empirical features better than the one that assumes rational expectations: it generates a positive one-year autoregressive coefficient, a negative five-year autoregressive coefficient, and the unit root test rejects a unit root. However, for the coastal group, the model does not generate enough magnitude of the one-year and five-year price change autoregressive
coefficients, and for the inland group, it matches closely to the one-year price change, but over generates the magnitude of the five-year price change autocorrelation.

The remainder of the paper is organized as follows. Section 2 discusses the related literature. Section 3 presents a model of house prices, solves the solutions separately under rational expectations and adaptive expectations, and provides impulse response analyses. Section 4 describes the data, estimation methodology, and reports the estimation results. Section 5 presents the simulation results showing how the model fits the empirical features of house prices. Section 6 concludes.

2.2 Related Literature

The discussion of the stationarity puzzle relates to papers that study time series aspects of house prices, especially cointegration analysis and error correction models. House prices and income are usually both treated as unit root processes in the housing literature. This unit root assumption is a prerequisite for applying cointegration analysis and error correction models; otherwise, if either house prices or income does not have a unit root, cointegration tests and error correction specifications would be invalid. Many papers apply cointegration analysis to test for stable long-run equilibrium relationships between house prices and fundamental variables such as income. If such a long-run relationship exists so that house prices and income are cointegrated, then they cannot drift away from each other over the long run. Therefore, a deviation from this long-run equilibrium can be viewed as an indicator of whether house prices are
overvalued or undervalued and the tendency to adjust back to the equilibrium relationship, and this is the idea in error correction models. Examples of applying cointegration analysis and error correction models to house prices include Abraham and Hendershott (1996), Malpezzi (1999), Capozza, Hendershott, and Mack (2004), Gallin (2006), Mikhed and Zemcik (2009), and Holly, Pesaran, and Yamagata (2010). In those papers, either they do not formally test for a unit root, or they do not include the recent housing bust period.\footnote{Mikhed and Zemcik (2009) is the only one among those papers that reports unit root tests for real house prices that include the bust period, and over the period 1978-2007 a unit root is rejected at the 5% level. But the authors then choose the sub-periods ending before 2006, and apply cointegration tests when a unit root is not rejected. This is an ad hoc method to treat house prices.}

However, Zhang, de Jong, and Haurin (2013) find a stationarity puzzle: there is strong evidence that house prices are trend stationary while income has a unit root. They first study a national log real home price index constructed by Robert Shiller over a 120-year period 1890-2011. The unit root tests results indicate that real house prices are trend stationary with structural breaks. They also apply the Pesaran (2007) CIPS panel unit root test to log real house prices for a metro area panel dataset and a state panel dataset. A unit root is statistically significantly rejected over 1975-2011. They further find that the reason for this disagreement with the literature is because the previous studies do not cover the full housing cycle of 2000-2010. In contrast to log real house prices, a panel unit root test for log real income over 1975-2010 cannot reject a unit root. Therefore, their paper poses a puzzle for housing models: how to explain different stationarity properties between real house price and real income in a typical urban model.
The issues of short-run momentum and long-run reversion of house price changes studied in this paper connect to empirical papers about house price fluctuations. Many papers empirically examine the short-run positive serial correlation and long-run mean reversion of house price changes. Case and Shiller (1989) use repeated sales data from Atlanta, Chicago, Dallas, and San Francisco over period 1970-1986. They find positive autocorrelation over a one year period for both real house price appreciation and after-tax excess housing returns. In a follow-up study (1990), Case and Shiller test a variety of other public economic variables and find that besides one-year lagged real house price appreciation, real income growth, population growth among age 25-40, and the ratio of construction costs to prices also have power to forecast house prices. Abraham and Hendershott (1996) use annual data for 30 metropolitan areas over the period 1977-1992 to investigate the determinants of real house price appreciation. The explanatory variables consist of three parts: the change in fundamental price, the lagged real house price appreciation with its coefficient called the “bubble builder,” and the deviation of house price from its equilibrium level in the previous period with its coefficient called the “bubble burster.” They find a positive bubble builder coefficient and a negative bubble burster coefficient. Moreover, the absolute values of these coefficients are higher for the coastal city group than for the inland city group.

Motivated by the finding in Abraham and Hendershott (1996) that serial correlation and mean reversion are significantly higher in absolute value for coastal areas than inland areas, Capozza, Hendershott, and Mack (2004) empirically study which variables explain the significant difference in the geographic patterns of
the “bubble builder” and the “bubble burster” coefficients found in Abraham and Hendershott’s (1996) error correction specification. Using a panel dataset of 62 MSAs from 1979-1995, they find that higher real income growth, higher population growth, and a higher level of real construction costs increase the “bubble builder” coefficient, while higher real income growth, a larger population, and a lower level of real construction costs increase the absolute value of the “bubble burster” coefficient. There is no theoretical model supporting their selection of explanatory variables, but they offer some reasonable justifications such as information and search costs, supply inelasticity, and market euphoria and expectation.

Our model is complementary to housing models that aim to provide theoretical answers to the momentum and reversion of house prices changes found in the above empirical studies. Despite of the large number of empirical papers, theoretical counterparts are relatively limited. The most relevant theoretical study to our paper is Glaeser et al. (2012). They present a dynamic spatial equilibrium model with rational expectations. Housing costs reflect the value of access to an area’s income and amenities. Construction is endogenously determined by marginal costs equaling the expected house prices. They divide MSAs into three regions: coastal, sunbelt, and interior, and estimate the parameters separately for each of those regions. When the model is used to fit the data, it has both successes and failures. Notably, it fails to explain the short term serial correlation of house price changes. The model always generates negative autocorrelation at the one year horizon, although the data in their sample reveals strong persistence (over a one year horizon the autocorrelation is 0.59-0.81 when income data is
from the Home Mortgage Disclosure Act (HMDA), and 0.53-0.70 when using the Bureau of Economic Analysis (BEA) income data). Their model succeeds qualitatively in explaining the mean reversion of price changes at a five-year horizon, although it often understates the extent of reversion, especially in coastal areas. Moreover, the model generates short term volatility of house price changes and construction, and positive serial correlation of construction quantities that are close to data, but under predicts the volatility over the five-year horizon. The failures and successes of Glaeser et al.’s model suggest that, “a perfectly rational model in which construction is not instantaneous and local income series themselves mean revert” neglects some important elements in explaining short term housing market dynamics, especially the short run persistence in house price changes.

Head, Lloyd-Ellis, and Sun (2013), use a search and matching model to explain the short-term movements of house prices and construction. In their setting, the demand for buying a house will continue for some periods after a positive demand shock because it takes time for the buyer to find a house through the matching process. Moreover, new houses need time to be built. Therefore, the ratio of buyers to sellers, “market tightness” as they call it, continues to go up for some time, and so the prices will not rise to the maximum immediately and then fall, as in a frictionless market, but keep rising for some periods. As the matching process gradually ends and new houses are built, prices will then begin to fall. In this way, the model can generate short run momentum and long run mean reversion for house price changes. The first-order autocorrelation of house price growth generated by their search model is around half of that in the data,
which means that search frictions alone are not sufficient for explaining the short term house price dynamics.

The comparison between rational expectations and adaptive expectations in our paper relates to a number of studies in the housing literature that document the important role of backward-looking expectation and speculation in explaining house price cycles. Malpezzi and Wachter (2005) develop and simulate a model with speculative backward looking expectations. Housing demand increases if house prices rise in the past. They find that speculation in housing demand plays an important role in generating property cycles when supply is inelastic. Their simulation results do not show significant house price cycle patterns if speculation is not incorporated in the model, but when speculation is added, strong boom and bust cycles are generated. But the role of speculation relates to supply conditions: no significant cycle is generated by speculation when supply is very elastic. Their model reveals that backward-looking expectations could be an important contributor to house price dynamics. Glaeser, Gyourko, and Saiz (2008) construct a model with extrapolative expectations and obtain the results that areas with more housing supply inelasticity have greater house price appreciations during housing booms. Their study reveals two important potential factors to explain housing bubbles: backward-looking expectations and supply inelasticity. A group of recent papers adopt heterogeneous expectation models to explain housing cycles. Some of the examples are Burnside, Eichenbaum, and Rebelo (2011), and Dieci and Westerhoff (2012). In these studies, some economic agents form their expectations extrapolatively while others expect house prices to revert to the fundamental values. The relative weights of these views
evolve over time. Housing cycles are generated as extrapolative expectations become dominant during booms while more people believe that house price will revert to fundamentals during busts.

2.3 Model

2.3.1 Setup

The model relies on three conditions: a spatial equilibrium condition, a housing supply condition, and a condition regarding how people form their expectations.

We assume migration across cities are costless. Therefore, households equalize utility across cities in equilibrium. This equalized utility is assumed to be a national utility level $\bar{U}$. This indifference across space holds for all periods. The no moving costs is a strong assumption that makes the model more tractable. We assume each household consumes one unit of housing, and therefore upsizing or downsizing the house is not considered here. The indirect utility of a household living in city $i$ during period $t$ is assumed to be

$$V_{it} = Income_{it} + Value\ of\ amenities_{it} - Housing\ costs_{it}. \hspace{1cm} (2.1)$$

This indirect utility expression is the same as in Glaeser et al.

Assume that in city $i$, the income $y_{it}$ is a unit root process $x_{it}$ plus an intercept and a linear time trend,

$$y_{it} = D_i + \mu_i t + x_{it},$$

where

$$x_{it} = x_{i,t-1} + \varepsilon_{it}, \text{ and } \varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + v_{it}. \hspace{1cm} (2.2)$$
In other words, $D_i$ is an area specific fixed effect, $\mu_i t$ is the time trend, and $x_{it}$ is a unit root process with first order autocorrelation in the error terms.

The value of a city’s amenities is measured by

$$\text{Amenities}_{it} = A_i - \alpha_i N_{it},$$

where $A_i$ is an area specific fixed effect for amenities such as weather, $N_{it}$ is the housing stock, and $\alpha_i$ is a disamenity coefficient measuring marginal congestion costs. The assumption $\alpha_i > 0$ implies that households prefer less crowded cities. Additional housing stock decreases amenities and decreases households’ indirect utility of living in the city.

We focus on homeowners so that housing costs are measured by owner costs. The “owner cost” term\textsuperscript{14} used by Hendershott and Slemrod (1983) and Poterba (1984), measures the annual flow of cost for owning a home with the specification,

$$OC_{it} = P_{it} c_{it}, \text{ where } c_{it} = [(r_{it} + \tau_{it}^p)(1 - \tau_{it}^y) + d_i + tc_i / n_i - \pi_{it}].$$

In this equation, $P_{it}$ is house value, $c_{it}$ is the user cost per unit expense on housing, $r_{it}$ is the mortgage interest rate, $\tau_{it}^p$ is the property tax rate, $\tau_{it}^y$ is the marginal income tax rate, $d_i$ is the sum of maintenance and depreciation costs, $tc_i / n_i$ is the transaction cost per unit value of the house divided by the expected length of stay, and $\pi_{it}$ is the expected rate of house price inflation. We assume for simplicity that the interest rate, the property tax rate, the marginal income tax rate, the maintenance and depreciation cost rate, and the transaction cost rate do not change over time. In other words, the term $(r_i + \tau_{it}^p)(1 - \tau_{it}^y) + d_i + tc_i / n_i$ does not change over time. In other words, the term $(r_i + \tau_{it}^p)(1 - \tau_{it}^y) + d_i + tc_i / n_i$ does not change over time.

\textsuperscript{14}Glaeser et al. abstract from income tax, property tax, maintenance costs, and transaction costs by assuming that the expected housing costs equal $P_{it} - \frac{EP_{i+1}^t}{1+r}$, where $r$ is the interest rate.
vary over time, but it can differ across cities. Therefore, we abstract from studying the effects of time-varying interest rates and taxes. Denote

$$\eta_i = (r_i + r_i^T)(1 - \tau_i^T) + d_i + tc_i/n_i.$$ (2.5)

The annual owner cost can be rewritten as

$$OC_{it} = P_{it} \left( \frac{\eta_i - E_tP_{i,t+1} - P_{it}}{P_{it}} \right) = (1 + \eta_i)P_{it} - E_tP_{i,t+1}. \quad (2.6)$$

Therefore, the utility flow of city \(i\) at period \(t\) is

$$V_{it} = y_{it} + A_i - \alpha_iN_{it} - (1 + \eta_i)P_{it} + E_tP_{i,t+1}. \quad (2.7)$$

(2.7) should equal the national utility level in equilibrium:

$$y_{it} + A_i - \alpha_iN_{it} - (1 + \eta_i)P_{it} + E_tP_{i,t+1} + w_{it} = U, \quad (2.8)$$

where \(w_{it}\) is the error term which captures the disturbances in the national utility level. Substituting (2.2) into (2.8), we get

$$D_i + x_{it} + \mu_i t - \alpha_iN_{it} - (1 + \eta_i)P_{it} + E_tP_{i,t+1} + w_{it} = U, \quad (2.9)$$

where \(D_i = \bar{D}_i + A_i\). Equation (2.9) is our spatial equilibrium equation. This equation implies that the difference in owner costs of housing between two cities must reflect differences in income or differences in amenities between these two areas.

The second element of the model is housing supply. We assume that the marginal cost of producing a unit of housing is

$$MC_{it} = c_i + c_{1i}I_{it}, \quad (2.10)$$

\text{15We try varying values of the parameter } \eta \text{ in the estimation, and estimation results do not change much regarding the choice of } \eta.
where $c_{1i} > 0$, and $I_{it}$ is the amount of new construction in period $t$. Denote $N_{it}$ as the housing stock at the beginning of period $t$. Then the change in the housing stock, assuming a depreciation rate of $\delta_i$, satisfies

$$N_{i,t+1} = I_{it} + (1 - \delta_i)N_{it}. \quad (2.11)$$

The condition $c_{1i} > 0$ implies that as construction increases, the marginal cost increases. This condition can be interpreted as diminishing returns due to scarcity of certain inputs such as land of producing housing.

We assume that the construction market is competitive, and that houses require one period to be built, so houses under construction in this period will be on the market next period, thus the decisions of builders at time $t$ are based on the expected price level for the next period. Therefore, in equilibrium, the costs of producing a unit of housing equal the current value of the expected house price in the next period:

$$E_tP_{i,t+1} = c_i + c_{1i}I_{it}. \quad (2.12)$$

Substituting (2.11) into (2.12), we get

$$E_tP_{i,t+1} = c_i + c_{1i}N_{i,t+1} - (1 - \delta_i)c_{1i}N_{it}. \quad (2.13)$$

Rearranging (2.13) and we get an equation in which the next period housing stock depends on the expected house price and the current housing stock:

$$c_{1i}N_{i,t+1} = -c_i + E_tP_{i,t+1} + (1 - \delta_i)c_{1i}N_{it}. \quad (2.14)$$

We add an error term $u_{it}$ into (2.14) to capture the effects of factors other than expected house price and current housing stock on the construction and therefore
the next period’s housing stock. For example, changes in the weather may delay construction. Therefore, we get the housing supply equation:

\[ c_{1i}N_{i,t+1} = -c_i + E_t P_{i,t+1} + (1 - \delta_i)c_{1i}N_{i,t} + u_{it}. \]  (2.15)

A model with only the spatial equilibrium equation (2.9) and the housing supply equation (2.15) is not identifiable because house price \( P \) and housing stock \( N \) are both endogenous variables and the expectation term \( E_t P_{i,t+1} \) is unobserved. Therefore, an additional equation regarding the behavior of the expectation \( E_t P_{i,t+1} \) is needed. We will consider two cases: rational expectations and adaptive expectations.

The above setup shares many aspects of Glaeser et al. (2012)’s framework. The main differences with their paper are the assumptions about income and about expectations, as mentioned in the introduction and literature review.

Our model treats income as an exogenous unit root process by Equation (2.2). To justify this assumption, we applied Pesaran (2007)’s CIPS panel unit root test to real per capita income for 363 MSAs over 1975-2010, and found that a unit root cannot be rejected. This finding is robust to varying lag length included. We then regressed the first difference of real income on its one-year lag value and two-year lag value, the coefficient of the former is significant and the coefficient of the latter is insignificant. These results indicate that a unit root process with an AR(1) error is an appropriate assumption for real income.

In contrast, Glaeser et al. treat income as endogenous. They use the term “wage” in the model setup, but use income data as the measure of wage in the
estimation, thus the variable “wage” is essentially income. They specify wage as
\[ Wage_{it} = W_{it} - d_i N_{it}, \]
where \( W_{it} \) is an exogenous productivity variable and \( N_{it} \) is population, and this variable is assumed to equal housing stock. They also assume that amenities change linearly in population:
\[ Amenities_{it} = A_{it} - a_i N_{it}. \]
Therefore, the parameter \( \alpha_i \) in (2.8) is the combined effects of population on wages and amenities \((d_i + a_i)\) in their model. They make several simplifying assumptions and set \( d_i = 0.1 \) and \( a_i = 0 \). Then they construct the productivity variable \( W_{it} \) from income and housing stock by \( W_{it} = Wage_{it} + d_i N_{it} \), assume \( W_{it} \) follows an ARMA(1,1) process, and estimate the parameters of the ARMA(1,1) process.

Our approach of treating income is an improvement to Glaeser et al.’s approach for two reasons. First, the choice of the parameter values for \( d_i \) and \( a_i \) in their approach is somewhat arbitrary. Moreover, they set the same values for \( d_i \) and \( a_i \) for all areas, thus cannot capture any geographic differences. In contrast, our approach estimates the effect of population on amenities (the disamenity coefficient \( \alpha_i \)). Also, the estimation can be implemented for different geographic groups of cities, and hence examine any difference in the disamenity coefficient \( \alpha_i \) for different cities. Second, our model examines how the time series properties of income are transmitted to house prices, and studies the stationarity puzzle: how a unit root income series is related to a trend stationary house price process. Glaeser et al. do not consider this issue in their model.
2.3.2 Trend-adjusted Equations

Before moving on to make any assumptions about expectations and solving the model, we eliminate the deterministic trend from the spatial equilibrium equation (2.9) and the supply equation (2.15). Detrending the equations will make the follow-up calculations cleaner. Details of the detrending calculation is in Appendix 1.

The deterministic trends for housing stock and house prices are:

\[
\tilde{N}_{it} = \left( \frac{\mu_i}{\eta_i \delta_i c_{i1} + \alpha_i} \right) t + \left( \frac{D_i + \mu_i - \bar{U} - \eta_i c_i}{\eta_i \delta_i c_{i1} + \alpha_i} + \frac{-(\eta_i - \delta_i) c_{i1} - \alpha_i^2}{(\eta_i \delta_i c_{i1} + \alpha_i)^2} \right) \mu_i , \quad (2.16)
\]

and

\[
\tilde{P}_{it} = \left( \frac{\mu_i \delta_i c_{i1}}{\eta_i \delta_i c_{i1} + \alpha_i} \right) t + c_i + (1 - \delta_i) c_{i1} \left( \frac{\mu_i}{\eta_i \delta_i c_{i1} + \alpha_i} \right) \quad (2.17)
\]

\[
+ \delta_i c_{i1} \left( \frac{D_i + \mu_i - \bar{U} - \eta_i c_i}{\eta_i \delta_i c_{i1} + \alpha_i} + \frac{-(\eta_i - \delta_i) c_{i1} - \alpha_i^2}{(\eta_i \delta_i c_{i1} + \alpha_i)^2} \right) \mu_i \right) . \quad (2.18)
\]

The detrended spatial equilibrium equation and housing supply equation are:

\[
(1 + \eta_i) p_{it} - E_t p_{i,t+1} = x_{it} - \alpha_i n_{it} + w_{it}, \quad (2.19)
\]

\[
c_{i1} n_{i,t+1} = E_t p_{i,t+1} + (1 - \delta_i) c_{i1} n_{it} + u_{it}. \quad (2.20)
\]

We will derive the expressions for house prices from the detrended system (2.19) and (2.20), under rational expectations and adaptive expectations, respectively.

2.3.3 Rational Expectations

We apply Blanchard and Kahn (1980) to show that the model has a unique non-explosive solution under rational expectations and we derive this solution.
The results are in Proposition 1. Appendix 2 describes Blanchard and Kahn (1980), and Appendix 3 provides the proof for Proposition 1.

**Proposition 1:** Assume our model of (2.19) and (2.20) under rational expectations have the parameter restrictions: $\alpha_i > 0$, $\eta_i > 0$, $1 > \delta_i \geq 0$, and $c_{1i} > 0$. We also require that people do not expect house prices and housing stock to exponentially explode in the future:

$$\forall t, \exists \left[ \begin{array}{c} \pi_t \\ \overline{p}_t \end{array} \right] \in R^2, \sigma_t \in R \text{ such that}$$

$$-(1 + i)^{\sigma_t} \left[ \begin{array}{c} \pi_t \\ \overline{p}_t \end{array} \right] \leq \left[ \begin{array}{c} E(n_{t+1}|\Omega_t) \\ E(p_{t+1}|\Omega_t) \end{array} \right] \leq (1 + i)^{\sigma_t} \left[ \begin{array}{c} \pi_t \\ \overline{p}_t \end{array} \right], \forall i \geq 0. \quad (2.21)$$

Then there exists a unique nonexplosive solution under rational expectations. Under this solution, the trend-adjusted house price and housing stock satisfy

$$n_{it} = n_{i0}, \text{ for } t = 0,$$

$$n_{it} = \phi_i n_{i,t-1} + \frac{1}{c_{1i}(\overline{\phi}_i - 1)} x_{i,t-1} + \frac{\rho_i \overline{\phi}_i}{c_{1i}(\overline{\phi}_i - 1)(\overline{\phi}_i - \rho_i)} \varepsilon_{i,t-1} + \frac{1 + \eta_i}{c_{1i} \overline{\phi}_i} u_{i,t-1}, \text{ for } t > 0, \quad (2.22)$$

$$p_{it} = -\frac{\alpha_i + (1 - \delta_i - \phi_i)c_{1i}}{1 + \eta_i} n_{it} + \frac{\overline{\phi}_i}{(1 + \eta_i)(\overline{\phi}_i - 1)} x_{it} + \frac{\rho_i \overline{\phi}_i}{(1 + \eta_i)(\overline{\phi}_i - 1)(\overline{\phi}_i - \rho_i)} \varepsilon_{it}$$

$$+ \frac{1}{1 + \eta_i} w_{it} + \frac{1 + \eta_i - \overline{\phi}_i}{(1 + \eta_i) \overline{\phi}_i} u_{it}, \text{ for } t \geq 0. \quad (2.23)$$

Proposition 1 reveals that house prices are characterized as a first order autoregressive process with errors containing both a stationary component and a
nonstationary component. The behavior of the model under rational expectations characterized by Proposition 1 will be studied in a later subsection by an impulse response analysis, to examine the effect of an income shock over time on the path of house prices.

### 2.3.4 Adaptive Expectations

Assume households and home builders form their expectations adaptively for the deviation from the deterministic trend of house prices:

\[
E_{t}p_{i,t+1} = E_{t-1}p_{it} + \lambda_{i}(p_{it} - E_{t-1}p_{it}) = \lambda_{i}p_{it} + (1 - \lambda_{i})E_{t-1}p_{it}, \text{ where } 0 < \lambda_{i} < 1. \tag{2.24}
\]

The reason for applying the adaptive formula to the trend-adjusted house price rather than the original series \( P_{it} \) is to avoid the systematic error made by agents under adaptive expectations when a trend exists. Agents adjust their expectations according to their previous period’s expectation error. The parameter \( \lambda_{i} \) is a measure of the adjustment speed. A small \( \lambda_{i} \) indicates that agents do not change their expectations much, while a \( \lambda_{i} \) that is close to 1 indicates that agents’ house price expectations for the next period is close to house prices in this period.

The adaptive expectation equation \( (2.24) \), the spatial equilibrium equation \( (2.19) \), and the housing supply equation \( (2.20) \) are the three equations from which we derive the solution.

Combining the supply equation \( (2.20) \) and the adaptive expectation equation \( (2.24) \) yields

\[
-\lambda_{i}p_{it} + c_{1i}n_{i,t+1} - (2 - \lambda_{i} - \delta_{i})c_{1i}n_{it} + (1 - \lambda_{i})(1 - \delta_{i})c_{1i}n_{i,t-1} = u_{it} - (1 - \lambda_{i})u_{i,t-1}. \tag{2.25}
\]
Combining the spatial equilibrium equation (2.19) and the adaptive expectation equation (2.24) yields

$$(1 + \eta_i - \lambda_i) p_{it} - (1 + \eta_i)(1 - \lambda_i) p_{i,t-1} + \alpha_i n_{it} - \alpha_i (1 - \lambda_i) n_{i,t-1} - x_{it} + (1 - \lambda_i) x_{i,t-1} = w_{it} - (1 - \lambda_i) w_{i,t-1}. \quad (2.26)$$

**Proposition 2:** Assume our model of (2.19) and (2.20) under adaptive expectations have the parameter restrictions: \( \alpha_i > 0, \eta_i > 0, 1 > \delta_i \geq 0, c_{1i} > 0, \) and \( 0 < \lambda_i < 1. \) We also assume that the parameters satisfy a stability condition

$$\lambda < \frac{(2 + 2\eta_i)(2 - \delta_i)c_{1i}}{(2 + \eta_i)(2 - \delta_i)c_{1i} + \alpha_i}. \quad (2.27)$$

Then there is a unique solution under adaptive expectations. Under this solution, the trend-adjusted house price and housing stock satisfy

$$-p_{it} + c_{1i} \frac{n_{i,t+1}}{\lambda_i} - \frac{(2 - \lambda_i - \delta_i)c_{1i}}{\lambda_i} n_{it} + \frac{(1 - \lambda_i)(1 - \delta_i)c_{1i}}{\lambda_i} n_{i,t-1} = \frac{1}{\lambda_i} u_{it} - \frac{1 - \lambda_i}{\lambda_i} u_{i,t-1}. \quad (2.28)$$

$$(1 + \eta_i - \lambda_i) p_{it} - (1 + \eta_i)(1 - \lambda_i) p_{i,t-1} + \alpha_i n_{it} - \alpha_i (1 - \lambda_i) n_{i,t-1} - x_{it} + (1 - \lambda_i) x_{i,t-1} = w_{it} - (1 - \lambda_i) w_{i,t-1}. \quad (2.29)$$

The derivation of the stability condition (2.27) and the proof for Proposition 2 is in Appendix 4.

Proposition 2 reveals that house prices are characterized as a second order autoregressive process with errors containing both a stationary component and a nonstationary component. The behavior of the model under adaptive expectations characterized by Proposition 2 will be studied in the next subsection by an
impulse response analysis, and is compared to the case under rational expectations.

### 2.3.5 Impulse Response

In this subsection, we study the path of trend-adjusted house prices following a one time income shock of 1000 dollars, for both the rational expectations and adaptive expectations cases.

For rational expectations, the impulse responses are graphed in Figure 2.3. The baseline parameter values are set to $\rho = 0.2$, $c_1 = 2$, and $\alpha = 0.1$, represented by the solid lines. We also vary these values one at a time to examine how the over time effect of income on house prices changes when the parameter changes. Under the baseline parameter values, house prices increase immediately at the time of the income shock, but gradually decline from that point. Intuitively, when at the time of the income shock, house prices will increase to reflect this positive shock. Because of the increase in house prices, new houses will start to be built, and as new constructions occur, house prices decline gradually. The dashed lines in the upper, middle, and lower subgraphs respectively represent cases where there is no persistence in income change ($\rho = 0$), where supply is more elastic ($c_1 = 0.5$), and where the negative effect of housing stock on households’ indirect utility is larger ($\alpha = 0.2$). Compared to the baseline case, house prices in all these dashed-line cases increase less and revert faster.

17 The parameter values chosen in the impulse response analyses are based on the estimation results in the later section.
on house prices are smaller, which are anticipated by forward-looking households and reflected in the lower house price rise and faster reversion. When housing supply is more elastic, more new homes will be constructed after a positive income shock, and hence moderate the responses of house prices. When the disamenity coefficient is larger, new construction has bigger effects on reducing house prices, resulting in weaker house price responses.

For adaptive expectations, the impulse responses are depicted in Figure 2.4. The baseline parameter values are set to $\rho = 0.2$, $c_1 = 2$, $\alpha = 0.1$, and $\lambda = 0.9$, represented by the solid lines. House prices shoot up immediately when hit by an income shock, as in the case of rational expectations. But different from rational expectations where house prices start to decrease right after the shock, house prices under adaptive expectations continue to increase in the next period, therefore generate momentum in house price appreciation. After one period, house prices begin to decline gradually and thus generate reversion. The dashed lines in the top three subgraphs represent the same parameter changes as for rational expectations. The dashed line in the lowest subgraph represents a case where agents adjust their expectations more slowly ($\lambda = 0.5$). Compared to the baseline case, dashed lines in the top three subgraphs indicate lesser or no momentum and faster reversion, and the dashed line in the lowest subgraph presents a smoother house price path.
2.3.6 Implications of the Impulse Response Analyses

The impulse response analyses presented in Figure 2.3 and Figure 2.4 reveal some information about the behavior of the model related to the issues of stationarity, short-run momentum, and long-run reversion. The figures show that, under both rational expectations and adaptive expectations, house prices rapidly increase when an income shock occurs and gradually decline in the long run as new construction occurs. Therefore, the long-run reversion is generated by the model.

Also, even though the over time effect of the one time shock on income is permanent because income is a unit root process, the effect of the shock on house prices diminishes over time as shown in the figures. Recall that Proposition 1 and Proposition 2 both reveal that house prices contain both a unit root component and a stationary component. The fact that house prices decline over time to a very low level in response to a one time shock indicates that the weighting of the unit root component is quite small compared to the stationary component. This provides an intuitive explanation of why a unit root is rejected for house prices.

The short-run momentum is the main disagreement of the model with rational expectations and the one with adaptive expectations. Under rational expectations, since people are forward looking, any future consequence of a positive income shock would be anticipated and incorporated into the initial house price change. Therefore, after a positive income shock, house price immediately rise to its maximum level and then gradually declines and thus short term momentum cannot be generated in this setting. However, under adaptive expectations,
people are backward looking instead of forward looking. Therefore, when a positive demand shock occurs, house prices rise in this period and people adjust their expectations to a higher level. In the second period after the shock, the higher expectations will be incorporated into the price level, and price momentum could be generated if the expectation effect is strong enough to outweighs the construction effect.

### 2.4 Estimation

#### 2.4.1 Data

The data we use are annual MSA level income, housing units, and house prices over 1980-2010. Income and house prices are converted to real terms by deflating them by the CPI-U. Levels of the variables rather than logarithms are used in order to be consistent with the model. We obtain the per capita personal income from the Bureau of Economic Analysis (BEA), then multiply it by 2.63 to get household income. House prices are derived from the Freddie Mac House Price Index (FMHPI). It is a constant quality index over time, but not over areas. December 2000 is the base month when house price index in every MSA equals 100. To transform the house price index to a constant quality index both over time and over MSAs, we multiply this index by a MSA level hedonic estimated house value for a national average characteristic house using census 2000 data.

---

18 1984 is the CPI base year.

19 We regress log house value on a number of independent variable including state dummies, MSA dummies, # bedroom (third-order polynomial), # total rooms (third-order polynomial), categorical dummies of built year (8 categories), and categorical dummies of acreage of house (2 categories).
The number of housing units is derived from the housing stock estimated by decadal census and the annual permits data from the Department of Housing and Urban Development (HUD) according to

\[ N_{t+j} = N_t + \sum_{k=0}^{j-1} Permits_{t+k} (N_{t+10} - N_t). \] (2.30)

The MSAs in our sample are separated into a coastal group and an inland group. MSAs which contain a county that is within 200km from either the Pacific or Atlantic are defined as coastal. Other MSAs are inland. By this definition, the sample contains 114 coastal MSAs and 242 inland MSAs.

Table 2.2 presents the summary statistics for real per capita income, real house prices, housing units, and the changes of these three variables. On average, coastal MSAs have higher income, higher house prices, and a larger housing stock. These MSAs also have faster growth on average in income, house price, and the housing stock.

2.4.2 Methodology

We set the parameter in the user cost term that represents the after tax per dollar cost of mortgage borrowing, property taxes, depreciation and maintenance, and transaction costs \( \eta = 0.06 \) set the housing depreciation rate parameter.

After obtaining the coefficients, we set the independent variables to their median values, and predict the MSA level log house price for a national average characteristic house. Then, we exponentiate the estimated “log price” and get the price index of a constant-quality house which is comparable cross-sectionally.

We set the values of the variables in the user cost term as follows: real mortgage rate=0.04 (the average of the conventional conforming 30-year fixed real mortgage rate from Freddie Mac is 4.7% over 1975-2010), marginal income tax rate of the typical home buyer=25%, property tax rate=0.02, annual maintenance and depreciation costs rate=2%, transaction costs are assumed to be zero because our model assumes no moving costs for households. Then we have \( \eta_i = (r_i + \tau_i^p)(1 - \tau_i^p) + d_i + tc_i/n_i = 0.065. \)
\(\delta = 0.02\), and estimate the parameters \(\rho, \mu, c_1, \alpha\), and \(\lambda\). In the model, these parameters can differ for each MSA. But due to limited observations for each MSA (1980-2010 annually), the MSAs are separated into two groups (coastal and inland). We assume that MSAs in the same group have the same parameter values, and we estimate the parameters separately for each group. We allow for city-specific fixed effects within each group. That is, for each group, we estimate a panel data model with fixed effects.

For the rational expectations case, we adopt the sequential two-step GMM estimation methodology following Glaeser et al., but use different moment conditions. The parameters that need to be estimated are divided into a income related vector of parameters denoted by \(\theta = (\mu, \rho)\), and a vector which contains the rest of the parameters denoted by \(\gamma = (c_1, \alpha)\). Parameters in \(\theta\) are estimated using moment conditions derived from the income process equation. Given the estimated \(\hat{\theta}\), \(\gamma\) are then estimated using moment conditions derived from equations in Proposition 1. Estimate of \(\hat{\gamma}\) is consistent, but its asymptotic variance must be adjusted to take into account that it depends on the first step estimation. Newey and McFadden (1994) and Hansen (2007) provide the formula for the adjusted asymptotic variance of the second step estimators.

For the adaptive expectations case, the income parameters \(\theta = (\mu, \rho)\) are estimated using moment conditions derived from the income process equation. The rest parameters \(\gamma = (c_1, \alpha, \lambda)\) are estimated using moment conditions derived from equations in Proposition 2. The moment conditions to estimate \(\gamma\) do

\footnote{We do sequential GMM estimation instead of combining all moment conditions together and estimate the parameters simultaneously because we got more stable estimation results for the income parameters using only the income process equation. The same reasoning applies to the adaptive expectations case below.}
not depend on the income parameters $\theta$. Therefore, we implement the normal two-step GMM estimation procedure separately for $\theta$ and $\gamma$. Because the variables in the propositions are in trend-adjusted form, we add the deterministic trends back to the equations and use these equations to construct moment conditions.

### 2.4.3 Moment Conditions under Rational Expectation

The moment conditions under the rational expectations are based on (2.22), (2.23) in Proposition 1, and the exogenous income process (2.2). The derivation of the moment conditions are in Appendix 5.

First stage moment conditions are:

$$m_1(\theta) = \begin{cases} \Delta y_{i,t} - \rho \Delta y_{i,t-1} - (1 - \rho) \mu \\ [\Delta y_{i,t} - \rho \Delta y_{i,t-1} - (1 - \rho) \mu] \Delta y_{i,t-1} \\ [\Delta y_{i,t} - \rho \Delta y_{i,t-1} - (1 - \rho) \mu] \Delta y_{i,t-2} \end{cases}. \quad (2.31)$$

Second stage moment conditions are:

$$m_R(\theta) = \begin{cases} \psi_{i,t} \\ \psi_{i,t} \Delta y_{i,t-4} \\ \psi_{i,t} \Delta P_{i,t-4} \\ \psi_{i,t} \Delta N_{i,t-4} \\ \zeta_{i,t} \\ \zeta_{i,t} \Delta y_{i,t-3} \\ \zeta_{i,t} \Delta P_{i,t-3} \\ \zeta_{i,t} \Delta N_{i,t-3} \end{cases}, \quad (2.32)$$

where $\psi_{i,t}$ and $\zeta_{i,t}$ are defined as

$$\psi_{i,t} = -(1 - \rho)(1 - \phi)K_1 + \Delta N_{i,t} - (\rho + \phi) \Delta N_{i,t-1} + \rho \phi \Delta N_{i,t-2}, \quad (2.33)$$

$^{22}$The two-step GMM means we take the identity matrix as the weighting matrix in the first step estimation, and in the second step estimation the weighting matrix is computed from the first step estimation.
and
\[ \zeta_{it} = -(1 - \rho)[S_1 + \frac{\alpha + (1 - \delta - \phi)c_1}{1 + \eta}K_1] + \Delta P_{it} - \rho \Delta P_{i,t-1} \]
\[ + \frac{\alpha + (1 - \delta - \phi)c_1}{1 + \eta}(\Delta N_{it} - \rho \Delta N_{i,t-1}). \tag{2.34} \]

### 2.4.4 Moment Conditions under Adaptive Expectations

The moment conditions under the adaptive expectations are based on (2.28), (2.29) in Proposition 2, and the exogenous income process (2.2). The derivation of the moment conditions are in Appendix 6.

First stage moment conditions are:
\[ m_1(\theta) = \begin{cases} 
\Delta y_{it} - \rho \Delta y_{i,t-1} - (1 - \rho)\mu \\
\Delta y_{it} - \rho \Delta y_{i,t-1} - (1 - \rho)\mu \\
\Delta y_{it} - \rho \Delta y_{i,t-1} - (1 - \rho)\mu 
\end{cases} \tag{2.35} \]

Second stage moment conditions are:
\[ m_A(\theta) = \left\{ \begin{array}{l}
\chi_{it} \\
\chi_{it} \Delta y_{i,t-4} \\
\chi_{it} \Delta P_{i,t-4} \\
\chi_{it} \Delta N_{i,t-4} \\
\kappa_{it} \\
\kappa_{it} \Delta y_{i,t-3} \\
\kappa_{it} \Delta P_{i,t-3} \\
\kappa_{it} \Delta N_{i,t-3}
\end{array} \right\} \tag{2.36} \]

where \( \chi_{it} \) is defined as
\[ \chi_{it} = -\Delta P_{i,t-1} + \frac{c_1}{\lambda} \Delta N_{it} - \frac{(2 - \lambda - \delta)c_1}{\lambda} \Delta N_{i,t-1} + \frac{(1 - \lambda)(1 - \delta)c_1}{\lambda} \Delta N_{i,t-2}, \tag{2.37} \]

and \( \kappa_{it} \) is defined as
\[ \kappa_{it} = (1 + \eta - \lambda)\Delta P_{it} - (1 + \eta)(1 - \lambda)\Delta P_{i,t-1} + \Delta N_{it} - \alpha(1 - \lambda)\Delta N_{i,t-1} - \Delta y_{it} + (1 - \lambda)\Delta y_{i,t-1}. \tag{2.38} \]
2.4.5 Estimation Results

Table 2.3 reports the estimates of the income trend parameter $\mu$ and the persistence coefficient of income changes $\rho$. The coastal group has a higher income trend than the inland group. The persistence in income changes are also higher for the coastal group: the persistence coefficient is 0.24 for the coastal group while it is not significantly different from zero for the inland group.

Table 2.4 and Table 2.5 report the estimates of the rest of the parameters for the rational expectation case and the adaptive expectation case. For both cases, the coastal group has a bigger $c_1$ and a smaller $\alpha$ than the inland group. The parameter $c_1$ is inversely related to the responsiveness of construction to an expected house price change. Under rational expectations, the estimate of $c_1$ equals 1.5 for the coastal group and is 0.6 for the inland group. Under adaptive expectations, $c_1$ is 3.6 for the coastal group and is 0.4 for the inland group. Thus, the estimation results indicate that the coastal group has more inelastically supplied housing. This finding is consistent with the literature that housing supply is more inelastic in coastal cities because coastal areas have less developable land and more stringent land regulation. The parameter $\alpha$ is the disamenity coefficient. The estimate of $\alpha$ is 0.1 for the coastal group and 0.2 for the inland group under rational expectations, and is 0.04 for the coastal group and 0.07 for the inland group under adaptive expectations. The estimation results of $\alpha$ indicate that as housing units grow in number, the amenities decrease faster in inland cities than in coastal cities. The city expansion can have negative amenity effects due to congestion and pollution, and it may also have positive amenity effects due to agglomeration. The city expansion may have a larger effect in both congestion and
agglomeration for a coastal area than for an inland area, resulting in a smaller net
disamenity effect for coastal cities.

The adaptive speed parameter $\lambda$ is bigger in the coastal group, suggesting that
households in the coastal areas adjust their expectations of house price faster,
possibly because households living in places with more volatile house price cy-
cles pay more attention to house price changes and therefore adjust their expec-
tations faster.

2.5 Simulation

To examine how well the model can match the empirical features of house
prices raised in the beginning of this paper: rejecting a unit root, positive auto-
correlation for one-year price change, and negative autocorrelation for five-year
price change, we simulate house price series for the coastal group and the inland
group under both rational expectations case and the adaptive expectations case.

The estimates of the variance for the errors ($v$, $u$, and $w$) are derived from (E.6),
(E.9), and (E.11) for the rational expectations case, and (E.6), (F.1), and (F.2) for the
adaptive expectations case. After obtaining the estimates for the parameters and
the variance of the errors, house price series are simulated for 100 periods. We
then calculate the first-order autoregressive coefficients for one-year house price
change and five-year house price change, as well as the t-statistic value for the
Augmented Dickey-Fuller (ADF) unit root test with the number of lags ranging
from 0 through 3. The simulations are repeated 10,000 times, and we calculate
the mean of the autoregressive coefficients and unit root test statistics.
Table 2.6 reports the simulation results. The actual autoregressive coefficients for one-year and five-year house price changes are 0.66 and -0.61 for the coastal group and 0.44 and -0.11 for the inland group. The model under rational expectations does not match the data well. The variance of the error $u$ is negative for the coastal group and thus a house price series cannot be generated. This failure suggests that the model with rational expectations is not a good setting for coastal cities, or that the assumption that the errors are independent is violated for coastal cities under rational expectations. For the inland group under rational expectations, the ADF unit root test significantly rejects a unit root. A negative autoregressive coefficient of -0.48 is generated for the five-year price changes, which over predicts the magnitude of the actual coefficient. The model under rational expectations fails to generate the positive autocorrelation for the one-year price changes.

The model under adaptive expectations fits the empirical features better. For both the coastal group and the inland group, the ADF unit root test rejects a unit root, and the model generates positive autocorrelation for one-year price changes and negative autocorrelation for five-year price changes. But for the coastal group, the model does not generate enough magnitude of the one-year and five-year autoregressive coefficients. The one-year price change autoregressive coefficient predicted by the model is 0.1 which is smaller than the actual value of 0.66, and the five-year price change autoregressive coefficient is -0.32 from the model, which is smaller in absolute value than the actual coefficient of -0.61 from the data. For the inland group, the model matches relatively closely
with the data for the one-year price change autoregressive coefficient (0.37 predicted by the model and 0.44 from the data), but over generates the magnitude of the five-year change autoregressive coefficient (-0.75 from the model and -0.11 from the data).

The above comparisons suggest that the model under rational expectations captures the negative autocorrelation of house price changes over the five-year horizon and the rejection of a unit root in house prices for the inland cities, but it completely fails in explaining the short run momentum of house price changes, which is consistent with the finding in Glaeser et al. When we change the assumption of expectations from rational to adaptive, the model’s ability of matching data improves. For both coastal and inland groups, the model rejects a unit root and generates a negative five-year autocorrelation. Moreover, the model succeeds in predicting a positive autocorrelation of house price changes over the one-year horizon.

2.6 Conclusion

This paper modifies the dynamic spatial equilibrium model in Glaeser et al. (2012) to study the stationarity of house prices, and the short-run momentum and the long-run reversion of house price changes. The model is solved and estimated separately under rational expectations and under adaptive expectations, and then predictions of the model are used to fit the data. The estimation results reveal that, under both rational expectations and adaptive expectations, coastal
areas have higher persistence in income growth, more housing supply inelasticity, and a smaller disamenity coefficient. Moreover, in the case of adaptive expectations, the coastal city group has a higher adjustment speed in the expectation formation.

The model under adaptive expectations matches the data better than the one with rational expectations. When expectations are assumed to be formed adaptively, the model succeeds in accounting for the rejection of a unit root and the positive one-year autocorrelation and negative five-year autocorrelation of house price changes. But the model underpredicts the magnitude of autocorrelations over both the one-year and five-year horizons for the coastal areas, and overpredicts the magnitude of five-year autocorrelation for the inland city group.
2.7 Figures and Tables

Figure 2.1: U.S. log real per capita income and log real house price index, 1975-2010

Notes: HPI is from FMHPI, and per capita income is from BEA. The vertical axis are scaled so that the two series cross each other in 1975.
Figure 2.2: Log of real house prices, 1975-2010

Notes: House prices are derived from FMHPI.
Figure 2.3: Impulse response to an income shock under rational expectations

Notes: Presented are the paths of de-trended house prices after a one time income shock of 1000 dollars. The solid lines are the baseline case with parameter values being set to $\rho = 0.2$, $c_1 = 2$, and $\alpha = 0.1$. 
Figure 2.4: Impulse response to an income shock under adaptive expectations

Notes: Presented are the paths of de-trended house prices after a one time income shock of 1000 dollars. The solid lines are the baseline case with parameter values being set to $\rho = 0.2$, $c_1 = 2$, $\alpha = 0.1$, and $\lambda = 0.9$. 
Table 2.1: CIPS panel unit root test, 1975-2010

<table>
<thead>
<tr>
<th></th>
<th>CADF(1)</th>
<th>CADF(2)</th>
<th>CADF(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>coastal real house price</td>
<td>-3.131***</td>
<td>-2.684**</td>
<td>-3.076***</td>
</tr>
<tr>
<td>inland real house price</td>
<td>-2.825***</td>
<td>-2.823***</td>
<td>-2.702***</td>
</tr>
<tr>
<td>coastal real income</td>
<td>-2.060</td>
<td>-1.922</td>
<td>-1.911</td>
</tr>
<tr>
<td>inland real income</td>
<td>-1.917</td>
<td>-1.851</td>
<td>-1.856</td>
</tr>
</tbody>
</table>

Notes: Truncated CIPS with an intercept and a trend are reported. 1% 5% and 10% critical values for T=20 and N=100 are -2.70, -2.57, -2.51, for N=200 are -2.65, -2.55, -2.49.
Table 2.2: Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>coastal</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>min</td>
<td>max</td>
<td>mean</td>
<td>s.d.</td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td>y</td>
<td>15803</td>
<td>3663</td>
<td>7688</td>
<td>38650</td>
<td>14249</td>
<td>2471</td>
<td>6332</td>
<td>31099</td>
</tr>
<tr>
<td>p</td>
<td>98588</td>
<td>55021</td>
<td>35252</td>
<td>557720</td>
<td>56112</td>
<td>16547</td>
<td>21408</td>
<td>172753</td>
</tr>
<tr>
<td>n</td>
<td>375421</td>
<td>828067</td>
<td>8257</td>
<td>7500000</td>
<td>191829</td>
<td>355931</td>
<td>11767</td>
<td>3800000</td>
</tr>
<tr>
<td>dy</td>
<td>187</td>
<td>463</td>
<td>-4408</td>
<td>5188</td>
<td>164</td>
<td>388</td>
<td>-8670</td>
<td>6457</td>
</tr>
<tr>
<td>dp</td>
<td>851</td>
<td>11798</td>
<td>-116216</td>
<td>69486</td>
<td>8</td>
<td>3099</td>
<td>-43321</td>
<td>31304</td>
</tr>
<tr>
<td>dn</td>
<td>5280</td>
<td>9061</td>
<td>-188</td>
<td>77787</td>
<td>2840</td>
<td>6713</td>
<td>-515</td>
<td>82627</td>
</tr>
</tbody>
</table>

Notes: Summary statistics for real income, real house prices, housing stock, and their changes.
Table 2.3: Income parameters

<table>
<thead>
<tr>
<th></th>
<th>$\mu$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal</td>
<td>488.78</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>(26.349)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>Inland</td>
<td>422.49</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(11.455)</td>
<td>(0.060)</td>
</tr>
</tbody>
</table>

Notes: Standard errors are in the parentheses.
Table 2.4: Rational expectations

<table>
<thead>
<tr>
<th></th>
<th>$c_1$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal</td>
<td>1.487</td>
<td>0.111</td>
</tr>
<tr>
<td></td>
<td>(0.215)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Inland</td>
<td>0.567</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.014)</td>
</tr>
</tbody>
</table>

Notes: Standard errors are in the parentheses.
Table 2.5: Adaptive expectations

<table>
<thead>
<tr>
<th></th>
<th>$c_1$</th>
<th>$\alpha$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal</td>
<td>3.557 (0.585)</td>
<td>0.041 (0.004)</td>
<td>0.969 (0.008)</td>
</tr>
<tr>
<td>Inland</td>
<td>0.428 (0.188)</td>
<td>0.065 (0.008)</td>
<td>0.661 (0.029)</td>
</tr>
</tbody>
</table>

Notes: Standard errors are in the parentheses.
<table>
<thead>
<tr>
<th></th>
<th>1-year</th>
<th>5-year</th>
<th>ADF(0)</th>
<th>ADF(1)</th>
<th>ADF(2)</th>
<th>ADF(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rational Coastal</td>
<td>-0.395</td>
<td>-0.476</td>
<td>-6.910</td>
<td>-5.074</td>
<td>-4.290</td>
<td>-3.844</td>
</tr>
<tr>
<td>Inland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive Coastal</td>
<td>0.105</td>
<td>-0.315</td>
<td>-2.276</td>
<td>-2.671</td>
<td>-2.672</td>
<td>-2.555</td>
</tr>
<tr>
<td>Inland</td>
<td>0.373</td>
<td>-0.754</td>
<td>-3.188</td>
<td>-5.581</td>
<td>-6.904</td>
<td>-6.587</td>
</tr>
</tbody>
</table>

Notes: The 5% and 1% critical values for the ADF unit root test are -1.95 and -2.60 (no intercept, no trend, T=100). The autoregressive coefficients for 1-year and 5-year house price changes from the data are 0.66 and -0.61 for the coastal group, and 0.44 and -0.11 for the inland group, over the period 1975-2010.
Chapter 3: House Price Expectations: Unbiasedness and Efficiency of Forecasters

3.1 Introduction

House price expectations formation is of vital importance in housing economics. Expectations of house price are obviously important when individuals make decisions about buying houses, when they decide whether to default on their mortgages, and when corporations make investment decisions in real estate markets.

However, the understanding of how house price expectations are formed is quite limited. There are essentially two ways to examine how expectations are formed. One way is based on the predictions from economics models combined with assumptions about the nature of expectations formation. But the problem with this method is that it is a joint test of the assumption about the expectations formation and the correctness of the model, and therefore it involves an identification problem. For example, a rejection of the predictions may be caused by the failure of the model, not the failure of the assumption about how expectations are formed. The second way to test expectations formation is directly using survey data. Although there are a large amount of papers studying expectation surveys
for major macroeconomic variables such as GDP and the inflation rate, studies of house price expectations are quite rare. This is mainly because expectations about house price changes did not receive much attention until the burst of the recent housing bubble. Surveys of house price expectations generally started to be collected around or after the peak of the bubble.

This paper is the first one to rigorously examine the rationality of house price expectations using survey data. Specifically, I test for the unbiasedness and efficiency of forecasters’ predictions about the percentage change in house prices using the Wall Street Journal economic forecasting survey. The house price change expectation data from the Wall Street Journal survey is a “three-dimensional” panel dataset and thus errors are correlated over three dimensions. I apply the econometrics methodology proposed by Davies and Lahiri (1995) to analyze this panel dataset.

The remainder of the paper is organized as follows. The next section describes the data. Then I present the methodology for the unbiasedness and efficiency tests. Next, I report the empirical results, and the last section concludes.

Pesaran and Weale (2006) provide a comprehensive review of theoretical and empirical work on survey expectations.

To the best of my knowledge, the only two papers that study house price expectation using survey data are Case, Shiller, and Thompson (2012) for U.S. housing markets and Howard and Karagedikli (2012) for New Zealand housing markets. However, the focus of their studies is different from here. Instead of testing unbiasedness and efficiency, they focus on descriptive analyses of survey answers.
3.2 Data

The dataset that I use is the Wall Street Journal economic forecasting survey (hereafter, WSJ survey)\textsuperscript{25} This survey collects, in the first half of each month, the predictions of several U.S. macroeconomic variables from 50 to 60 forecasters. This is a panel dataset, but is unbalanced because forecasters enter and leave the survey, or fail to submit answers. Since August 2006, the survey has asked the forecasters to predict the annual percentage change of the U.S. Federal Housing Finance Agency (FHFA) house price index for the current and the next year\textsuperscript{26}

I include in the dataset 6 target years over 2007-2012, 24 forecast horizons ranging from 24 months to 1 month, and 47 forecasters who submitted answers for at least 50\% of all the possible observations. This yields a total of 4,925 observations. The dates of the forecasts submitted are from August 2006 through December 2012 (the survey was not conducted in several months during this period).

\textsuperscript{25}The WSJ survey data can be found at http://online.wsj.com/public/resources/documents/informationflash08.html?project=EFORECAST07

\textsuperscript{26}The WSJ survey online data source does not indicate which FHFA index forecasters were asked to predict, but it provides the “actual” percentage changes for each year. Comparing the “actual” percentage changes with all the indices available on the FHFA website, I find that the seasonally unadjusted quarterly purchase-only index is the closest. For year 2007 to 2012, the “actual” percentage changes provided on the WSJ survey website (as on 6/30/2013) are -2.4, -9.68, -2.12, -4.26, -2.38, 5.45, and the percentage changes of the seasonally unadjusted quarterly purchase-only index are -2.4, -9.65, -2.08, -4.2, -2.35, and 5.47 (based on the data downloaded from the FHFA website on 3/3/2013). Therefore, when I mention the “actual” house price index in the remainder of the paper, I refer to the seasonally unadjusted quarterly purchase-only index.

Note that the FHFA indices may have minor difference, depending on the date the data is downloaded from the website. This is because the FHFA indices are constructed by a repeated-sale methodology. Therefore, as new transactions occur and are matched with previous transactions on the same property, these new transactions will be included in the dataset used to construct the repeated-sale indices, and hence the FHFA indices are under constant minor revision.
In the appendix, I describe three other surveys in U.S. that include house price expectations, and explain why the WSJ survey is chosen in this paper.

3.3 Econometrics Methodology

The implementation of the unbiasedness test and the efficiency test requires a model of the forecast errors and an estimate of the covariance matrix. Given the estimated error covariance matrices, I then implement unbiasedness and efficiency tests.

3.3.1 Forecast Error Covariance Matrix

By the nature of the house price question in the survey, this panel dataset has a “three-dimensional” forecast error structure proposed by Davies and Lahiri (1995). The first dimension of error correlation is due to the fact that all forecasters would be affected at the same time by aggregate shocks. The second source of error correlation is from shocks that affect forecasting errors for the same target year at different forecasting horizons. For example, a person's forecast errors of the annual percentage change for year 2009 made at October 2009 and November 2009 are both affected by monthly shocks in November and December 2009. The third dimension of the error correlation is caused by monthly shocks that are common to adjacent target years. For example, a person's forecast errors of the annual percentage change in price for year 2009 made at December 2009 and for year 2010 made at December 2009 are both affected by the shock occurs in December 2009.
I adopt the econometric methodology developed in Davies and Lahiri (1995) to decompose the forecast errors. Following their notation, there are \( N \) individuals, \( T \) target years, and \( H \) forecast horizons. Denote \( F_{ith} \) to be the forecast made by individual \( i \), for year \( t \), at \( h \) months before year \( t \) ends. The forecast data are compiled in the vector \( F' = (F_{11H}, \ldots, F_{111}, F_{12H}, \ldots, F_{121}, \ldots, F_{1TH}, \ldots, F_{1T1}, \ldots, \ldots, F_{N1H}, \ldots, F_{N11}, F_{N2H}, \ldots, F_{N21}, \ldots, F_{NTH}, \ldots, F_{NT1}) \). That is, the data are sorted first by individual forecasters, then by target years, and last by forecast horizons with the forecast horizons being sorted in descending order. Denote \( A_t \) to be the actual house price percentage change for year \( t \). The forecast errors are decomposed as

\[
A_t - F_{ith} = \phi_i + \lambda_{th} + \varepsilon_{ith},
\]

(3.1)

\[
\lambda_{th} = \sum_{j=1}^{h} u_{tj}.
\]

(3.2)

In the above equations, \( \phi_i \) is individual specific bias, \( \varepsilon_{ith} \) is idiosyncratic error, and \( \lambda_{th} \) is aggregate shock that is the accumulation of the monthly shocks \( u_{tj} \) that occur over the span of \( h \) months prior to the end of year \( t \). \( \varepsilon_{ith} \) and \( \lambda_{th} \) are uncorrelated, \( \varepsilon_{ith} \) is white noise across all dimensions with \( E(\varepsilon_{ith}^2) = \sigma_{\phi_i}^2 \), and \( u_{tj} \) is white noise with \( E(u_{tj}^2) = \sigma_u^2 \).

A positive \( \phi_i \) indicates that the individual is persistently underestimating house price changes and a negative \( \phi_i \) indicates that she is persistently overestimating it, after taking account of the aggregate shocks. The monthly shocks \( u_{tj} \) could be caused by events such as the Federal Reserve unexpectedly reducing interest rates. The idiosyncratic error \( \varepsilon_{ith} \) could be the result of errors in information collection, forecasting and calculation techniques, or private information.
The covariance between two forecast errors is therefore

\[
\text{cov}(A_{t_1} - F_{i_1t_1h_1}, A_{t_2} - F_{i_2t_2h_2})
\]

(3.3)

\[
= \text{cov}(\lambda_{t_1h_1} + \varepsilon_{i_1t_1h_1}, \lambda_{t_2h_2} + \varepsilon_{i_2t_2h_2})
\]

\[
= \text{cov}(\sum_{j_1=1}^{h_1} u_{t_1j_1} + \varepsilon_{i_1t_1h_1}, \sum_{j_2=1}^{h_2} u_{t_2j_2} + \varepsilon_{i_2t_2h_2})
\]

\[
= \begin{cases} 
\sigma^2_{\varepsilon_i} + h\sigma^2_u, & \forall i_1 = i_2 = i, t_1 = t_2, h_1 = h_2 = h, \\
\min(h_1, h_2)\sigma^2_u, & \forall i_1 \neq i_2, t_1 = t_2, \text{ or } i_1 = i_2, t_1 = t_2, h_1 \neq h_2, \\
\min(h_1, h_2 - 12)\sigma^2_u, & \forall t_2 = t_1 + 1, h_2 > 12, \\
0, & \text{otherwise}. 
\end{cases}
\]

Applying the above equation to my case, the forecast error covariance matrix \((\Sigma)\) can be written as

\[
\Sigma = \begin{bmatrix}
A_1 & B & \cdots & B \\
B & A_2 & \cdots & B \\
\vdots & \vdots & \ddots & \vdots \\
B & B & \cdots & A_N
\end{bmatrix}_{N \times N},
\]

where

\[
A_i = \sigma^2_{\varepsilon_i}I_{TH} + B,
\]

\[
B = \begin{bmatrix}
b & c & 0 & 0 & \cdots & 0 & 0 \\
c' & b & c & 0 & \cdots & 0 & 0 \\
0 & c' & b & c & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & c' & b & c \\
0 & \cdots & 0 & c' & b
\end{bmatrix}_{T \times T},
\]

\[
b = \sigma^2_u \begin{bmatrix}
24 & 23 & 22 & \cdots & 2 & 1 \\
23 & 23 & 22 & \cdots & 2 & 1 \\
22 & 22 & 22 & \cdots & 2 & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
2 & 2 & 2 & \cdots & 2 & 1 \\
1 & 1 & 1 & \cdots & 1 & 1
\end{bmatrix}_{H \times H}.
\]

\(^{27}N\) is 47, \(T\) is 6, and \(H\) is 24 below, which are different from Davies and Lahiri, and therefore, the matrices \(b\) and \(c\) below are different from theirs.
Therefore, the forecast error covariance matrix is fully characterized by \( N + 1 \) parameters \( \sigma_u^2 \) and \( \sigma_{\varepsilon_i}^2 \). To estimate this matrix (\( \Sigma \)), Davies and Lahiri (1995) propose the consistent estimates:

\[
\frac{1}{TH} \sum_{t=1}^{T} \sum_{h=1}^{H} (A_t - F_{ith}) = \hat{\phi}_i, \tag{3.5}
\]

\[
\frac{1}{N} \sum_{i=1}^{N} (A_t - F_{ith} - \hat{\phi}_i) = \hat{\lambda}_{th}, \tag{3.6}
\]

\[
A_t - F_{ith} - \hat{\phi}_i - \hat{\lambda}_{th} = \hat{\varepsilon}_{ith}. \tag{3.7}
\]

Consistent estimates of \( \sigma_{\varepsilon_i}^2 \) are obtained by regressing \( \hat{\varepsilon}_{ith}^2 \) on \( N \) individual dummies because \( E(\hat{\varepsilon}_{ith}^2) = \sigma_{\varepsilon_i}^2 \), and a consistent estimate of \( \sigma_u^2 \) is obtained by regressing \( \hat{\lambda}_{th}^2 \) (a \( TH \times 1 \) vector) on a horizon index ranging from 24 to 1 because \( E(\hat{\lambda}_{th}^2) = h\sigma_u^2 \).

Because this dataset is unbalanced, the data matrix and forecast error covariance matrix are compressed by deleting each row and column if the corresponding observation in the forecast vector is missing. These compressed matrices are used in the regressions below.
3.3.2 Test for Unbiasedness

With the consistent estimate of the forecast error covariance $\Sigma$ mentioned above, the test for unbiasedness is to run an OLS regression of Equation (3.1) to get estimates of the $\phi_i$’s. The covariance of the estimators is given by the formula

$$(Z'Z)^{-1}Z'\Sigma Z(Z'Z)^{-1}$$

where $Z$ is the matrix of regressors.

3.3.3 Test for Efficiency

The test for efficiency is determining if the forecast error and variables in the information set known by the forecasters at the time the forecast is made are correlated. That is, the forecasters are efficient if those available information cannot improve forecast accuracy. Specifically, rejecting the null hypothesis of $\delta = 0$ in the regression below indicates a rejection of the efficiency hypothesis

$$A_t - F_{ith} = \delta X_{t,h+1} + \phi_i + \lambda_{th} + \varepsilon_{ith},$$

(3.8)

where the $X_{t,h+1}$ represents information known by the forecaster when she makes the forecast $F_{ith}$, and $X_{t,h+1}$ contains any publicly available economic variables or previous forecasts made by the forecaster.

Davies and Lahiri (1995) propose to take the first difference of (3.8) to eliminate the individual dummies $\phi_i$ and get the following regression equation for the efficiency test:

$$F_{ith} - F_{i,t,h+1} = -\delta (X_{t,h+1} - X_{t,h+2}) + u_{t,h+1} - \varepsilon_{ith} + \varepsilon_{i,t,h+1},$$

(3.9)

where $(X_{t,h+1} - X_{t,h+2})$ should be uncorrelated with the error term $u_{t,h+1} - \varepsilon_{ith} + \varepsilon_{i,t,h+1}$. 
To estimate $\delta$ in (3.9), I first need to estimate the error covariance matrix for (3.9).

A typical element in the covariance matrix $\Omega$ in (3.9) is

$$cov(u_{t_1,h_1+1} - \varepsilon_{i_1,t_1,h_1} + \varepsilon_{i_1,t_1,h_1+1}, u_{t_2,h_2+1} - \varepsilon_{i_2,t_2,h_2} + \varepsilon_{i_2,t_2,h_2+1}) = \begin{cases} 
\sigma^2_u + 2\sigma^2_{\varepsilon_i}, & \forall i_1 = i_2 = i, t_1 = t_2, h_1 = h_2, \\
\sigma^2_u, & \forall i_1 \neq i_2, t_1 = t_2, h_1 = h_2, \\
-\sigma^2_{\varepsilon_i}, & \forall i_1 = i_2 = i, t_1 = t_2, |h_1 - h_2| = 1, \\
0, & \text{otherwise.}
\end{cases}$$

(3.10)

Therefore, the covariance matrix $\Omega$ can be written as

$$\Omega_{(NTH \times NTH)} = \begin{bmatrix} A_1 & B & \ldots & B \\
B & A_2 & \ldots & B \\
\vdots & \vdots & \ddots & \vdots \\
B & B & \ldots & A_N \end{bmatrix}_{N \times N},$$

where

$$A_i_{(TH \times TH)} = 2\sigma^2_{\varepsilon_i}I_{TH} + B = \sigma^2_uI_{TH},$$

$$B = \sigma^2_uI_{TH},$$

$$C_{H \times H} = -\sigma^2_{\varepsilon_i} \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\
1 & 0 & 1 & \ldots & 0 & 0 \\
0 & 1 & 0 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 1 & 0 & \ldots & 0 & 1 \\
0 & \ldots & 1 & 0 & \ldots & 0 \end{bmatrix}_{H \times H}.$$  

(3.11)

After obtaining the consistent estimates of $\sigma^2_{\varepsilon_i}$ and $\sigma^2_u$ as previously mentioned, I obtain a consistent estimate of $\Omega$. I then apply OLS to (3.9) with the covariance of the estimators given by the formula $(Z'Z)^{-1}Z'\Omega(Z'Z)^{-1}$ where $Z$ is the matrix of regressors.
3.4 Empirical Results

Figure 3.1 provides a descriptive view of the forecasters’ performance. The graphs show the box-and-whisker plots of the forecasts of house price changes in each month for every target year. The bottom and top of the box are the first and third quartiles, and the band inside the box is the median. The vertical lines extending from the box are called whiskers and cover most or all the remaining data. The upper (lower) whisker is restricted to the upper (lower) quartile plus (minus) 1.5 times the interquartile range. Dots outside the whiskers are outliers.

Three patterns appear here. First, the forecasters seem generally conservative when the aggregate shocks are not controlled. During the 2007-2011 period when house price is declining, the predictions of house price changes are higher than the actual changes, except for the predictions for target year 2009 made during year 2009. During this period, the forecasters are overly pessimistic about 2009 house price changes. During 2012, house price changes become positive, but the forecasters predicted lower house price appreciation. Second, as forecasting horizons approach zero, the forecast error generally diminishes, although they still deviate from the actual values for some years. Third, boxes for target years 2007 to 2009 are bigger than boxes for target years 2011 and 2012, indicating that the disagreement among forecasters was larger during the housing bust, and their disagreement was smaller when the housing market reached the bottom and recovered.
3.4.1 Unbiasedness

Table 3.1 provides estimates of the variance of the idiosyncratic error $\sigma^2_{\varepsilon_i}$, the individual bias $\phi_i$, and the standard errors of $\hat{\phi}_i$ for the 47 forecasters. The estimate of the variance of the monthly aggregate shock $\sigma^2_u$ is 0.84. Note that, when I take account of the aggregate shocks, the majority of the forecasters have negative bias, meaning that their predictions of house price changes are systematically higher than the actual changes. Only 9 out of the 47 forecasters have biases that are statistically significant, and the bias of all of these 9 forecasters are negative, that is, they are persistently predicting too high house price changes.

The estimates of $\sigma^2_{\varepsilon_i}$ and $\sigma^2_u$ reveal some information about the sources of forecasting error, that is, the relative contributions of the idiosyncratic error and the aggregate shock to the forecasting error. According to Equation (3.1) and (3.2), the variance of the accumulated aggregate shock decreases as the forecast horizon decreases (24 times $\sigma^2_u$ when the horizon is 24 and just $\sigma^2_u$ when the horizon is 1). In my result, the average of the estimates of the variance of the idiosyncratic error $\sigma^2_{\varepsilon_i}$ over $i$ is 5.71, the variance of the accumulated aggregate shock is 20.16 (24 times 0.84) when it is 24 months prior to the end of the target year and it decreases to 0.84 when it is only one month before the target year ends. Therefore, about 78 percent ($20.16/(5.71+20.16)$) of the average variance of the forecasting error is contributed by the accumulated aggregated shock when it is 24 months before the target year ends, and the contribution of aggregate shock one month before the target year ends decreases to only 13 percent ($0.84/(5.71+0.84)$).
3.4.2 Efficiency

I run separate regressions for (3.9), including one of the following variables as the exogenous regressor \((X_{t,h+1} - X_{t,h+2})\) at each time: (1) the individual’s forecast revision, lagged two months; (2) change in the percentage change of the monthly FHFA house price index over the past 12 months, lagged three months; (3) change in the conventional conforming 30-year fixed mortgage rate released in the last week each month, lagged two months; (4) change in NAHB/Wells Fargo national Housing Market Index, lagged two months; and (5) change in housing starts, lagged two months.

The above five variables are chosen as the latest information that are “available to the forecasters \(h+1\) months prior to the end of the year \(t\),” as in Davies and Lahiri (1995), so they are contemporaneously uncorrelated with the error term. The sources and descriptions of these five variables are listed below.

(1) Because the individual’s own previous forecast should be in her information set, I use the two months lagged forecast revision \((F_{i,t,h+2} - F_{i,t,h+3})\) as an exogenous regressor \((X_{t,h+1} - X_{t,h+2})\). I do not use the change in one period lagged forecast revision \((F_{i,t,h+1} - F_{i,t,h+2})\) because it is correlated with the error terms \((F_{i,t,h+1} is correlated with \(\varepsilon_{i,t,h+1}\)). Therefore, \((F_{i,t,h+1} - F_{i,t,h+2})\) is not a valid exogenous regressor.

(2) The monthly FHFA house price index is released at the end of two months later.\(^{28}\) For example, January’s index is released in late March. Therefore, when forecasters make predictions in early April, January’s house price index is in their information set.

\(^{28}\)I use seasonally unadjusted national purchase-only index.
(3) The data of the conventional conforming 30-year fixed mortgage rate is released by Freddie Mac weekly on Thursday.\textsuperscript{29} I use the change in the mortgage rate released on the last Thursday lagged two months as the exogenous regressor. I do not use the rate released late last month because although it is in the forecasters’ information set, it is correlated with $u_{t,h+1}$ in the error term.

(4) The NAHB/Wells Fargo national Housing Market Index (HMI) is a seasonally adjusted series derived from a monthly survey of NAHB members.\textsuperscript{30} It reflects builders’ views of housing market conditions. Each month’s index is released around the 16th-20th of the same month, and I use the change in the HMI lagged two months as the exogenous regressor, due to the same reason as in (3).

(5) The housing starts measure is a monthly data which is released around the 16th-20th of the next month.\textsuperscript{31} When forecasters answer the survey in, say, early April, the housing starts in February is in their information set, so I use the change in housing starts lagged two month as the regressor.

Table 3.2 reports the separate regressions for the efficiency test using each of the above regressors. None of the regression can reject the efficiency hypothesis. This result indicates that none of the available information improve the forecast accuracy. The forecasters fully incorporate this information when they make predictions.

I further conduct a joint test of efficiency by estimating equation \textsuperscript{(3.9)}, including all the variables (1) to (5) in the regression at the same time. Results are

\textsuperscript{29}http://www.freddiemac.com/pmms/
\textsuperscript{30}http://www.nahb.org/reference_list.aspx?sectionID=134
\textsuperscript{31}I use seasonally adjusted housing starts series. The number of housing units are measured in thousands. The data can be found at http://www.census.gov/construction/nrc/historical_data/
presented in Table 3.3. None of the coefficients are significantly from zero. The null hypothesis that the five coefficients are jointly zero cannot be rejected by the $\chi^2$ statistic. Therefore, the efficiency hypothesis cannot be rejected.

### 3.4.3 Excluding Year 2012

I also conduct the unbiasedness test and the efficiency test limited to target years 2007 to 2011, excluding 2012. As shown in Figure 3.1, 2012 is the only year that has positive house price changes, and is the only year that the forecasts are lower than the actual value.

Table 3.4 presents the unbiasedness test results for year 2007 to 2011. A noticeable point is that many more forecasters have statistically significant bias than the case that includes 2012, and all those biases are negative. Now 25 out of the 47 forecaster are systematically over predicting house price changes from 2007 through 2011.

Table 3.5 reports the efficiency test results that include only one regressor in each regression. Table 3.6 reports the joint efficiency test results that include all the regressors at the same time. Similar to the case that includes 2012, none of the coefficients is significant, and the $\chi^2$ statistic indicates that we cannot reject the null hypothesis that the coefficients are jointly zero. Therefore, efficiency cannot be rejected, indicating that the forecasters fully incorporate this information when they made predictions during 2007-2011.

---

$^{32}$Under the null hypothesis $H_0 : \delta = (\delta_1, ..., \delta_5)' = 0$, we have $\hat{\delta}'[(Z'Z)^{-1}Z'\Omega Z(Z'Z)^{-1}]^{-1}\hat{\delta}$ being approximately $\chi^2(5)$. 

3.5 Conclusions

This paper examines the unbiasedness and efficiency of forecasters when they predict house price changes. Using the Wall Street Journal economic forecasting survey that covers 2007-2012, and implementing the econometric methodology proposed in Davies and Lahiri (1995) to deal with a “three-dimensional” panel dataset, I find that, after controlling for aggregate shocks, 9 out of the 47 forecasters have statistically significant biases, and the biases are all negative, indicating that they persistently predict higher than the actual house price changes. For the efficiency test, I examine whether the following information can improve forecast accuracy: the forecaster’s own forecast lagged two months; change in FHFA monthly house price index over the past 12 months lagged three months; mortgage rate lagged two months; NAHB/Wells Fargo national Housing Market Index lagged two months; and housing starts lagged two month. The hypothesis of efficiency cannot be rejected in any case, indicating that the forecasters have fully incorporated these information when they make predictions. If the target year 2012 is excluded, 25 out of the 47 forecasters have significant biases and all these biases are negative. The efficiency tests results are similar to the case that includes 2012, that is, the hypothesis of efficiency cannot be rejected.
3.6 Figures and Tables

Figure 3.1: Box-and-Whisker plots for each target year

Notes: The solid lines are the actual annual percentage changes of house price (FHFA). The graphs show the distribution of forecasts made in each month. The bottom and top of the box are the first and third quartiles, and the band inside the box is the median. The upper (lower) whisker is restricted to the upper (lower) quartile plus (minus) 1.5 times the interquartile range. Dots outside the whiskers are outliers.
Table 3.1: Test for Unbiasedness, 2007-2012

<table>
<thead>
<tr>
<th>Forecaster</th>
<th>Institute</th>
<th>$\sigma^2_{\varepsilon_i}$</th>
<th>$\phi_i$</th>
<th>s.e. of $\phi_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bart van Ark</td>
<td>The Conference Board</td>
<td>6.21</td>
<td>0.53</td>
<td>1.52</td>
</tr>
<tr>
<td>Paul Ashworth</td>
<td>Capital Economics</td>
<td>6.60</td>
<td>0.20</td>
<td>1.39</td>
</tr>
<tr>
<td>Nariman Behravesh</td>
<td>Global Insight</td>
<td>6.79</td>
<td>1.02</td>
<td>1.36</td>
</tr>
<tr>
<td>Richard Berner</td>
<td>Morgan Stanley</td>
<td>6.81</td>
<td>0.47</td>
<td>1.62</td>
</tr>
<tr>
<td>Ram Bhagavatula</td>
<td>Combinatorics Capital</td>
<td>9.54</td>
<td>0.64</td>
<td>1.38</td>
</tr>
<tr>
<td>Jay Brinkmann</td>
<td>Mortgage Bankers Association</td>
<td>3.72</td>
<td>-0.61</td>
<td>1.49</td>
</tr>
<tr>
<td>Joseph Carson</td>
<td>AllianceBernstein</td>
<td>2.45</td>
<td>-1.40</td>
<td>1.34</td>
</tr>
<tr>
<td>Mike Cosgrove</td>
<td>Econoclast</td>
<td>3.18</td>
<td>-1.13</td>
<td>1.35</td>
</tr>
<tr>
<td>Lou Crandall</td>
<td>Wrightson ICAP</td>
<td>2.77</td>
<td>-2.10</td>
<td>1.34</td>
</tr>
<tr>
<td>J. Dewey Daane</td>
<td>Vanderbilt University</td>
<td>5.68</td>
<td>-1.61</td>
<td>1.35</td>
</tr>
<tr>
<td>Richard DeKaser</td>
<td>National City Corporation</td>
<td>5.20</td>
<td>-3.17</td>
<td>1.64*</td>
</tr>
<tr>
<td>Douglas Duncan</td>
<td>Mortgage Bankers Association</td>
<td>3.67</td>
<td>-0.52</td>
<td>1.34</td>
</tr>
<tr>
<td>Stephen Gallagher</td>
<td>Societe Generale</td>
<td>6.60</td>
<td>-4.52</td>
<td>1.69***</td>
</tr>
<tr>
<td>Ethan S. Harris</td>
<td>Lehman Brothers</td>
<td>5.36</td>
<td>-1.02</td>
<td>1.46</td>
</tr>
<tr>
<td>Maury Harris</td>
<td>UBS</td>
<td>11.60</td>
<td>0.16</td>
<td>1.43</td>
</tr>
<tr>
<td>Tracy Herrick</td>
<td>The Private Bank</td>
<td>11.80</td>
<td>-0.04</td>
<td>1.38</td>
</tr>
<tr>
<td>Stuart Hoffman</td>
<td>PNC Financial Services Group</td>
<td>1.91</td>
<td>-0.30</td>
<td>1.48</td>
</tr>
<tr>
<td>Gene Huang</td>
<td>FedEx Corp.</td>
<td>5.21</td>
<td>-1.31</td>
<td>1.36</td>
</tr>
<tr>
<td>William B. Hummer</td>
<td>Wayne Hummer Investments</td>
<td>26.04</td>
<td>0.15</td>
<td>1.42</td>
</tr>
<tr>
<td>Dana Johnson</td>
<td>Comerica Bank</td>
<td>3.29</td>
<td>-4.40</td>
<td>1.55***</td>
</tr>
<tr>
<td>Bruce Kasman</td>
<td>JP Morgan Chase &amp; Co.</td>
<td>3.68</td>
<td>-0.86</td>
<td>1.35</td>
</tr>
<tr>
<td>Paul Kasriel</td>
<td>The Northern Trust</td>
<td>11.10</td>
<td>0.20</td>
<td>1.50</td>
</tr>
<tr>
<td>Joseph LaVorgna</td>
<td>Deutsche Bank Securities, Inc.</td>
<td>13.35</td>
<td>-1.09</td>
<td>1.68</td>
</tr>
<tr>
<td>Edward Leamer</td>
<td>UCLA Anderson Forecast</td>
<td>4.03</td>
<td>0.02</td>
<td>1.34</td>
</tr>
<tr>
<td>John Lonski</td>
<td>Moody’s Investors Service</td>
<td>2.27</td>
<td>-1.12</td>
<td>1.34</td>
</tr>
<tr>
<td>Dean Maki</td>
<td>Barclays Capital</td>
<td>2.92</td>
<td>-1.04</td>
<td>1.40</td>
</tr>
<tr>
<td>David Malpass</td>
<td>Encima Global LLC</td>
<td>3.52</td>
<td>-2.04</td>
<td>1.60</td>
</tr>
<tr>
<td>Jim Meil/ Tianlun Jian</td>
<td>Eaton Corp.</td>
<td>4.82</td>
<td>-1.51</td>
<td>1.35</td>
</tr>
</tbody>
</table>
Table 3.2: Test for Unbiasedness, 2007-2012 II

<table>
<thead>
<tr>
<th>Forecaster</th>
<th>Institute</th>
<th>$\sigma_{\hat{\epsilon}_i}^2$</th>
<th>$\hat{\phi}_i$</th>
<th>s.e. of $\hat{\phi}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark Nielson</td>
<td>MacroEcon Global Advisors</td>
<td>7.05</td>
<td>-2.55</td>
<td>1.47*</td>
</tr>
<tr>
<td>Michael P. Niemira</td>
<td>Intl Council of Shopping Centers</td>
<td>4.05</td>
<td>-2.14</td>
<td>1.35</td>
</tr>
<tr>
<td>Nicholas S. Perna</td>
<td>Perna Associates</td>
<td>2.68</td>
<td>-0.28</td>
<td>1.33</td>
</tr>
<tr>
<td>Joel Prakken</td>
<td>Macroeconomic Advisers</td>
<td>3.97</td>
<td>-0.48</td>
<td>1.47</td>
</tr>
<tr>
<td>Arun Raha</td>
<td>Econ and Revenue Forecast</td>
<td>4.83</td>
<td>-1.71</td>
<td>1.77</td>
</tr>
<tr>
<td>David Resler</td>
<td>Nomura Securities Intl Inc.</td>
<td>4.66</td>
<td>-1.10</td>
<td>1.39</td>
</tr>
<tr>
<td>John Ryding</td>
<td>Bear Stearns &amp; Co. Inc.</td>
<td>3.06</td>
<td>-0.83</td>
<td>1.35</td>
</tr>
<tr>
<td>Ian Shepherdson</td>
<td>High Frequency Economics</td>
<td>3.40</td>
<td>0.01</td>
<td>1.39</td>
</tr>
<tr>
<td>John Silvia</td>
<td>Wachovia Corp.</td>
<td>4.07</td>
<td>-1.14</td>
<td>1.35</td>
</tr>
<tr>
<td>Allen Sinai</td>
<td>Decision Economics Inc.</td>
<td>4.37</td>
<td>-0.16</td>
<td>1.35</td>
</tr>
<tr>
<td>James F. Smith</td>
<td>Western Carolina Univ &amp; Parsec Financial Mgmt</td>
<td>5.34</td>
<td>-5.90</td>
<td>1.35***</td>
</tr>
<tr>
<td>Sung Won Sohn</td>
<td>Hanmi Bank</td>
<td>4.67</td>
<td>-1.51</td>
<td>1.39</td>
</tr>
<tr>
<td>Stephen Stanley</td>
<td>Pierpont Securities</td>
<td>2.83</td>
<td>-2.29</td>
<td>1.36*</td>
</tr>
<tr>
<td>Susan M. Sterne</td>
<td>Economic Analysis</td>
<td>7.24</td>
<td>-3.38</td>
<td>1.37**</td>
</tr>
<tr>
<td>Diane Swonk</td>
<td>Mesirow Financial</td>
<td>4.56</td>
<td>-0.26</td>
<td>1.34</td>
</tr>
<tr>
<td>Brian S. Wesbury</td>
<td>First Trust Advisors, L.P.</td>
<td>2.32</td>
<td>-2.75</td>
<td>1.34**</td>
</tr>
<tr>
<td>William T. Wilson</td>
<td>Keystone Business Intelligence</td>
<td>6.82</td>
<td>-3.06</td>
<td>1.67*</td>
</tr>
<tr>
<td>David Wyss</td>
<td>Standard and Poor's</td>
<td>8.37</td>
<td>-0.40</td>
<td>1.59</td>
</tr>
<tr>
<td>Lawrence Yun</td>
<td>National Association of Realtors</td>
<td>3.90</td>
<td>-2.17</td>
<td>1.43</td>
</tr>
</tbody>
</table>

Notes: Listed are the estimates of the variance of the idiosyncratic error $\sigma_{\hat{\epsilon}_i}^2$, the individual bias $\hat{\phi}_i$, and the standard errors of $\hat{\phi}_i$ for the 47 forecasters. Since Feb 2012, Arun Raha moved job and joined Jim Meil. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.
Table 3.3: Test for Efficiency, 2007-2012

<table>
<thead>
<tr>
<th>Regressor</th>
<th>$-\delta$</th>
<th>s.e. of $\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Forecast revision lagged 2 months</td>
<td>-0.045</td>
<td>0.046</td>
</tr>
<tr>
<td>(2) Change in percentage change in FHFA index over the past 12 months lagged 3 months</td>
<td>0.163</td>
<td>0.115</td>
</tr>
<tr>
<td>(3) Change in mortgage rate lagged 2 months</td>
<td>0.448</td>
<td>0.422</td>
</tr>
<tr>
<td>(4) Change in HMI lagged 2 months</td>
<td>0.035</td>
<td>0.045</td>
</tr>
<tr>
<td>(5) Change in housing starts lagged 2 months</td>
<td>0.0013</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

Notes: The coefficient and its standard error are obtained from running separate regressions for (3.9), including only one of the variables (1)-(5) as the exogenous regressor $(X_{t,h+1} - X_{t,h+2})$ at each time.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast revision lagged 2 months</td>
<td>-0.051</td>
<td>0.046</td>
</tr>
<tr>
<td>Change in percentage change in FHFA index over the past 12 months lagged 3 months</td>
<td>0.141</td>
<td>0.13</td>
</tr>
<tr>
<td>Change in mortgage rate lagged 2 months</td>
<td>0.476</td>
<td>0.494</td>
</tr>
<tr>
<td>Change in HMI lagged 2 months</td>
<td>0.017</td>
<td>0.05</td>
</tr>
<tr>
<td>Change in housing starts lagged 2 months</td>
<td>0.0004</td>
<td>0.002</td>
</tr>
<tr>
<td>$\chi^2$ statistic</td>
<td></td>
<td>3.856</td>
</tr>
</tbody>
</table>

Notes: Coefficient and their standard errors are obtained from running the regression (3.9), including all the variables (1) to (5) in the regression at the same time. The 1%, 5%, and 10% critical values for the $\chi^2(5)$ are 15.086, 11.070, and 9.236.
<table>
<thead>
<tr>
<th>Forecasters</th>
<th>Institute</th>
<th>$\sigma^2_{\varepsilon_i}$</th>
<th>$\phi_i$</th>
<th>s.e. of $\phi_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bart van Ark</td>
<td>The Conference Board</td>
<td>6.46</td>
<td>-0.42</td>
<td>1.31</td>
</tr>
<tr>
<td>Paul Ashworth</td>
<td>Capital Economics</td>
<td>7.46</td>
<td>-0.80</td>
<td>1.18</td>
</tr>
<tr>
<td>Nariman Behravesh</td>
<td>Global Insight</td>
<td>4.11</td>
<td>0.57</td>
<td>1.14</td>
</tr>
<tr>
<td>Richard Berner</td>
<td>Morgan Stanley</td>
<td>6.91</td>
<td>0.32</td>
<td>1.28</td>
</tr>
<tr>
<td>Ram Bhagavatula</td>
<td>Combinatorics Capital</td>
<td>10.39</td>
<td>0.06</td>
<td>1.17</td>
</tr>
<tr>
<td>Jay Brinkmann</td>
<td>Mortgage Bankers Association</td>
<td>4.10</td>
<td>-1.87</td>
<td>1.28</td>
</tr>
<tr>
<td>Joseph Carson</td>
<td>AllianceBernstein</td>
<td>2.84</td>
<td>-2.39</td>
<td>1.14**</td>
</tr>
<tr>
<td>Mike Cosgrove</td>
<td>Econoclast</td>
<td>3.30</td>
<td>-2.52</td>
<td>1.13**</td>
</tr>
<tr>
<td>Lou Crandall</td>
<td>Wrightson ICAP</td>
<td>3.28</td>
<td>-3.35</td>
<td>1.13***</td>
</tr>
<tr>
<td>J. Dewey Daane</td>
<td>Vanderbilt University</td>
<td>4.85</td>
<td>-3.24</td>
<td>1.14***</td>
</tr>
<tr>
<td>Richard DeKaser</td>
<td>National City Corporation</td>
<td>3.86</td>
<td>-3.17</td>
<td>1.25**</td>
</tr>
<tr>
<td>Douglas Duncan</td>
<td>Mortgage Bankers Association</td>
<td>4.19</td>
<td>-1.77</td>
<td>1.13</td>
</tr>
<tr>
<td>Stephen Gallagher</td>
<td>Societe Generale</td>
<td>5.42</td>
<td>-4.52</td>
<td>1.30***</td>
</tr>
<tr>
<td>Ethan S. Harris</td>
<td>Lehman Brothers</td>
<td>5.49</td>
<td>-1.73</td>
<td>1.15</td>
</tr>
<tr>
<td>Maury Harris</td>
<td>UBS</td>
<td>12.39</td>
<td>-0.25</td>
<td>1.17</td>
</tr>
<tr>
<td>Tracy Herrick</td>
<td>The Private Bank</td>
<td>13.46</td>
<td>-0.73</td>
<td>1.18</td>
</tr>
<tr>
<td>Stuart Hoffman</td>
<td>PNC Financial Services Group</td>
<td>2.03</td>
<td>-1.04</td>
<td>1.24</td>
</tr>
<tr>
<td>Gene Huang</td>
<td>FedEx Corp.</td>
<td>6.18</td>
<td>-2.36</td>
<td>1.15**</td>
</tr>
<tr>
<td>William B. Hummer</td>
<td>Wayne Hummer Investments</td>
<td>27.26</td>
<td>-0.74</td>
<td>1.23</td>
</tr>
<tr>
<td>Dana Johnson</td>
<td>Comerica Bank</td>
<td>2.45</td>
<td>-4.60</td>
<td>1.20***</td>
</tr>
<tr>
<td>Bruce Kasman</td>
<td>JPMorgan Chase &amp; Co.</td>
<td>4.05</td>
<td>-1.81</td>
<td>1.13</td>
</tr>
<tr>
<td>Paul Kasriel</td>
<td>The Northern Trust</td>
<td>10.99</td>
<td>-0.07</td>
<td>1.20</td>
</tr>
<tr>
<td>Joseph LaVorgna</td>
<td>Deutsche Bank Securities, Inc.</td>
<td>16.08</td>
<td>-2.29</td>
<td>1.53</td>
</tr>
<tr>
<td>Edward Leamer</td>
<td>UCLA Anderson Forecast</td>
<td>4.58</td>
<td>-0.88</td>
<td>1.14</td>
</tr>
<tr>
<td>John Lonski</td>
<td>Moody's Investors Service</td>
<td>2.55</td>
<td>-2.35</td>
<td>1.13**</td>
</tr>
<tr>
<td>Dean Maki</td>
<td>Barclays Capital</td>
<td>2.11</td>
<td>-2.06</td>
<td>1.17*</td>
</tr>
<tr>
<td>David Malpass</td>
<td>Encima Global LLC</td>
<td>3.94</td>
<td>-3.34</td>
<td>1.33**</td>
</tr>
<tr>
<td>Jim Meil/ Tianlun Jian</td>
<td>Eaton Corp.</td>
<td>5.63</td>
<td>-2.52</td>
<td>1.14**</td>
</tr>
</tbody>
</table>
### Table 3.6: Test for Unbiasedness, 2007-2011 II

<table>
<thead>
<tr>
<th>Forecasters</th>
<th>Institute</th>
<th>$\sigma^2_{\epsilon_i}$</th>
<th>$\phi_i$</th>
<th>s.e. of $\hat{\phi}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark Nielson</td>
<td>MacroEcon Global Advisors</td>
<td>6.87</td>
<td>-4.27</td>
<td>1.25***</td>
</tr>
<tr>
<td>Michael P. Niemira</td>
<td>Intl Council of Shopping Centers</td>
<td>3.56</td>
<td>-3.62</td>
<td>1.13***</td>
</tr>
<tr>
<td>Nicholas S. Perna</td>
<td>Perna Associates</td>
<td>2.86</td>
<td>-1.39</td>
<td>1.12</td>
</tr>
<tr>
<td>Joel Prakken</td>
<td>Macroeconomic Advisers</td>
<td>4.17</td>
<td>-1.39</td>
<td>1.21</td>
</tr>
<tr>
<td>Arun Raha</td>
<td>Econ and Revenue Forecast</td>
<td>4.97</td>
<td>-2.92</td>
<td>1.42***</td>
</tr>
<tr>
<td>David Resler</td>
<td>Nomura Securities Intl Inc.</td>
<td>5.08</td>
<td>-2.28</td>
<td>1.16**</td>
</tr>
<tr>
<td>John Ryding</td>
<td>Bear Sterns &amp; Co. Inc.</td>
<td>3.11</td>
<td>-2.31</td>
<td>1.14**</td>
</tr>
<tr>
<td>Ian Shepherdson</td>
<td>High Frequency Economics</td>
<td>3.77</td>
<td>-0.79</td>
<td>1.14</td>
</tr>
<tr>
<td>John Silvia</td>
<td>Wachovia Corp.</td>
<td>4.33</td>
<td>-2.47</td>
<td>1.14**</td>
</tr>
<tr>
<td>Allen Sinai</td>
<td>Decision Economics Inc.</td>
<td>4.96</td>
<td>-1.31</td>
<td>1.14</td>
</tr>
<tr>
<td>James F. Smith</td>
<td>Western Carolina Univ &amp; Parsec Financial Mgmt</td>
<td>4.64</td>
<td>-7.58</td>
<td>1.13***</td>
</tr>
<tr>
<td>Sung Won Sohn</td>
<td>Hanmi Bank</td>
<td>4.97</td>
<td>-2.88</td>
<td>1.17**</td>
</tr>
<tr>
<td>Stephen Stanley</td>
<td>Pierpont Securities</td>
<td>3.10</td>
<td>-3.56</td>
<td>1.14***</td>
</tr>
<tr>
<td>Susan M. Sterne</td>
<td>Economic Analysis</td>
<td>8.15</td>
<td>-4.69</td>
<td>1.16***</td>
</tr>
<tr>
<td>Diane Swonk</td>
<td>Mesirow Financial</td>
<td>4.78</td>
<td>-1.52</td>
<td>1.13</td>
</tr>
<tr>
<td>Brian S. Wesbury</td>
<td>First Trust Advisors, L.P.</td>
<td>2.69</td>
<td>-3.74</td>
<td>1.13***</td>
</tr>
<tr>
<td>William T. Wilson</td>
<td>Keystone Business Intelligence</td>
<td>5.39</td>
<td>-3.06</td>
<td>1.28**</td>
</tr>
<tr>
<td>David Wyss</td>
<td>Standard and Poor's</td>
<td>7.35</td>
<td>-0.50</td>
<td>1.23</td>
</tr>
<tr>
<td>Lawrence Yun</td>
<td>National Association of Realtors</td>
<td>4.69</td>
<td>-3.44</td>
<td>1.22***</td>
</tr>
</tbody>
</table>

Notes: Listed are the estimates of the variance of the idiosyncratic error $\sigma^2_{\epsilon_i}$, the individual bias $\phi_i$, and the standard errors of $\hat{\phi}_i$ for the 47 forecasters. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.
### Table 3.7: Test for Efficiency, 2007-2011

<table>
<thead>
<tr>
<th>Regressor ((X_{t,h+1} - X_{t,h+2}))</th>
<th>(-\delta)</th>
<th>s.e. of (\delta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Forecast revision lagged 2 months</td>
<td>-0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>(2) Change in percentage change in FHFA index over the past 12 months lagged 3 months</td>
<td>0.17</td>
<td>0.11</td>
</tr>
<tr>
<td>(3) Change in mortgage rate lagged 2 months</td>
<td>0.46</td>
<td>0.36</td>
</tr>
<tr>
<td>(4) Change in HMI lagged 2 months</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>(5) Change in housing starts lagged 2 months</td>
<td>0.002</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notes: The coefficient and its standard error are obtained from running separate regressions for (3.9), including only one of the variables (1)-(5) as the exogenous regressor \((X_{t,h+1} - X_{t,h+2})\) at each time.
Table 3.8: Joint Test for Efficiency, 2007-2011

<table>
<thead>
<tr>
<th>Term</th>
<th>coeff.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast revision lagged 2 months</td>
<td>-0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Change in percentage change in FHFA index</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>over the past 12 months lagged 3 months</td>
<td>0.48</td>
<td>0.44</td>
</tr>
<tr>
<td>Change in mortgage rate lagged 2 months</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>Change in HMI lagged 2 months</td>
<td>0.0005</td>
<td>0.002</td>
</tr>
<tr>
<td>Change in housing starts lagged 2 months</td>
<td>0.0005</td>
<td>0.002</td>
</tr>
<tr>
<td>$\chi^2$ statistic</td>
<td>4.971</td>
<td></td>
</tr>
</tbody>
</table>

Coefficient and their standard errors are obtained from running the regression (3.9), including all the variables (1) to (5) in the regression at the same time. The 1%, 5%, and 10% critical values for the $\chi^2(5)$ are 15.086, 11.070, and 9.236.
Bibliography


Appendix A: Deterministic trends and trend-adjusted equations

In this appendix, we derive the deterministic trends for house price and housing stock, and derive the trend-adjusted spatial equilibrium equation and housing supply equation.

Detrending is done by solving the equations without randomness for house price and housing stock.

Therefore, we set $x_{it} = w_{it} = u_{it} = 0$, implying that $E_t P_{i,t+1} = P_{i,t+1}$ because there is no randomness. Combining equations (2.9) and (2.15), we get

$$c_{1i} \hat{N}_{i,t+1} - [\alpha_i + (2 + \eta_i - \delta_i) c_{1i}] \hat{N}_{i,t} + (1 + \eta_i) (1 - \delta_i) c_{1i} \hat{N}_{i,t-1} = U - D_i + \eta_i c_i - \mu_i t \quad (A.1)$$

The solution to (A.1) that is linear in time is given by

$$\hat{N}_{it} = \left( \frac{\mu_i}{\eta_i \delta_i c_{1i} + \alpha_i} \right) t$$

$$+ \left( \frac{D_i + \mu_i - U - \eta_i c_i}{\eta_i \delta_i c_{1i} + \alpha_i} + \frac{-(\eta_i - \delta_i) c_{1i} - \alpha_i}{(\eta_i \delta_i c_{1i} + \alpha_i)^2} \mu_i \right) . \quad (A.2)$$

Substituting (A.2) into (2.15) with $E_t P_{i,t+1}$ replaced by $P_{i,t+1}$, we find that the deterministic component of house price equals

$$\hat{P}_{it} = \left( \frac{\mu_i \delta_i c_{1i}}{\eta_i \delta_i c_{1i} + \alpha_i} \right) t$$

113
Now define the detrended house price and detrended house stock as
\[ p_{it} = P_{it} - \hat{P}_{it}, \text{ and } n_{it} = N_{it} - \hat{N}_{it}. \] (A.4)
Substituting (A.4) into (2.9) and (2.15) to substitute for \( P_{it} \) and \( N_{it} \), we find
\[
D_i + \mu_i t - \alpha_i \hat{N}_{it} - (1 + \eta_i) \hat{P}_{it} + \hat{P}_{i,t+1} + [x_{it} - \alpha_i n_{it} - (1 + \eta_i) p_{it} + E_t h_{i,t+1}] = U, \] (A.5)
\[
E_t p_{i,t+1} + \hat{P}_{i,t+1} = c_i + c_{1i} \hat{N}_{i,t+1} - (1 - \delta_i) c_{1i} \hat{N}_{it} + [c_{1i} n_{i,t+1} - (1 - \delta_i) c_{1i} n_{it}]. \] (A.6)
Since \( \hat{P}_{it} \) and \( \hat{N}_{it} \) by definition satisfy
\[
D_i + \mu_i t - \alpha_i \hat{N}_{it} - (1 + \eta_i) \hat{P}_{it} + \hat{P}_{i,t+1} = U,
\]
\[
\hat{P}_{i,t+1} = c_i + c_{1i} \hat{N}_{i,t+1} - (1 - \delta_i) c_{1i} \hat{N}_{it}, \] (A.7)
(A.5) and (A.6) simplify to
\[
(1 + \eta_i) p_{it} - E_t p_{i,t+1} = x_{it} - \alpha_i n_{it} + w_{it}, \] (A.8)
\[
c_{1i} n_{i,t+1} = E_t p_{i,t+1} + (1 - \delta_i) c_{1i} n_{it} + u_{it}. \] (A.9)
Appendix B: Blanchard and Kahn (1980)

The model in Blanchard and Kahn is:

\[
\begin{bmatrix}
  X_{t+1} \\
  tP_{t+1}
\end{bmatrix}
= A \begin{bmatrix}
  X_t \\
  P_t
\end{bmatrix}
+ \gamma Z_t, X_{t=0} = X_0, \quad (B.1)
\]

where \( X_t \) is an \((n \times 1)\) vector of predetermined variables at \( t \); \( P_t \) is an \((m \times 1)\) vector of non-predetermined variables at \( t \); \( Z_t \) is a \((k \times 1)\) vector of exogenous variables; \( tP_{t+1} \) is the agents’ expectation of \( P_{t+1} \) held at \( t \); \( A, \gamma \) are \((n+m) \times (n+m)\) and \((n+m) \times k\) matrices.

The rational expectation defined in Blanchard and Kahn is

\[
_{t}P_{t+1} = E(P_{t+1}|\Omega_t), \quad (B.2)
\]

where \( E(\cdot) \) is the mathematical expectation operator; \( \Omega_t \) is the information set at \( t \); \( \Omega_t \supseteq \Omega_{t-1} \).

The non-explosion assumption in their paper is

\[
\forall t, \exists Z_t \in \mathbb{R}^k, \theta_t \in \mathbb{R} \text{ such that}
\]

\[
-(1+i)\theta_t Z_t \leq E(Z_{t+i}|\Omega_t) \leq (1+i)\theta_t Z_t, \forall i \geq 0. \quad (B.3)
\]

Also, they require that the expectations of \( X_t \) and \( P_t \) do not explode:

\[
\forall t, \exists \begin{bmatrix}
  X_t \\
  P_t
\end{bmatrix} \in \mathbb{R}^{n+m}, \sigma_t \in \mathbb{R} \text{ such that}
\]
$-(1 + i)^\sigma_t \left[ \frac{X_t}{P_t} \right] \leq \left[ \frac{E(X_{t+i} | \Omega_t)}{E(P_{t+i} | \Omega_t)} \right] \leq (1 + i)^\sigma_t \left[ \frac{X_t}{P_t} \right], \forall i \geq 0. \quad \text{(B.4)}$

The non-explosion condition rules out exponential growth of expectations held at time $t$.

Under the above settings, Blanchard and Kahn's Proposition 1 gives the condition for the existence of a unique solution for (B.1): if the number of eigenvalues of $A$ in (B.1) outside the unit circle is equal to the number of non-predetermined variables, then there exists a unique solution.

Their paper also provides the expression of this solution for the general case, as well as the case with one predetermined and one non-predetermined variable, which is described below. Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \gamma = \begin{bmatrix} \gamma_1 \\ (1 \times k) \end{bmatrix},$$

$$| \lambda_1 | < 1, \quad | \lambda_2 | > 1,$$

let $\lambda_1, \lambda_2$ be the eigenvalues of $A$, and define $\mu \equiv (\lambda_1 - a_{11}) \gamma_1 - a_{12} \gamma_2$. Then, there exists a unique solution and is given by

$$X_t = X_0, \text{ for } t = 0, \quad (B.5)$$

$$= \lambda_1 X_{t-1} + \gamma_1 Z_{t-1} + \mu \sum_{i=0}^{\infty} \lambda_2^{-i-1} E(Z_{t+i-1} | \Omega_{t-1}), \text{ for } t > 0,$$

$$P_t = a_{12}^{-1} [(\lambda_1 - a_{11}) X_t + \mu \sum_{i=0}^{\infty} \lambda_2^{-i-1} E(Z_{t+i} | \Omega_t)], \text{ for } t \geq 0. \quad \text{(B.6)}$$

Their paper writes $\mu \equiv (\lambda_1 - a_{11}) \lambda_1 - a_{12} \lambda_2$, and we believe it is a typo.
Appendix C: Proof of Proposition 1

To apply Blanchard and Kahn to our model, substitute (2.20) into (2.19) and get

\[ c_{1i}n_{i,t+1} = (1 + \eta_i)p_{it} + [\alpha_i + (1 - \delta_i)c_{1i}]n_{it} - x_{it} - w_{it} + u_{it}. \quad (C.1) \]

Equations (2.19) and (C.1) can then be rewritten as

\[
\begin{bmatrix}
    n_{i,t+1} \\
    E_t(p_{i,t+1})
\end{bmatrix} =
\begin{bmatrix}
    \frac{\alpha_i + (1 - \delta_i)c_{1i}}{c_{1i}} & \frac{1 + \eta_i}{c_{1i}} \\
    \alpha_i & 1 + \eta_i
\end{bmatrix}
\begin{bmatrix}
    n_{it} \\
    p_{it}
\end{bmatrix}
+ \begin{bmatrix}
    \frac{1}{c_{1i}}(-x_{it} - w_{it} + u_{it}) \\
    -x_{it} - w_{it}
\end{bmatrix}, \quad (C.2)
\]

where \( n_t \) is predetermined at time \( t \), \( p_t \) is non-predetermined at time \( t \), and \( x_t, w_{it}, u_{it} \) are exogenous variables. This is the form of (B.1).

The corresponding non-explosion assumption in our model is

\[
\forall t, \exists \begin{bmatrix}
   \pi_t \\
   \bar{p}_t
\end{bmatrix} \in \mathbb{R}^2, \sigma_t \in \mathbb{R} \text{ such that}
\]

\[
-(1 + \delta)t \begin{bmatrix}
   \pi_t \\
   \bar{p}_t
\end{bmatrix} \leq \begin{bmatrix}
   E(n_{t+1}|\Omega_t) \\
   E(p_{t+1}|\Omega_t)
\end{bmatrix} \leq (1 + \delta)t \begin{bmatrix}
   \pi_t \\
   \bar{p}_t
\end{bmatrix}, \forall i \geq 0. \quad (C.3)
\]

This means that people do not expect house price and house stock to exponentially explode in the future.

Denote

\[
A = \begin{bmatrix}
   \frac{\alpha_i + (1 - \delta_i)c_{1i}}{c_{1i}} & \frac{1 + \eta_i}{c_{1i}} \\
   \alpha_i & 1 + \eta_i
\end{bmatrix} \quad (C.4)
\]
then the eigenvalues of $A$ can be found by solving the equation below:

$$0 = \det(A - \lambda I)$$

$$= \lambda^2 - \frac{\alpha_i + (2 + \eta_i - \delta_i) c_{11} i}{c_{11}} \lambda + \frac{(1 + \eta_i)(1 - \delta_i) c_{11}}{c_{11}} = f(\lambda). \quad (C.5)$$

Assume $\alpha_i > 0$, $\eta_i > 0$, $1 > \delta_i \geq 0$, and $c_{11} > 0$, then it can be verified that

$$f(0) = (1 + \eta_i)(1 - \delta_i) > 0, \quad f(1) = -\frac{\delta_i \eta_i c_{11} + \alpha_i}{c_{11}} < 0. \quad (C.6)$$

Therefore, the matrix $A$ has two real eigenvalues, one of them is between 0 and 1, and the other is larger than 1. Denote them as $0 < \phi < 1$ and $\tilde{\phi} > 1$. Thus, the number of eigenvalues of $A$ outside the unit circle is equal to the number of non-predetermined variable $p_{it}$ (both equal one). Applying Proposition 1 in Blanchard and Kahn, our model has a unique solution.

Next, we derive the solution.

The $a_{11}, a_{12}, \gamma_1, \gamma_2, \lambda_1, \lambda_2$ and $Z_{t+i}$ in (B.5) and (B.6) correspond to our model as

$$\frac{\alpha_i + (1 - \delta_i) c_{11} - c_{21}}{c_{11}}, \frac{1 + \eta_i}{c_{11}}, (1, 0), (0, 1), \phi, \tilde{\phi}, i$$

and \[ \begin{bmatrix} \frac{1}{c_{11}} (-x_{it} - w_{it} + u_{it}) \\ -x_{it} - w_{it} \end{bmatrix}. \] Therefore,

$$\mu = \left( -\frac{\alpha + (1 - \delta - \phi) c_1 - c_2}{c_1}, \frac{1 + \eta}{c_1} \right).$$

Hence,

$$\mu E(Z_{t+i}|\Omega_t) = \frac{\alpha + (1 - \delta - \phi) c_1 - c_2}{c_1^2} E_t(x_{t+i} + w_{t+i} - u_{t+i}) + \frac{1 + \eta}{c_1} E_t(x_{t+i} + w_{t+i})$$

$$= \frac{\alpha + (2 + \eta - \delta - \phi) c_1 - c_2}{c_1^2} E_t(x_{t+i} + w_{t+i}) - \frac{\alpha + (1 - \delta - \phi) c_1 - c_2}{c_1^2} E_t u_{t+i}$$

$$= \frac{\tilde{\phi}}{c_1} E_t(x_{t+i} + w_{t+i}) + \frac{1 + \eta - \tilde{\phi}}{c_1} E_t u_{t+i}. \quad (C.7)$$
We have

\[ E_t w_{t+i} = E_t u_{t+i} = 0, \text{ for } i > 0, \]

and

\[ E_t x_{t+i} = x_t + \frac{\rho(1 - \rho^i)}{1 - \rho} \varepsilon_t, \text{ for } i > 0. \]

Thus,

\[
\begin{align*}
\mu \sum_{i=0}^{\infty} \phi^{-i-1} E(Z_{t+i}|\Omega_t) & = \phi^{-1} \sum_{i=0}^{\infty} \phi^{-i-1} \left( x_t + w_t \right) + \sum_{i=1}^{\infty} \phi^{-i-1} \phi \left( x_t + \frac{\rho(1 - \rho^i)}{1 - \rho} \varepsilon_t \right) \\
& = \frac{1}{c_1} x_t + \frac{1}{c_1} w_t + \frac{1 + \eta - \phi}{c_1 \phi} u_t + \frac{1}{c_1 (\phi - 1)} x_t + \frac{\rho \phi}{c_1 (\phi - 1) (\phi - \rho)} \varepsilon_t - 1.
\end{align*}
\]

(C.8)

Therefore, (B.5) becomes

\[
n_t = \phi n_{t-1} + \frac{1}{c_1} (-x_{t-1} - w_{t-1} + u_{t-1}) + \frac{1}{c_1} x_{t-1} + \frac{1}{c_1} w_{t-1} + \frac{1 + \eta - \phi}{c_1 \phi} u_{t-1} + \frac{1}{c_1 (\phi - 1)} x_{t-1} + \frac{\rho \phi}{c_1 (\phi - 1) (\phi - \rho)} \varepsilon_{t-1}
\]

\[ = \phi n_{t-1} + \frac{1}{c_1 (\phi - 1)} x_{t-1} + \frac{\rho \phi}{c_1 (\phi - 1) (\phi - \rho)} \varepsilon_{t-1} + \frac{1 + \eta}{c_1 \phi} u_{t-1}, \text{ for } t > 0,
\]

and (B.6) becomes

\[
p_t = \frac{c_1}{1 + \eta} \left[ + \frac{1}{c_1} x_t + \frac{1}{c_1} w_t + \frac{1 + \eta - \phi}{c_1 \phi} u_t + \frac{1}{c_1 (\phi - 1)} x_t + \frac{\rho \phi}{c_1 (\phi - 1) (\phi - \rho)} \varepsilon_t \right]
\]

\[ = - \alpha + (1 - \delta - \phi) c_1 - c_2 \frac{n_t}{1 + \eta} + \frac{\phi}{(1 + \eta)(\phi - 1)} x_t + \frac{\rho \phi}{(1 + \eta)(\phi - 1)(\phi - \rho)} \varepsilon_t + \frac{1}{1 + \eta} w_t + \frac{1 + \eta - \phi}{(1 + \eta) \phi} u_t.
\]

Q.E.D.
Appendix D: Proof of Proposition 2 and derivation of the stability condition

(2.19) and (2.20) imply that (the same as (C.1))

\[ c_{1i}n_{i,t+1} = (1 + \eta_i)p_{it} + [\alpha_i + (1 - \delta_i)c_{1i} - c_{2i}]n_{it} - x_{it} - w_{it} + u_{it}. \]  

(D.1)

Plug (D.1) into (2.25), we get

\[ (1 + \eta_i - \lambda_i)c_{1i}n_{i,t+1} + \{\lambda_i\alpha_i - (1 + \eta_i - \lambda_i)[(1 - \delta_i)c_{1i} - c_{2i}] \]
\[ - (1 + \eta_i)(1 - \lambda_i)c_{1i}n_{it} + (1 + \eta_i)(1 - \lambda_i)[(1 - \delta_i)c_{1i} - c_{2i}]n_{i,t-1} \]
\[ = \lambda_ix_{it} + \lambda_iw_{it} + (1 + \eta_i - \lambda_i)u_{it} - (1 + \eta_i)(1 - \lambda_i)u_{i,t-1}. \]  

(D.2)

Using lag operator \( L \), the above equation can be rewritten as

\[ [(1 + \eta_i - \lambda_i)c_{1i} + \{\lambda_i\alpha_i - (1 + \eta_i - \lambda_i)[(1 - \delta_i)c_{1i} - c_{2i}] - (1 + \eta_i)(1 - \lambda_i)c_{1i}\} L \]
\[ + (1 + \eta_i)(1 - \lambda_i)[(1 - \delta_i)c_{1i} - c_{2i}][L^2]n_{i,t+1} = \lambda_ix_{it} + \lambda_iw_{it} + (1 + \eta_i - \lambda_i)u_{it} - (1 + \eta_i)(1 - \lambda_i)u_{i,t-1}. \]  

(D.3)

Under the stability condition (2.27), the roots of the polynomial \( f(L) \) lie outside the unit circle, where

\[ f(L) = (1 + \eta - \lambda)c_{1i} + \{\lambda\alpha - (1 + \eta - \lambda)[(1 - \delta)c_{1i} - c_{2i}] - (1 + \eta)(1 - \lambda)c_{1i}\} L \]
\[ + (1 + \eta)(1 - \lambda)[(1 - \delta)c_{1i} - c_{2i}]L^2. \]  

(D.4)

120
Then the polynomial $f(L)$ in (D.3) is invertable, and this guarantees the existence of a unique particular solution for (D.2).

We derive the stability condition \[(2.27)\] under the adaptive expectations.

To ensure that (D.2) do not explode, we need to require the roots of

$$g(y) = y^2 + \frac{\lambda \alpha - (1 + \eta - \lambda)\{(1 - \delta)c_1 - c_2\} - (1 + \eta)(1 - \lambda)c_1}{(1 + \eta - \lambda)c_1}y$$

$$+ \frac{(1 + \eta)(1 - \lambda)\{(1 - \delta)c_1 - c_2\}}{(1 + \eta - \lambda)c_1}$$

(D.5)

to lie within the unit circle.

Enders (2010) provides the stability conditions for a quadratic equation of

$$y^2 - a_1y - a_2 = 0.$$ \[(D.6)\]

There are three cases depending on the discriminant

$$d = (a_1)^2 + 4a_2.$$ \[(D.7)\]

In case 1, $(a_1)^2 + 4a_2 > 0$, there are two distinct real roots, and the conditions for these two roots to lie within the unit circle are $a_1 + a_2 < 1$ and $a_2 - a_1 < 1$.

In case 2, $(a_1)^2 + 4a_2 = 0$, there are repeated real roots, and the condition is $|a_1| < 2$. In case 3, $(a_1)^2 + 4a_2 < 0$, the roots are imaginary, and the condition is $-a_2 < 1$ ($a_2 < 0$).

We apply these results to (D.5).

$$a_1 = -\frac{\lambda \alpha - (1 + \eta - \lambda)\{(1 - \delta)c_1 - c_2\} - (1 + \eta)(1 - \lambda)c_1}{(1 + \eta - \lambda)c_1},$$ \[(D.8)\]

$$a_2 = -\frac{(1 + \eta)(1 - \lambda)\{(1 - \delta)c_1 - c_2\}}{(1 + \eta - \lambda)c_1},$$ \[(D.9)\]

\[ d = \frac{\lambda \alpha - (1 + \eta - \lambda)[(1 - \delta)c_1 - c_2] - (1 + \eta)(1 - \lambda)c_1^2}{(1 + \eta - \lambda)^2c_1^2 - 4(1 + \eta)(1 - \lambda)(1 + \eta - \lambda)c_1[(1 - \delta)c_1 - c_2]} \]  

(D.10)

Case 3.

\[ -a_2 - 1 = \frac{(1 + \eta)(1 - \lambda)[(1 - \delta)c_1 - c_2] - (1 + \eta - \lambda)c_1}{(1 + \eta - \lambda)c_1} \]

\[ = \frac{-\eta \lambda c_1 - (1 + \eta)(1 - \lambda)(\delta c_1 + c_2)}{(1 + \eta - \lambda)c_1} < 0, \]

so in case 3, \(-a_2 < 1\) and therefore the roots are inside the unit circle.

Case 2. The stability condition is \(-2 < a_1 < 2\).

\[ a_1 - 2 = \frac{-\lambda \alpha - \eta \lambda c_1 - (1 + \eta - \lambda)(\delta c_1 + c_2)}{(1 + \eta - \lambda)c_1} < 0. \]

Therefore it remains to verify that \(a_1 + 2 > 0\). From (D.8) we need to verify that

\[ a_1 + 2 = \frac{-\lambda \alpha + (1 + \eta - \lambda)[(1 - \delta)c_1 - c_2] + (1 + \eta)(1 - \lambda)c_1 + 2(1 + \eta - \lambda)c_1}{(1 + \eta - \lambda)c_1} > 0. \]

(D.11)

If \(a_1 \geq 0\), the condition \(a_1 + 2 > 0\) is satisfied. If \(a_1 < 0\), because \(d = (a_1)^2 + 4a_2 = 0\), we have \(|a_1| = \sqrt{-4a_2}\), that is \(-a_1 = \sqrt{-4a_2}\), hence \(a_1 = -2\sqrt{-a_2}\). Therefore,

\[ -\lambda \alpha - (1 + \eta - \lambda)[(1 - \delta)c_1 - c_2] - (1 + \eta)(1 - \lambda)c_1 \]

\[ = -2\sqrt{(1 + \eta)(1 - \lambda)[(1 - \delta)c_1 - c_2]} \]

rearranging,

\[ -\lambda \alpha + (1 + \eta - \lambda)[(1 - \delta)c_1 - c_2] + (1 + \eta)(1 - \lambda)c_1 \]

\[ = -2\sqrt{(1 + \eta)(1 - \lambda)[(1 - \delta)c_1 - c_2]}(1 + \eta - \lambda)c_1. \]

Plug the above equation into the numerator of (D.11), we get the following inequality that needs to be verified

\[-2\sqrt{(1 + \eta)(1 - \lambda)[(1 - \delta)c_1 - c_2]}(1 + \eta - \lambda)c_1 + 2(1 + \eta - \lambda)c_1 > 0, \]

122
which is equivalent to
\[(1 + \eta - \lambda)c_1 > (1 + \eta)(1 - \lambda)[(1 - \delta)c_1 - c_2],\]
which is equivalent to
\[\eta\lambda c_1 + (1 + \eta)(1 - \lambda)(\delta c_1 + c_2) > 0,\]
which satisfies. Therefore, in case 2 we have \(-2 < a_1 < 2\), and thus the roots are inside the unit circle.

Case 1. We need to verify that \(a_1 + a_2 < 1\) and \(a_2 - a_1 < 1\).
\[
a_1 + a_2 - 1 = \frac{-\lambda \alpha + (1 + \eta - \lambda)[(1 - \delta)c_1 - c_2] + (1 + \eta)(1 - \lambda)c_1}{(1 + \eta - \lambda)c_1} - \frac{(1 + \eta)(1 - \lambda)[(1 - \delta)c_1 - c_2]}{(1 + \eta - \lambda)c_1} - 1 \tag{D.12}
\]
\[
a_2 - a_1 - 1 = \frac{\lambda \alpha - [(1 + \eta)(1 - \lambda) + (1 + \eta - \lambda)][(1 - \delta)c_1 - c_2 + c_1]}{(1 + \eta - \lambda)c_1} - \frac{(1 + \eta)(1 - \lambda)c_1}{(1 + \eta - \lambda)c_1} - 1
\]
\[
= \frac{\lambda \alpha - [(2 + 2\eta) - (2 + \eta)\lambda][(2 - \delta)c_1 - c_2]}{(1 + \eta - \lambda)c_1} - \frac{(1 + \eta - \lambda)c_1}{(1 + \eta - \lambda)c_1}
\]
\[
= \frac{(2 + \eta)[(2 - \delta)c_1 - c_2] + \alpha}{(1 + \eta - \lambda)c_1} - \frac{(2 + 2\eta)[(2 - \delta)c_1 - c_2]}{(1 + \eta - \lambda)c_1}
\]
\[
= \frac{(2 + \eta)[(2 - \delta)c_1 - c_2] + \alpha}{(1 + \eta - \lambda)c_1} - \frac{(2 + 2\eta)[(2 - \delta)c_1 - c_2]}{(1 + \eta - \lambda)c_1}
\]

Hence, to require \(a_2 - a_1 < 1\) is equivalent to require
\[
\{(2 + \eta)[(2 - \delta)c_1 - c_2] + \alpha\} \lambda - (2 + 2\eta)[(2 - \delta)c_1 - c_2] < 0,
\]
which is equivalent to

\[
\lambda < \frac{(2 + 2\eta)[(2 - \delta)c_1 - c_2]}{(2 - \eta)[(2 - \delta)c_1 - c_2] + \alpha} \equiv \bar{\lambda}.
\] (D.14)

Thus, in case 1, for the roots to be inside the unit circle, we require that \( \lambda < \bar{\lambda} \).

In sum, when \( d \leq 0 \), the roots would be inside the unit circle; when \( d > 0 \), a condition of \( \lambda < \bar{\lambda} \) is required for the roots to lie inside the unit circle.
Appendix E: Moment conditions for rational expectations

Simplify the notation of (2.16) and (2.18) as

\[ \hat{N}_{it} = K_{0i} + K_1 t \]  
\[ \hat{P}_{it} = S_{0i} + S_1 t. \]  
(E.1)

The intercepts have subscript \(i\) because we allow for city fixed effects. The time trend coefficients do not have subscript \(i\) because cities within the same group are assumed to have the same parameters. Substitute \(n_{it} = N_{it} - \hat{N}_{it}, \ p_{it} = P_{it} - \hat{P}_{it}\) and \(y_{it} = \bar{D}_i + \mu t + x_{it}\) into (2.22) and (2.23), and together with (2.2), we have the following three equations based on which the moment conditions will be derived:

\[
Z_i - \left[ (1 - \phi) K_1 - \frac{\mu}{c_1(\phi - 1)} \right] t + N_{it} - \phi N_{i,t-1} - \frac{1}{c_1(\phi - 1)} y_{i,t-1} = \frac{1}{c_1(\phi - 1)(\phi - \rho)} \epsilon_{i,t-1} + \frac{1 + \eta}{c_1(\phi - 1)} u_{i,t-1}, \tag{E.2}
\]

\[
Q_i - [S_1 + \frac{\alpha + (1 - \delta - \phi)c_1}{1 + \eta} K_1 - \frac{\bar{\phi} \mu}{(1 + \eta)(\phi - 1)(\phi - \rho)}] t + P_{it} + \frac{\alpha + (1 - \delta - \phi)c_1}{1 + \eta} N_{it} - \frac{1}{1 + \eta} w_{it} + \frac{1 + \eta - \phi}{(1 + \eta)(\phi - 1)} u_{it}, \tag{E.3}
\]

\[
y_{it} = \bar{D}_i + \mu t + x_{it}, x_{it} = x_{i,t-1} + \epsilon_{it} + \epsilon_{it} = \rho \epsilon_{i,t-1} + v_{it}. \tag{E.4}
\]

where \(Q_i\) and \(Z_i\) are fixed effects.
Take first difference of \( y_{it} = \overline{D}_i + \mu t + x_{it} \), we eliminate the fixed effects and convert the unit root series into a stationary series:

\[
\Delta y_{it} = \mu + \varepsilon_{it}.
\] (E.5)

We have

\[
\Delta y_{it} - \rho \Delta y_{i,t-1} - (1 - \rho)\mu = v_{it}.
\] (E.6)

Take the first difference of (E.2),

\[
- \left[ (1 - \phi)K_1 - \frac{\mu}{c_1(\overline{\phi} - 1)} \right] + \Delta N_{it} - \phi \Delta N_{i,t-1} - \frac{1}{c_1(\overline{\phi} - 1)} \Delta y_{i,t-1} - \frac{\rho \overline{\phi}}{c_1(\overline{\phi} - 1)(\overline{\phi} - \rho)} \Delta \varepsilon_{i,t-1} + \frac{1 + \eta}{c_1 \phi} \Delta u_{i,t-1}.
\] (E.7)

Because

\[
\Delta \varepsilon_{it} = \rho \Delta \varepsilon_{i,t-1} + \Delta v_{it},
\] (E.8)

thus applying the Cochrane-Orcutt transformation yields

\[
- (1 - \rho) \left[ (1 - \phi)K_1 - \frac{\mu}{c_1(\overline{\phi} - 1)} \right] + \Delta N_{it} - \rho \Delta N_{i,t-1} - \phi (\Delta N_{i,t-1} - \rho \Delta N_{i,t-2}) - \frac{\rho \overline{\phi}}{c_1(\overline{\phi} - 1)(\overline{\phi} - \rho)} \Delta v_{i,t-1} + \frac{1 + \eta}{c_1 \phi} (\Delta u_{i,t-1} - \rho \Delta u_{i,t-2}).
\] (E.9)

Rearrange and substitute (E.6) in, we get

\[
- (1 - \rho)(1 - \phi)K_1 + \Delta N_{it} - (\rho + \phi) \Delta N_{i,t-1} + \rho \phi \Delta N_{i,t-2} = \frac{v_{i,t-1}}{c_1(\overline{\phi} - 1)} + \frac{\rho \overline{\phi}}{c_1(\overline{\phi} - 1)(\overline{\phi} - \rho)} \Delta v_{i,t-1} + \frac{1 + \eta}{c_1 \phi} (\Delta u_{i,t-1} - \rho \Delta u_{i,t-2}).
\] (E.10)
Then apply the Cochrane-Orcutt transformation and plug (E.6) in, we get

\[- (1 - \rho) [S_1 + \frac{\alpha + (1 - \delta - \phi)c_1}{1 + \eta} K_1] + \Delta P_{it} - \rho \Delta P_{i,t-1} + \frac{\alpha + (1 - \delta - \phi)c_1}{1 + \eta} (\Delta N_{it} - \rho \Delta N_{i,t-1}) \]

\[= \frac{\bar{\phi} v_{it}}{(1 + \eta)(\bar{\phi} - 1)} + \frac{\rho \bar{\phi}}{(1 + \eta)(\bar{\phi} - 1)(\bar{\phi} - \rho)} \Delta v_{it} + \frac{1}{1 + \eta} (\Delta w_{it} - \rho \Delta w_{i,t-1}) + \frac{1 + \eta - \bar{\phi}}{(1 + \eta) \bar{\phi}} (\Delta u_{it} - \rho \Delta u_{i,t-1}). \]

(E.11)

Based on (E.6), (E.9), and (E.11), the moment conditions are

First stage:

\[m_1(\theta) = \begin{cases} 
\Delta y_{it} - \rho \Delta y_{i,t-1} - (1 - \rho) \mu \\
[\Delta y_{it} - \rho \Delta y_{i,t-1} - (1 - \rho) \mu] \Delta y_{i,t-1} \\
[\Delta y_{it} - \rho \Delta y_{i,t-1} - (1 - \rho) \mu] \Delta y_{i,t-2} 
\end{cases} \]

(E.12)

Second stage:

\[m_R(\theta) = \begin{cases} 
\psi_{it} \\
\psi_{it} \Delta y_{i,t-4} \\
\psi_{it} \Delta P_{i,t-4} \\
\psi_{it} \Delta N_{i,t-4} \\
\zeta_{it} \\
\zeta_{it} \Delta y_{i,t-3} \\
\zeta_{it} \Delta P_{i,t-3} \\
\zeta_{it} \Delta N_{i,t-3}
\end{cases}, \]

(E.13)

where \(\psi_{it}\) and \(\zeta_{it}\) are defined as

\[\psi_{it} = -(1 - \rho)(1 - \phi) K_1 + \Delta N_{it} - (\rho + \phi) \Delta N_{i,t-1} + \rho \phi \Delta N_{i,t-2}, \]

(E.14)

and

\[\zeta_{it} = -(1 - \rho)[S_1 + \frac{\alpha + (1 - \delta - \phi)c_1}{1 + \eta} K_1] + \Delta P_{it} - \rho \Delta P_{i,t-1} + \frac{\alpha + (1 - \delta - \phi)c_1}{1 + \eta} (\Delta N_{it} - \rho \Delta N_{i,t-1}). \]

(E.15)
Appendix F: Moment conditions for adaptive expectations

Substitute \( n_{it} = N_{it} - \hat{N}_{it} \), \( p_{it} = P_{it} - \hat{P}_{it} \) and \( y_{it} = D_{i} + \mu t + x_{it} \) into (2.2), (2.28) and (2.29), we have the three equations based on which the moment conditions will be derived. One of them are the same as (E.4). The other two are (after taking the first difference):

\[
- \Delta P_{i,t-1} + \frac{c_1}{\lambda} \Delta N_{it} - \frac{(2 - \lambda - \delta)c_1}{\lambda} \Delta N_{i,t-1} + \frac{(1 - \lambda)(1 - \delta)c_1}{\lambda} \Delta N_{i,t-2} = \frac{1}{\lambda} \Delta u_{i,t-1} - \frac{1 - \lambda}{\lambda} \Delta u_{i,t-2},
\]

(F.1)

and

\[
(1 + \eta - \lambda)\Delta P_{it} - (1 + \eta)(1 - \lambda)\Delta P_{i,t-1} + \alpha \Delta N_{it} - \alpha(1 - \lambda)\Delta N_{i,t-1} - \Delta y_{it} + (1 - \lambda)\Delta y_{i,t-1} = \Delta w_{it} - (1 - \lambda)\Delta w_{i,t-1}.
\]

(F.2)

Therefore, the first stage moment conditions are the same as the ones in the rational expectations case. The second stage moment conditions are

\[
A(\theta) = \begin{cases} 
\chi_{it} \\
\chi_{it}\Delta y_{i,t-4} \\
\chi_{it}\Delta P_{i,t-4} \\
\chi_{it}\Delta N_{i,t-4} \\
\kappa_{it} \\
\kappa_{it}\Delta y_{i,t-3} \\
\kappa_{it}\Delta P_{i,t-3} \\
\kappa_{it}\Delta N_{i,t-3} 
\end{cases},
\]

(E.3)

where \( \chi_{it} \) is defined as

\[
\chi_{it} = -\Delta P_{i,t-1} + \frac{c_1}{\lambda} \Delta N_{it} - \frac{(2 - \lambda - \delta)c_1}{\lambda} \Delta N_{i,t-1} + \frac{(1 - \lambda)(1 - \delta)c_1}{\lambda} \Delta N_{i,t-2},
\]

(F.4)
and $\kappa_{it}$ is defined as

$$
\kappa_{it} = (1+\eta-\lambda)\Delta P_{it} - (1+\eta)(1-\lambda)\Delta P_{i,t-1} + \alpha \Delta N_{it} - \alpha(1-\lambda)\Delta N_{i,t-1} - \Delta y_{it} + (1-\lambda)\Delta y_{i,t-1}.
$$

(E.5)
In the appendix, I describe three other surveys about or that include house price expectations in U.S., and explain why I chose the WSJ survey to study. The “House Price Expectations Survey” conducted by Pulsenomics surveys a panel of around 100 to 110 economists and industry professionals in the housing field. The individuals are asked to predict the annual house price percentage changes (“on a Q4-over-preceding Q4 basis”) of the S&P/Case-Shiller U.S. national home price index, for the current year and the next five years. This survey began in May 2010 and was conducted monthly through Dec 2010. Beginning at 2011, the survey has been collected quarterly. Case and Shiller conduct surveys in the year 1988 and then annually from 2003 to 2012, for recent homebuyers in four U.S. cities: Boston, Los Angeles, San Francisco, and Milwaukee. The number of survey answers returned are 886 in 1988, 705 in 2003, but declines to 328 in 2012. The relevant questions in their survey that are closely related to my study are the two questions asking about the expectation of future house price changes. Question 6: “How much of a change do you expect there to be in the value of your home over the next 12

35This survey can be found at https://pulsenomics.com/Home-Price-Expectations.html
months?” Question 7: “On average over the next ten years how much do you expect the value of your property to change each year?”

Figure 5 in Case and Shiller’s paper plots actual and expected house price changes and shows that the respondents in their survey (the recent homebuyers) are much more optimistic than the forecasters in the WSJ survey. Their figure shows that the only negative one-year expectations for house price changes occur in 2008 in SF, Boston, and LA. In Milwaukee, there is never a negative expectation. This indicates that, either forecasters are making better predictions, or the recent homebuyers in their sample are not a random sample of the population. Rather, people who buy homes are generally optimistic about housing markets. Case and Shiller also run regressions testing the hypothesis of rational expectations. But since individuals are predicting their own property’s value changes, only the mean expectations for each city in each year can be used. This gives only 9 observations for each of the 4 cities. They implement an efficiency test by regressing the actual subsequent one-year home price change on the expectation of one-year home price change, the lagged actual own-city one-year home price change, and the lagged actual U.S. one-year home price change. Both of the last two variables have insignificant coefficients, and therefore “this confirms the rational expectations for the 12-month forecasts. Respondent do appear to incorporate this other information in making the 12-month forecasts.”
The third survey is the Surveys of Consumers conducted by the Survey Research Center at the University of Michigan.\footnote{Its information can be found at http://press.sca.isr.umich.edu/ Its login is at http://www.sca.isr.umich.edu/} 500 individuals are randomly selected each month. Starting from May 2007, one-year and five-year house price expectations changes have been asked in the survey: “By about what percent do you expect prices of homes like yours in your community to go (up/down), on average, over the next 12 months?” “..., over the next 5 years or so?”

The advantages of the WSJ survey compared to the above three surveys are the following.

First, the WSJ survey has the largest number of observations. It is monthly and begins four year before the Pulsenomics survey, which is quarterly. The Case-Shiller survey and Michigan survey are not panel datasets, and the questions are about the individual's own house or houses in her community, so to test for expectation formations, only aggregated mean value can be used. This yields 36 annual observations for the Case-Shiller survey, and around 70 monthly observations for the Michigan survey.\footnote{The Michigan/Reuters survey also releases mean values of the expectation of house price changes for four regions, that would yield 280 observations.}

The second advantage is that the WSJ survey is a panel dataset. The tests for individuals’ expectation formation using a panel dataset are more reliable than using the aggregated mean values. The rationality of the mean of forecasts does not imply the rationality of the individual forecasts. For example, individuals
may have different forecasts with some being positively biased and others being negatively biased, but the mean of these individual forecasts may be unbiased when they offset each other. Moreover, a panel dataset allows for controlling the aggregate shocks, which is impossible using the aggregated mean value data.