

2016

Three Essays on Asset Price Forecasting

Zachary McGurk

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

McGurk, Zachary, "Three Essays on Asset Price Forecasting" (2016). *Graduate Theses, Dissertations, and Problem Reports*. 6198.

<https://researchrepository.wvu.edu/etd/6198>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Three Essays on Asset Price Forecasting

Zachary McGurk

Dissertation submitted
to the College of Business and Economics
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Economics

Arabinda Basistha, Ph.D., Chair

Jack Dorminey, Ph.D.

Stratford Douglas, Ph.D.

Brad Humphreys, Ph.D.

Adam Nowak, Ph.D.

Department of Economics

Morgantown, West Virginia

2016

Keywords: Housing inflation, Capital flows, Exchange rate, Prediction, Investor sentiment, supervised learning, social media.

Copyright©2016 Zachary McGurk

Abstract

Three Essays on Asset Price Forecasting

Zachary McGurk

This dissertation is a collection of essays examining current issues in asset price forecasting. The first chapter of this dissertation discusses the relationship between investor sentiment and excess stock return. This essay takes a novel approach in estimating investor sentiment use social media posts. In this study, I construct daily, equity-specific, investor sentiment indexes from Twitter and test the efficient market theory. We use a multinomial inverse regression to build the dictionary of relevant words and phrases for construction of the indexes. We find that our investor sentiment measure has a positive and statistically significant effect on individual stock returns. These findings are robust to different models and specifications. Chapter 2 examines the ability of international sector predictors to forecast US housing price inflation. Under floating exchange rate regimes, the Dornbusch model predicts shocks to domestic or foreign economies will be reflected in exchange rates. When exchange rates are fixed, shocks are likely to affect the net foreign asset holdings. In this study, I examine the role of the exchange rates and the net change in foreign asset holdings in improving US real estate inflation forecasts. I conduct in-sample and out-of-sample comparison of forecasting models relative to an autoregressive baseline model. I find that inclusion of foreign sector variables can improve the US real estate inflation forecasts by up to 40 percent. This improvement is mostly driven by changes in the net foreign asset holdings at longer horizons. The results are robust to samples at the metropolitan level although with different gains. Chapter 3 continues with this line research. Here, I determine the ability of net capital inflows from regions to forecast US housing inflation. Over the last decade, there has been a high correlation between balance of payment measures (Current Account deficits and Net Financial Accounts). The international finance theory has focused on determining the cause of this relationship. Specifically, this theory has found that deregulation in credit markets, accommodative US monetary policy, and fixed exchange rates caused US housing prices and balance of payments measures to move together. In 2015, BEA released new estimates of balance of payments measures in line with international standards, such that now bilateral financial account data has been created. In this study, I use a number of components from bilateral financial account data, to forecast US housing prices. Further, to empirically test the implications of the international finance theory, I use factor analysis methods to create an bilateral financial account index to forecast US housing prices. Overall, I find that many of these measures are able to produce improved forecasts of up to 50 percent.

Acknowledgments

Over the past four years I have received support and encouragement from a number of individuals. I would like to thank my dissertation committee of Arabinda Basistha, Stratford Douglas, Brad Humphreys, Jack Dorminey and Adam Nowak for all the support and guidance in this process. I am particularly thankful for my dissertation chair, Arabinda Basistha, who spent countless hours meeting with me to improve the quality of my research. I am also very grateful to Adam Nowak who has been a great coauthor and mentor. His help on my dissertation is invaluable. I am grateful for the numerous opportunities Brad Humphreys gave to present my research and improve my presentation skills.

I finally would like to thank all of the other graduate students in my cohort, particularly Danko Tarabar, and Imran Arif, who have been a constant source of support and guidance. Without the other graduate student there is no way I could have passed my comprehensive exams and survive graduate school.

Contents

1	Investor Sentiment and Stock Returns	1
1.1	Introduction	1
1.2	Literature Review	3
1.3	Methodology and Data	5
1.3.1	Tokenization Approach	5
1.3.2	Multinomial Distribution	6
1.3.3	Supervised Learning	8
1.3.4	Bag of Words Approach	9
1.3.5	Timing and Stock Returns	11
1.4	Data and Descriptive Statistics	11
1.5	Cross-Sectional Results	12
1.5.1	Robustness By Size	13
1.6	Forecasting	14
1.6.1	Forecasting Results	15
1.6.2	Forecasting by Market Capitalization	16
1.7	Conclusion	17
1.8	Tables and Figures	17
2	US Real Estate Inflation Prediction: Exchange Rates and Net Foreign Assets	41
2.1	Introduction	41
2.2	Literature Review	43
2.2.1	Finance Factors	43
2.2.2	Macroeconomic Conditions	44
2.2.3	International Factors	44
2.3	Theoretical Considerations	45
2.4	Data	46

2.4.1	S&P/Case-Shiller Home Price	47
2.4.2	Exchange rates and The Net Change in Financial Accounts	47
2.4.3	Descriptive Statistics	48
2.5	Methodology	49
2.6	In-Sample Results	52
2.6.1	National Level	52
2.6.2	Metropolitan Area	53
2.7	Out-of-Sample Results	54
2.7.1	National Level	54
2.7.2	Metropolitan Area	56
2.8	Conclusion	57
2.9	Tables and Figures	58
2.10	Appendix	73
3	Bilateral Capital Inflow and US Housing Inflation Prediction	74
3.1	Introduction	74
3.2	Literature Review	77
3.2.1	In-sample Forecasting Literature	77
3.2.2	Out-of-Sample Forecasting Literature	77
3.2.3	Housing Prices and Capital Inflows	78
3.3	Theory	79
3.4	Data	81
3.4.1	Repeat Sales Housing Price Indexes	81
3.4.2	Bilateral Capital Inflows	83
3.4.3	Descriptive Statistics	85
3.5	Forecasting Methods	86
3.5.1	Empirical Forecasting Model	86
3.6	In-Sample Results	90
3.6.1	National Level In-Sample Results	90
3.6.2	Metropolitan Area Level In-Sample Results	91
3.7	Out-Of-Sample Results	92
3.7.1	National Level Out-Of-Sample Results	92
3.7.2	Metropolitan Area Level Out-Of-Sample Results	94
3.8	Combination Methods	96
3.8.1	Forecast Combination	96

3.8.2	Diffusion Indexes	98
3.9	Combination Results	99
3.9.1	Out-of-Sample National Level	100
3.9.2	Out-of-sample Metropolitan Level	101
3.10	Conclusion	103
3.11	Tables and Figures	104
3.12	Appendix	129

List of Figures

1.1	SP-500 and Cumulative Sentiment Movement	18
2.1	National Housing Price Inflation	59
2.2	Housing Price Inflation and Exchange Rate	60
2.3	Housing Price Inflation and Net Financial Account	61
2.4	Squared Forecast Errors (2Q): AR Model and NFA Model	62
2.5	Squared Forecast Errors (4Q): AR Model and NFA Model	63
3.1	IMF De Facto Exchange Rate Regimes	105
3.2	BEA Defined Regions	105
3.3	BEA Defined Regions-OPEC	106
3.4	Short Term Interest Rates: Saudi Arabia and US	106
3.5	US National Level Housing Inflation	107
3.6	East Coast Metropolitan Area Housing Inflation	108
3.7	West Coast Metropolitan Area Housing Inflation	109
3.8	Non-Coast Metropolitan Area Housing Inflation	110
3.9	Regional Bilateral Capital Inflow	111
3.10	National Level Housing Inflation and 4th Lag of Capital Inflows	112
3.11	East Coast Housing Inflation and 4th Lag of Capital Inflows	113
3.12	West Coast Housing Inflation and 4th Lag of Capital Inflows	114
3.13	Non-Coast Housing Inflation and 4th Lag of Capital Inflows	115
3.14	Squared Forecasting Errors: AR(2) and Capital Inflows (2Q)	116
3.15	Squared Forecasting Errors: AR(2) and Capital Inflows (4Q)	117
3.16	Squared Forecasting Errors: AR(2) and Combination Methods (2Q)	118
3.17	Squared Forecasting Errors: AR(2) and Combination Methods (4Q)	119

List of Tables

1.1	Examples of Labeled Positive and Negative Tweets: Supervised Method	19
1.2	Estimated Coefficients Supervised Method	20
1.3	Finance Dictionary	21
1.4	Examples of Labeled Positive and Negative Tweets: Bag of Words	22
1.5	Descriptive Statistics	23
1.6	Cross Section Excess Returns: Unigram	24
1.7	Cross Section Excess Returns: Bigram	25
1.8	Cross Section Excess Returns: Bag of Words	26
1.9	Cross Section Excess Returns Deciles: Bigram (1-5)	27
1.10	Cross Section Excess Returns Deciles: Bigram (6-10)	28
1.11	Cross Section Excess Returns Deciles: Unigram (1-5)	29
1.12	Cross Section Excess Returns Deciles: Unigram (6-10)	30
1.13	Cross Section Excess Returns Deciles: Dictionary (1-5)	31
1.14	Cross Section Excess Returns Deciles: Dictionary (6-10)	32
1.15	Forecasting Excess Returns: Bigram	33
1.16	Forecasting Excess Returns: Unigram	34
1.17	Forecasting Excess Returns: Dictionary	34
1.18	Forecasting Excess Returns: Bigram Decile (2-5)	35
1.19	Forecasting Excess Returns: Bigram Decile (6-10)	36
1.20	Forecasting Excess Returns: Unigram Decile (2-5)	37
1.21	Forecasting Excess Returns: Unigram Decile (6-10)	38
1.22	Forecasting Excess Returns: Dictionary Decile (2-5)	39
1.23	Forecasting Excess Returns: Dictionary Decile (6-10)	40
2.1	Descriptive Statistics	64
2.2	In-sample National Level Predictability Test	65
2.3	In-sample Metropolitan Level Predictability Test(1Q)	66

2.4	In-sample Metropolitan Level Predictability Test (2Q)	67
2.5	In-sample Metropolitan Level Predictability Test (4Q)	68
2.6	National level: Out-of-Sample Country-wise Forecasting Performance	69
2.7	Metropolitan Level: Out-of-Sample Country-wise Forecasting Performance (1Q)	70
2.8	Metropolitan Level: Out-of-Sample Country-wise Forecasting Performance (2Q)	71
2.9	Metropolitan Level: Out-of-Sample Country-wise Forecasting Performance (4Q)	72
2.10	Housing Inflation Measures and The Predictor Set	73
3.1	Descriptive Statistics	120
3.2	In-sample National Level Predictability Test	120
3.3	In-sample Metropolitan Level Predictability Test(2Q)	121
3.4	In-sample Metropolitan Level Predictability Test (4Q)	122
3.5	National level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance	123
3.6	Metropolitan Level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance (2Q)	124
3.7	Metropolitan Level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance (4Q)	125
3.8	National level: Out-of-Sample Combination Methods Forecasting Performance	126
3.9	Metropolitan Level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance (2Q)	127
3.10	Metropolitan Level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance (4Q)	128
3.11	Data Source	129
3.12	Predictor Definition	129
3.13	In-sample National Level Predictability Test	130
3.14	In-sample Metropolitan Level Predictability Test(2Q)	131
3.15	In-sample Metropolitan Level Predictability Test(4Q)	132
3.16	Out-of-Sample National Level Predictability Test	133
3.17	Out-of-Sample (MSFE Ratio) Metropolitan Level (2Q)	134
3.18	Out-of-Sample (MSFE Ratio) Metropolitan Level(4Q)	135

Chapter 1

Investor Sentiment and Stock Returns

1.1 Introduction

Efficient Market Theory predicts rational investors (with perfect information) choose asset portfolios which diversify away idiosyncratic risk. Under this theory, the cross section of asset returns is only determined by market returns. In contrast, the Behavioral Finance Theory rather assumed that some investors act “irrationally” by making investment decisions partially based on their feelings on specific assets. ¹ Further, if information is assumed to imperfect and transaction costs are positive, the cross section of asset returns then are not only a function of market returns but also investor sentiment. ² The empirical finance literature has found little evidence for the Efficient Market Theory ((Case and Shiller (1988), Baker and Wurgler (2006, 2007)).

Investor sentiment is difficult to measure, and most studies have been limited to overall market measures. Due to data limitations, the behavioral finance literature has relied on macro-level estimates of sentiment using widely published consumer confidence measures Schmeling (2009) or factor model methods Baker and Wurgler (2006, 2007).

Market sentiment measures are limited, and asset specific investor sentiment may be more informative. Baker and Wurgler (2006) discusses that due to imperfect information about smaller firms, any new information causes investors to engage in irrational speculative trading. Market sentiment will not necessarily capture this speculative feeling if the firms are relatively small.

Further, the previously literature has relied on the effect of market sentiment on stock returns for

¹Irrationality is any investment decision that does not conform to the efficient market hypothesis.

²Investor sentiment is defined as the underlying feeling investors have towards assets.

timer horizons of a month or greater (Baker and Wurgler (2006, 2007); Schmeling (2009)). If asset markets are at least some part efficient (i.e. Investor sentiment does not determine a portion of stocks), and information is randomly dispersed, then markets should be the least efficient in the very short run.

To account for both of these critiques, our paper utilizes textual analysis from *Twitter* in order to estimate a daily measure of investor sentiment and test a number of hypothesis. We employ two unique methods to estimate investor sentiment. The first uses counts of predefined positive and negative words to estimate investor sentiment. This method, which has been widely used in the literature (Loughran and McDonald (2011)), creates a dictionary of finance words which we believe explain the underlying feelings of investor. This method does have its limitations, particularly each word in the dictionary has to be assumed to indicate the same amount of sentiment information and only the words which we predefine are the only words which matter in determining sentiment. To account for these limitations, we use a textual analysis method described in Taddy (2013a). Here a training set of tweets are determined to either convey positive, neutral, or negative sentiment. These are then used to predict the sentiment information from all remaining tweet. There is no predefined dictionary, and here individual words are not assumed to be related the same sentiment information.

To test the empirical validity of either theory we determine the relationship between individual stock sentiment and the cross-section stock returns. For robustness, we test if this relationship is similar across firm size. Finally, we determine if investor sentiment is useful in forecasting excess returns at the market level and by firm size.

With the increase use of social media like *Twitter*, or Facebook, the opinions and sentiment of investors can now be estimated for individual assets at high frequency. Anecdotal evidence has shown individual Twitter posts (tweets) to have an effect on specific stock returns. On January 10, 2011, *Business Insider* reported Hip Hop artist, 50-Cent (Curtis Jackson), tweeted

HNHI is the stock symbol for TVG there launching 15 different products. they are no joke get in now.

The article goes on to state:

In the three months to the end of September, the company was operating at a loss with cash of just \$198,000 and a deficit of \$3.3m. Then, on November 23, it said it would offer 180m shares to the public at a price of just 17 cents... trading under the stock name HNHI was worth just 4 cents each. Spurred by the tweet, the stock took off. It hit nearly 50 cents on Monday, before closing at 39 cents.³

By the end of the month, the stock was up to \$1.68. This price increase was relatively short lived.

³<http://www.businessinsider.com/50-cent-tweeting-about-hnhi-2011-1>

In early May, 50-cent terminated his relationship with this company, and the stock dropped in value to \$0.1.⁴

Other examples have shown Twitter has had a market-wide impact. On September 21st, 2015 presidential hopeful Hillary Clinton tweeted:

Price gouging like this in the specialty drug market is outrageous. Tomorrow I'll lay out a plan to take it on...

In response, the iShares Nasdaq Biotechnology ETF dropped nearly 5%.⁵

The social media platform Twitter is used by over 500 million users who express opinions and thoughts on a number of different subject matters including equity prices. Further, an individual can reference specific stocks by affixing a '\$' before the stock symbol in a tweet. This allows all Twitter users to search for tweets discussing a particular stock. Also, this allows researchers to collect tweets supplied by individuals specific to a stock.

Overall we find a relationship between stock returns and investor sentiment. An increase in positive sentiment related to an increase in daily stock returns. Using out-of-sample forecasting tests, we find investor sentiment is marginally able to produce more accurate forecasts compared to a mean only model. Thus, our results indicate that individuals on Twitter are relaying stale information as opposed to providing novel insights.

The remainder of the paper proceeds as follows. Section 2 details the relevant literature, Section 3 describes the methodology and data, Section 4 discusses the results, and Section 5 concludes.

1.2 Literature Review

Empirical research has tested the role of investor sentiment in determining cross sectional stock returns. Baker and Wurgler (2006, 2007) find sentiment is negatively correlated with future returns for those assets that are difficult to value or difficult to arbitrage. Schmeling (2009), Chung et al. (2012) and Baker et al. (2012) find similar results for stock market indexes at the global and national level. Other studies have emphasized the transmission mechanism between sentiment and stock returns. Tetlock (2007) and Dougal et al. (2012) finds sentiment derived from *Wall Street Journal* columnists can be used to predict stock price movements. Most closely related to the study at hand, Da et al. (2011), Chen et al. (2014), Da et al. (2015), and Bartov et al. (2015) find crowd-sourced information from the Internet also has predictive power.

⁴See: <http://www.timothysykes.com/2011/05/rapper-50-cents-headphones-sleek-by-50-canceled-h-h-imports-inc-hnhi-stock-in-freefall/>

⁵<http://money.cnn.com/2015/09/21/investing/hillary-clinton-biotech-price-gouging/index.html>

One point of contention in such research on investor sentiment is the actual measure of sentiment. Baker and Wurgler (2006) go so far as to state that there are no perfect or uncontroversial measures of sentiment. To gauge investor sentiment, we aggregate information from 2.5 million tweets that explicitly reference a stock symbol. Each tweet contains character data that must be processed into a numeric variable that measures sentiment. Two procedures have been used in the textual analysis literature both of which treat the text as a collection of exchangeable *tokens*; a token can be either a word or a phrase. The first approach is the so-called *bag of words* approach that requires the researcher to specify a dictionary of positive and negative tokens, (Loughran and McDonald (2011, 2014); Chen et al. (2014)). From this dictionary, text can be scored based on the relative frequency of positive and negative tokens. To produce reliable results, it is necessary to use a dictionary tailored for the task at hand. For example, Loughran and McDonald (2011) find that 10-k specific dictionaries outperform the more general Harvard-IV-4 TagNeg dictionary of positive and negative tokens.

The second approach does not require the researcher to explicitly specify any prior beliefs regarding the positivity or negativity of individual tokens but rather uses labeled text to identify relevant tokens (Taddy (2013b,a); Mitra and Gilbert (2014)). We refer to this methodology as the *tokenization* approach and use a Bayesian method termed Multinomial Inverse Regression (MNIR), developed in Taddy (2013b), to fit our model. Using labeled data to identify relevant parameters is commonly referred to in the statistical literature as supervised learning.⁶ In the context of the task at hand, the researcher must label a sample set of texts as either positive, neutral, or negative.⁷ From this scored set of texts, the researcher can 1) identify which tokens are associated with positive, neutral or negative texts, 2) estimate weights that indicate the relative positivity or negativity associated with each token, and 3) use the predicted values to create sentiment scores for each tweet.

The tokenization approach requires an indicator variable for each token. As such, the number of variables can grow large for even a modest number of observations which necessitates the need for some dimension reduction. We choose MNIR for two specific reasons. First, MNIR is a dimension reduction method that links token frequencies to class labels. Exploiting this linkage is in contrast to principal components analysis which would not use information in labels but only information contained in the covariance matrix of indicator variables. Second, MNIR uses sparsity-inducing priors in which many coefficients are set equal to 0. By doing so, the MNIR selects a set of empirically relevant tokens. This proves incredibly beneficial for the bag of words approach as results from MNIR can be used to build dictionaries or validate existing dictionaries. The sparse solution of MNIR stands in contrast to principal components analysis where factors are estimated as linear combinations of all tokens. As

⁶Unsupervised learning would refer to the situation where we would infer labels for all tweets without having any available labeled data.

⁷Alternative approaches can be taken that would allow for continuous classifications

noted in the literature, interpreting such loadings can be incredibly difficult.

1.3 Methodology and Data

1.3.1 Tokenization Approach

The goal of any textual analysis is to accurately score a given body of text. The following section details the specific procedures we use to score the tweets. As mentioned above, we compare the bag-of-words methodology to the tokenization approach of [Taddy \(2013b\)](#).

Each tweet contains 140 characters or less. Each tweet can be indexed by date, t , and company, j . On each date, there are $i = 1, \dots, N_{jt}$ tweets that discuss company j . In the following discussion, we abstract from ijt subscripts and instead use $n = 1, \dots, N$ when referring to any of the N tweets.

To indicate that an individual is discussing a company, he or she will append a \$ before the ticker symbol. For instance, when discussing Exxon (Ticker symbol: XOM) and Haliburton (Ticker symbol: HAL) an individual will use \$*HAL* in the tweet. For example:

$$\textit{sold } \$XOM \textit{ shares and bought } \$HAL \textit{ shares} \tag{1.1}$$

In addition to indicating what company is being discussed, each x_n contains words that describe sentiment regarding company j . That is, we can describe x_n as a set of W_n tokens. When the tokens are single words, also called unigrams, dropping subscripts on W_n we write:

$$x_n = \{w_n^1, w_n^2, \dots, w_n^{W_n}\} \tag{1.2}$$

For the example in Eq 1.1, we have

$$x_n = \{\textit{sold}, \$XOM, \textit{shares}, \textit{and}, \textit{bought}, \$HAL, \textit{shares}\} \tag{1.3}$$

We also process the tweets to distil the relevant information using conventional methods in textual analysis. First, we convert all text to lowercase characters. Second, we remove all stop words. Stop words are common words that do not directly indicate sentiment including prepositions, articles, pronouns, etc. Because the tweets frequently indicate the direction of the stock (up or down), we modify a list of stop words from the *SnowballC* package in R to retain finance-specific words. Third, remaining words are stemmed to retain a root form of the word. For example, *buying*, *bought*, *buy* are converted to *buy*. Fourth, for reasons discussed below, we convert the stock symbol of the reference company to *COMPANY* and all other stock symbols in the tweet to *OTHERCOMPANY*. By doing so, we can properly associate

the sentiment with the stock. Fifth, we drop all numbers from the tweet. In unreported results, we found our quantitative results were not substantially impacted by this procedure. However, the list of significant tokens is impacted by this procedure, as expected.

Following the above procedure, if the reference company is Haliburton, the tweet in Eq 1.1, would now read:

$$x_n = \{sold, OTHERCOMPANY, shares, bought, COMPANY, shares\} \quad (1.4)$$

From the tweet in Eq 1.1, it is clear that the individual has a positive outlook on Haliburton and a negative outlook on Exxon. From the tweet in Eq 1.4, it is clear that the individual is bearish on the *OTHERCOMPANY* and bullish on the *COMPANY*.

The above tweet highlights the role of word ordering in tweets. For instance, if we were to switch *sold* and *bought* it would appear as if the individual is bearish on Exxon and bullish on Haliburton. A simple way to control for ordering is to use two word phrases, or bigrams, as the tokens. Using the bigram approach, the above tweet would then become⁸

$$x_n = \{sold.OTHERCOMPANY, OTHERCOMPANY.shares, shares.bought, bought.COMPANY, COMPANY.shares\} \quad (1.5)$$

From the above collection of bigrams, there is no confusion as to the sentiment of the reference stock. When a tweet does not make reference to other companies, there is generally no risk of confusion as to the sentiment. Therefore, it is not clear whether or not the unigram or bigram method will perform better.

1.3.2 Multinomial Distribution

We use a collection of $k = 1, \dots, K$ unique tokens collected from the N tweets. The W_n tokens in x_n are represented as draws from a multinomial distribution with unknown token probabilities. In other words, we draw W_n tokens from the K possible tokens, where each token is drawn with probability q_{nk} . When using the multinomial distribution, we assume the tokens in x_n are exchangeable.

For each token, define the indicator variable $z_n^k = 1$ if token k is in the set x_n and $z_n^k = 0$ otherwise. Associated with these indicator variables is the vector z_n , the $K \times 1$ vector of these indicator variables. We assume each tweet can be classified into one of 3 distinct, unordered categories: *positive*, *negative* or *neutral*. Without loss of generality, define the class of each tweet as $y_n \in \{1, 2, 3\}$. In what follows,

⁸We use the convention where different words in a bigram are separated by a "."

it will be useful to keep track of y_n by defining the 3×1 vector v_n where the y_n element is equal to 1, and all other elements are equal to 0.

Of course, this classification scheme is by no means exhaustive. There exist an innumerable number of ways to classify each tweet based on various criteria. It should be emphasized that, although the categories appear to have an inherent ordering, the procedure used in this paper does not require such an ordering. Further, we only classify tweets into the *positive* and *negative* categories. Because of this, the *neutral* category might also be thought of as a *neither* category.

Conditional on y_n and W_n , we assume that x_n is drawn from a multinomial distribution. Using notation similar to Taddy (2013b), the vector of token counts, z_n , is distributed as a multinomial random variable $z_n \sim MN(q_n, W_n)$ where

$$q_{nk} = \frac{e^{\eta_{nk}}}{\sum_l e^{\eta_{nl}}} \quad (1.6)$$

$$\eta_{nk} = \alpha_k + v_n \psi_k + \epsilon_{nk}$$

The above equation requires a discussion in order to identify the relevance of each parameter. Conditional on y_n , q_{nk} indicates the probability that token k will be in x_n . α_k is a token specific parameter. When token k appears more frequently across all classes, α_k will be large. ψ_k is a 3×1 vector of parameters specific to token k that indicates the class-specific relative frequency of token k . The product $v_n \psi_k$ is the y_n element of ψ_k ; the y^{th} element of ψ_k will be positive whenever token k appears more frequently in tweets of class y compared to other classes. The term ϵ_{ij} is an i.i.d. error term.

Given the parameters ψ_k , we can summarize all relevant information in x_n using the linear combination $s_n^T = W_n^{-1} \Psi' z_n$ where $\Psi' = [\psi_1, \dots, \psi_K]$. Taddy (2013b) shows that this linear combination is a sufficient reduction where by the distribution of y_n given the 3×1 vector s_n^T is independent of the distribution of y_n given the entire $K \times 1$ vector z_n . For the purposes at hand, s_n^T can be interpreted as a measure of sentiment. In a regression context, s_n^T can be used as an explanatory variable in a parsimonious regression in lieu of using the potentially thousands of regressors in z_n^T . At its core, the above procedure is a method to reduce the dimension of the token vector z_n .

As mentioned above, the total number of unique tokens can be quite large. Any model with a large number of parameters runs the risk of over-fitting. In our estimation, we experimented with various token sets and settled on a modest cutoff using only the 3,000 most frequent tokens when estimating the model. In unreported results, the coefficient estimates were robust to token sets as large as 5,000. To prevent this, Taddy (2013b) uses a Laplace prior for the ψ_k and selects the vector $\hat{\Psi}$ as the vector which maximizes the posterior likelihood given the priors. Because Laplace priors are used, the resulting $\hat{\Psi}$ is sparse with many elements equal to 0. In this sense, the procedure is both variable selection and

coefficient estimation. From a predictive standpoint, only tokens with non-zero coefficients in $\hat{\Psi}$ are relevant when predicting the class of the tweet.

1.3.3 Supervised Learning

To estimate Eq 1.6, it is necessary to use a training set of tweets labeled as either negative, neutral or positive. Taddy (2013b) uses restaurant ratings labeled by multiple users. In contrast, Taddy (2013a) examines political sentiment towards presidential candidates on Twitter, which does not contain labels. To score the tweets, Taddy (2013a) employed multiple individuals using workers from the Amazon Mechanical Turk service to read and score 2,242 tweets.

Labeling tweets that reference stocks is not as straightforward as labeling tweets that reference political candidates. Most tweets manifest a clear sentiment. Examples include *\$XOM looks like a buy* and *bearish on \$XOM*. However, due to esoteric finance vocabulary, it is possible that a randomly selected individual will miss subtleties associated with payoffs that can lead to an incorrect labeling. Such instances frequently occur when discussing options. For example, *bought May puts* and *bought May calls* should be classified as negative and positive, respectively, although this distinction is not apparent to an individual without an understanding of finance terminology. Taddy (2013a) takes several steps to pre-screen the workers to ensure the fidelity of their labeling scheme. However, due to cost considerations and conversations with several individuals in the crowd-sourcing industry, we concluded that using the Amazon Mechanical Turk service was not viable.

Instead, both authors manually labeled using a random sample of 3000 tweets to be used as the training set.⁹ Similar to the arbitrary notions expressed in Baker and Wurgler (2006) regarding the measurement of sentiment, it is possible that our categories and labeling procedure are also arbitrary and subjective. To mitigate such problems, we undertook several steps to produce results that were both objective and thorough. First, a minimum observation period of 30 seconds was required before scoring was allowed to ensure a sufficient amount of time was spent analyzing the tweet. Second, a repeated sampling procedure was performed to ensure consistent labeling by author. Third, some tweets were scored by each author to mitigate any individual bias. Any tweets that were had multiple labels by the same individual or across individuals were discarded. This resulted in a collection of 2919 labeled tweets or an error rate of approximately 2.7%. An examination of these discarded tweets indicate human error and not any discrepancy in the perceived sentiment of the tweet.

As a preview of the results, Table 1.1 provides a snapshot of the ten most negative and most positive tweets based on s_{ijt}^T . The original tweet is shown where the reference stock symbol has been removed and replaced with *COMPANY*. In most cases, the individual is stating the position (long or short) and

⁹The Authors are more than happy to provide the labeled training set upon request

in some cases the type of security (call, put or stock).

Table 1.2 shows the largest coefficients estimated for the unigram and bigram methods for both positive and negative sentiment. Panel A shows the coefficient of the tokens most related to positive sentiment. Here both buy and long seem to be important as having some of the largest coefficients. For Unigrams, some of the most interesting important tokens are green and strong. In general, these are tokens which would not seem to be as important. For the positive Bigram, biotechmoney COMPANY is another token which off hand may not seem that important to sentiment.

Panel B shows the estimated coefficient for the negative sentiment. Unlike the positive sentiment, all of these tokens seem to be standard finance words or phrases which are would correspond to negative sentiment. Tokens like short, sell, and weak all seem to indicate negative feelings towards assets.

1.3.4 Bag of Words Approach

The bag of words approach begins with a dictionary of positive and negative words. We first create a dictionary of unigrams we believe Twitter users would use when referring to stocks. This dictionary is built using unigram counts from the labeled data from the procedure above. The finance dictionary is given in Table 1.3, with the list of positive and negative words. We favor a dictionary with fewer words as we find dictionaries with more words tend to produce worse results.

These words can indicate either company fundamentals, trading positions, or technical indicators. Next, we divide this dictionary into positive and negative words. The Positive Word List (PWL) is a list of words or phrases that indicate the Twitter user is positive about the stock. The Negative Word List (NWL) is like wise a list of words or phrases that indicate negative sentiment. We write PWL and NWL as mathematical sets:

$$PWL = (PWL_1, PWL_2, \dots) = (bought, bullish, breakout, \dots) \quad (1.7)$$

$$NWL = (NWL_1, NWL_2, \dots) = (old, bearish, dismal, \dots) \quad (1.8)$$

Using PWL and NWL , three measures of sentiment are estimated. First, positive sentiment is the total count of positive words in each tweet. Positive sentiment for a given tweet is written as:

$$Ps_n^B = \#(x_n \cap PWL) \quad (1.9)$$

Here, $\#(X)$ is the number of words in the set X . Second, negative sentiment is the total count of negative words in each tweet. Third, overall sentiment is estimated positive sentiment subtracted by

negative sentiment. Negative sentiment for a given tweet is written as:

$$Ns_n^B = \#(x_n \cap NWL) \tag{1.10}$$

We determine overall sentiment of each tweet by counting the total number of words in PWL minus the total number of words in NWL . The sentiment for a given tweet is written as:

$$s_n^B = \#(x_n \cap PWL) - \#(x_n \cap NWL) \tag{1.11}$$

If there are more positive words than negative words in x_n , then $s_n^B > 0$ and vice-versa.

Differences and similarities between the bag of words approach and the tokenization approach can be found by comparing s_n^B and s_n^T . Each measure of sentiment is a linear combination of the vector z_n . However for the bag of words approach, positive words are given a weight +1 and negative words are given a -1. In the tokenization approach, tokens are assigned weights based on the strength of their association with negative and positive tweets. Of course, it is possible to estimate weights for words in the dictionary using the Eq 1.6 or ex-ante specify weights. However, we find that the words in the dictionary are almost always selected when using a much larger set of tokens.

For comparison, Table 1.1 provides of the 10 most negative and positive tweets based on s_n^T and Table 1.4 shows the 5 most negative and positive tweets for s_n^B . The original tweet is shown where the reference stock symbol has been removed and replaced with *COMPANY*. In most cases, the individual is stating the position (long or short) and in some cases the type of security (call, put or stock).

The supervised method seems to be more accurate are estimating the sentiment of the tweet. Specifically, the bag of words approach mislabels the 2nd most positive tweet as positive. This tweet is more likely displaying negative sentiment. The text :

$$Nervous abt buying COMPANY... \tag{1.12}$$

implies the investor is unsure the price of that stock will increase.

The supervised method also can be superior from a theoretical stand point. The bag of words approach puts equal weight on each token and do not allow for the difference. Specifically, it assumes that each token defined in our finance dictionary conveys equal information about sentiment. For example, the tokens “bought” and “great” provide equal sentiment information. In contrast, the supervised method differs in that it weights each token by likelihood that the specific tweet shows positive, negative, or neutral sentiment information.

1.3.5 Timing and Stock Returns

Social media is unique in that users can provide information at any time of day; however, trading on the NYSE begins 9:30 AM and ends 4 PM. Our analysis is interested in asking if information from social media can explain variation in excess stock returns. Because of this, it is necessary to be precise when indicating the start and end of a period. We use two breakpoints in order to segment the 24 hour day into three segments. The two breakpoints indicate the opening and closing of the NYSE. The day is then split up into three periods defined by:

- Pre-Market: 00:00:00-09:29:59
- Market: 09:30:00-15:59:59
- Post-Market: 14:00:00-23:59:59

Using this segmentation, we regress excess returns on the sentiment calculated within each time segment. That is, we regress excess return on pre-market sentiment, market sentiment and post-market sentiment. We are interested in whether or not sentiment is correlated with current or future stock returns. We estimate:

$$ER_{jt} = \alpha + \gamma S_{jt} + u_{jt} \quad (1.13)$$

Where ER_{jt} is the excess return of the security calculated as the raw return minus the return on the S&P 500, and S_{jt} is the average sentiment for stock j at time t calculated by averaging the individual s_{ijt} across all $i = 1, \dots, N_{jt}$.

1.4 Data and Descriptive Statistics

The data was collected using the *Twitter* package in R beginning August 20, 2012, and ending June 12, 2013. Data was collected daily by searching Twitter for tweets discussing stocks in the Russell 5000. The search program allowed us to search for the 5,000 most recent tweets from the past seven days for any number of stocks. At the start of the data collection, Twitter did not limit the total number of searches. However, by mid-2013, Twitter began to limit the number of downloads, and data collection was effectively terminated. Over this period, we collected 3,941,149 unique tweets over 296 days discussing 4,972 unique stocks.

Figure 1.1 shows the movement of cumulative investor sentiment and the SP-500. The ¹⁰ There seems to be a relationship, albeit a weak one, between the SP-500 and cumulative sentiment. From September

¹⁰cumulative sentiment is the cumulative value of the positive bigram sentiment

to November, both the SP-500 and cumulative sentiment measure move in the same direction. Starting in mid-November, the SP-500 again started to increase and followed this pattern until May 2013. The cumulative sentiment differs and declined until March 2013 and then finally increased. Even though these two measures do not seem to follow each other during the mid-November to March, there seems to be a relationship for the other time periods.

The period between mid-November and March also saw a number of external events which likely had a negative impact on the market, specifically the US Government “Shutdown.”¹¹ Investors during this period were known to being less optimistic and finally “climbing over the wall of worry” preceding the recovery of the great recession.

Table 1.3 shows some descriptive statistics. The number of total observations is 116,649. This counts for the total number of average sentiment for each firm (which were tweeted about) for each day. This relates to on average 23 days per firm and 792 tweets for each firm. This also corresponds to around 13,314 tweets about firms per day.

The first portion shows the descriptive statistics for the investor sentiment measures. The bag of words measure, called dictionary, has a mean of 1.873 which corresponds to an average positive investor sentiment during this period. The minimum is -2,989 and maximum is 1,156.

The descriptive statistics from bigram and unigrams are similar. The negative.bigram.z has a mean of -0.449, while the negative.unigram has a mean of -0.482. The positive.bigram.z has a mean of -1.576, while the positive.unigram has a mean -1.596. This would imply by that the tokens only containing one word may overestimate the relative sentiment compared to a token containing two words. Not only do the bigram measures have lower means than the unigram, but these investor sentiment measures also have lower standard deviations.

The bottom portion shows descriptive statistics on the specific firms themselves. These include volume, price, return, book to market value and market value. The average return is 0.403. The minimum and maximum returns are -49.624 and 49.383. Market value is calculated by multiplying volume by price (prc). As a note, market value is in tens thousands. The average market value is about 154 million.

1.5 Cross-Sectional Results

Table 1.6 and 1.8 shows the cross sectional results for the supervised and bag of words methods. These results show evidence that investor sentiment is related to excess returns. The positive bigram measure performs the best and produces significant results for all timings at the 1% significance level. The coefficient are positive and range from 0.177 to 0.360. The neutral measure also performs well and is

¹¹<http://www.usatoday.com/story/money/markets/2012/12/31/stocks-post-doubledigit-gains-in-2012/1800223/>

significant at the 1% level for all timings except pre-market. The coefficients are largest and range from 0.582 to 0.797. Finally, the negative measure performs the worst and it is only significant for the 10% level, for the pre-market and post market timing with coefficients of 0.326 and 0.168.

The results for the unigram measures, found in Table 1.6, are similar to the bigram measures. The positive unigram measure is significant for all timings and at the 1% level. Coefficients range from 0.254 to 0.641. For the negative and neutral method all but the pre-market are significant at the 1% level. The coefficients for negative measure ranges from 0.624 to 0.822 and for the neutral measure coefficients range from 0.483-0.684.

Table 1.8 shows the results for the bag of words method. The results are similar to both supervised methods and bag of words measure is significant for all, market and post-market at the 1% level. The coefficients here are relatively smaller than any of the supervised methods, ranging from 0.003-0.006.

These results show that investor sentiment is an important determinant for determining excess returns. All investor sentiment measures are at least significant for one timing. This provides evidence against the standard efficient market hypothesis in favor of the behavioral finance theory.

There is also evidence that the supervised method is superior to the bag of words method and seems to be economically significant. In general, there is a difference in the way each coefficient is estimated. For the bag of words method, with an increase in the average number of net positive words by one, excess returns increases by the coefficient. The supervised method is interpreted as with an increase in the average sentiment by one, excess returns increases by the coefficient. The supervised methods coefficients are harder to interpret. Although the difference between the two shows there seems to be some difference, if not simply by differences in the precision on the investor sentiment measure. Finally, there is relatively no difference in results from timing. Each performs equally as well.

1.5.1 Robustness By Size

Baker and Wurgler (2006, 2007) detail that due to limited information about smaller less known stock are particularly influenced by investor sentiment. The idea is that any information released about these stocks will cause investors speculate and drive excess returns to be positive or negative. This would imply the relationship between excess returns and investor sentiment should be stronger for smaller firms. To test this, we divide the sample by decile market capitalization and estimate Eq 1.13.¹²

The results for the bigram supervised method are found in Table 1.9 - 1.10. The first and sixth decile do not have significant results. Only the second, ninth, and tenth decile have all bigram investor sentiment measures which are significant. The negative measure produces a positive coefficient for only the second decile with a large coefficient of 12.936, and it is significant at the 5% level. For the third

¹²As a note, due to the change in the daily value of market capitalization, some stocks would be part of different decile at different dates.

decile through the tenth decile, the coefficient is negative, and the absolute value declines as the decile increases range from -16.60 for the third decile to -0.294 for the tenth decile.

The positive measure is similar where for the second decile, the measure is significant at the 1% level, and it has a coefficient of -17.367. The third -tenth decile are all positive and range from 8.085 to 0.281. Again a similar decline in the coefficient value.

Table 1.11 - 1.12 show the results for the unigram measure. Similarly, many of supervised measures are significant where only the first, sixth, and seventh decile do not have any significant measures. These also see a steady decline in coefficients.

Table 1.13 - 1.14 show the bag of words method decile results. The change in coefficients does not follow as strict of a pattern as the supervised method, although the coefficient seems to decrease in size as the decile increases. The second decile has the largest significant coefficient of 0.157 where as the tenth decile has the smallest significant coefficient of 0.003 (both at the 1% significant level).

The relative size of the coefficients of lower decile market capitalization can be explained through a number of ways. First, this results fall in line with the behavioral finance theory of Baker and Wurgler (2006) where smaller firms are disproportionately effected by investor sentiment. Limited information about these firms causes investors to speculate when any information is released. The conflicting signs of the supervised methods for the second and third decile are interesting. The negative sign from the positive sentiment for both the bigram and unigram measure could be explained by firms which are relatively small most likely that positive sentiment is believed to be overly optimistic causing investors to sell. Second, these result could be caused by a limited number of observations. Fewer tweets are available for the smaller size firms, so individual tweets have larger weight. A few extremely positive or negative tweets can determine this relationship.

1.6 Forecasting

Given that investor sentiment is useful in determining excess returns, we further test the ability of our investor sentiment measures to forecast excess returns. To do this, we employ both in-sample and out-of-sample analysis. We specifically estimate: ¹³

$$ER_{j,t+1} = \alpha + \gamma S_{j,t} + u_{j,t+1} \quad (1.14)$$

The out-of-sample analysis is a forecasting technique where the sample is split into two portions by time. The first portion is used to estimate the parameters. The estimated parameters are then used to forecast the second portion. The out-of-sample analysis allows the researcher to determine how well the

¹³Due to data limitations, we only use a one day forecast horizon

model forecasts in the past. To estimate forecast accuracy the ratio of the mean squared forecast errors (MSFE) between the non-predictable model and Eq 1.14 is taken.

In contrast, in-sample analysis takes advantage of the full data set to estimate the relationship over the whole period. A simple t-test is used to determine if the investor sentiment measure can provide some predictability. If any measure is significant, we can say that that investor sentiment measure is significant. To compare this added predictability of these measures we take the ratio of mean squared errors (MSE) between the non-predictable model and Eq 1.14.

When the MSE and MSFE ratios are less than one, we can say that Eq 1.14 produces a more accurate forecast. The close to zero these ratios are the more accurate are the forecasts from the investor sentiment model.

There has been much discussion in the econometric theory in which method (in-sample or out-of-sample) is superior in determining the relatively forecast ability of models. In general, this literature has focused on the cases when the results from in-sample and out-of-sample differ (Inoue and Kilian (2005); Clark and McCracken (2005b); Clark (2004)). These works show some conflicting results. Due to this, we only focus on results which match, in-sample and out-of-sample.

1.6.1 Forecasting Results

Tables 1.15 - 1.17 show the in-sample and out-of-sample forecasting results for only a one day ahead forecast. For the in-sample analysis, t-test and F-test, and MSE ratio are used to determine added forecast accuracy over the constant model. Only the MSFE ratio is shown for the out-of-sample result. The split sample date is 3-February-2013. This relates to a split sample ratio of around 0.9.¹⁴

Table 1.15 shows the forecasting results for the bigram model. The in-sample analysis shows Lag.positive.z significant but only at the 10% level. No other predictors (besides the constant) are significant. F-test shows that these predictors are jointly significant at the 1% level. Finally, this model can produce more accurate forecasts compared to the constant model with increases in accuracy by 0.05 percent for in-sample analysis. For out-of-sample the bigram model is less accurate compared the constant model.

Table 1.16 shows the in-sample and out-of-sample results for the unigram model. Similar to the bigram model, where only the lag.positive.unigram.z is significant at the 5% level. No other predictors are significant. The F-test shows that these predictors are jointly significant at the 1% level. Again the MSE ratio shows that the unigram model produces a more accurate forecast by around 0.05 percent. The in-sample and out-of-sample results match. The out-of-sample gains are about 0.01%.

Table 1.17 show the forecasting results for the dictionary model. This model preforms poorly in-sample and out-of-sample. Lag.dictionary is not significant. This model can produce a lower MSE ratio

¹⁴Only the all sample timing results are shown, given the limited difference between the timing.

than the constant model again with an increase in forecast accuracy by 0.05 percent. The out-of-sample MSFE ratio is also greater than one with decrease in forecast accuracy by about 0.05 percent.

Overall, the in-sample and out-of-sample results only match for the unigram model. This gives some evidence that the investor sentiment can predict future excess returns. Given the large volatility of stock returns/ excess returns, these results do not imply that the unigram model can produce relatively accurate daily forecasts. It does mean that it can do it better than the constant model and shows that daily excess return has some predictable portion.

1.6.2 Forecasting by Market Capitalization

Continuing with the idea that smaller firms may be asymmetrically effected by investor sentiment provided by [Baker and Wurgler \(2006\)](#), we determine how well does our investor sentiment measures can forecast excess returns by market capitalization deciles. We again use in-sample and out-of-sample analysis and focus on the matching results.

Table 1.18 - 1.21 show the results from the supervised methods. While Table 1.3 shows the results for the bag of words methods.

For the Bigram measure, the lags of the investor sentiment measures are only significant for the third, fifth, and eighth and ninth decile. While the in-sample and out-of-sample forecast accuracy gains only match for the second, fifth and sixth decile. The out-of-sample accuracy gain is only around 0.1%.

The Unigram measure only has significant predictors for the fifth, eighth, and tenth decile. The in-sample MSE and out-of-sample MSFE ratios match for the fifth-seventh deciles. The biggest gain is for the fifth decile of around 1%.

The bag of word measure is only significant for third- fifth and tenth decile, in in-sample testing. The in-sample MSE ratio and out-of-sample MSFE ratio are below one for these decile, showing an increase in forecast accuracy. The out-of-sample gains in forecast accuracy range from 0.2-0.4%, with the largest gains for the 3rd decile.

The most interesting part of these results are that each investor sentiment measure differs in which decile it produces significant results and by gains in forecast accuracy. The gains may seem rather small, but due to this being daily forecasts, these are substantial over monthly or quarterly horizons. For example, the increase in forecast accuracy for the fifth decile from the unigram model would be around 34% over a month horizon. Overall, there is some strong evidence that not only is investor sentiment an important determinant of excess returns, it is also able to predict excess returns.

1.7 Conclusion

The standard finance theory predicts excess returns should not be a function of any variable. This result breaks down when investors do not have perfect information and act irrationally. There has been little empirical evidence that these assumptions hold. The behavioral finance has explained mis-pricing of assets through investor sentiment. We overall find evidence of investor sentiment playing a role in determining the cross section of stocks returns.

Adding more to the empirical evidence against the classic finance theory, we estimate a daily firm-specific investor sentiment measure of investor sentiment. We improve upon the classic dictionary method which gives equal weight to each word, by using supervised sentiment measures. With both methods, we find that overall investor sentiment is driving overall cross section of excess returns. This relationship seems to be the strongest for the smallest least known firms. Thus providing additional evidence for these firms being particularly vulnerable to investor sentiment.

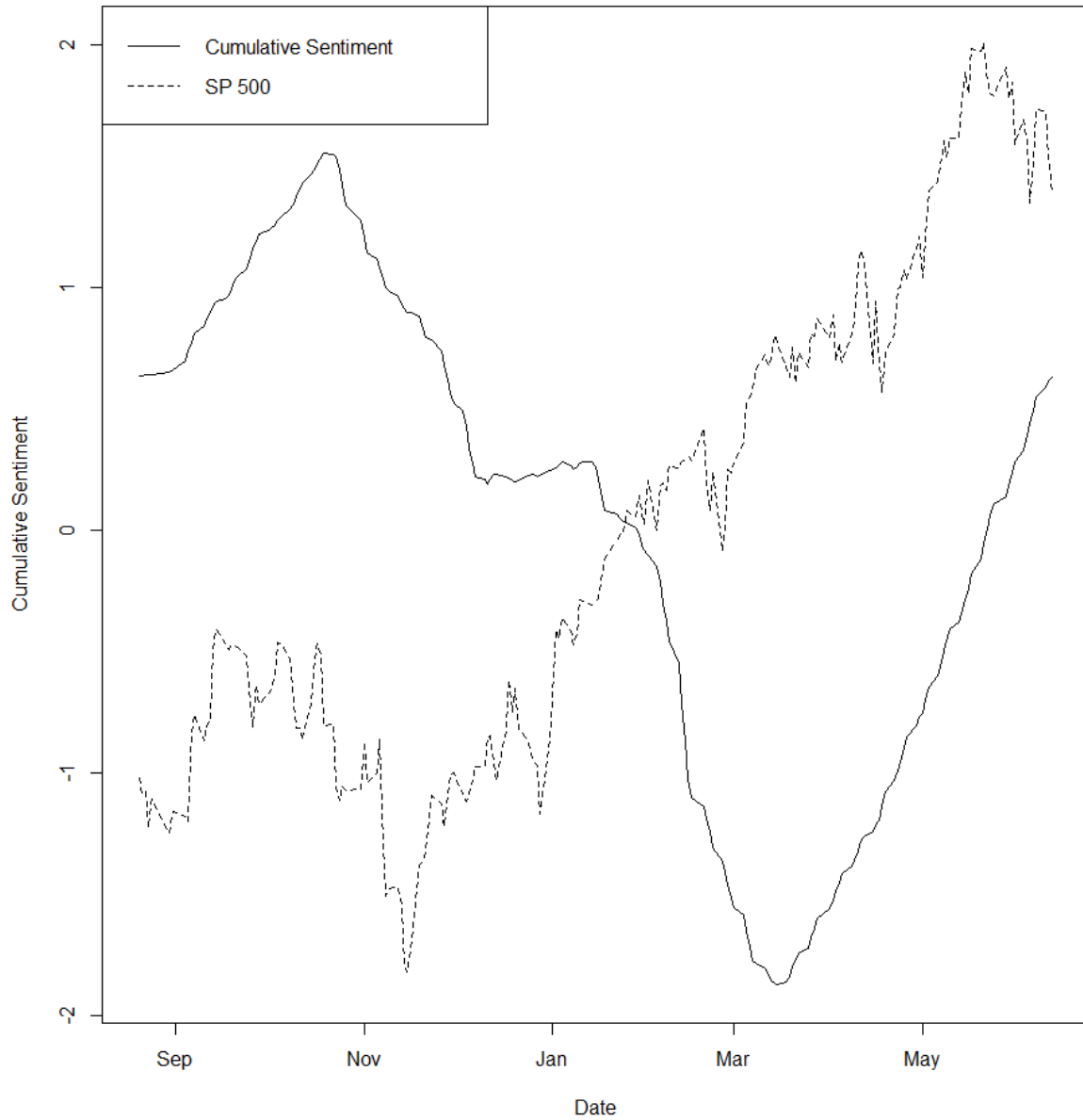
We also find some evidence that investor sentiment can produce more accurate forecasts compared to a constant only model. Increases in forecast accuracy for the overall market may be relatively small but over longer time horizons these may be substantial. Further gains are found depending on the size.

With the increased use of social media, more information on the individual feeling of investors will become more available. This would allow for future researchers to understand the longer run implications of individual firm investor sentiment. In fact with the Library of Congress currently cataloging all tweets a better understanding of this relationship can be determined.

With a complete twitter data, further works can test the role of each the sources of investor sentiment, listed in [Avery and Chevalier \(1999\)](#), on excess returns. Specifically, the number of followers or re-tweets may be useful in determining the effect of expert opinion caused investor sentiment on excess returns.

1.8 Tables and Figures

Figure 1.1: SP-500 and Cumulative Sentiment Movement



Note: Figure 1.1 shows the cumulative sentiment measure against the SP-500 Returns for 2012-2013.

Table 1.1: Examples of Labeled Positive and Negative Tweets: Supervised Method

Panel A: Positive Tweets

- 1: rt @reddogt3live: nice to see the treasure for trash trade inching higher. still long some COMPANY as it still has upside room. entries and exits matter
 - 2: rt @redacre: COMPANY upgraded by mizuho securities from neutral to buy @chasingthealpha see? #pumpers - COMPANY was a screaming buy at \$25 ; below
 - 3: bought COMPANY \$92.5 jan calls...looks like its about to break
 - 4: pr: special dividend declaration alert for COMPANY (expect long).
 - 5: bought COMPANY at \$3.55 and then again at \$2.74. sitting over \$9 now. #letitride
 - 6: rt @sciencetrader: things i like with good long term charts that are high but have room to run: \$xlb, \$itb, \$xhb, COMPANY, \$vz, \$oih, \$lmt, \$wfm,(\$amzn needs time)
 - 7: bought some COMPANY and \$vvus today
 - 8: paper bought 5000 COMPANY jan 2014 40 calls for \$.88
 - 9: bought COMPANY jan 2014 95 calls for 1.33 #optionsaction
 - 10: @jimcramer hi j, bought COMPANY jan 600 calls in oct. when stock was at 610. down now, time decay a worry, do i ride to the end?
-

Panel B: Negative Tweets

- 1: i am still short COMPANY and \$pkx. getting tough to stay short as the major averages and sector indexes are oversold near term.
 - 2: sold some COMPANY march 85 calls for \$0.73
 - 3: sold COMPANY stock at 628.50 for +6.00 – hedge now removed. 20 sma test ; rejection was the trigger.
 - 4: good time to reset the portfolio...sold COMPANY and \$scsco holdings...time to find the next investments #retirement #dividends
 - 5: sold all COMPANY for \$1.5k profit. what an amazing move today.
 - 6: sold those COMPANY jan calls, +16.66%, ffv, -5.2%.
 - 7: sold COMPANY at \$4.87 from \$4.80 entry. small gain
 - 8: sold my COMPANY mar \$5 calls at \$1.2 (from \$1.4) - no need to be a hero holding thru briefing docs tomorrow
 - 9: thinking of shorting COMPANY ... just keeps going down
 - 10: sold more COMPANY +60%
-

Note: Table 1.1 provides a snapshot of the ten most negative and most positive tweets based on s_{ijt}^T .

Table 1.2: Estimated Coefficients Supervised Method

Panel A: Positive			
Unigram		Bigram	
highs	1.1	buy COMPANY	1.3
buy	1	long COMPANY	1.1
long	1	biotechmoney COMPANY	1
added	0.9	look OTHERCOMPANY	1
positive	0.8	special dividend	1
green	0.7	call options	0.9
strong	0.7	eye COMPANY	0.9
dividend	0.6	look COMPANY	0.9
breaking	0.5	looks good	0.9
Panel B: Negative			
Unigram		Bigram	
lower	0.7	morning COMPANY	1.4
weak	0.7	taking COMPANY	1.4
profits	0.8	downgrades COMPANY	1.5
seeing	0.7	sell half	1.5
bearish	1.3	PRICE puts	1.8
short	1.4	short term	1.8
downgrades	1.5	sell COMPANY	1.9
sell	1.5	short COMPANY	1.9
shorted	1.7	drops COMPANY	2

Note: Table 1.2 shows the largest word coefficients from the supervised method.

Table 1.3: Finance Dictionary

Panel A: Positive Dictionary

buy, buying, bought, long, bull, bearish, good, acceptable, excellent, exceptional, favorable, great, positive, awesome

Panel B: Negative Dictionary

added, sell, selling, sold, short, bear, bullish, bad, atrocious, awful, crummy, dreadful, lousy, poor, rough, sad, unacceptable, blah, bummer, downer, garbage, gross, imperfect, inferior, junky, abominable, amiss, crappy, cruddy, dissatisfactory, erroneous, fallacious, faulty, godawful, inadequate, substandard, unsatisfactory, shitty

Note: Table 1.3 shows positive and negative finance dictionary used in estimating the bag of words investor sentiment measure.

Table 1.4: Examples of Labeled Positive and Negative Tweets: Bag of Words

Panel A: Positive Tweets

- 1: @JEFFREYJKEITH no mans land when in doubt buy time. I'm going long cover me . Jan 600/700 bull call spread at 52.5 COMPANY
 - 2: Nervous abt buying COMPANY at 545ish then buy deep money long dated calls. Too expensive? Then buy call spread (lower cost+limit prof)
 - 3: Very tempted to buy some COMPANY... havent bought a non-financial long time. Other than \$YNGFF.yuck
 - 4: Not 2 long ago, everyone was saying COMPANY was a buy at \$550 and change; goin to \$600, then \$1,000.00 millions of people bought . They lost.
 - 5: Everyone has been saying buy COMPANY and shit, well based on my charting skills, and bearish look on \$GOOG now is time to long COMPANY 430.58
-

Panel B: Negative Tweets

- 1: sold \$tza jul \$11 ns contracts for 1.04, basis .84. added to COMPANY jul short \$6 puts by selling .65. added to \$wlt, sold \$cprx
 - 2: also added to COMPANY am & sold 1/2 at 6.38 for (avg breakeven) & pondered last 15 mins before selling other 1/2 now at 6.52 for nice gain.
 - 3: @tlmontana @mwonder74 i had sold my COMPANY shares & replaces them w/ short puts over \$34. i think dips buyable. sell otm puts.
 - 4: path as i did with COMPANY .. best way sell to open, sold jan \$5 calls for \$0.25 ... will keep premium or short it at \$5 on opex
 - 5: positions. added to \$npcuf. holding \$exel, exel nov calls, small \$ziop & \$ttnp. short \$osir. sold COMPANY on bounce; will re-enter after nce.
-

Note: Table 1.4 shows provides a snapshot of the most negative and most positive tweets based on the bag of words method.

Table 1.5: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Dictionary	116,649	1.873	15.921	-2,989	1,156
buy.sell	116,649	0.959	7.248	-192	731
long.short	116,649	-0.001	6.656	-268	234
negative.bigram.z	116,649	-0.449	0.170	-2.205	-0.340
neutral.bigram.z	116,649	-0.095	0.131	-1.778	-0.0001
positive.bigram.z	116,649	-1.576	0.270	-1.696	1.036
negative.unigram.z	116,649	-0.482	0.218	-2.623	-0.301
neutral.unigram.z	116,649	-0.103	0.150	-2.037	-0.0001
positive.unigram.z	116,649	-1.596	0.388	-2.365	1.819
prc	116,649	38.826	58.048	-123.235	2,766.000
vol	116,649	4,031,242.000	10,589,024.000	0	463,491,000
ret	116,649	0.403	3.756	-49.624	49.383
book.value	116,649	7,584.627	22,426.350	-8,341.000	236,956.000
market.value	116,649	15,463.110	39,799.860	0.018	499,821.000

Note: Table 1.5 shows the descriptive statistics for a number of financial variables and investor sentiment measures.

Table 1.6: Cross Section Excess Returns: Unigram

	<i>Dependent variable:</i>			
	Excess.Return			
	All	Pre-Market	Market	Post-Market
negative.unigram.z	0.691*** (0.071)	0.191 (0.127)	0.624*** (0.108)	0.822*** (0.090)
neutral.unigram.z	0.684*** (0.079)	-0.109 (0.134)	0.483*** (0.116)	0.682*** (0.105)
positive.unigram.z	0.621*** (0.038)	0.254*** (0.067)	0.546*** (0.057)	0.641*** (0.049)
Observations	116,649	34,342	62,740	76,778
R ²	0.003	0.001	0.002	0.003
Adjusted R ²	0.003	0.0005	0.002	0.003
Residual Std. Error	3.657	3.391	4.157	3.961
F Statistic	111.363***	6.526***	36.217***	74.722***

Note: Table 1.6 shows the cross section unigram results *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.7: Cross Section Excess Returns: Bigram

	<i>Dependent variable:</i>			
	excess.return			
	All	Pre-Market	Market	Post-Market
negative.bigram.z	0.120 (0.080)	0.326* (0.168)	0.115 (0.116)	0.168* (0.099)
neutral.bigram.z	0.797*** (0.088)	-0.215 (0.152)	0.582*** (0.127)	0.794*** (0.116)
positive.bigram.z	0.308*** (0.048)	0.326*** (0.104)	0.360*** (0.067)	0.177*** (0.064)
Observations	116,649	34,342	62,740	76,778
R ²	0.001	0.0004	0.001	0.001
Adjusted R ²	0.001	0.0003	0.001	0.001
Residual Std. Error	3.660	3.391	4.159	3.965
F Statistic	40.246***	4.036***	15.983***	21.486***

Note: Table 1.7 shows the cross section bigram results *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.8: Cross Section Excess Returns: Bag of Words

	<i>Dependent variable:</i>			
	excess.return			
	All	Pre-Market	Market	Post-Market
Dictionary	0.003*** (0.001)	0.001 (0.002)	0.005*** (0.002)	0.006*** (0.001)
Constant	0.293*** (0.011)	0.075*** (0.018)	0.312*** (0.017)	0.417*** (0.014)
Observations	116,649	34,342	62,740	76,778
R ²	0.0002	0.00001	0.0001	0.0003
Adjusted R ²	0.0002	-0.00002	0.0001	0.0002
Residual Std. Error	3.662	3.391	4.161	3.966
F Statistic	20.224***	0.348	8.979***	19.407***

Note: Table 1.8 shows the cross section dictionary results. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.9: Cross Section Excess Returns Deciles: Bigram (1-5)

Decile	<i>Dependent variable:</i>				
	excess.return				
	1	2	3	4	5
negative.bigram.z	-6.677 (16.966)	12.936** (5.692)	-16.680*** (4.998)	-3.147** (1.530)	-4.315*** (1.009)
neutral.bigram.z	1.289 (1.293)	-0.665* (0.374)	0.658** (0.278)	0.110 (0.119)	0.124 (0.089)
positive.bigram.z	-8.059 (20.177)	-17.367*** (5.771)	8.085 (5.526)	2.152 (1.510)	3.235*** (0.952)
Constant	5.439*** (1.277)	1.128*** (0.285)	3.038*** (0.323)	0.942*** (0.119)	0.826*** (0.083)
Observations	198	1,246	1,362	3,074	5,521
R ²	0.082	0.018	0.050	0.004	0.007
Adjusted R ²	0.068	0.015	0.048	0.003	0.006
Residual Std. Error	10.003	8.326	7.460	6.257	5.979
F Statistic	5.789***	7.522***	23.710***	4.429***	12.801***

Note: Table 1.9 shows the cross section decile bigram results. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.10: Cross Section Excess Returns Deciles: Bigram (6-10)

Decile	<i>Dependent variable:</i>				
	excess.return				
	6	7	8	9	10
negative.bigram.z	0.046 (0.719)	-1.382** (0.597)	-0.412*** (0.160)	-1.107*** (0.140)	-0.294*** (0.048)
neutral.bigram.z	0.017 (0.057)	0.018 (0.046)	0.031 (0.027)	0.048** (0.020)	0.026*** (0.009)
positive.bigram.z	-0.155 (0.750)	1.187* (0.615)	0.271* (0.156)	0.969*** (0.138)	0.281*** (0.049)
Constant	0.587*** (0.054)	0.433*** (0.044)	0.283*** (0.027)	0.197*** (0.019)	0.086*** (0.008)
Observations	8,259	12,083	17,304	22,083	40,942
R ²	0.0002	0.001	0.001	0.004	0.001
Adjusted R ²	-0.0002	0.001	0.001	0.004	0.001
Residual Std. Error	4.765	4.457	3.539	2.801	1.706
F Statistic	0.430	3.035**	4.922***	30.447***	14.750***

Note: Table 1.10 shows the cross section decile bigram results. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.11: Cross Section Excess Returns Deciles: Unigram (1-5)

Decile	<i>Dependent variable:</i>				
	excess.return				
	1	2	3	4	5
negative.unigram.z	-2.007 (19.867)	14.230*** (3.711)	-12.288*** (4.389)	-1.607 (1.049)	-2.321*** (0.671)
neutral.unigram.z	-0.430 (1.079)	-0.339 (0.343)	0.690** (0.303)	0.269** (0.130)	0.248*** (0.092)
positive.unigram.z	-13.498 (22.213)	-17.780*** (3.661)	2.523 (4.512)	0.544 (0.970)	1.197** (0.582)
Constant	5.819*** (1.153)	0.824*** (0.289)	3.239*** (0.310)	0.953*** (0.120)	0.825*** (0.084)
Observations	198	1,246	1,362	3,074	5,521
R ²	0.077	0.026	0.047	0.005	0.006
Adjusted R ²	0.062	0.024	0.045	0.004	0.006
Residual Std. Error	10.034	8.292	7.471	6.256	5.981
F Statistic	5.364***	11.014***	22.362***	4.854***	11.601***

Note: Table 1.11 shows the cross section decile unigram results. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.12: Cross Section Excess Returns Deciles: Unigram (6-10)

Decile	<i>Dependent variable:</i>				
	excess.return				
	6	7	8	9	10
negative.unigram.z	-0.504 (0.465)	-0.010 (0.464)	-0.181** (0.084)	-0.493*** (0.082)	-0.144*** (0.029)
neutral.unigram.z	0.094 (0.059)	-0.009 (0.046)	-0.037 (0.027)	0.040** (0.019)	0.017** (0.009)
positive.unigram.z	0.429 (0.488)	-0.210 (0.466)	0.045 (0.086)	0.352*** (0.078)	0.132*** (0.029)
Constant	0.578*** (0.054)	0.442*** (0.044)	0.282*** (0.027)	0.197*** (0.019)	0.086*** (0.008)
Observations	8,259	12,083	17,304	22,084	40,942
R ²	0.0004	0.0003	0.001	0.003	0.001
Adjusted R ²	0.0001	0.0001	0.001	0.003	0.001
Residual Std. Error	4.764	4.458	3.539	2.803	1.707
F Statistic	1.218	1.209	4.888***	20.693***	9.613***

Note: Table 1.12 shows the cross section decile unigram results. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.13: Cross Section Excess Returns Deciles: Dictionary (1-5)

Decile	<i>Dependent variable:</i>				
	excess.return				
	1	2	3	4	5
dictionary	0.223 (0.174)	0.157*** (0.051)	0.028 (0.048)	0.020 (0.018)	0.043*** (0.010)
Constant	2.710*** (0.738)	0.373 (0.242)	1.481*** (0.210)	0.818*** (0.114)	0.652*** (0.082)
Observations	198	1,246	1,362	3,074	5,521
R ²	0.008	0.008	0.0002	0.0004	0.003
Adjusted R ²	0.003	0.007	-0.0005	0.0001	0.003
Residual Std. Error	10.345	8.363	7.646	6.268	5.989
F Statistic	1.648	9.448***	0.338	1.168	16.854***

Note: Table 1.13 shows the cross section decile dictionary results. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.14: Cross Section Excess Returns Deciles: Dictionary (6-10)

Decile	<i>Dependent variable:</i>				
	excess.return				
	6	7	8	9	10
dictionary	0.008** (0.004)	-0.003 (0.006)	0.00001 (0.001)	0.010*** (0.002)	0.003*** (0.0005)
Constant	0.562*** (0.053)	0.413*** (0.041)	0.271*** (0.027)	0.178*** (0.019)	0.078*** (0.009)
Observations	8,259	12,083	17,304	22,084	40,942
R ²	0.001	0.00003	0.000	0.001	0.001
Adjusted R ²	0.0004	-0.0001	-0.0001	0.001	0.001
Residual Std. Error	4.763	4.459	3.540	2.805	1.706
F Statistic	4.459**	0.349	0.00005	29.871***	57.378***

Note: Table 1.14 shows the cross section decile dictionary results. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.15: Forecasting Excess Returns: Bigram

	<i>Dependent variable:</i>
	excess.return
lag.negative.bigram.z	-0.056 (0.075)
lag.neutral.bigram.z	-0.008 (0.011)
lag.positive.bigram.z	0.134* (0.075)
Constant	0.286*** (0.011)
Observations	112,072
R ²	0.001
Adjusted R ²	0.0005
Residual Std. Error	3.600
In-sample MSE Ratio	0.99950
Out-of-sample MSFE Ratio	1.0008
F Statistic	18.900***

Note: Table 1.15 shows forecasting results for Bigrams. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.16: Forecasting Excess Returns: Unigram

	<i>Dependent variable:</i>
	excess.return
lag.negative.unigram.z	-0.013 (0.043)
lag.neutral.unigram.z	0.010 (0.011)
lag.positive.unigram.z	0.092** (0.043)
Constant	0.286*** (0.011)
Observations	112,072
R ²	0.001
Adjusted R ²	0.0005
Residual Std. Error	3.600
In-sample MSE Ratio	0.99950
Out-of-sample MSFE Ratio	0.99984
F Statistic	18.838***

Note: Table 1.16 shows forecasting results for unigrams. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.17: Forecasting Excess Returns: Dictionary

	<i>Dependent variable:</i>
	excess.return
lag.dictionary	0.001 (0.001)
Constant	0.284*** (0.011)
Observations	112,072
R ²	0.00001
Adjusted R ²	-0.00000
Residual Std. Error	3.601
In-sample MSE Ratio	0.99950
Out-of-sample MSFE Ratio	1.0056
F Statistic	0.699

Note: Table 1.17 shows forecasting results for dictionary. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.18: Forecasting Excess Returns: Bigram Decile (2-5)

Decile	<i>Dependent variable:</i>			
	excess.return			
	2	3	4	5
lag.negative.bigram.z	-2.645 (5.568)	-9.811* (5.016)	-0.597 (1.528)	2.882*** (1.006)
lag.neutral.bigram.z	0.188 (0.361)	0.210 (0.292)	-0.040 (0.120)	-0.229*** (0.089)
lag.positive.bigram.z	7.342 (5.687)	12.246** (5.543)	1.691 (1.508)	-1.600* (0.949)
Constant	-0.094 (0.274)	1.036*** (0.329)	0.689*** (0.119)	0.587*** (0.083)
Observations	1,246	1,362	3,074	5,521
R ²	0.018	0.004	0.005	0.009
Adjusted R ²	0.015	0.002	0.004	0.009
Residual Std. Error	8.327	7.637	6.256	5.972
F Statistic	7.441***	1.909	4.856***	17.063***
In-sample MSE Ratio	0.9823431	0.9958010	0.9952773	0.9908070
Out-of-sample MSFE Ratio	0.996000	1.0193278	1.032612	0.9959577

Note:

*p<0.1; **p<0.05; ***p<0.01

Note: Table 1.18 shows forecasting results for decile Bigram. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.19: Forecasting Excess Returns: Bigram Decile (6-10)

Decile	<i>Dependent variable:</i>				
	excess.return				
	6	7	8	9	10
lag.negative.bigram.z	0.170 (0.716)	0.117 (0.595)	-0.170 (0.159)	-0.190 (0.140)	0.027 (0.049)
lag.neutral.bigram.z	-0.032 (0.057)	-0.030 (0.046)	0.038 (0.027)	0.007 (0.020)	-0.008 (0.008)
lag.positive.bigram.z	0.137 (0.746)	0.362 (0.612)	0.339** (0.156)	0.244* (0.138)	-0.021 (0.049)
Constant	0.538*** (0.054)	0.339*** (0.044)	0.261*** (0.027)	0.192*** (0.019)	0.090*** (0.008)
Observations	8,259	12,083	17,304	22,084	40,942
R ²	0.001	0.001	0.001	0.0005	0.00005
Adjusted R ²	0.001	0.001	0.001	0.0003	-0.00003
Residual Std. Error	4.762	4.456	3.538	2.806	1.707
F Statistic	3.372**	5.938***	6.581***	3.328**	0.645
In-sample MSE Ratio	0.9987761	0.9985275	0.9988602	0.9995480	0.9999527
Out-of-sample MSFE Ratio	0.9990061	1.0000213	1.002264	1.001222	1.0001544

Note: Table 1.19 shows forecasting results for decile Bigram. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.20: Forecasting Excess Returns: Unigram Decile (2-5)

Decile	<i>Dependent variable:</i>			
	excess.return			
	2	3	4	5
lag.negative.unigram.z	2.093 (3.700)	-3.129 (4.505)	0.377 (1.049)	2.854*** (0.670)
lag.neutral.unigram.z	0.033 (0.340)	0.076 (0.312)	0.113 (0.131)	-0.128 (0.092)
lag.positive.unigram.z	2.519 (3.631)	4.806 (4.621)	0.782 (0.970)	-1.330** (0.581)
Constant	-0.101 (0.280)	1.203*** (0.316)	0.683*** (0.120)	0.545*** (0.084)
Observations	1,246	1,362	3,074	5,521
R ²	0.017	0.002	0.005	0.010
Adjusted R ²	0.015	-0.0005	0.004	0.010
Residual Std. Error	8.330	7.646	6.255	5.969
F Statistic	7.165***	0.786	5.077***	18.741***
In-sample MSE Ratio	1.010210	1.0245226	1.064432	0.9938456
Out-of-samlpe MSFE Ratio	0.9829884	0.9982670	0.9950631	0.9899118

Note: Table 1.20 shows forecasting results for decile Unigram. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.21: Forecasting Excess Returns: Unigram Decile (6-10)

Decile	<i>Dependent variable:</i>				
	excess.return				
	6	7	8	9	10
lag.negative.unigram.z	0.063 (0.464)	0.007 (0.463)	-0.032 (0.084)	-0.002 (0.082)	-0.013 (0.029)
lag.neutral.unigram.z	0.090 (0.059)	-0.035 (0.046)	0.045* (0.027)	-0.0001 (0.019)	-0.017** (0.008)
lag.positive.unigram.z	0.261 (0.487)	0.483 (0.464)	0.202** (0.086)	0.059 (0.079)	0.019 (0.029)
Constant	0.533*** (0.054)	0.337*** (0.044)	0.261*** (0.027)	0.192*** (0.019)	0.089*** (0.008)
Observations	8,259	12,083	17,304	22,084	40,942
R ²	0.002	0.002	0.001	0.0003	0.0002
Adjusted R ²	0.001	0.001	0.001	0.0002	0.0001
Residual Std. Error	4.761	4.456	3.538	2.807	1.707
F Statistic	4.340***	6.363***	6.404***	2.552*	2.062
In-sample MSE Ratio	0.9979596	0.9997497	1.001762	1.001413	1.0001930
Out-of-samlpe MSFE Ratio	0.9984254	0.9984221	0.9988908	0.9996534	0.9998489

Note: Table 1.21 shows forecasting results for decile Unigram. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.22: Forecasting Excess Returns: Dictionary Decile (2-5)

Decile	<i>Dependent variable:</i>			
	excess.return			
	2	3	4	5
lag.dictionary	-0.030 (0.051)	0.105** (0.048)	0.042** (0.018)	-0.029*** (0.010)
Constant	0.559** (0.244)	1.419*** (0.210)	0.802*** (0.114)	0.749*** (0.082)
Observations	1,246	1,362	3,074	5,521
R ²	0.0003	0.004	0.002	0.001
Adjusted R ²	-0.001	0.003	0.001	0.001
Residual Std. Error	8.394	7.634	6.263	5.994
F Statistic	0.355	4.834**	5.295**	7.759***
In-sample MSE Ratio	1.002135	0.9957453	1.001730	0.9985194
Out-of-sample MSFE Ratio	0.9997145	0.9964584	0.9982795	0.9985960

Note: Table 1.23 shows forecasting results for decile Dictionary. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 1.23: Forecasting Excess Returns: Dictionary Decile (6-10)

Decile	<i>Dependent variable:</i>				
	excess.return				
	6	7	8	9	10
lag.dictionary	0.005 (0.004)	-0.002 (0.006)	0.001 (0.001)	0.002 (0.002)	0.001** (0.0005)
Constant	0.567*** (0.053)	0.411*** (0.041)	0.270*** (0.027)	0.189*** (0.019)	0.085*** (0.009)
Observations	8,259	12,083	17,304	22,084	40,942
R ²	0.0002	0.00001	0.00003	0.0001	0.0001
Adjusted R ²	0.0001	-0.0001	-0.00002	0.00001	0.0001
Residual Std. Error	4.764	4.459	3.540	2.807	1.707
F Statistic	1.538	0.134	0.600	1.253	5.895**
In-sample MSE Ratio	1.0002121	1.0008117	1.018050	1.000756	0.9998346
Out-of-sample MSFE Ratio	0.9998138	0.9999889	0.9999653	0.9999433	0.9998560

Note: Table 1.22 shows forecasting results for decile Dictionary. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Chapter 2

US Real Estate Inflation Prediction: Exchange Rates and Net Foreign Assets

2.1 Introduction

The last decade saw an unprecedented rise and fall of US housing prices nationally. Figure 2.1, shows the movement of housing price inflation from 1991-2014. With this increased volatility, there has been renewed interest in forecasting real estate inflation. Accurate forecasts of housing prices are not only important for real estate market participants but for the macroeconomy as a whole. Directly, forecasts play a role in the rent or buy decision on the demand side and the sell or rent decision for suppliers. With much of US housing being financed, forecasts are an input in bank mortgage loan decisions. Further, the Federal Housing Finance Agency (FHFA) and the Federal Housing Administration (FHA) use forecasts to determine policy. Specifically, the FHA provides mortgage insurance where forecasts are used to determine these insurance premiums. Finally, the statements by the last three Federal Reserve chairs indicate that the Federal Reserve closely watched the housing markets.

In a conference call of the Federal Open Market Committee on January 16, 2009 Vice Chairman Donald Kohn states:

I remind ourselves that the most frequent criticism we get now is that we paid too much attention to the forecast for consumer price inflation in 2003-04, allowing a housing and credit bubble to develop. I am not convinced that we should lean against the wind of asset bubbles with monetary policy, but we do need to be careful that an inflation objective does

not constrain our ability to address financial instability when necessary.¹

Given the need for better forecasts, the recent academic literature has focused on developing forecasting models using US domestic conditions and lags of real estate inflation (Ghysels et al. (2012); Rapach et al. (2007); Rapach and Strauss (2009)). Major macroeconomic variables like output, inflation, interest rates and labor market conditions have found the most use as predictors. These studies find modest gains in forecast accuracy over autoregressive models using in-sample and out-of-sample evidence. The search for possible predictors has not been exhaustive, external sector variables have largely been excluded.

Through interconnected global financial and goods markets, the external sector is likely to provide useful information in forecasting real estate inflation regardless of the exchange rate regime. The Dornbusch (1976) overshooting model predicts, under a floating regime any shocks to a domestic or foreign economy would first pass through to exchange rates. Under a fixed regime, exchange rates would not be informative predictors. When shocks to either domestic or foreign economies happen, foreign economies have to trade their own currency to keep the exchange rates fixed. Given the possible importance of foreign influence in US real estate markets and the high correlation between current account deficits and housing prices during the last decade, other theoretical literature has focused on determining the relationship between international capital flows and real estate returns. (Bernanke (2005)) “global saving glut” hypothesis explains that increased savings globally led to an inflow of capital into US asset markets, thus decreasing real interest rates and increasing demand for housing. Ferrero (2015) shows that due to the relaxing of borrowing constraints in the US, accommodative monetary policy and fixed foreign exchange rate regimes have led to the housing market bubble. Finally, Favilukis et al. (2010, 2014) show that the relationship between housing prices and international capital flows are purely coincidental and housing prices were driven by financial market liberalization while net capital inflows were driven by foreign government regulatory, reserve currency and economic policy motives.

The goal of this paper is test the implication of these theories on forecasting housing price inflation and answer the question, does the inclusion of external sector variables in US real estate forecasting models provide added predictability?

In this study, I determine the added forecast accuracy of models containing exchange rate returns, and the net change in US financial account (NFA) over a simple autoregressive model. I use both in-sample and out-of-sample analysis of short-run forecasts to determine if these additional predictors produce better forecasts for the national and metropolitan area housing price inflation level. I find the NFA model is able to produce the most accurate forecasts across housing price levels and forecast horizons with the largest increase in forecast accuracy being around 40 percent. While for exchange rates I find the that Norwegian model is able to produce improved forecasts for the national and metropolitan area

¹Source: <http://www.federalreserve.gov/monetarypolicy/files/FOMC20090116confcall.pdf>

level with the largest improvement of around 25 percent.

The rest of this paper is organized as follows. Section 2 describes the background literature. Section 3 discusses theoretical considerations. Section 4 provides detailed discussion of the data. Section 5 shows the in-sample and out-of-sample methodology. Section 6 & 7 provide in-sample and out-of-sample results for both the national and metropolitan area level and Section 8 concludes.

2.2 Literature Review

The recent forecasting literature has focused on two possible sets of domestic predictors. The first, finance factors, which are likely to affect the asset use of housing and the second, macroeconomic conditions, which are likely to affect the consumption use of housing. Most of these studies have focused on using in-sample evidence of added predictably.

This industry level search has been yielded mixed results, with limited evidence of more accurate forecasts over autoregressive models. Further works like [Case and Shiller \(1989\)](#) and [Gu \(2002\)](#), find that lags of housing price inflation are important predictors. [Ghysels et al. \(2012\)](#) notes that due to how the repeat sales housing price measures are created lags should be an important part of the relevant predictor set.

2.2.1 Finance Factors

The efficient market hypothesis predicts that excess returns from an asset are only a function of market returns. Due to high transaction costs and inability to short housing, market efficiency may not be reasonable for housing inflation. [Case and Shiller \(1989\)](#); [Campbell and Vuolteenaho \(2004\)](#); [Ghysels et al. \(2012\)](#) find that stock market returns can not predict housing price inflation.

Valuation ratios have long been used in the financial forecasting literature to predict asset returns. In the case of real estate inflation this relates to the rent-price ratio. If rents are high relative to the price then it would be expected that future demand for housing would increase as renters become homeowners. It also can be seen through investors choice of purchasing investment properties. If rents are high compared to prices then there is a high return on investing in real estate. So demand for real estate would increase. [Campbell and Shiller \(1988\)](#) define this a log difference between rent and real estate prices. [Campbell et al. \(2009\)](#) employ this in a vector autoregression (VAR) setting and find that this measure has some predictive ability. [Ghysels et al. \(2012\)](#) results differ and find added no predictability in models with both lagged inflation and the rent-price ratio.

In all, there is has been some conflicting evidence for financial predictors providing marginal increases in forecast accuracy compared to simple autoregressive models.

2.2.2 Macroeconomic Conditions

US macroeconomic conditions are important determinants of housing demand and supply. The literature has included a number of different macroeconomic variables (interest rates, inflation, industrial production, and credit). By using in-sample analysis, [Ghysels et al. \(2012\)](#) find only interest rates produce marginally more accurate forecasts compared to an autoregressive model. Other predictors, like monetary policy and money supply have also found limited use. These may also be important in determining the demand and supply of housing. For out-of-sample testing, [Rapach and Strauss \(2009\)](#) finds consumer confidence, PCE deflator, and population produce more accurate forecasts ranging from 1% to 70% across the largest US states for a 4 and 8 quarter forecast horizon. Although none of these predictors are able to produce more accurate forecasts across all states.

Lower mortgage rates may stimulate housing demand. Due to uncovered interest rate parity (UIP) theory, differences in interest rates between two countries are explained by expected exchange rate movements. Thus, domestic interest rates may also have an international component. Although, exchange rates may have some advantages over interest rates. First, if the only difference in interest rates is expected exchange rates, then actually exchange rates may provide more forecasting information when exchange rates differ from expected. Second, interest rates are limited in that they have a lower bound. Particularly, during the later half of the last decade US monetary policy drove short-term interest rates toward zero. When interest rates are near the lower bound UIP may not hold. For example say the short-term domestic and foreign interest rate are effectively zero. The difference between these does not necessarily imply that the expected change in exchange rates would be zero. Exchange rates do not have a zero bound and thus would not suffer from such a problem. [Chinn and Zhang \(2015\)](#) show through a standard New Keynesian DSGE that the UIP does not hold when interest rates are low for long periods of time.

2.2.3 International Factors

Finally the most recent literature has focused on the role of which current international factors play in determining current real estate inflation. [Favilukis et al. \(2012\)](#) finds no role for foreign capital flows in determining housing prices for both a panel of eleven countries and the US individually. [Aizenman and Jinjara \(2009\)](#), in a panel of thirty-six countries, attempt to determine the relationship between current account balances and housing prices and find that a one standard deviation increase in lagged current accounts are associated with an increase in housing prices by 10%. [Fratzscher et al. \(2010\)](#) study how changes in housing prices effect current accounts and find that housing price shocks account for 32% in movements of U.S. trade balances. [Sá and Wieladek \(2015\)](#) use a global VAR estimate to isolate shocks from “global saving glut”, external monetary policy, easing of borrowing constraints and US monetary

policy. This study find all shocks led to significant changes in housing prices and specifically the “global saving glut” shocks increase US housing prices. [Cesa-Bianchi et al. \(2015\)](#) find that foreign capital flows through global liquidity have a particularly larger impact of emerging market economies (EME) housing prices compared to advanced economies (AE). Where a one percent shock to global liquidity increases housing prices by 0.1 percent over 3 quarters for AE’s and by 0.3 percent for EME’s.

International factors has found a place in the inflation forecasting literature. Specifically, [D’Agostino and Surico \(2009\)](#) find that inflation forecasting models which contain measures of global liquidity produce more accurate forecasts. Here autoregressive models are compared against models with global liquidity. If global liquidity matters in forecasting inflation, there is a strong possibility that it matters in forecasting real estate.

2.3 Theoretical Considerations

The [Dornbusch \(1976\)](#) overshooting model is quite useful in explaining the theoretical background of this forecasting exercise. This model predicts that with an unanticipated shock to money supply, changes in exchange rates “overshoot,” where the short term change in exchange rates are larger in magnitude than in the long run. In a broader context, this model gives concrete theory of how changes in exchange rates can predict changes in housing prices. This model relates exchanges rates with financial and goods markets. With the dual use of real estate changes in these markets are likely to affect housing demand and supply in both the short run and long run. For the short term, this model shows how interest rates are first affected by changes in exchange rates. So with changes in interest rates, mortgage rates are likely to change. This affects the cost of financing housing. Higher mortgage costs would then correspond to a higher cost of financing real estate causing demand to decrease. In the long run, changes in goods markets are also likely to affect housing prices. With changes in the prices of good markets the real cost of home building may increase.

The source of the shock, whether domestic or foreign does not matter in determining the relationship between exchange rates and housing prices. Specifically a shock to the US economy is likely to have an effect on exchange rates and housing prices. From 2004-2007 the Federal funds rates dramatically rose from around 1% to 5%, to combat the financial crisis in 2007-2008 the Federal Reserve dropped interest rates to around zero.² During this same time period, the trade-weighted dollar exchange rate sequentially appreciated and depreciated.³

Surprisingly changes in the 4 quarter lag in exchange rates line up with housing price inflation. Figure [2.2](#) shows the growth rate of the trade weighted dollar exchange rate and Case-Shiller composite housing

²<https://research.stlouisfed.org/fred2/series/DFP>

³The trade-weighted dollar exchange rate is an exchange rates weighted by trade balances with the US.

price inflation from 2001-2014.⁴ While exchange rates seem to more volatile, the trade-weighted dollar exchange rate and housing price inflation seem to move together. What is particularly interesting is the sharp depreciation of the dollar in 2008 corresponds to a rise in housing prices in 2009.

Finally, changes in exchange rates may directly or indirectly affect US housing prices. The direct effect works through changes in demand from international buyers. The National Association of Realtors (NAR) reports that around 7 percent of total US real estate sales were contributed to international buyers during 2014.⁵ So when there is an unanticipated positive monetary shock in foreign economy, foreign demand for US housing would decrease due to the relative increase in the cost of US housing for international purchasers. The full extent of the decrease in foreign demand would be limited by the overshooting mechanism.

Further changes in exchange rate can have indirect effects on housing demand. When the price of imports rise, the amount of goods and services Americans can buy decrease. With this decrease housing demand may decrease. Americans who own foreign assets may be affected in the opposite way. When the dollar depreciates relative to other currencies, the wealth of Americans owning foreign assets would increase, thus giving rise to an increase in housing demand. Overall the effects of these changes may be small, but when there is an unanticipated global money shock, the accumulative effect may be substantially large.

Favilukis et al. (2010, 2012); Ferrero (2015); Justiniano et al. (2015); Sá and Wieladek (2015) develop DSGE models which account for the correlation of current account deficits and housing prices during the past decade. Favilukis et al. (2010, 2012) show these are coincidentally related and there is no causal relationship between international capital flows and US housing prices. Ferrero (2015); Justiniano et al. (2015); Sá and Wieladek (2015) show that the effect of foreign capital flows on US housing prices work through a number of channels. The first, through the “global saving glut” hypothesis, an increase in world savings led to capital inflow for the US and appreciation of the dollar. The second, increases in housing prices and current account deficits are symptoms of both accommodative US monetary policy and relaxed borrowing constraints. Ferrero (2015); Sá and Wieladek (2015) specifically details how when foreign economies follow a fixed exchange rate this relationship is stronger.

2.4 Data

Creating aggregate housing price indexes are more complex than other measures of aggregate prices. For example, the Consumer Price Index (CPI) tracks the prices of a basket of goods and services over time.

⁴Due to the number of exchange rates and measures of housing prices, only one measure of housing price and exchange rate is shown for brevity.

⁵This accounts for \$92 Billion out \$1.2 Trillion in real estate sales. Source: <http://www.realtor.org/reports/profile-of-international-home-buying-activity>

The goods and services may differ in where and when they are sold but are usually of the same quality. For example, a “Hostess Twinkie” sold in Morgantown, WV is assumed to be the same quality as one sold in Providence, RI. Real estate is inherently different, where the overall quality of available housing depends on economic conditions. To account for these quality differences repeat-sales indexes, like the S&P/Case-Shiller Home Price indexes have been constructed.⁶

2.4.1 S&P/Case-Shiller Home Price

This index employs the methodology detailed in [Case and Shiller \(1987\)](#) and tracks repeat sales of single-family homes.⁷ By tracking repeat sales, they are able to better account for quality differences. Individual housing prices are collected from city or county deed records which usually includes the sales price, address, and owner. This measure is estimated by using the Robust Interval and Value-Weighted Arithmetic Repeat Sales algorithm. Past sale prices are regressed against the sale price of the specific house after the initial period.

This measure is available for 20 metropolitan areas, large cities, and nationally. Sales of new homes are not included. For this study, I use the national and the ten metropolitan area composite measure.⁸ The two national level Case-Shiller measures are identified as Case-Shiller and Composite.

The metropolitan area Case-Shiller indexes are estimated using the same procedure and data as the national level Case-Shiller housing price index but instead the data is grouped by metropolitan area. To be consistent with the Case-Shiller composite measure, I only include the 10 metropolitan areas used in estimating this index. These 10 metropolitan areas are San Diego, Los Angeles, Las Vegas, San Francisco, Miami, Washington, DC, New York, Boston, Chicago, and Denver. They are labeled as their abbreviated metropolitan area name and Case.

2.4.2 Exchange rates and The Net Change in Financial Accounts

I use 12 different exchange rates and the net change in financial accounts (NFA) as predictors for housing price inflation. These 12 countries represent over 75% of all bilateral trade.^{9,10}

The countries used in the sample have a relatively diverse range of policies. The exchange rate regimes for these countries range from free floating to fixed. Some countries have also changed currencies, like Brazil, and exchange rate regimes over the same period. Finally, some countries have changed economic systems during the sample period as is the case for Russia, which for the first three observations in the

⁶Table 2.10 in the appendix details all variables used, and their source. Data is quarterly and ranges from Q1:1991-Q1:2014. Variables are the quarterly growth rate estimated by log differences.

⁷For exact methodology, See: <http://us.spindices.com/index-family/real-estate/sp-case-shillerforexactmethodology>

⁸The other 20 metropolitan area composite is limited to data only available after 2000.

⁹Source: <https://www.census.gov/foreign-trade/statistics/highlights/top/top1507cm.html>

¹⁰Large oil producing countries are excluded due to the relationship between their exchange rates and commodity prices.

sample followed a communist economic system and was part of the Soviet Union.¹¹ A composite index exchange rate is used for Euro Zone countries prior to its official adoption in 1999. The Euro exchange rate is a rather problematic measure, where a number of countries joined the monetary union during the sample.^{12 13}

Overall these countries in this sample cover a wide range of possible exchange rate outcomes due to policy and monetary shocks. The exchange rates are labeled by their associated country names.

To proxy for net capital inflows, I use the NFA measure estimated by the Bureau of Economic Analysis. This measure is comprised of net foreign direct investment, net portfolio investment, and net reserve accounts. The change in this measure would show the inflow or outflow of capital of the US in each quarter. The literature has focused on the use of the current account deficits instead of the NFA measure (Fratzscher et al. (2010); Sá and Wieladek (2015)). These should be equal given the balance of payments measure. Favilukis et al. (2012) although notes the current account deficit measure is limited in that it does not account for the valuation effect by only using asset purchases data.

2.4.3 Descriptive Statistics

Table 2.1 details descriptive statistics for national housing price inflation measures and selected predictors. The first two columns show the mean and standard deviation for the Case-Shiller and Case-Shiller composite measures and selected predictors. The third and fourth columns show the correlation between the housing price measures and the correlation between the 4th lag of the predictors and the national housing price measure.

The Case-Shiller composite inflation has a higher mean and standard deviation compared to the other national housing price measure. This relates to a higher risk and return for housing in cities. Both measures are highly correlated with a coefficient of 0.951.

For exchange rates, a negative mean relates to an average depreciation of the dollar compared to the other currency. The Brazilian exchange rates has the highest mean of 11.051 and standard deviation of 27.582. This is something to be expected given Brazil revalued its currency a number of times in the beginning of the sample in an effort to limit high inflation.

The Mexican exchange rate has the lowest mean of -1.608. While the Euro has the lowest absolute value of the mean of 0.021. Singapore has lowest standard deviation out of these exchange rates of 2.323. This is likely due to it being relatively fixed during the sample period. In general, countries with higher inflation have higher average rates of appreciation, and higher standard deviations. While

¹¹In fact for the first three observations Russian growth in exchange rate takes a value of zero, this is due to the currency of the Soviet Union not being publicly traded and was used in all members of the Soviet Union.

¹²Starting in 2001, Greece added the Euro, in 2007 Slovenia, Cyprus and Malta in 2008, Slovakia in 2009, Estonia in 2011 and Latvia in 2014.

¹³The exchange rates used are Norway, Mexico, Chile, Brazil, Russia, India, Australia, Euro Area, Sweden, Singapore, China, and Canada.

countries which employ a fixed exchange rate or are more stable have lower rates of appreciation, and lower standard deviations.

The NFA measure is the only variable in the whole data set which is not a growth rate. Because of this it is not comparable to any of the other predictors. The mean value is negative which relates to an overall inflow of capital over the sample. The 4th lag of the NFA has the highest correlation for the Case-Shiller National housing inflation measures with 0.320 and the composite measure with 0.239. The Brazilian exchange rate is the lowest correlated with the Case-Shiller national measure with a coefficient of -0.003. While Mexican exchange rate is the lowest correlated with the composite measure with a coefficient of -0.012.

2.5 Methodology

In this section I detail the standard forecasting methods used to obtain my empirical results. This is further divided into forecasting models estimated, and in-sample and out-of-sample testing. In my analysis, I test 12 models against 12 different measure of housing prices for the national and metropolitan area levels. Eq 2.1,2.2 show my empirical models.

$$h_{i,t+\tau} = H_{i,t}\beta + u_{t+\tau} \quad (2.1)$$

$$h_{i,t+\tau} = H_{i,t}\beta + P_{k,t}\gamma + u_{t+\tau} \quad (2.2)$$

Eq 2.1 is the restricted model and Eq 2.2 is the unrestricted model. h_i denotes the i measure of housing price inflation, and $H_{i,t}$ is a vector of two lags of the i housing price. τ is the forecast horizon either 1,2, or 4 quarters ahead. k is the specific additional predictor and $P_{k,t}$ denotes four lags of exchange rates or the NFA.

In-Sample Model

In-sample nested model testing is a simple F-test of joint significance. This tests if the additional variables parameter vector in each nested model are jointly equal to zero. For my in-sample analysis, I test if the parameter vector, γ , in Eq 2.2 is zero.

Out-Of-Sample Estimation

All out-of-sample analysis start out by dividing the time series into two parts, in-sample and out-of-sample. The first part, in-sample, is denoted as P which is equal to 1, 2, ..., P . The out-of-sample is

denoted as R which is equal to $P + \tau + 1, P + \tau + 2, \dots, T + \tau - 1, T + \tau$. The ratio of in-sample to out-of-sample or $\frac{P}{R}$ is denoted π . If the sample contains 100 observations, $\tau = 1$ and $\pi = 1$, in-sample would be $P = 1, 2, 3, \dots, 50$ and $R = 51, 52, \dots, 100$.

The choice of sample split is key in making correct analysis. The literature has been relatively sparse about the optimal sample split. [Clark et al. \(2013\)](#) note that there is a trade-off between more in-sample observations and more out-of-sample observations. Where more in-sample observation allows for better parameter estimates and thus allowing for better forecasts. More out-of-sample observations gives more information on the accuracy of the forecasts. [Rossi and Inoue \(2012\)](#) finds the optimal split being relatively large in nested models (such as this). [Rossi and Inoue \(2012\)](#) develop a optimal sample split test where the null hypothesis is no difference in forecast accuracy between sample splits. In these studies it is assumed the search for the optimal value of π is extensive meaning that all possible values of π are considered. For most researchers an extensive search is rather time consuming and infeasible. So a general rule of thumb is having a value of π greater than or equal to one is best. For my estimation, I use a π value of approximately 1. Due to the relatively small sample size of real estate inflation a relatively large value of π cannot be festively estimated.

Forecast and the associated parameters can be estimated either once (fixed), updated as more data is added (recursive), or with a fixed window (rolling). The literature has focused on recursive or rolling.¹⁴ I apply the recursive forecasting method to estimate model parameters. Eq 2.3 shows the recursive method with a quadratic forecasting loss function.

$$\hat{\beta}_t, \hat{\gamma}_t = \underset{\beta, \gamma}{\operatorname{argmin}} \sum_{s=1}^{t-\tau} (h_{i,s+\tau} - (H_{i,s}\beta + P_{k,s}\gamma))^2 \quad (2.3)$$

Once forecasts are estimated, the mean squared forecast error (MSFE) is used to determine accuracy (Eq 2.4).

$$MSFE = (P - \tau + 1)^{-1} \sum_{t=P+\tau}^T (h_{i,s+\tau} - \hat{h}_{i,s+\tau})^2 \quad (2.4)$$

Out-Of-Sample Forecast Evaluation

When two forecast are to be compared, the out-of-sample forecasting literature has relied on the difference in MSFE to determine which forecasts are most accurate. This has ranged from simple comparisons in MSFE to more complicated statistical tests. The MSFE ratio, which is calculated by dividing the MSFE of the unrestricted model by the MSFE of the restricted model is a simple statistic to determine which model produces a more accurate forecast. This statistic is defined in Eq 2.5.

¹⁴Recursive and rolling methods are usually better with long time series and instability

$$MSFE_{ratio} = \frac{MSFE_{unrestricted}}{MSFE_{restricted}} \quad (2.5)$$

When the MSFE ratio is less than one, the unrestricted model produces a more accurate forecast. If it is greater than one, the restricted model produces a more accurate forecast. When these are equal, both models are equally predictive.

Early works by [Diebold and Mariano \(1995\)](#); [West \(1996\)](#) develop the first statistical tests for out-of-sample forecast evaluation. These tests are limited to non-nested model comparisons. Further works by [Clark and McCracken \(2001, 2005a\)](#); [McCracken \(2007\)](#) develop a number of nested model forecasting evaluation tests. One of the most used of these tests is a test of equal mean square error, where the mean of squared forecasting error is compared between the restricted and unrestricted models. Where the null hypothesis is that both models have equal MSFE's.

Eq 2.6 describes this test statistic. $\hat{d}_{t+\tau}$ is the difference in squared errors between the restricted model and the unrestricted model and $\hat{\sigma}_2^2$ is the estimated variance of the unrestricted model.

$$MSFE - F = \frac{(P)^{-\frac{1}{2}} \sum_{t=R}^{T-1} \hat{d}_{t+\tau}}{\hat{\sigma}_2^2} \quad (2.6)$$

This test is one-sided where positive values relate to more accurate forecasts. When the difference in squared forecasting errors is negative the restricted model would then produce lower MSFE's.

For my out-of-sample analysis, I split the sample in around half for all forecast horizons.¹⁵ The forecasted portion contains most of the housing market bubble and bust. I estimate the forecasts using the recursive forecasting method. I also provide the MSFE ratio and conduct the MSFE F-test.¹⁶

Which is better to use? Out-of-Sample or In-Sample

It is the general perception that out-of-sample testing provides better inference about which models or predictors produce more accurate forecasts. Out-of-Sample Test allows the forecaster to actually judge how well the models forecast over a portion of the data and help guide forecasting models in the future. In-sample testing may produce spurious results, due to over-fitting or parameter instability. [Clark \(2004\)](#); [Clark and McCracken \(2005b\)](#) find using Monte Carlo simulations that out-of-sample testing is superior under these conditions and can guard against over-fitting and parameter instability.

In contrast to this evidence, [Inoue and Kilian \(2005\)](#) find that out-of-sample evidence is likely to be subjected to the same data mining problems as in-sample results. If one chooses the best model only based on out-of-sample accuracy, [Inoue and Kilian \(2005\)](#) shows there is no advantage to out-of-sample

¹⁵Depending on the forecast horizon, the same may have an odd number of observations. When that is the case, I use one more for the forecasting sample.

¹⁶Critical values are obtained from the Monte Carlo simulations from [Clark and McCracken \(2001\)](#).

testing over in-sample testing.

So given this conflicting evidence, it is best to use both in-sample and out-of-sample analysis and focusing on results which match. When the results are the same, it is likely there is some added predictability from the additional predictors.

2.6 In-Sample Results

Table 2.2-2.5 show the in-sample results for the national and metropolitan area housing price levels. Only the F-test statistics are shown.¹⁷ With a higher F-stat, the additional predictors are able to produce more accurate forecasts.

2.6.1 National Level

Table 2.2 shows the national level in-sample results for the three forecast horizons. The Norwegian, Chilean, Indian, Australian, Singaporean, Canadian, and NFA model produce significant results for both national measures for all forecast horizons. Specifically, the Chilean and NFA perform the best with the additional predictors significant at the 1% level.

For the one quarter, hereafter 1Q, forecast horizon, the F-stat for the Case-Shiller national measure ranges from 0.088 for the Chinese model to 5.053 for the Chilean model. The composite has a similar range, with the Australian model having the largest F-stat of 5.615 and the for the Chinese model with a value of 0.086. The Russian, Brazilian, Euro, and Chinese model performs the worst and are unable to produce significant results for any measure.

The results for the two quarter, hereafter 2Q, forecast horizon are similar. The Russian and Chinese model perform the worst with the lowest F-stat for both measures. While the NFA model performs the best with a F-stat of 7.714 for the Case-Shiller national measure and 11.485 for the composite measure. The Brazilian, Russian, and Chinese models are unable to produce significant results for any of the housing price measures.

Finally, for the four quarter, hereafter 4Q, forecast horizon, the NFA model performs the best, with F-stats of 10.884 and 20.966. While, the Brazilian model performs the worst. For this forecast horizon, fewer models are able to produce significant results. Where the Mexican, Brazilian, Russian, Euro, and Chinese models are unable to produce any significant results. The NFA measure has more forecasting ability as the forecast horizon increases. With the biggest gain for the 4Q forecast horizon. Overall, this model performs the best. This provides initial evidence that international capital flows play a role

¹⁷Showing the coefficients would make the tables unreadable. The goal of this exercise to determine which international predictors are able to produce better forecasts for in-sample analysis

in determining housing prices. Particularly these results are similar to [Fratzcher et al. \(2010\)](#); [Sá and Wieladek \(2015\)](#); [Cesa-Bianchi et al. \(2015\)](#) which all find foreign capital flows to be significant.

For the in-sample results, highly volatile exchange rates (Brazilian and Russian) and those that follow a fixed regime (Chinese) are unable to produce more accurate forecasts. This would be expected given the limited forecasting information under these conditions.

2.6.2 Metropolitan Area

Table 2.3 shows the results for the 1Q forecast horizon. The Norwegian, Chilean, Australian, Canadian and NFA perform relatively well. All of these are significant for at least 8 out of the 10 metropolitan areas. Specifically, the NFA and Chilean model produces the best forecasts for 3 metropolitan areas each. The NFA model is able to produce the largest F-stat for San Diego, Las Vegas, and Miami. The Chilean model produces the largest F-stat for Denver, Washington, DC and Chicago. Similar to the national results, the Brazilian, Russian and Chinese models perform as well or worse than the simple autoregressive model.

Table 2.4 shows the results for the 2Q forecast horizon. The Norwegian, Chilean, Australian, and NFA models performs relatively well, where all of these have significant results for all metropolitan area measures. The NFA model produces the highest F-stat for six of the ten metropolitan areas. Specifically the NFA model outperforms all other models for Los Angeles, San Diego, Las Vegas, and Miami where the F-stat here is over double F-stat of all others. The largest F-stat overall is 12.153 for the San Diego measures. The Norwegian model is able to produce the largest F-stat for Denver, and Boston. The Brazilian, Russian and Chinese models are unable to produce significant results for any of the metropolitan measures.

Table 2.5 shows the results for the 4Q forecast horizon. The Indian, and NFA models perform the best and produce significant results for all metropolitan areas. The NFA model has the highest F-stat for eight out of the ten metropolitan areas. These range from about two to five times larger than the second best model. The largest F-stat is 21.818 for San Diego. The Indian model is able to produce the largest F-stat for Denver and Boston. The Mexican, Brazilian, Russian, and Chinese models are unable to produce any significant results.

The results overall are relatively similar to the in-sample national level. The NFA model is the best with the biggest gains coming from the 2Q and 4Q forecast horizons. The strongest increase is for San Diego and other west coast metropolitan areas. The fixed or highly volatile exchange rates (Brazil, Russia, and Chinese) perform rather poorly. At least for both in-sample national and metropolitan area levels, the external sector is able to forecast better than the autoregressive model.

2.7 Out-of-Sample Results

Out-of-Sample analysis may provide some additional support for the in-sample results. This methodology is less likely to find models significant when the predictors are explaining noise rather than signal. Finally, it allows us to better determine what models have actually produced better forecasts in the past.

This section shows the out-of-sample results for the national, and metropolitan area housing price inflation. Specifically, the MSFE ratio and MSFE F-test of the baseline AR(2) model and the 12 nested exchange rate models and the NFA model are shown.¹⁸

2.7.1 National Level

Table 2.6 shows the out-of-sample results for the national level. In general, the out-of-sample results echo the in-sample results.

For the 1Q forecast horizon, the Norwegian, Mexican, Chilean, Australian, Swedish, Singaporean and Canadian models have significant forecasts for both national housing price inflation measures. For the Case-Shiller national measure the Norwegian model produces the best forecasts with an 8.3% increase in forecast accuracy.¹⁹ The Canadian model performs similarly as well for this measure with increases in accuracy of 6.5% respectively. For the Case-Shiller composite measure the Chilean model has the most accurate forecast with a 12.4% increase in forecast accuracy. The Australian, Norwegian and Canadian models also produce relatively more accurate forecasts with gains in accuracy of 10%. The Brazilian, Russian, and Chinese models are unable to produce any significant forecasts. The Brazilian model performs the worst out of these and increases MSFE by about 4-5% for each measure.

For the 2Q forecast horizon, the Norwegian, Chilean, Indian, Australian, Euro, Swedish, Singaporean, Canadian, and NFA models are able to produce significant forecasts for both national housing price inflation measures. The Singaporean model produces the most accurate forecasts for the Case-Shiller measure with an increase of 10%.

The NFA model provides the largest gain in forecast accuracy of 23.3% for the Case-Shiller composite measure. Figure 2.4 shows the 2 quarter forecast horizon square forecast errors for both the autoregressive and NFA models for the composite measure. Most of the gain in forecast accuracy comes for the forecasts between 2007-2011. This time period particularly covers the large decline in housing prices due to the housing price inflation burst.

The Chilean and Norwegian model also performs rather well with increases of accuracy of 8.2% and 8% for the Case-Shiller national measure and 17.4% and 13.6% for the composite measure. The Brazilian, Russian and Chinese models are unable to produce significant forecasts for either measure. The Brazilian

¹⁸A small MSFE ratio will relate to being able to reject the null of the MSFE-F test

¹⁹The increase in forecast accuracy is just the percentage decrease in MSFE compared from the base model to the other. It is calculated by subtracting the MSFE ratio by one.

model produces the worst forecasts for both measures with an decrease in forecast accuracy of 4.2% for the Case-Shiller measure and 6.1% for the other.

Finally for the 4Q forecast horizon, The Norwegian, Chilean, Indian, Australian, Swedish, Singaporean, and NFA models produce significant forecasts for both national housing price inflation measures. The NFA model is able to produces the most accurate forecasts. For the Case-Shiller national measure NFA produces 14.7% more accurate forecasts and for the composite measure it produces 43.6% more accurate forecasts. Figure 2.5 shows the 4Q forecast horizon square forecast errors for both the autoregressive and NFA models for the composite measure. Like the 2Q forecast horizon most of the gain in forecast error is found from 2007-2011. The Indian and Chilean models also perform well with increases of 7.8% and 13.5% for the overall Case-Shiller measure and 21.4% and 15.1% for the composite measure. The Brazilian, Chinese and Mexican models forecasts are not significant for either measures of national housing prices. The Brazilian model does particularly poorly and decreases accuracy by about 5% for both measures.

The out-of-sample results only slightly differ from in-sample results. In general, models that do well out-of-sample also do well in-sample. For the 1Q forecast horizon, out-of-sample and in-sample differs in which model is most accurate.

Given the in-sample and out-of-sample results some generalizations can be made. Countries with floating and stable exchange rates generally produce more accurate forecasts compared to fixed or volatile currencies. Evidence is given by, the Norwegian, Canadian, and Australian models perform well compared to the Chinese, and Brazilian models. For the Brazilian exchange rate, the noise may overwhelming the signal. Such that forecasting information is very limited. The Chinese exchange rate is known to relatively fixed to the dollar where changes in exchange rates is most likely due to change in policy.

The Norwegian model performs relatively well out of all the floating stable currencies. It is most likely due to Norway's closeness with US asset markets that is leading to this. Norway has the largest sovereign wealth fund with a value of \$882 Billion.²⁰ This value by itself cannot directly explain the increased forecast ability. This \$822 Billion is spread out across 75 different countries with the US investments accounting for around 30%. These investments are divided among equity (62.5%), fixed income (35.3%), and real estate investments (2.3%).²¹ When the Norwegian sovereign wealth fund purchases US assets, these would cause the Norwegian Krone to depreciate (higher demand for dollars and higher supply of Kroner on currency exchange markets). This then would have a number of direct and indirect effects on US housing markets. First, with the purchase of US assets, the value of these assets would increase and then leading to increase in the value of American portfolios. Through the wealth effect, demand for US housing would also increase. At the same time, with an increase in the value of the dollar, the relative

²⁰<http://www.swfinstitute.org/sovereign-wealth-fund-rankings/>

²¹All information on Norwegian Sovereign Wealth Fund was obtained from <http://www.nbim.no/en/>

costs of US housing would increase for Norwegians, decreasing demand. Finally, with the increased value of the dollar the value of Norwegian oil exports would increase (due to the oil prices being denoted in dollars) lead to an increase in Norwegian income. With this increase Norwegians may increase demand for US goods and services or US housing. It is most likely that the shocks which affect Norwegian exchange rates also affect US housing.

The NFA model performs significantly better than any other model. Particularly, this model produced more accurate forecast during the housing bust. This measure would pick up the changes in currency holdings of countries that employ fixed rate regimes. During this period, the dollar rapidly depreciated and appreciated against other currencies. This is mostly due to financial crisis and the Federal Reserve response. Countries who followed a fixed exchange rate regime had to change their dollar holding in response. So the NFA would also have to change. This is the theory provided in [Justiniano et al. \(2015\)](#). This measure would fully encompass changes in the value of Norwegian sovereign wealth fund through changes in portfolio holdings.

Finally, these results provide show capital flows as providing limited forecasting gains for the housing price boom but large increases for the housing price bust. This is inconsistent with both the “global saving glut” hypothesis, and the financial market liberalization theory provided in [Ferrero \(2015\)](#). While the forecasting gains during the bust period is inconsistent with the theory in [Favilukis et al. \(2010, 2012\)](#). Given that empirical relationships are inherently unstable ([Stock and Watson \(2001\)](#)), it may be the relationship between housing and international capital flows are coincidental for the boom period, while both are symptoms of global economic conditions for the bust period. In any case, further theoretical research needs to be conducted that can account for both of these results.

2.7.2 Metropolitan Area

Table [2.7](#) shows the 1Q forecast horizon, the Norwegian model produces significant results for all metropolitan areas. Although it is only able to produce the most accurate forecast for Boston with an increase in forecast accuracy of 23.4 %. The Canadian, Australian, and NFA models are able to produce the most accuracy forecasts for 2 metropolitan areas each with increases ranging from 10-15%. The Mexican, Chilean, and Indian models produce the most accurate for the remaining three metropolitan areas. Like the national level, the Brazilian, and Russian models perform relatively poorly with significant results for only two out of the ten measures. In general, these results are relatively similar to the in-sample analysis.

Table [2.8](#) shows the 2Q forecast horizon, here the Norwegian, Chilean, Australian, Euro, and Singaporean models produce significant results for at least nine of the metropolitan areas. The NFA model performs the best and it is able to produce the most accurate forecasts for San Francisco, Los Angeles,

San Diego, and Miami with an increase in accuracy of 26.4%, 14.9%, 26.4%, and 18.4%. The Norwegian and Chilean models performs the second best. The Norwegian model produces the most accurate forecasts for New York, and Boston with an increase of 12.2%, and 23.3% respectively. The Chilean produces the most accurate forecasts for Washington, DC and Chicago with an increase of 13.1% and 21.1%. Finally, the Euro model produces the most accurate forecast for Las Vegas with an increase of 9.5%. The Brazilian, and Chinese models again perform the worst, and only able to produce significant results for only one metropolitan area.

Table 2.9 shows the 4Q forecast horizon, the Indian, Singaporean and NFA models are able to produce significant results for all metropolitan areas. The NFA model performs the best and produces the most accurate forecast for seven out of the ten metropolitan areas. This model produces relatively large gains in forecast accuracy all at least of 13%. For San Diego, San Francisco, and Los Angeles, it is even greater with gains of 40%. The Indian model also performs well and it is able to produces the most accurate forecasts for Denver and Boston with increases of 23.1% and 31.1%. Finally, the Chilean model produces the most accurate forecast for New York with gains of 22.9%. These results are completely in-line with the in-sample results. Like the other two forecast horizons, the Brazilian and Chinese models perform the worst and unable to produce a single significant result.

Like the results for the national level, the NFA produces the most accurate forecast. With the strongest gains coming from the 2Q and 4Q forecasts. The gain does not seem to be just located in one metropolitan area, this model broadly increases forecast accuracy. It is likely external sector is related to overall national housing conditions. This result is very similar to the best exchange rate models where the best models perform well overall.

2.8 Conclusion

The numerous theoretical literature has found a role of the external sector in determining housing prices. The housing price forecasting literature in contrast has largely focused on determining which domestic predictors are most important. This study provides evidence that exchange rates and foreign capital flows are valid additional predictors in forecasting housing price inflation.

Given the unprecedented decline in housing prices across the US due to the financial crisis in 2007, more accurate forecasting of housing prices have become more important. When housing forecasts are wrong, forecasts for a number of other different macroeconomic variables are also likely to be wrong. This could lead to incorrect policy choices by the Federal Reserve or the FHA, and FHFA.

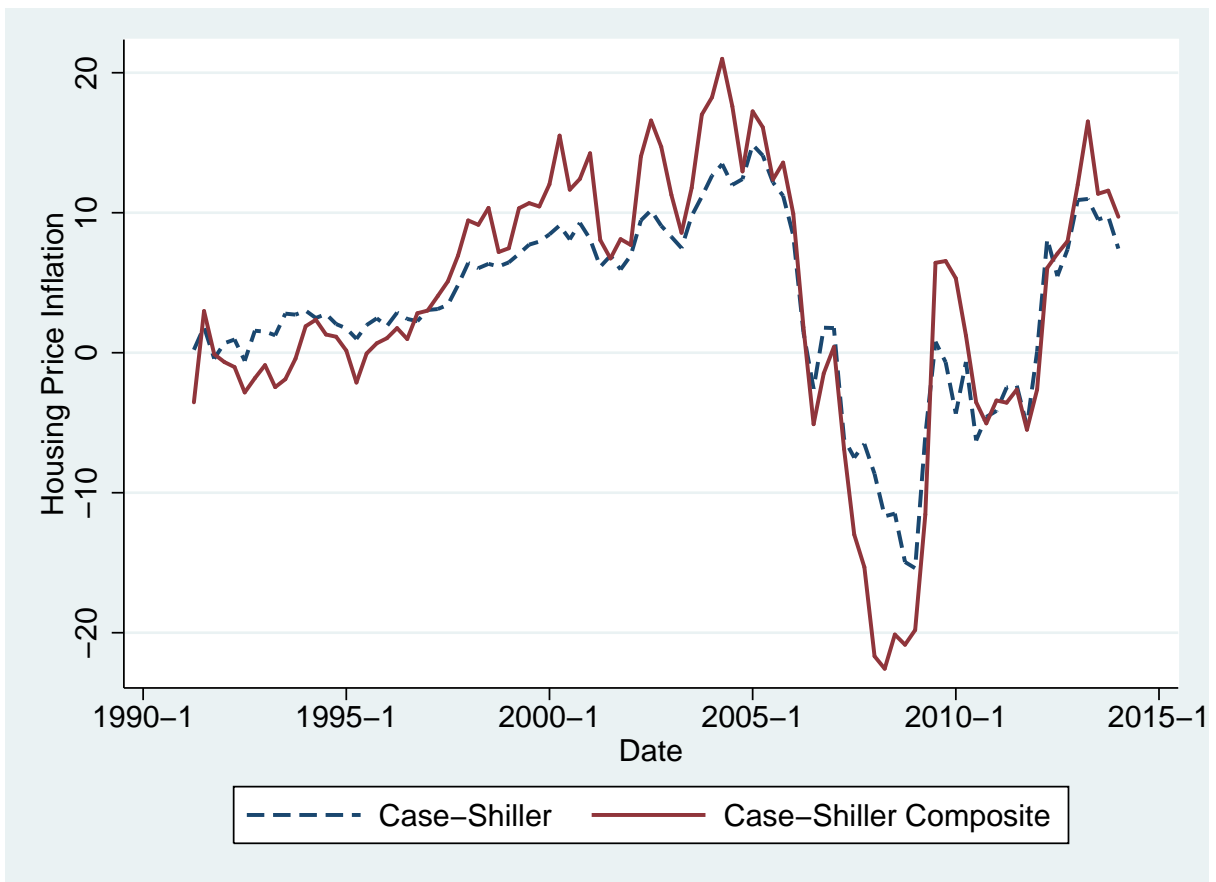
I find there is evidence for international factors providing increased forecast accuracy over a simple autoregressive model. While most models only reduce forecasting errors by only 5%, some are able to

reduce forecasting errors by up to 40%. The Norwegian, and NFA have the widest added predictability across all levels. The countries which employ fixed exchange rates produce the worst forecasts. The added predictability is not uniform across all measures and metropolitan areas. The Case-Shiller composite measure seems to be the most affected by international factors and sees the largest gains in forecast accuracy. Additionally, this work provides some evidence against the major competing theoretical housing price models.

Future forecasting works, may employ a number of different additional international predictors to further determine the role foreign factors have in forecasting housing price inflation. Particularly, future works need to determine which portion of the net change in financial accounts is driving this added predictability. This may provide evidence to which theoretical housing price model is correct.

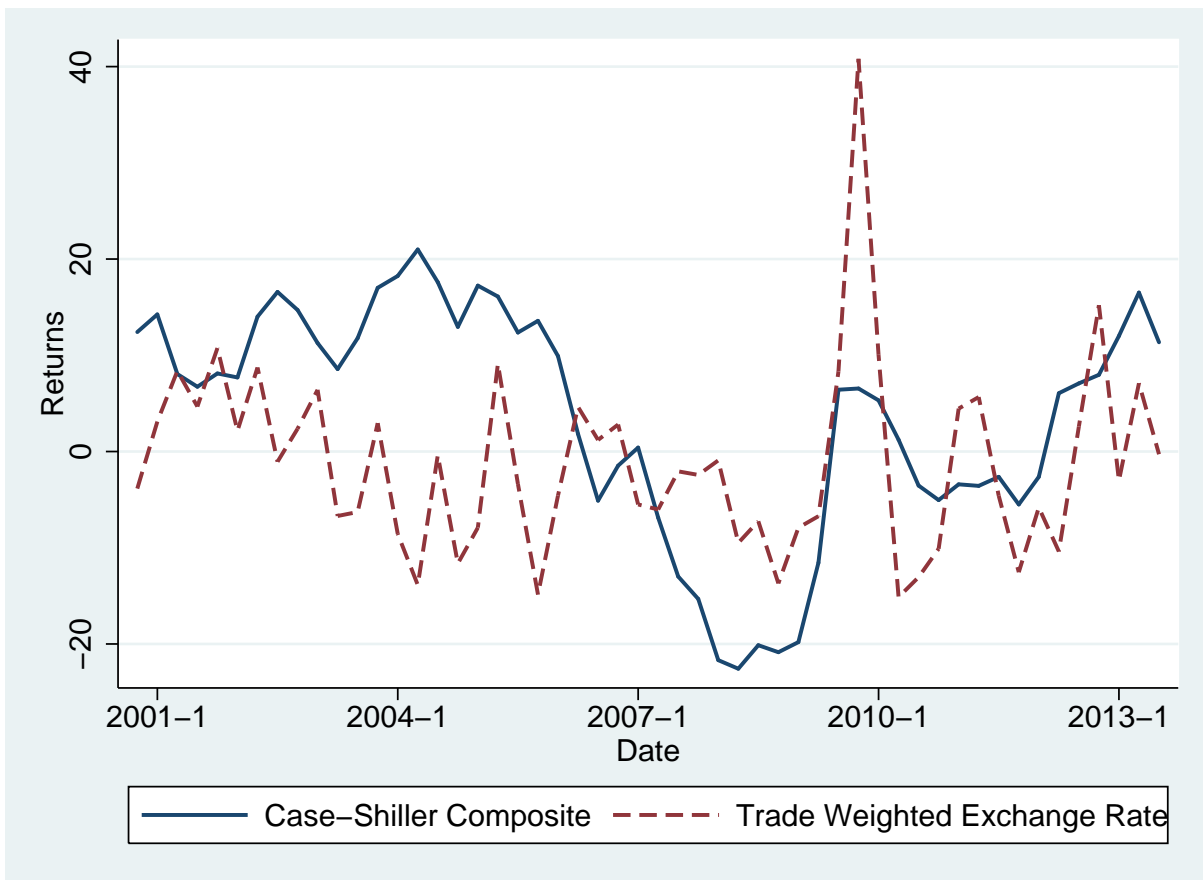
2.9 Tables and Figures

Figure 2.1: National Housing Price Inflation



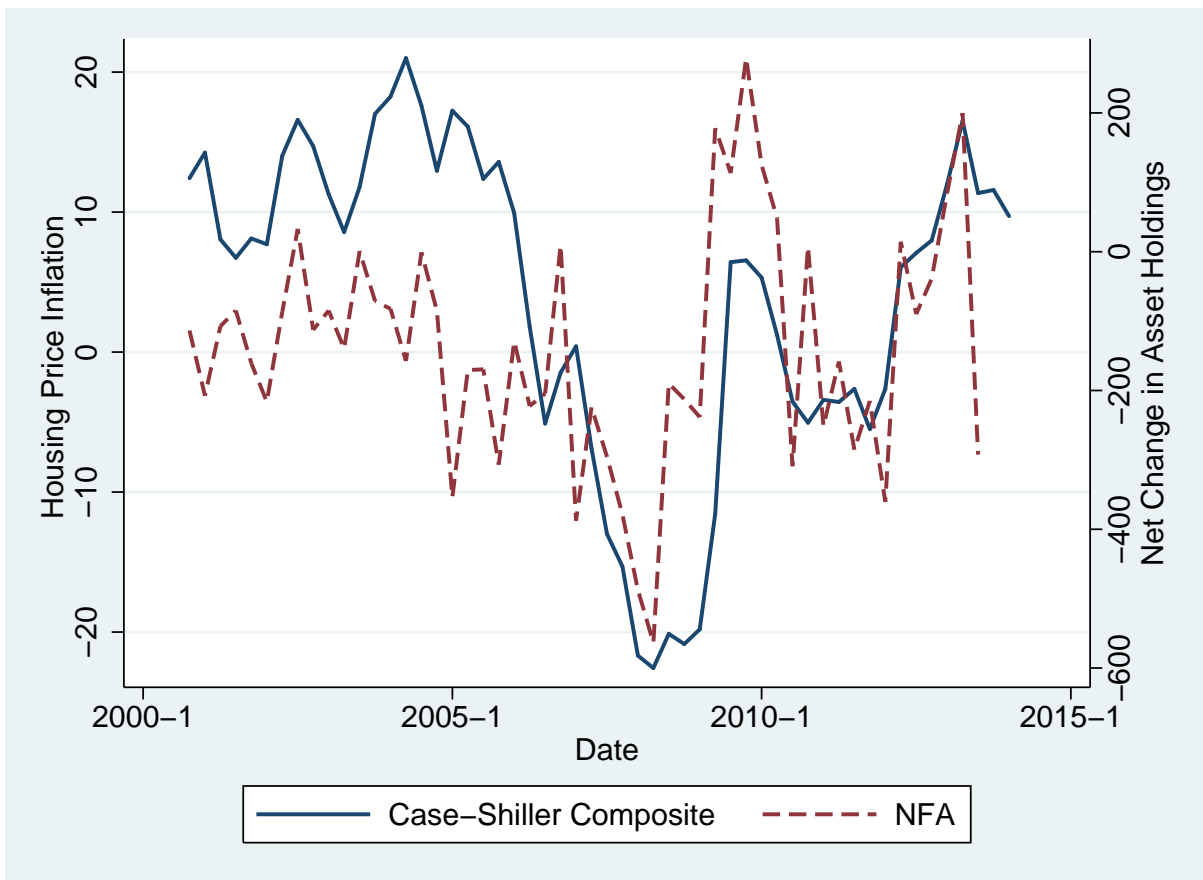
Note: Figure 2.1 shows the movement of the Case-Shiller and Case-Shiller Composite housing price inflation measure from Q2:1991-Q1:2014.

Figure 2.2: Housing Price Inflation and Exchange Rate



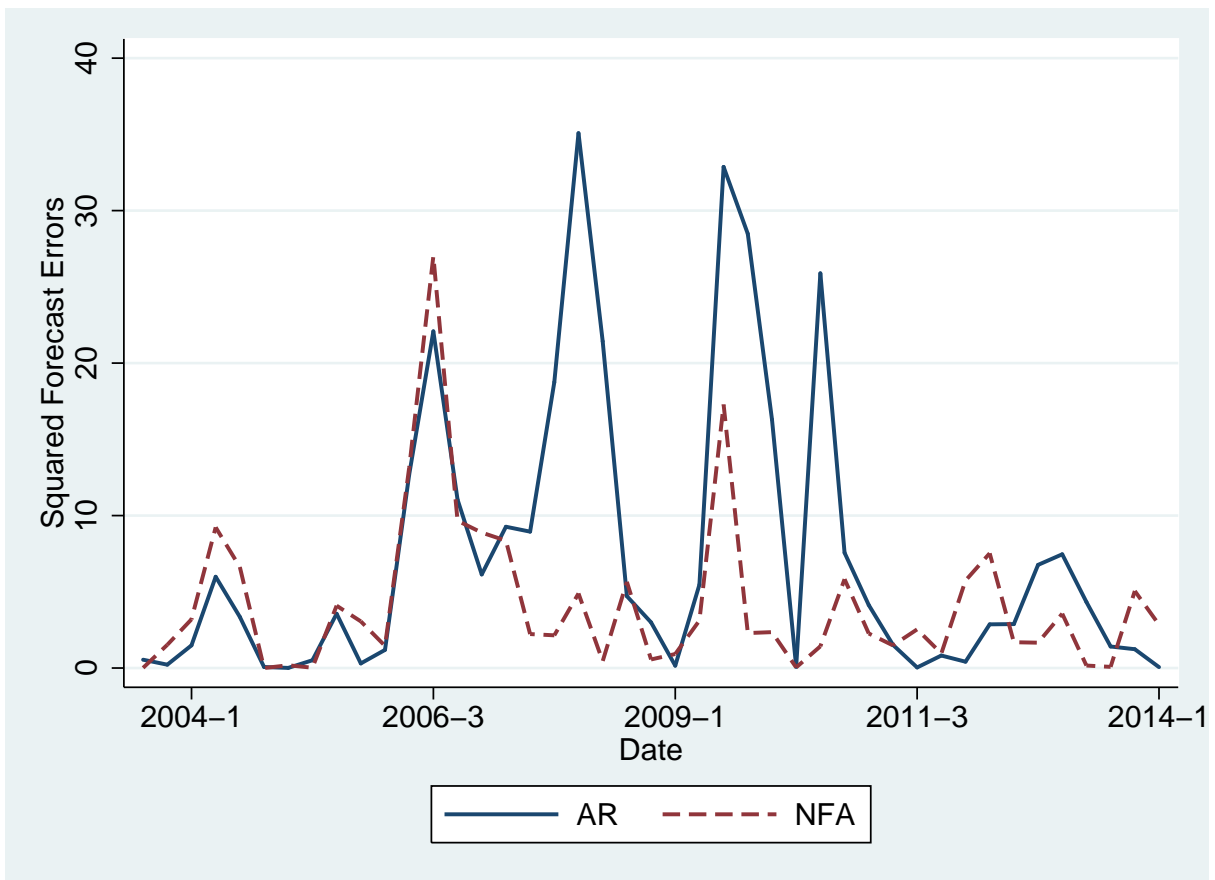
Note: Figure 2.2 shows the movement of the Case-Shiller Composite and the fourth quarter lag of the US Trade Weighted exchange rate from Q2:2000-Q1:2014.

Figure 2.3: Housing Price Inflation and Net Financial Account



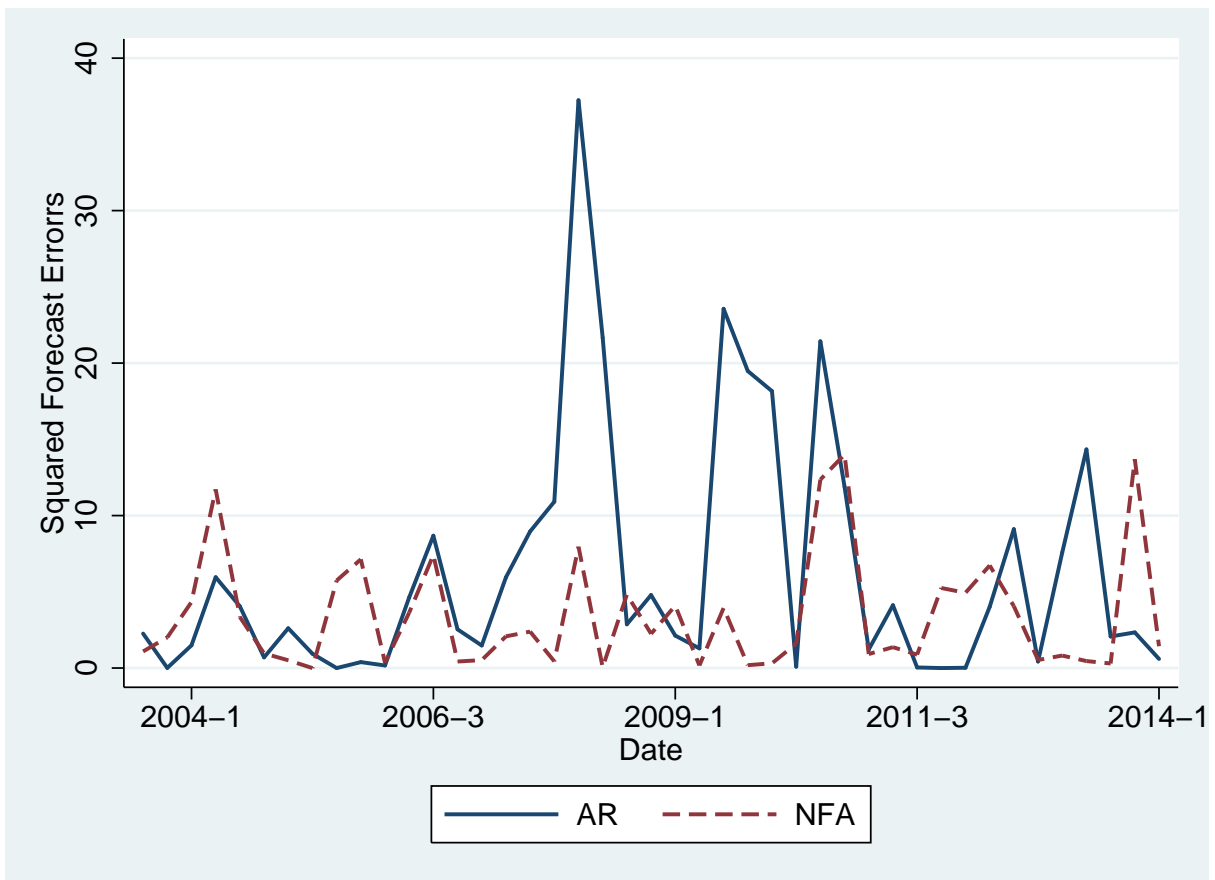
Note: Figure 2.3 shows the movement of the Case-Shiller Composite and the fourth quarter lag of the Net Change in Financial Accounts (NFA) from Q2:2000-Q1:2014.

Figure 2.4: Squared Forecast Errors (2Q): AR Model and NFA Model



Note: Figure 2.4 shows the 2Q horizon squared forecast error for the Case-Shiller Composite national housing price inflation for the AR(2) and the net change in financial accounts (NFA) models.

Figure 2.5: Squared Forecast Errors (4Q): AR Model and NFA Model



Note: Figure 2.5 shows the 4Q horizon squared forecast error for the Case-Shiller Composite national housing price inflation for the AR(2) and the net change in financial accounts (NFA) models.

Table 2.1: Descriptive Statistics

	Mean	Std Dev	Corr: Case	Corr: Composite
Case	0.892	1.669	1	0.951
Composite	1.0272	2.450	0.951	1
Norway	0.018	4.816	0.159	0.081
Mexico	-1.608	7.080	-0.046	-0.012
Chile	0.509	5.226	0.104	0.107
Brazil	11.051	27.582	-0.003	-0.059
Russia	6.756	17.837	0.051	-0.130
India	1.292	3.886	-0.098	-0.031
Australia	0.155	5.055	0.065	0.051
Euro	0.021	4.329	-0.144	-0.092
Sweden	0.137	5.287	-0.123	-0.055
Singapore	-0.344	2.323	0.118	0.050
China	0.166	4.360	0.039	0.055
Canada	-0.051	3.147	0.100	0.064
NFA	-114.167	135.346	0.320	0.239

Note: Table 2.1 shows the descriptive statistics for the Case-Shiller national level housing price Inflation measures, selected exchange rates and the net change in financial account (NFA). The first column shows the mean. The second column the standard deviation. The third and fourth column show their correlation of the 4th lag of the predictors with the two national level housing price inflation measures and the correlation between the two housing inflation measures.

Table 2.2: In-sample National Level Predictability Test

	1 Quarter Horizon		2 Quarter Horizon		4 Quarter Horizon	
	Case	Composite	Case	Composite	Case	Composite
Norway	2.69**	3.936***	3.653***	5.382***	2.245*	3.561**
Mexico	4.252***	3.99***	2.617**	2.872**	1.113	1.126
Chile	5.053***	5.277***	5.84***	6.719***	4.633***	5.804***
Brazil	0.474	0.542	0.215	0.343	0.143	0.13
Russia	0.185	0.249	0.186	0.396	0.153	0.46
India	2.522**	2.675**	3.914***	4.845***	5.923***	7.116***
Australia	3.917***	5.615***	3.589***	5.864***	2.546**	3.637***
Euro	1.374	1.842	1.674	2.342*	1.155	1.562
Sweden	2.219*	2.212*	2.236*	3.035**	1.62	2.385*
Singapore	2.43*	2.468*	3.919***	3.965***	2.69**	3.284**
China	0.088	0.086	0.193	0.236	0.087	0.242
Canada	3.781***	5.176***	4.061***	4.777***	3.148**	3.477**
NFA	3.582***	4.71***	7.714***	11.485***	10.884***	20.966***

Note: Table 2.2 shows the national level in-sample results for selected exchange rates and the net change in financial accounts (NFA) models and the three forecast horizons. The numbers represent the F-stat. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 2.3: In-sample Metropolitan Level Predictability Test(1Q)

	SFCase	LACase	NYCase	SDCase	DenCase	LVCcase	MiCase	DcCase	ChiCase	BosCase
Norway	3.373**	3.17**	2.147*	3.754***	5.454***	2.164*	2.262*	3.019**	4.014***	5.731***
Mexico	3.804***	2.232*	2.654**	0.503	1.714	0.556	2.085*	2.398*	4.251***	1.467
Chile	4.626***	2.805**	2.177*	1.717	6.507***	2.156*	2.093*	4.855***	7.459***	3.563***
Brazil	1.55	0.271	0.575	0.196	0.454	0.243	0.328	0.101	1.431	0.176
Russia	0.29	0.373	0.204	1.091	2.008	0.374	2.724**	0.462	1.476	1.027
India	3.063**	1.398	0.838	2.612**	4.339***	1.695	2.179*	4.241***	2.758**	4.023***
Australia	5.124***	5.414***	2.72**	2.044*	5.318***	3.911***	2.44*	3.499**	6.474***	4.195***
Euro	1.585	2.382*	1.078	1.892	2.565**	2.916**	1.247	1.83	2.53**	2.62**
Sweden	1.415	2.263*	1.39	1.685	3.602***	3.003**	1.996	2.053*	3.095**	2.511**
Singapore	3.234**	2.273*	1.326	2.502**	2.837**	1.38	1.011	2.593**	3.249**	4.513***
China	0.215	0.339	0.411	0.153	0.066	0.351	0.089	0.13	0.436	0.208
Canada	4.753***	3.464**	4.672***	1.573	2.87**	1.099	2.195*	3.05**	7.013***	3.784***
NFA	3.211**	5.146***	1.998	6.518***	2.009	5.48***	5.995***	4.188***	2.04*	2.937**

Note: Table 2.3 shows the 1 quarter forecast horizon metropolitan level in-sample results for selected exchange rates and the net change in financial accounts (NFA) models. The numbers represent the F-stat. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 2.4: In-sample Metropolitan Level Predictability Test (2Q)

	SFCase	LACase	NYCase	SDCase	DenCase	LVCCase	MiCase	DcCase	ChiCase	BosCase
Norway	5.532***	4.076***	3.409**	5.061***	7.264***	2.402*	2.808**	4.319***	5.667***	7.581***
Mexico	2.646**	1.565	2.242*	0.678	1.666	0.854	2.093*	2.212*	3.151**	1.618
Chile	5.461***	4.204***	4.017***	4.168***	5.092***	3.936***	2.449*	6.03***	8.593***	6.026***
Brazil	0.925	0.204	0.199	0.361	0.273	0.176	0.12	0.068	0.661	0.401
Russia	0.451	0.388	0.312	0.946	1.725	0.314	0.448	0.214	0.804	0.5
India	4.574***	3.075**	1.763	3.208**	4.668***	1.469	3.668***	4.893***	4.731***	4.996***
Australia	4.582***	5.373***	3.548**	2.599**	5.408***	3.731***	3.92***	4.041***	6.279***	5.523***
Euro	2.839**	2.637**	1.222	2.816**	3.237**	2.292*	1.239	1.732	3.29**	2.868**
Sweden	2.417*	2.872**	1.935	2.58**	4.845***	2.479*	1.541	2.094*	4.372***	3.042**
Singapore	4.515***	3.101**	3.473**	2.945**	4.171***	1.986	1.672	4.253***	4.122***	6.14***
China	0.359	0.49	0.084	0.127	0.145	0.195	0.141	0.311	0.199	0.14
Canada	4.946***	2.285*	4.535***	1.916	4.176***	0.931	2.919**	4.763***	7.142***	5.407***
NFA	7.674***	10.418***	4.419***	12.153***	3.635***	9.616***	10.012***	8.214***	6.569***	4.579***

Note: Table 2.4 shows the 2 quarter forecast horizon metropolitan level in-sample results for selected exchange rates and the net change in financial accounts (NFA) models. The numbers represent the F-stat. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 2.5: In-sample Metropolitan Level Predictability Test (4Q)

	SFCase	LACase	NYCase	SDCase	DenCase	LVCase	MiCase	DcCase	ChiCase	BosCase
Norway	2.99**	2.492**	4.111***	3.136**	3.03**	1.12	1.569	1.516	3.96***	4.279***
Mexico	0.315	0.7	1.407	0.416	1.04	0.846	0.621	0.657	1.142	1.349
Chile	3.488**	4.123***	6.846***	4.887***	2.855**	3.144**	1.604	3.282**	4.366***	5.883***
Brazil	0.283	0.145	0.19	0.078	0.358	0.18	0.044	0.083	0.342	0.1
Russia	0.664	0.415	0.586	0.629	1.313	0.188	0.897	0.268	0.464	0.551
India	5.604***	5.337***	5.56***	5.77***	5.241***	2.966**	4.467***	4.436***	7.253***	8.505***
Australia	1.66	2.97**	3.887***	1.772	3.112**	0.801	2.437*	1.355	3.999***	4.002***
Euro	1.788	1.351	1.579	1.249	1.452	0.423	0.642	0.493	2.583**	1.947
Sweden	1.633	1.941	2.453*	1.639	2.771**	0.61	1.017	0.643	3.163**	2.047*
Singapore	3.241**	2.285*	5.804***	2.864**	2.341*	1.949	1.37	1.691	3.914***	4.009***
China	0.233	0.392	0.16	0.24	0.196	0.157	0.072	0.299	0.016	0.297
Canada	2.411*	1.618	6.862***	2.378*	2.798**	1.7	1.627	1.981	5.792***	4.996***
NFA	13.235***	18.745***	9.418***	21.818***	4.774***	16.847***	12.107***	10.176***	13.241***	8.145***

Note: Table 2.5 shows the 4 quarter forecast horizon metropolitan level in-sample results for selected exchange rates and the net change in financial accounts (NFA) models. The numbers represent the F-stat. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 2.6: National level: Out-of-Sample Country-wise Forecasting Performance

		1 Quarter Horizon		2 Quarter Horizon		4 Quarter Horizon	
		Composite	Case	Composite	Case	Composite	Case
Norway	MSFE Ratio	0.907	0.917	0.864	0.92	0.905	0.972
	MSFE F-test	4.51***	3.996**	6.765***	3.739**	4.489***	1.221*
Mexico	MSFE Ratio	0.931	0.965	0.953	1.008	0.99	1.015
	MSFE F-test	3.236**	1.614*	2.098**	-0.337	0.428	-0.647
Chile	MSFE Ratio	0.876	0.959	0.826	0.918	0.849	0.865
	MSFE F-test	6.221***	1.881**	9.072***	3.861**	7.62***	6.703***
Brazil	MSFE Ratio	1.056	1.039	1.061	1.042	1.063	1.05
	MSFE F-test	-2.344	-1.644	-2.486	-1.738	-2.534	-2.061
Russia	MSFE Ratio	0.997	0.989	0.989	1.001	0.982	0.998
	MSFE F-test	0.135	0.509	0.497	-0.044	0.805*	0.098
India	MSFE Ratio	0.995	0.984	0.913	0.975	0.786	0.922
	MSFE F-test	0.2	0.717*	4.092**	1.105*	11.692***	3.627**
Australia	MSFE Ratio	0.9	0.977	0.914	0.962	0.953	0.98
	MSFE F-test	4.864***	1.052*	4.021**	1.716*	2.133**	0.89*
Euro	MSFE Ratio	1	0.963	0.962	0.956	0.982	0.999
	MSFE F-test	0.017	1.689*	1.714*	1.964**	0.778*	0.055
Sweden	MSFE Ratio	0.975	0.953	0.934	0.941	0.938	0.966
	MSFE F-test	1.147*	2.192**	3.062**	2.689**	2.821**	1.522*
Singapore	MSFE Ratio	0.937	0.953	0.889	0.904	0.908	0.958
	MSFE F-test	2.969**	2.179**	5.396***	4.57***	4.381***	1.901**
China	MSFE Ratio	1.016	1.012	1.016	1.016	1.018	1.026
	MSFE F-test	-0.706	-0.518	-0.695	-0.663	-0.746	-1.079
Canada	MSFE Ratio	0.898	0.935	0.929	0.93	0.991	0.963
	MSFE F-test	4.997***	3.047**	3.311**	3.217**	0.378	1.632*
NFA	MSFE Ratio	1.031	1.04	0.767	0.962	0.564	0.853
	MSFE F-test	-1.318	-1.688	13.047***	1.679*	33.185***	7.4***

Note: Table 2.6 shows selected national level out-of-sample results for exchange rates and the net change in financial accounts (NFA) and the three forecast horizons. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 2.7: Metropolitan Level: Out-of-Sample Country-wise Forecasting Performance (1Q)

		SFCase	LACase	NYCase	SDCase	DenCase	LVCCase	MICase	DCCase	CHICase	BOSCase
Norway	MSFE Ratio	0.961	0.938	0.94	0.906	0.904	0.952	0.965	0.924	0.93	0.766
	MSFE F-test	1.766*	2.913**	2.796**	4.558***	4.67***	2.214**	1.574*	3.599**	3.336**	13.411***
Mexico	MSFE Ratio	0.76	0.958	0.962	1.033	0.958	1.019	0.972	0.978	0.903	0.975
	MSFE F-test	13.93***	1.916**	1.757*	-1.399	1.948**	-0.828	1.257*	0.994*	4.747***	1.106*
Chile	MSFE Ratio	0.768	0.989	1.043	1.024	0.853	1.04	1.037	0.9	0.837	0.934
	MSFE F-test	13.294***	0.51	-1.804	-1.043	7.584***	-1.711	-1.563	4.888***	8.542***	3.119**
Brazil	MSFE Ratio	0.927	1.042	1.066	1.056	1.017	1.034	1.028	1.052	0.974	1.144
	MSFE F-test	3.458**	-1.762	-2.734	-2.336	-0.722	-1.453	-1.191	-2.188	1.169*	-5.526
Russia	MSFE Ratio	0.99	0.993	1.002	0.966	0.971	1.002	0.931	1.003	0.956	0.98
	MSFE F-test	0.44	0.326	-0.096	1.531*	1.293*	-0.102	3.265**	-0.138	2.037**	0.89*
India	MSFE Ratio	0.85	1.081	1.042	0.963	0.922	0.991	1.048	0.889	0.992	0.901
	MSFE F-test	7.786***	-3.279	-1.77	1.669*	3.701**	0.385	-2.032	5.499***	0.348	4.821***
Australia	MSFE Ratio	0.789	0.909	0.927	1.097	1.015	0.89	1.028	0.961	0.828	0.843
	MSFE F-test	11.74***	4.383***	3.487**	-3.894	-0.643	5.462***	-1.204	1.781*	9.114***	8.165***
Euro	MSFE Ratio	1.026	0.978	0.997	1.01	0.96	0.897	0.983	0.964	0.949	0.945
	MSFE F-test	-1.127	0.968*	0.144	-0.416	1.815**	5.068***	0.783*	1.637*	2.376**	2.551**
Sweden	MSFE Ratio	0.991	0.985	0.966	1.003	0.909	0.917	0.953	0.961	0.932	0.915
	MSFE F-test	0.4	0.67*	1.534*	-0.153	4.411***	3.991**	2.194**	1.778*	3.224**	4.081**
Singapore	MSFE Ratio	0.904	0.947	0.993	0.952	0.915	0.977	1.003	0.958	0.945	0.868
	MSFE F-test	4.657***	2.472**	0.309	2.195**	4.08**	1.036*	-0.119	1.91**	2.573**	6.681***
China	MSFE Ratio	0.997	0.997	1.019	1.011	1.026	0.991	1.012	1.018	1.013	1.043
	MSFE F-test	0.136	0.136	-0.838	-0.479	-1.124	0.397	-0.533	-0.776	-0.577	-1.811
Canada	MSFE Ratio	0.885	0.989	0.871	1.088	1.183	1.036	1.085	0.98	0.814	0.917
	MSFE F-test	5.694***	0.489	6.514***	-3.575	-6.796	-1.546	-3.437	0.903*	10.063***	3.984**
NFA	MSFE Ratio	0.995	1.052	1.186	0.906	1.253	1.006	0.902	1.126	1.131	1.095
	MSFE F-test	0.208	-2.181	-6.91	4.542***	-8.892	-0.247	4.786***	-4.909	-5.094	-3.826

Note: Table 2.7 shows selected 1 quarter forecast horizon metropolitan area out-of-sample results for exchange rates and the net change in financial accounts (NFA) and the three forecast horizons. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 2.8: Metropolitan Level: Out-of-Sample Country-wise Forecasting Performance (2Q)

		SFCase	LACase	NYCase	SDCase	DenCase	LVCCase	MICase	DCCase	CHICase	BOSCase
Norway	MSFE Ratio	0.815	0.909	0.885	0.87	0.794	0.923	0.984	0.911	0.878	0.767
	MSFE F-test	9.774***	4.282***	5.572***	6.448***	11.17***	3.606**	0.678*	4.201**	5.954***	13.092***
Mexico	MSFE Ratio	0.788	0.98	0.961	1.005	0.957	1.011	0.972	0.978	0.947	0.973
	MSFE F-test	11.58***	0.858*	1.764*	-0.219	1.95**	-0.469	1.22*	0.988*	2.417**	1.175*
Chile	MSFE Ratio	0.809	0.927	0.941	0.945	0.899	0.949	1.006	0.869	0.789	0.87
	MSFE F-test	10.135***	3.407**	2.674**	2.526**	4.833***	2.29**	-0.247	6.465***	11.496***	6.415***
Brazil	MSFE Ratio	0.977	1.057	1.094	1.046	1.032	1.031	1.03	1.053	1.023	1.121
	MSFE F-test	1.023*	-2.31	-3.699	-1.903	-1.317	-1.303	-1.234	-2.182	-0.978	-4.647
Russia	MSFE Ratio	0.972	0.993	0.992	0.974	0.966	0.99	1.002	1.002	0.99	0.984
	MSFE F-test	1.233*	0.313	0.361	1.14*	1.523*	0.432	-0.067	-0.071	0.427	0.687*
India	MSFE Ratio	0.774	1.002	0.99	0.936	0.898	1.015	0.961	0.881	0.899	0.856
	MSFE F-test	12.536***	-0.071	0.414	2.954**	4.874***	-0.624	1.726*	5.811***	4.844***	7.229***
Australia	MSFE Ratio	0.862	0.917	0.9	1.098	0.872	0.913	0.973	0.97	0.912	0.838
	MSFE F-test	6.898***	3.893**	4.752***	-3.829	6.317***	4.077**	1.21*	1.317*	4.152**	8.33***
Euro	MSFE Ratio	0.959	0.958	0.987	0.969	0.932	0.905	1.006	0.982	0.916	0.954
	MSFE F-test	1.831**	1.905**	0.561*	1.388*	3.133**	4.524***	-0.25	0.799*	3.961**	2.094**
Sweden	MSFE Ratio	0.929	0.955	0.944	0.965	0.849	0.916	1.004	0.977	0.884	0.9
	MSFE F-test	3.28**	2.006**	2.574**	1.564*	7.62***	3.962**	-0.185	1.006*	5.625***	4.754***
Singapore	MSFE Ratio	0.813	0.928	0.894	0.938	0.858	0.958	0.983	0.897	0.949	0.831
	MSFE F-test	9.897***	3.359**	5.12***	2.838**	7.133***	1.884**	0.76*	4.939***	2.326**	8.729***
China	MSFE Ratio	0.974	0.998	1.051	1.014	1.027	1.006	1.011	1.013	1.022	1.058
	MSFE F-test	1.151*	0.079	-2.068	-0.602	-1.117	-0.277	-0.471	-0.54	-0.922	-2.364
Canada	MSFE Ratio	0.857	1.089	0.878	1.084	0.971	0.999	1.008	0.947	0.824	0.864
	MSFE F-test	7.203***	-3.497	5.953***	-3.315	1.268*	0.04	-0.329	2.383**	9.209***	6.762***
NFA	MSFE Ratio	0.736	0.851	1.067	0.736	1.084	0.942	0.816	0.929	0.891	1.026
	MSFE F-test	15.434***	7.533***	-2.709	15.452***	-3.332	2.656**	9.725***	3.263**	5.262***	-1.091

Note: Table 2.8 shows selected 2 quarter forecast horizon metropolitan area out-of-sample results for exchange rates and the net change in financial accounts (NFA) and the three forecast horizons. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 2.9: Metropolitan Level: Out-of-Sample Country-wise Forecasting Performance (4Q)

		SFCase	LACase	NYCase	SDCase	DenCase	LVCCase	MICase	DCCase	CHICase	BOSCase
Norway	MSFE Ratio	0.889	0.99	0.903	0.948	0.945	0.998	1	0.996	0.92	0.9
	MSFE F-test	5.364***	0.432	4.604***	2.35**	2.509**	0.098	0.003	0.166	3.73**	4.761***
Mexico	MSFE Ratio	0.976	0.994	0.979	1.006	0.972	0.999	1.027	1	1.011	0.965
	MSFE F-test	1.038*	0.273	0.929*	-0.248	1.231*	0.034	-1.117	-0.014	-0.454	1.539*
Chile	MSFE Ratio	1.062	0.99	0.771	0.996	0.94	0.952	1.006	0.948	0.93	0.814
	MSFE F-test	-2.526	0.438	12.799***	0.194	2.761**	2.183**	-0.242	2.371**	3.238**	9.855***
Brazil	MSFE Ratio	1.019	1.074	1.094	1.063	1.001	1.011	1.028	1.044	1.03	1.138
	MSFE F-test	-0.808	-2.973	-3.685	-2.56	-0.028	-0.458	-1.177	-1.812	-1.265	-5.203
Russia	MSFE Ratio	0.983	0.994	0.97	0.982	0.971	0.997	0.979	0.992	1.001	0.976
	MSFE F-test	0.746*	0.277	1.337*	0.784*	1.281*	0.129	0.925*	0.363	-0.023	1.053*
India	MSFE Ratio	0.674	0.827	0.836	0.802	0.769	0.934	0.885	0.874	0.834	0.689
	MSFE F-test	20.832***	9.004***	8.439***	10.609***	12.884***	3.062**	5.586***	6.206***	8.558***	19.38***
Australia	MSFE Ratio	0.9	0.973	0.9	1.105	0.944	1.046	0.97	1.031	0.935	0.941
	MSFE F-test	4.773***	1.181*	4.796***	-4.093	2.547**	-1.873	1.349*	-1.312	2.984**	2.698**
Euro	MSFE Ratio	0.97	1.024	1.011	1.043	1.018	1.017	1.019	1.035	0.952	0.985
	MSFE F-test	1.343*	-0.997	-0.464	-1.787	-0.742	-0.698	-0.809	-1.446	2.15**	0.668*
Sweden	MSFE Ratio	0.963	1	0.919	1.015	0.937	1.005	1.008	1.014	0.919	0.969
	MSFE F-test	1.638*	0.009	3.774**	-0.626	2.905**	-0.218	-0.322	-0.588	3.8**	1.386*
Singapore	MSFE Ratio	0.942	0.978	0.81	0.953	0.926	0.954	0.984	0.974	0.914	0.874
	MSFE F-test	2.669**	0.948*	10.115***	2.127**	3.413**	2.053**	0.709*	1.132*	4.064**	6.203***
China	MSFE Ratio	1.003	1.011	1.056	1.01	1	1.005	1.027	1.015	1.031	1.054
	MSFE F-test	-0.134	-0.448	-2.282	-0.446	0.016	-0.225	-1.142	-0.636	-1.29	-2.188
Canada	MSFE Ratio	0.951	1.218	0.796	1.112	0.951	1.028	1.038	1.058	0.89	0.936
	MSFE F-test	2.203**	-7.682	11.036***	-4.327	2.228**	-1.181	-1.572	-2.362	5.338***	2.95**
NFA	MSFE Ratio	0.562	0.602	0.846	0.558	0.871	0.757	0.791	0.869	0.741	0.983
	MSFE F-test	33.497***	28.44***	7.842***	34.031***	6.362***	13.819***	11.375***	6.499***	15.022***	0.744*

Note: Table 2.9 shows selected 4 quarter forecast horizon metropolitan area out-of-sample results for exchange rates and the net change in financial accounts (NFA) and the three forecast horizons. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

2.10 Appendix

Table 2.10: Housing Inflation Measures and The Predictor Set

Measures	Description	Source
Case-Shiller Home Price Index	Quarterly inflation for seasonalized Case-Shiller indexes	Federal Reserve Bank of St. Louis FRED database
Exchange Rates	Dollar Exchange Rates for 22 different countries	Federal Reserve Bank of St. Louis FRED database
NFA	The seasonalized net quarterly change in the value of foreign owned US assets	Federal Reserve Bank of St. Louis FRED database

Chapter 3

Bilateral Capital Inflow and US Housing Inflation Prediction

3.1 Introduction

The housing market bubble and bust of the last decade has increased uncertainty and decreased predictability of housing inflation. There has been a desire in academic and industry circles to develop better forecasting models. To increase forecast accuracy, the academic literature has focused on using non-linear forecasting modeling, employing structural breaks and regime shifting, to increase forecast accuracy. This has yielded limited improvement in forecast accuracy (up to 15 percent) and often rely on unrealistic assumptions. ¹

The recent international finance theory has focused on the relationship between capital inflows and US housing prices. Ferrero (2015) determines this relationship is due to a mixture of relaxation of borrowing constraints, accommodative US monetary policy, and fixed foreign exchange rates. Other work has focused on the direct impact of capital inflows on US housing prices. In a 2005 speech, Former Federal Reserve Chairman Ben Bernanke specifically states: ²

... the rapid increase in the U.S. current account deficit between 1996 and 2000 was fueled to a significant extent both by increased global saving (global saving glut) and the greater interest on the part of foreigners in investing in the United States....the key asset-price effects of the global saving glut appear to have occurred in the market for residential investment, as low mortgage rates have supported record levels of home construction and strong gains in housing prices.

¹Some of these works assume knowledge of a structural break before it occurring. In general, it is very unlikely to know a structural break will occur before it occurs.

²See: Bernanke (2005)

Ben Bernanke implies that at least a portion of the increase in housing inflation before 2005 was caused by foreign capital. Further in [Bernanke \(2005\)](#) states that it is capital inflows from Asia and Latin America are the likely cause. Empirical evidence has found similar results with the global saving glut channel having an impact on US housing returns ([Sá and Wieladek \(2015\)](#); [Cesa-Bianchi et al. \(2015\)](#)).

Forecasts of US housing inflation are important for a number of reasons. Given the impact of the housing price bubble and burst and the resulting recession, national housing markets have a significant relationship with the macro economy and financial markets. So changes in the US housing market can be informative in understanding business cycle conditions. This is particularly useful for policy makers like the Federal Reserve, who follow housing inflation forecasts carefully.

Although not the goal of this paper, increased use of the international sector conditions and their relationships with the US macroeconomy can be quite useful in policy determination. Policy makers like the FDIC now generally follow housing price fluctuations to determine the risk to banks which a number of assets in housing.³ More accurate forecasts can lead to more appropriate policy.

Similar to [McGurk \(2016\) \(Ch.2\)](#), this work combines the forecasting techniques from the housing inflation forecasting literature and the recent international finance theory to include foreign capital inflows to predict US housing inflation. US net financial accounts, in general, show net capital inflow into the from all countries to the US minus outflows. Using this broad aggregate does not allow forecaster to employ fully or test the full international finance theory for a number of reasons. [Ferrero \(2015\)](#) states that capital inflows from emerging Asian economies and among oil producers may have the strongest relationship with US housing prices.⁴ Further, in a Brooking Institute Blog in April 2015 Ben Bernanke states:

My conclusion was that a global excess of desired saving over desired investment, emanating in large part from China and other Asian emerging market economies and oil producers like Saudi Arabia, was a major reason for low global interest rates.

5

Given this statement, it would imply that country or regional level capital inflows should be a better forecaster compared to the broad aggregate capital inflow.

Figure 3.1 shows the a map of countries exchange rate policy in 2006⁶. Most countries follow some form of managed exchange rates, to accurately determine the forecasting ability from these theories lower level financial account data is needed. Another issue from [McGurk \(2016\)\(Ch.2\)](#) is that models which

³<https://www.fdic.gov/news/conferences/housing.html>

⁴Given the large capital flows and exchange rate pegs to the dollar.

⁵See: <http://www.brookings.edu/blogs/ben-bernanke/posts/2015/04/01-why-interest-rates-low-global-saving-s-glut>

⁶See: <https://www.imf.org/external/np/mfd/er/2006/eng/0706.htm>

contain the net financial account only produces more accurate forecast when housing inflation changes directions (going from negative to positive and vice versa).⁷

In August 2015, the Bureau of Economic Analysis (BEA) changed their estimation methods for financial accounts in-line with international standards.⁸ Due to this bilateral financial account data is now available for both regions and around 20 different countries.⁹ With this the international finance theory can be tested more accurately in a forecasting setting.

This study asks the question if international capital flows provide useful forecasting information for US housing inflation: (1) What set of assets are important in forecasting and from what regions does the largest gain come from? (2) Do models containing net financial account data perform well when housing inflation does not change direction? (3) With so many new predictors available do predictor/forecast combination methods produce more accurate forecasts?

In this study, I determine the added forecast accuracy of models containing regional, bilateral capital inflow data from a number of possible assets produce more accurate forecasts over a baseline model AR(2) model.

I use in-sample and out-of-sample forecasting tests to determine the marginal forecast accuracy of these models for short run periods (2 and 4 quarter ahead) and national and metropolitan area level housing inflation. I, further, compare models containing overall capital inflow measure against forecast combinations and diffusion indexes created using dynamic factor model methods.

Overall, I find that capital inflow for a number of assets in the middle east can produce a marginally better forecast for both housing inflation levels with the largest gains coming at around fifty percent. I also find some of the forecast combination are able to produce better forecasts across the entire forecast period, with gains of around 50 percent. Diffusion indexes are also able to increase forecast accuracy but depend on the method used, with gains up to 60 percent.

The rest of this paper is organized as follows. Section 2 describes the background literature. Section 3 discusses theoretical considerations. Section 4 provides a detailed discussion of the data. Section 5 shows the in-sample and out-of-sample methods. Section 6 & 7 provide in-sample and out-of-sample results for both the national and metropolitan area level. Section 8 & 9 detail the combination methods and empirical results. Finally, Section 10 concludes.

⁷See: Ch.2 Figure 2.5

⁸See: http://www.bea.gov/scb/pdf/2014/03%20March/0314_restructuring_the_international_economic_accounts.pdf

⁹Many of the countries available are considered to have free floating exchange rate policy and a few are part of those which explain the “global savings glut” hypothesis.

3.2 Literature Review

The most recent academic literature has focused on models of developing housing inflation models using non-linear models and mainly domestic predictors. Many of these works find some improvement over simple autoregressive models.

3.2.1 In-sample Forecasting Literature

Housing inflation forecasting has had a long empirical history. [Case and Shiller \(1989\)](#) finds stock market return (S&P 500) are useful in determining housing inflation, but lags of housing inflation are better predictors. It has been suggested that this is likely due to the way that the repeat sales indexes are estimated ([Ghysels et al. \(2012\)](#)).

[Ghysels et al. \(2012\)](#), in a review of the existent literature, concurs with [Case and Shiller \(1989\)](#) that autoregressive models are useful in forecasting. This work also finds that very few macroeconomic predictors are useful in forecasting. Specifically, interest rates of a number of different terms are jointly significant in in-sample tests. Given monetary policy during the great recession, it is unlike that interest rates or interest rate spreads would provide any forecast-able information from after 2008. Changes in interest rates would be more reflective of changes in US monetary policy which may in itself be a response to current macroeconomic conditions.

3.2.2 Out-of-Sample Forecasting Literature

While these earlier works have focused on determining forecast accuracy using in-sample tests, newer works have focused on employing out-of-sample forecast analysis.

[Rapach et al. \(2007\)](#); [Rapach and Strauss \(2009\)](#) are some of the most relevant works which use out-of-sample analysis in determining housing inflation forecast accuracy. These works both compare the forecast accuracy of an autoregressive model against a number autoregressive distributed lag models (ARDL). This is done at the for a number US states at the monthly frequency. The ARDL models are nested models containing lags of macroeconomic indicators, regional indicators, and state-specific indicators. Overall these works both finds that a number of ARDL models perform well for some specific US states but are unable to produce better forecasts across all states. [Rapach and Strauss \(2009\)](#) goes further to solve this problem by combining the forecasting from these models. In general, the results show that while the forecast combination methods are unable to produce the most accurate forecasts compared to the individual models, they can produce significant forecast across the board.

Further works have simply used non-linear forecasting techniques to try to improve forecasts over AR models. [Balcilar et al. \(2015\)](#); [Kouwenberg and Zwinkels \(2014\)](#) both use a smooth transition

autoregressive (STAR) model to try to produce more accurate forecasts against an AR model. [Balcilar et al. \(2015\)](#) looks at monthly data for the national and four census region. In general, this work finds that there is relatively no difference between the forecast accuracy of the AR and STAR models. These results are similar across the national level and census regions. [Kouwenberg and Zwinkels \(2014\)](#) differs in that it looks at quarterly data for only the national level. Forecasts are short term with one through four quarters ahead. This work finds some improvement from the STAR model compared to autoregressive moving average (ARMA) and vector error correction (VEC) models, with increases in forecast accuracy of 2 percent and 5 percent.

[Plakandaras et al. \(2015\)](#) also continues with this push to estimate new forecasting models. This work instead of forecasting housing inflation, this work tries to forecast the housing price index. They use a new smoothing algorithm, Ensemble Empirical Mode Decomposition, then to determine the forecasting model employing the Elastic Net approach (a standard shrinkage method). Finally, they fit a support vector regression model for forecasting. The predictor set only includes domestic predictors which include real GDP per capita, interest rates, stock price, real construction costs, unemployment rate, inflation rate, real oil price and finally, the ratio of the budget deficit. The data ranges from 1890-2010. This work differs in a number of different ways. First, this, in general, is able to predict the recent large increase and fall in housing prices. Second, this work differs in that forecasts are longer term (over a year ahead). Third, this spans 120 years. There have been many structural changes in the macro economy, specifically with the larger importance of globalization. It is unlikely that most of the change in housing prices during this period is due to the global savings glut.

McGurk (2016) (Ch.2) following with the AR and ARDL comparison in [Rapach and Strauss \(2009\)](#), finds that with a simple inclusion of net financial account in a forecasting model is able to produce up to 50 percent increase for short term forecast horizons. This has been the only work which uses international predictors in this setting.

Finally, [D'Agostino and Surico \(2009\)](#) in a forecasting exercise looking at inflation overall compares an AR model to one with a measure of international credit. Overall, this work finds that international credit is a marginally better predictor. International credit is estimated by taking the average of M2 money growth for a number of larger countries. If inflation and housing inflation are related, international predictors should be important for forecasting housing inflation.

3.2.3 Housing Prices and Capital Inflows

To back up some of the theoretical literature ([Ferrero \(2015\)](#)) there has been a number of empirical studies which try to determine the relationship between US housing prices and capital inflows. Although these are not necessarily looking at the forecastability of capital inflows of US housing inflation, it does

show that these are important in determining housing prices.

[Favilukis et al. \(2012\)](#) estimates the role current foreign capital inflows have in determining US housing prices. This work does this by estimating a panel of eleven foreign countries and the US. In general, the results show there to be no relationship. This does not necessarily mean that capital inflows are not important for the US. It implies that in this panel, current capital inflows seem to be insignificant. Past capital inflows may be important rather than current and also at the same time, the relationship between housing inflation and capital inflows may be different in the US compared to other countries.

[Aizenman and Jinjarak \(2009\)](#) in a similar study, looks a dynamic panel of thirty-six countries, to determine the relationship between current account balances and housing inflation. This differs in that, it first past lags of housing and current account balances matter. Specifically, a one standard deviation increase in lagged current accounts are associated with an increase in housing prices by 10 percent. Again, this is another work that assumes that the relationship between housing prices and current account balances is similar for all countries in the panel.

[Fratzcher et al. \(2010\)](#) differs from the other works, where instead of looking at how capital inflows/current account balances affect housing prices, it looks at how housing prices affect trade balances. In general, this study finds that shocks to housings prices can account for 32 percent of the movement of U.S. trade balances.

[Sá and Wieladek \(2015\)](#) use a global VAR estimate to isolate shocks from “global saving glut”, external monetary policy, easing of borrowing constraints and US monetary policy. This specifically, tests the theory from [Ferrero \(2015\)](#). Overall, this work shows that shocks to all these move housing inflation but specifically the “global saving glut” shocks increase US housing prices by more than the other shocks.

Finally, [Cesa-Bianchi et al. \(2015\)](#) looks at the difference in the effect from capital inflows on housing inflation from advanced economies (AE) and emerging market economies (EME). Overall, this work finds that capital inflows through global liquidity have a larger impact on EME housing prices compared to advance AE. Where a one percent shock to global liquidity increases housing prices by 0.1 percent over 3 quarters for AE’s and by 0.3 percent for EME’s.

3.3 Theory

[Ferrero \(2015\)](#) develops a dynamic stochastic general equilibrium (DSGE) model that attempts to explain the relationship between US housing prices and current account deficits over the last decade. ¹⁰

[Ferrero \(2015\)](#) differs from other theoretical studies where domestic factors are the main cause of this relationship and foreign factors play a minor role in amplifying this effect. In fact, prior works,

¹⁰Due to balance of payments, current accounts deficits and capital inflows should be equal

specifically Favilukis et al. (2010) develops a DSGE model where current accounts and housing inflations are only coincidentally related.

Ferrero (2015) assumes that changes in housing demands are caused by preference shocks and deregulation in credit markets. Including an assumption of accommodative US monetary policy and foreign managed/fixed exchange rates. The paper tries to account for a number of "facts" which precluded the great recession and its aftermath.¹¹ These facts include high housing inflation and lower world interest rates.

This work develops a two-country model where one country is a net exporter with a fixed exchange rate with the other country and the country has deregulation in credit markets and follows an accommodative (accommodative to inflation) monetary policy rule. With maximizing rational agents, leads to a solution to the model. Preference shocks are feed into the model. In general, this results in similar movements in current accounts and housing prices, as seen in the US economy.

Preference shocks are only able to account for changes in housing demand. It does not account for lower world interest rates. The general intuition works in this way. First, a positive preference shock increases demand for housing. Deregulation in domestic credit markets cause borrowing to be less costly, further increasing housing demand. Second, with accommodative US monetary policy to keep exchange rates fixed, foreign economies also have to have an accommodative monetary policy. With the increase in money domestically and abroad, world interest rates then decrease. Further, with interconnected financial markets, domestic borrowers financing housing purchases look abroad for extra funds. This can be understood as the global savings glut hypothesis, as described in Bernanke (2005).

In reality, to combat the 2001 recession the Federal Reserve drastically reduced the Federal Funds rate. With foreign exchange rate policy in emerging economies in Asia and Oil producing countries, to keep exchange rates fixed, these countries also followed similar policies. Figure 3.4 shows short-term interest rates for Saudi Arabia and the US Federal funds rate. Saudi Arabia is assumed to a fixed exchange rate with the US. The figure shows that these interest rates seem to move together. This would at least provide some anecdotal evidence that this Saudi Arabia follows similar policies as the US.

Although this theory shows similar dynamics as the US housing price bubble and bust, it does not specifically have a role for foreign investors purchasing US housing. Such that foreign investors play a role in the supply of credit rather than demanding US housing. This assumption is in contrast with reality. The National Association of Realtors (NAR) survey which finds foreign investors account for 7 percent of US housing sales in 2014.¹² Further, this survey has found that foreign demand has been increasing over the last decade.

¹¹Other assumptions are nominal price and wage rigidities causing "sticky" prices.

¹²See: <http://www.realtor.org/sites/default/files/reports/2015/2015-profile-of-international-home-buying-activity-2015-06-18.pdf>

Economic freedom may be another important channel. Investors who are looking to keep their assets safe may purchase US housing. In this case, safe refers does not refer to high return low risk. It refers to the keeping assets safe from government seizure. If investors purchase non-physical assets, like stocks, governments in these countries can seize them. While it is unlikely, they will be able to seize real estate in other countries. This is a safer way for investors to protect asset holding.

3.4 Data

The section details the estimation of national and metropolitan area housing price indexes, and the advantages and disadvantages of S&P/Case-Shiller Home Price indexes compared to alternative measures. This section also discusses the BEA's new bilateral capital inflows data. Finally, there is a description of descriptive statistics.

3.4.1 Repeat Sales Housing Price Indexes

Aggregate housing price indexes are estimated in a similar fashion to other aggregate prices, like the Consumer Price Index (CPI). The CPI is a measure of the cost of a specific “basket” of goods and services, where in most cases only differ where and when items are sold. Similar to the CPI, Housing price indexes track the sales price of a “basket” of houses over time.

Even though the similarity, repeat housing price indexes and CPI differ in a number of ways. For the CPI, it is assumed that specific goods are of the same quality over time. For example, a Honda CRV sold in Morgantown; WV is assumed to be of the same quality of a Honda CRV sold in Pittsburgh, PA. Housing is inherently different. Specifically, each home have a different quality level.

Further, it is assumed that CPI basket does not differ in quality due to macroeconomic conditions. In contrast, the overall quality of the housing stock may differ based business cycles. All else equal, when the value of all housing is going up (due to favorable economic conditions), owners of low-quality housing may be more willing to sell. This causes the overall stock of housing to be of lower quality during periods of high economic growth compared to recessions.

Due to these complexities, housing price indexes have to in some way account for changes in quality. To control for quality, [Case and Shiller \(1987\)](#) develops a method for estimating a repeat sales housing price index. In a repeat housing price index, the housing basket only consists of existing homes sold more than once. The specific measure estimated by [Case and Shiller \(1987\)](#) has been continued by the S&P/Case-Shiller Home Price indexes.¹³

Other repeat sales indexes have been estimated. These measures only differ based on how data is

¹³For exact methodology, See: <http://us.spindices.com/index-family/real-estate/sp-case-shillerforexactmethodology>

collected. The Federal Housing Finance Agency (FHFA) also produces a repeat sale housing price index. This differs from the S&P/Case-Shiller Home Price index by using home sales data from Fannie Mae and Freddie Mac. This measure does cover more of the geographic country compared to the S&P/Case-Shiller Home Price index, but it would not include the sales of homes to most international buyers.¹⁴

The S&P/Case-Shiller Home Price index is available for the national and metropolitan area level. For national level, there are two indexes. The first, labeled Case-Shiller, is an overall aggregate of homes across the entire US. The second, labeled Composite, is a composite measure of ten metropolitan areas across the US, four on the east coast, four on the west coast and two in the middle. The four east coast are Miami, Washington, DC, New York, Boston. The four west coast are San Diego, Los Angeles, Las Vegas, and San Francisco. Finally, the two in the middle of the country, Chicago, and Denver.

For the metropolitan area, I use individual metropolitan areas from the Composite index. These are labeled by their city abbreviation and Case. So for example, the housing price index for Washington, DC would be labeled DCCase.

To estimate housing inflation, I take the log difference of the individual housing price index. Data is quarterly and ranges from Q2:2003-Q3:2015.¹⁵

Figure 3.5-3.8 show housing inflation at the national level and metropolitan area levels. Metropolitan areas are divided into three groups by geographical location, East Coast (Miami, Washington, DC, New York, and Boston), West Coast (San Diego, Los Angeles, Las Vegas, and San Francisco) and Non-Coast (Chicago and Denver).

Figure 3.5 shows the two national level Case-Shiller housing inflation measure ranging from Q2:2003-Q3:2015. As seen from the figures, housing price inflation increased from 2003-2006 covering the housing bubble. Starting in 2007, we see housing inflation to go below zero and finally housing inflation going positive again in 2010. Housing inflation after this point was below zero from late 2010 to late 2012. From 2013-2015, housing inflation has been positive.¹⁶ In general, housing inflation and therefore housing prices decreased more in cities compared to the country as a whole. Besides the period from mid-2007-mid-2009, housing inflation in metropolitan areas and the national level have been similar.

Figure 3.6 shows Case-Shiller Housing Inflation for metropolitan areas located on the East Coast. Housing inflation here seems to follow the national level. With housing inflation being positive before the national housing market bubble burst. Miami saw the largest decrease in housing inflation compared to the others. This fall was larger than the national composite measure. Washington, DC also saw the second largest decline, but more in-line with the national level composite housing inflation measure. Finally, New York and Boston did not have as a large decrease in housing inflation as the national level

¹⁴The NAR survey notes most international buyer purchase homes outright, with no US financing

¹⁵Source of data and description is located in Table 3.11-3.12 of the appendix

¹⁶With the exception of the Composite housing inflation dipping below zero for one-quarter in late 2014

measure of Miami and Washington, DC.

Figure 3.7 shows Case-Shiller Housing Inflation for metropolitan areas located on the West Coast. Again, the West Coast metropolitan areas also saw a large decline in housing inflation during the bubble burst. Unlike the East Coast metropolitan areas, all cities saw a large decline in housing inflation, similar to Miami. Las Vegas's decline lasted longer, where housing inflation finally going positive by late 2012.

Figure 3.8 shows the Case-Shiller housing inflation for Chicago and Denver. Denver is quite different from the other metropolitan areas where it saw only small increases before the bubble burst and smaller declines afterward. Chicago, in contrast, looks similar to New York and Boston saw a modest/large decline in housing prices.

3.4.2 Bilateral Capital Inflows

In August 2015, the Bureau of Economic Analysis (BEA) restructured its balance of payments data. This was done to be in-line with international standards. With the balance of payments now following international standards, bilateral capital flows have been estimated. The international trade literature has widely employed bilateral trade statistics. While the international finance literature has also relied on current account balances and overall aggregate financial accounts as a measure of capital inflows. Due to the new data more detailed and bilateral capital inflows is now available.

This new bilateral capital flow comes in two geographic levels, individual country, and regional level. This country level data is limited in a number of ways. First, there are only a few number of countries.¹⁷ Second, for many of these countries data is not complete, and miss observations.¹⁸ Third, to study the implications of Ferrero (2015) regarding capital inflows from countries with managed exchange rates from Asia and oil-producing countries, the sample of countries is not very informative. A majority of individual countries would not be considered in either group.¹⁹ Due to these reasons, the use of regional bilateral capital inflows is more informative and it is used instead of country level data.

Figure 3.2 is a map showing BEA defined regions. These regions are Latin America, Euro Area, non-Euro Europe, Asia and Pacific, Africa, and the Middle East.²⁰ Figure 3.3 shows countries that are or have been members of the international organization, Organization of the Petroleum Exporting Countries (OPEC). Finally, Figure 3.1 show the 2006 IMF classification of exchange rate regimes. Exchange rate regimes range from completely independent, like most advanced economies, to currency substitution (where the country uses it's domestic currency in parallel with a foreign currency). Independent exchange

¹⁷Countries available are Belgium, France, Germany, Italy, Luxembourg, Netherlands, United Kingdom, Canada, Argentina, Brazil, Mexico, Venezuela, Australia, China, Hong Kong, India, Japan, Korea, Singapore, Taiwan, and South Africa.

¹⁸Majority of this data is available. Some are hidden to protect information to some specific firms

¹⁹Only China, Singapore, Taiwan, Hong Kong, and Venezuela would be considered part of this group

²⁰While these "regions" are primarily defined by geography, although non-Euro Europe does contain countries located in Asia. These Asian countries considered part of Europe are primarily countries which were former republics of the USSR.

rate regimes are ones where there is no perceived targeting of other exchange rates. ²¹ Comparing Figure 3.2 and 3.1 gives an idea which regions would most likely be able to forecast housing inflation. Latin American exchange rate regimes seem to have most independent exchange rate regimes, with the largest economies, Mexico, Brazil, and Chile, following free floating exchange rates. Non-Euro Europe is the region that has the second most independent exchange rates with majority of countries following a managed floating exchange rate regime. The Asia and Pacific region is third most independent, with a mixture of both managed floating exchange rate regimes (India) and fixed peg exchange rate regime (China). Africa is different than all of the other areas with a mixture of all exchange regimes. As a note, based on theory and anecdotal evidence, it would be expected that African capital inflows would produce forecasting information. Particularly, due to a number of smaller economies and relatively low capital inflows into the US, and solid outflow due to international aid. Finally, the Middle East has a majority of countries which have a pegged exchange rate regime. It is expected that the greatest gains in forecast accuracy would come from middle east capital inflows.

Not only did the restructuring of balance of payments data provide regional capital inflows data, but it also provided less aggregated capital inflow asset classes. So we also have overall region capital inflow data but a number of asset classes. Regional asset and liability data is available for outward direct investment, portfolio investments, equity and investment fund shares, short and long term Debit securities, Other investment assets, currency and deposits, loans and trade credit and advances.

It is likely that capital inflows from some assets would have a different impact on financial market and US housing prices. For example, direct investments, equity and investment fund shares, and portfolio investment are likely to have a larger impact on producers. With the case of equity and investment fund shares capital inflows increases leads to higher stock prices. Through the income effect leads to higher consumption and larger demand for housing. This idea is consistent with the “Global Savings Glut” hypothesis. So not only does the question become what region does capital inflows matter but also what assets inflows matter in forecasting US housing inflation. At this point, due to interconnecting financial markets, I will be agnostic and refrain from speculating which asset class would be the most important in forecasting US housing inflation. Depending on the region and the specific policies, different assets may be important in forecasting US housing inflation.

Data ranges from Q2:2003-Q3:2015. Regional capital inflows are available for quarterly and annual frequency and non-seasonalized. To account for this I use the X11 method to seasonally adjust the data.

²² Capital accounts is provided as both assets and liabilities. To get a measure of capital inflows for each asset class, net asset purchases are subtracted from net liabilities. This provides a better measure of capital inflows. If it were the case that net asset purchases were equal to net liabilities the total amount

²¹See Ghosh et al. (2002) for further discussion

²²See: <http://www.okstate.edu/sas/v8/saspdf/ets/chap21.pdf>

of capital inflowing would be zero. For example, say those from Latin America purchased \$150 worth of assets from the US, while Americans purchased \$100 worth of assets from Latin America. While at the same time, Latin Americans received \$150 in loans from the US and Americans received \$100 in loans from Latin America. Net asset purchases and net liabilities in Latin America from the US would be both \$50. The true extra capital in the US would be \$0. As in the net inflow of capital would be zero.

Figure 3.9 shows capital inflows from three regions, Non-Euro Europe, Latin America, and OPEC countries from Q2:2003 to Q3:2015.²³ Non-Euro Europe has the largest movements from outflows to inflow compared to the other regions. This is likely due to these countries having larger economies compared to the other regions. During this period, Latin America asset inflows are mostly negative. Finally, OPEC seems to stay around zero to slightly negative capital inflow. At this level, there seems to be no difference before the great recession and afterward.

Figure 3.10-3.13 show the 4th lag of the sum of the overall capital inflow from Latin America, non-Euro Europe, and Asia and Pacific and current housing inflation. Figure 3.10 shows the regional capital and the Case-Shiller Composite. These series seem to move together. Housing inflation seems to decline at both the same time as capital inflows. Following with the theory provided from Bernanke (2005) and Ferrero (2015), the Federal Reserve increased the federal funds rate before the housing market crash. Finally, when the housing bubble burst, the Federal Reserve lowered the federal funds rate to around zero. All else equal, with this monetary policy we would see the dollar appreciate and then depreciate against all other currencies. Due to fixed exchange rate regimes, exchange rates would not change. This would likely cause a change in capital flows. Capital inflows during this period would decrease and then increase. Which is similar to what this graph shows, all though not perfectly. This relationship is messy when other countries monetary policy responds to their own macroeconomic conditions.

Figure 3.11 - 3.13 show the 4th lag of the sum of the overall capital inflow from Latin America, non-Euro Europe, and Asia and Pacific and current metropolitan area housing inflation. Figure 3.11 shows the graph for the east coast, Figure 3.12 shows the west coast metropolitan area, and Figure 3.13 shows the non-coast metropolitan areas. In general, all metropolitan areas, capital inflows, and housing inflation seem to move together. Boston housing price inflation and capital inflows seem to have the strongest relationship.

3.4.3 Descriptive Statistics

Table 3.1 shows descriptive statistics for select regional capital inflows and the Case-Shiller national level housing inflation measures.²⁴ The first column shows the mean, second column the standard deviation, and the third and fourth columns show the correlation with the two national housing price measures.

²³As a note, Ecuador and Venezuela are both members of OPEC and Latin America.

²⁴Other capital inflows are shown in table 2.10

The Composite measure has the highest mean out of the national level housing inflation measures and a higher standard deviation. The mean is around 10 percent higher but with around a 50 percent higher standard deviation.

For the capital inflows, Middle East (ME) Direct Investment has the largest inflow. Latin America (LA) Inflow has the largest outflow. Non-Euro Europe (NEE) Long-Term has a lower mean amount of capital inflow, it also has the highest standard deviation, which is similar to NEE Other Investment.

Liabilities for ME Asset, AP Investment Asset, are unavailable. With the number of financial transactions taking place between the US and the Asian and Pacific region (this region has some of the world's largest economies, China, Japan, India, Indonesia, and Australia) it is hard to estimate how large liabilities would be. In any case, it still has a large amount of asset inflows. Its standard deviation is similar to the capital inflows.

Both national housings inflation measures are highly correlated with each other, with a correlation coefficient of 0.95. This is something to be expected given the national Case-Shiller measure containing a large portion of the housing basket from the Case-Shiller Composite measure. For the rest of 3rd and 4th column, these are the correlation between current housing inflation and the 4th lag of the regional capital inflow measures. Out of these, the 4th lag of NEE Long Term is the most correlated with housing inflation with a correlation coefficient of 0.241 and 0.293 respectively. ME Asset and ME Long-Term are the second most correlated with housing inflation. ME Asset is negatively correlated with a correlation coefficient of 0.141 and 0.201. The negative relationship would explicitly mean that larger capital inflows are related to lower housing inflation. This relationship may be explained by the fluctuations of the oil prices. Oil prices saw a large increase, then declined and finally increased during the great recession and its aftermath. The ME Long Term is positively correlated with a coefficient of 0.155 and 0.218.

3.5 Forecasting Methods

This section discusses the specific forecasting methods used in this study. This includes in-sample and out-of-sample analysis and their advantages and disadvantages.

3.5.1 Empirical Forecasting Model

In my analysis, I employ nested model forecast comparisons of two national level and 10 metropolitan area housing inflation models using 67 capital inflow predictors. Eq 3.1,3.2 show my empirical models.

$$h_{i,t+\tau} = \sum_{k=0}^1 h_{i,t-k} \beta_k + u_{i,t+\tau} \quad (3.1)$$

$$h_{i,t+\tau} = \sum_{k=0}^1 h_{i,t-k} \beta_k + u_{i,t+\tau} \sum_{k=0}^3 p_{j,t-k} \gamma_k + u_{i,t+\tau} \quad (3.2)$$

Eq 3.1 details the restricted model. This is simply an AR(2) model. h_i is the housing inflation measure, where i denotes the specific national or metropolitan area measure. τ is the forecast horizon either 2, or 4 quarters ahead.

Eq 3.2 is the unrestricted model. k is the specific additional predictor and $p_{j,t}$ denotes the specific capital inflow.

In-Sample Analysis

In-sample analysis is simply the standard tests of significant (t-test) or joint significance (F-test) using all observations to estimate parameters. In nested model comparison, the addition predictor(s) are tested to if it(they) are significantly different from zero.

Overall, I test if the additional predictors (p_j) have coefficients ($\gamma_{j,k}$) that are different from zero, or if $\gamma_{j,1} = \gamma_{j,2} = \gamma_{j,3} = \gamma_{j,4} = 0$

Out-Of-Sample Analysis

Out-of-sample analysis is a forecasting technique where instead of using all observations to estimate parameters, the researcher splits the sample into two portions and uses one portion to estimate the parameters and the other portion to forecast.

Assuming a time series is T in length and ranges from $1, 2, 3, \dots, T$. The first portion, in-sample, is of length of P and ranges from $1, 2, \dots, P$. The second portion, out-of-sample, is of length R and ranges from $P + \tau + 1, P + \tau + 2, \dots, T + \tau - 1, T + \tau$.

In this study, the sample has a length of $T = 50$. The in-sample portion, has a length of $P = 30$ and the out-of-sample portion has a length of $R = 10$ or 12 (depending on the forecast) horizon. So P ranges from $Q2 : 2004, Q3 : 2004, \dots, Q1 : 2013$ and R ranges from $Q3 : 2013, \dots, Q3 : 2015$ for the 2 quarter forecast horizon.

As a note the sample split is denoted π and is equal to $\frac{R}{P}$. The optimal value of π has been the subject of study by the recent theoretical econometric literature. Works like [Clark et al. \(2013\)](#), and [Rossi and Inoue \(2012\)](#) have found that a ratio should be greater than one. This is rather unfeasible in this study due to the short time series. In fact of a number of the most recent housing inflation literature ([Plakandaras et al. \(2015\)](#)) use a π value around $\frac{1}{4}$ similar to this study.

The associated parameters can be estimated three different ways. First, parameters can be estimated once (fixed). This is the best solution when it can be assumed that parameters do not change during the sample. Second, through a rolling window, (rolling). This method takes in new information after

each forecast has been realized and drops information at the same time. This is a superior method if the empirical relationship is not constant during the time series. Finally, parameters use all information before forecast being made (recursive). Again new information is added after the forecast is realized. This is superior when the sample size is smaller or the parameter estimate is unstable. As a note, if the empirical relationship is stable and parameter estimated does not change, then the forecast accuracy from the fixed, rolling, and recursive methods would be the same.

In this study, I employ the recursive forecasting method to estimate model parameters. Eq 3.3 shows the recursive method with a quadratic forecasting loss function for the unrestricted model. ²⁵

$$\hat{\beta}_t, \hat{\gamma}_t = \underset{\beta, \gamma}{\operatorname{argmin}} \sum_{s=1}^{t-\tau} (h_{i,s+\tau} - (\sum_{k=0}^1 h_{i,t-k} \beta_k + \sum_{k=0}^3 p_{j,t-k} \gamma_k))^2 \quad (3.4)$$

After all forecasts are estimated, the mean squared forecast error (MSFE) is used to determine accuracy (Eq 3.5).

$$MSFE = (P - \tau + 1)^{-1} \sum_{t=P+\tau}^T (h_{i,s+\tau} - \hat{h}_{i,s+\tau})^2 \quad (3.5)$$

Out-Of-Sample Forecast Evaluation

When there are two or more possible forecasting models, comparisons in MSFE can be used to determine forecast accuracy. In this discussion, I will only focus on two forecasting model accuracy comparison.

²⁶ The two standard methods of comparing forecast accuracy are simple statistics, MSFE ratio, and statistical tests, MSFE F-test.

First, the MSFE ratio is calculated by simply taking a ratio of MSFE's. Normally, the MSFE of the model in interest model is divided by the baseline model MSFE (in this study is the AR(2)). Eq 3.6 shows how this statistic is estimated.

$$MSFE_{ratio} = \frac{MSFE_{unrestricted}}{MSFE_{restricted}} \quad (3.6)$$

This statistic ranges from zero to positive infinity ($0 \geq MSFE_{ratio} \geq \infty$). When the value of this statistic is greater than one, the baseline model is more accurate, as it has a lower MSFE. In the same way, when the MSFE ratio is lower than one, the other model is more accurate. We can also tell by what percentage the unrestricted model is more accurate than the baseline model. This is simply done

²⁵For the restricted model the recursive method is

$$\hat{\beta}_t = \underset{\beta, \gamma}{\operatorname{argmin}} \sum_{s=1}^{t-\tau} (h_{i,s+\tau} - (\sum_{k=0}^1 h_{i,t-k} \beta_k))^2 \quad (3.3)$$

²⁶The out-of-sample literature has provided a number methods to compare forecast accuracy across a number of models. See: Clark et al. (2013)

by taking the subtracting one from the MSFE ratio and multiplying by 100.

If the MSFE ratio is near 1 (0.90-1), we are unsure which is more accurate. It could simply be characteristic of the data we are drawn. Due to this, a number of statistical tests have been developed. In general, we want to test specifically if two competing models are equally accurate in forecasting at the population level. The standard way to test is to determine if the difference in squared errors between model are different from zero. ²⁷ The null hypothesis is one where MSFE's are equal.

$$MSFE - F = \frac{(P)^{-\frac{1}{2}} \sum_{t=R}^{T-1} \hat{d}_{t+\tau}}{\hat{\sigma}_2^2} \quad (3.7)$$

Eq 3.7 describes the MSFE F-test statistic. \hat{d} is the difference in squared errors between the baseline model and the other model. While $\hat{\sigma}_2^2$ is the estimated variance of the unrestricted model.

This is a right sided test. Intuitively, this makes sense. If the difference between squared errors is positive in time, t . This would mean that the baseline model was less accurate.

In a majority of the out-of-sample econometric theory, getting the correct critical values has been the goal (Clark and McCracken (2001, 2005a); McCracken (2007)).

Diebold and Mariano (1995); West (1996) develop a number of tests for non-nest out-of-sample forecast comparison. These find under this cause the asymptotic distribution is normal, leading to easy estimation. Clark and McCracken (2001, 2005a), in contrast, find that nested models forecast comparison have a non-normal asymptotic distribution. As the value of π goes to ∞ the asymptotic distribution does not approach normal. The distribution and the critical values depend on the sample split, estimation method, and a number of restrictions. Clark and McCracken (2001) provides Monte Carlo estimations of the asymptotic distribution and the associated critical values for a number of different sample splits. These are used in this study.

Relative Benefits of Out-of-Sample and In-Sample Analysis

The discussion at this point has been about how in-sample and out-of-sample estimation are performed.

There has been little discussion about the relative merits of each. The question becomes which forecasting method is better at determining forecasting accuracy.

On the surface, the out-of-sample analysis allows the researcher to determine how well the specific model has actually forecasted in the past. Where in-sample only determines the relationship over the whole sample. Further, the in-sample analysis may also suffer from spurious results. Due to taking in the whole sample, models which in away include many predictors may be over-fit. The real problem develops when the model actually predicts the error rather than the series being forecasted (signal vs.

²⁷Another test if a linear combination of the two forecasts can produce a more accurate forecast, this is also known as an Encompassing test.

noise).

Another issue develops when there is structural instability in the estimation. Say, for example; there is a structural break halfway in the series. If there is a strong relationship between a specific predictor and the forecasted variable in the first portion of the sample and no relationship past that. With in-sample analysis, we may find that this predictor is significant. Out-of-sample analysis likely finds that predictors are poor in both of these cases.

Clark (2004); Clark and McCracken (2005b) tests the reliability of out-of-sample and in-sample analysis under over-fitting and parameter instability using Monte Carlo simulations. Overall, these studies find that out-of-sample analysis performs better under over-fitting and parameter instability compared to in-sample analysis.²⁸

Inoue and Kilian (2005), along the same reasoning, tests whether in-sample or out-of-sample have more power, under data mining and parameter instability. Using econometric theory, this study finds that first in-sample analysis is more powerful, due to larger sample size. Second, out-of-sample analysis can have suffered from the same over-fitting problems as in-sample analysis. In general, Inoue and Kilian (2005) shows there is no advantage to out-of-sample testing over in-sample testing.

These studies focus on the cases when in-sample and out-of-sample results differ. Due to the conflicting evidence to accurately determine forecast accuracy, it is best to use both in-sample and out-of-sample analysis and only focusing on matching results.

3.6 In-Sample Results

Tables 3.2- 3.4 show the national level and metropolitan area level in-sample results for the 2 and 4 quarter forecast horizon. As detailed in the methods section, F-tests are most commonly used to determine marginal predictability of additional predictors for in-sample forecast comparison. For these tables only F-statistics are shown with levels of significance denoted by *.²⁹

3.6.1 National Level In-Sample Results

Table 3.2 shows the national level in-sample results for selected bilateral capital inflow predictors both 2 quarter and 4 quarter forecast horizons.

The in-sample results indicate that bilateral capital inflows are important predictors of US housing inflation. A number of these predictors, NEE Long Term, ME Asset, ME Equity Shares, and OPEC Equity Shares, are significant for both national housing inflation measures and forecast horizons. Although

²⁸By perform better, I specifically mean that out-of-sample is better able to determine if the model actually produces a more accurate forecast.

²⁹Other results can be found in appendix (Table 3.13-3.15)

only the ME Asset measure that is significant at the 1 % level for both measures and forecast horizons.

For the 2 quarter forecast horizon, the ME Asset model performs the best by having the highest F-stat for both national housing inflation measure at the 1 % significance level. The NEE Long Term and OPEC Equity Shares models also perform well with both being significant at least for the 10 % level for both national level measures. No other measures for the 2 quarter forecast horizon are significant at the 1 % significance level. NEE Other Investment, LA Inflow, AP Investment Asset, and ME Long-Term are unable to produce significant results for any of the national level measures.

For the 4 quarter forecast horizon, the ME Asset model again performs the best, with the predictors being significant at the 1 % level for both national measures. The NEE Long Term, ME Direct Investment, ME Equity Shares, ME Short Term, and OPEC Equity Shares models are also statistically significant for both national level measures. Like the 2 quarter forecast horizon, the NEE Other Investment, LA Inflow, and AP Investment Asset models are unable to produce significant results. Finally, the LA Inflow model performs the worst, with the lowest F-stats for both national measures.

The asset purchases from middle eastern countries overall perform the best while the specific components (ME Direct Investment, ME Direct Investment, ME Short Term, and ME Long Term) do not produce more accurate forecasts.

The results from ME Equity Shares and OPEC Equity Shares are relatively similar. These both contain much of the same capital inflows data. This is to be expected due to many of the same countries located in the Middle East are also members of OPEC. ³⁰

3.6.2 Metropolitan Area Level In-Sample Results

Table 3.3 and 3.4 show the metropolitan area level housing inflation for the 2 quarter and 4 quarter forecast horizon. As the results for the national level, the ME Asset model is significant for all metropolitan areas and forecast horizons, many at the 1 % significance level. The LA Inflow and AP Investment Asset models perform relatively poorly. The AP Investment Asset model is unable to produce any significant results. The LA Inflow model only produces significant results for the 2 quarter forecast horizon and the Denver and Boston metropolitan areas.

Table 3.3 shows the results for the 2 quarter forecast horizon. The NEE Long Term, ME Asset, and OPEC Equity Shares perform relatively well and produces significant results for at least six of the ten metropolitan areas. Further, the ME Asset model produces the largest F-stat for eight out of the ten metropolitan areas. The NEE Long Term model produces the largest F-stat for Las Vegas. The ME Equity Shares produces the largest F-stat for the Boston metropolitan areas. The NEE Other Investment and AP Investment Asset models are unable to produce any significant results.

³⁰Iraq, Kuwait, Qatar, Saudi Arabia, and the United Arab Emirates are both part the middle east and part of OPEC.

The San Francisco, San Diego, Chicago, and Boston metropolitan areas have the largest number of predictors significant with six out of ten each. The Washington, DC and Denver metropolitan area are next with five. Las Vegas and Miami are the worst with only two models significant.

Table 3.4 shows the results for the 4 quarter forecast horizon. Like the two quarter forecast horizon, the NEE Long Term and ME Asset model perform relatively well and produces significant results for at least nine out of the ten metropolitan areas. The ME Equity Shares and ME Short Term also do well and are able to produce significant results for seven out of ten metropolitan areas. The LA Inflow and AP Investment Asset models perform the worst, and unable to produce significant results for any metropolitan areas.

Again, the ME Asset model produces the largest F-stats for seven out of the ten metropolitan areas. For the remaining, the ME Short Term produces the largest F-stat for Denver, NEE Long Term produces the largest for Las Vegas and finally, the ME Direct Investment produces the largest F-stat for the Boston metropolitan area.

The San Francisco, Los Angeles, Denver, and Boston metropolitan areas have the largest number of significant models with six each. Again the Washington, DC metropolitan area produces similar results with five significant models. The New York, Las Vegas, and Miami perform the worst with three each.

3.7 Out-Of-Sample Results

Tables 3.5- 3.7 shows the national level and metropolitan area level in-sample results for the 2 and 4 quarter forecast horizon. These results show both the MSFE ratio between the baseline model and the specific capital inflow model and the MSFE F-stat. As noted before, MSFE ratio's with a value less than one denote improved forecast accuracy. For the MSFE F-test, the level of significance is denoted by *.

31

3.7.1 National Level Out-Of-Sample Results

Table 3.5 shows the out-of-sample results for the national level 2 and 4 quarter forecast horizons. Unlike the in-sample results, very few models are able to produce more accurate forecasts compared to the restricted model. Only the NEE Long Term model is able to produce significant results across all national housing inflation measures and forecast horizons.

For the 2 quarter forecast horizon, The NEE Long Term, and LA Inflow models are significant for both national housing inflation measures, at the 1 % significance level. Conversely, the NEE Other Investment, LA Inflow, AP Investment Asset, ME Asset, ME Direct Investment, and ME Long Term

³¹Other results can be found in the appendix (Table 3.16-3.18)

models are unable to produce any significant results. For the Case-Shiller measure, the NEE Long Term and LA Inflow model are able to produce a 14 percent increase in forecast accuracy over the baseline model. For the Composite measure, the NEE Long Term model produces the most accurate forecasts, with an increase in forecast accuracy of 26 percent. The ME Short Term and ME Long Term models perform the worst and decrease forecast accuracy by around 20-30 percent for both measures. These results conflict with the in-sample testing. Only the NEE Long Term model can produce significant results for the 2 quarter forecast horizon, in-sample and out-of-sample.

For the 4 quarter forecast horizon, the NEE Long Term, NEE Other Investment, ME Equity Shares, and OPEC Equity Shares model are able to produce significant results for both national housing inflation measures. The NEE Long Term and ME Equity Shares models produce the largest gain in accuracy for the Case-Shiller measure, of around 35 percent. The OPEC Equity Shares also performs well with an increase in accuracy of 25 percent. For the Composite measure, the ME Equity Shares, and OPEC Equity Shares model perform the best and are able to increase forecast accuracy by 47-54 percent. The NEE Long Term and LA Inflow models also perform well with increases of 30 percent. The results here are relatively consistent with in-sample results. The LA Inflow model differ between in-sample and out-of-sample testing.

Based on both in-sample and out-of-sample results, the asset class and region seems to matter in forecasting US national level housing inflation. The net purchases of long-term assets from non-EU Europe produce the best forecast accuracy across forecast horizons. These results are consistent with the international finance theory provided by the “global glut savings” hypothesis and Ferrero (2015). As shown by Figure 3.2 and 3.1, many countries which are located in the BEA defined non-EU Europe followed a managed floating exchange rate in 2006. Changes in capital flows may be either response to Federal Reserve policy or less risk associated with longer-term US assets. Although forecast exercises, in general, do not determine causality, these results provide additional evidence against the theory in Favilukis et al. (2010, 2014).

Figure 3.14 and 3.15 show a graph of the squared forecasting errors of both the baseline AR(2) and the European Long Term Capital inflow and Latin American Capital inflow. Figure 3.14 shows the 2 quarter forecast horizon, while Figure 3.15 shows the 4 quarter forecast horizon.

For Figure 3.14, there does not seem to be many periods which are able to produce more accurate forecasts. For the Case-Shiller measure, both Capital inflows are able to produce slightly better forecasts for Q1:2013-Q3:2013. The Composite measure is different where the European long-term inflow seems to produce a slightly better forecast across the whole forecasting period (except for Q3:2013).

Looking at Figure 3.15, some improvement is shown for either Capital inflow measure and the national level Case-Shiller measure. Specifically, the European Long-term model has a lower squared error for

specifically from Q4:2012-Q4:2013, and Q2:2014- Q4:2014. In most of these time periods, national level housing inflation changed direction (from increasing, to decreasing). This is similar to the results in McGurk (2016), which shows that Capital inflows perform the best when the housing inflation changes direction. The Composite sees even a larger different between the two models from the Q4:2012-Q4:2013. This model also has substantial gains at the end of the out-of-sample forecasts with a large gain from Q2:2014-Q1:2015. The Latin American Capital inflow differs produces similar forecasts for both national housing inflation measures. These also would relate to the poor MSFE ratio.

3.7.2 Metropolitan Area Level Out-Of-Sample Results

Table 3.6 and 3.7 shows the out-of-sample results for the metropolitan area level 2 and 4 quarter forecast horizons. Unlike the other housing inflation levels, there is limited consistency across forecast horizons.

Table 3.6 shows the results for the 2 quarter forecast horizon. The NEE Long Term model outperforms all others and produces six significant results out of ten. These include the four west coast metropolitan areas (San Francisco, Las Vegas, Los Angles, and San Diego) and Chicago and Boston. These slightly differ from the in-sample results where the results from San Francisco, San Diego, Los Angles, and Chicago are consistent across out-of-sample and in-sample results. The NEE Long Term model produces the largest or second largest gain in forecast accuracy for San Francisco (38 percent gain), Los Angles (24 percent gain), San Diego (24 percent gain), Las Vegas (18 percent gain), and Boston (11 percent gain).

The results from the LA Inflow and ME Asset differ between in-sample and out-of-sample results. Specifically, the ME Asset does very well in-sample and produces the largest F-stat for many of the predictors. While for the out-of-sample results the ME Asset model is only able to produce significant results for two metropolitan areas (New York and Chicago). For New York, the ME Asset model produces the most accurate forecast with an increase of 11 percent. For Chicago, the ME Asset model is able to increase forecast accuracy of around 15 percent.

The LA Inflow model, in contrast, performs relatively well for the out-of-sample test and poorly for in-sample. For the out-of-sample test, the LA Inflow model is able to produce a more accurate forecast for Los Angles, New York, San Diego, Las Vegas and Boston. Further, this model produces the most accurate forecast for Los Angles(38 percent gain), San Diego (24 percent), and Boston(16 percent gain). At the same time, this model only produces significant results for Denver and Boston in-sample.

For the three other metropolitan areas, Denver, Washington, DC, and Chicago, the ME Short Term, AP Investment Asset, and ME Direct Investment models produce the largest gains of 35 percent, 9 percent, and 30 percent. Finally, no models are able to produce more accurate forecast compared to the baseline model for the Miami metropolitan area.

The number of additional significant predictors for each metropolitan areas is lower than for the in-sample results. San Diego has the most significant models with 5, followed by four from Los Angeles and Chicago. Besides Miami, the Washington, DC and San Francisco metropolitan areas also are poorly predicted by the models and only have two significant models each.

Table 3.7 shows the results for the 4 quarter forecast horizon. The ME Equity Shares model here performs the best, and it is able to produce significant results for all metropolitan areas except for Miami. This is relatively consistent with the in-sample results where only the results from New York and Las Vegas differ.

The ME Equity Shares produces the largest or second largest gain in forecast accuracy for a number of metropolitan areas (Los Angeles (38 percent gain), Denver (34 percent gain), Chicago (31 percent gain), and Boston (17 percent gain)). The NEE Long Term model also performs relatively well, and it is able to produce significant results for six of the metropolitan areas. This is similar to in-sample results, where the NEE Long Term model is able to produce significant results for nine of the metropolitan areas. Only the results from San Francisco, Los Angeles, and Miami differ. The NEE Long Term model produces the largest or the second largest increase of forecast accuracy for New York (33 percent gain), Las Vegas (21 percent gain), and Washington, DC (22 percent gain).

The results from the LA Inflow significantly differ from in-sample and out-of-sample tests. Specifically, the LA Inflow finds significant results for six metropolitan areas for out-of-sample results while for in-sample it can not produce a single significant result. The opposite is the case for the ME Asset where in-sample is significant for all metropolitan areas and insignificant for all measures for out-of-sample. Finally, the AP Investment Asset and ME Long Term models are unable to produce any significant results for this forecast horizon.

Again, the number of significant models for each metropolitan areas differ from the in-sample results. The San Diego metropolitan area again has the most with 5, followed by 4 for Las Vegas and Chicago. Again the Miami measure has no models which produce more accurate forecasts compared to the baseline model. San Francisco and Washington, DC also see the lowest gains from the models and only have two significant models each.

Across both forecast horizons, the NEE Long Term model seems to perform the best for metropolitan areas both in-sample and out-of-sample. This model as well at the national level. This results can be easily explained by the "global glut savings" hypothesis.

There is a number of reasons why the NEE long-term capital inflows should be related to US housing. First, many of the countries in the Non-EU Europe region are oil producers (Norway, Russia, and Kazakhstan). Like the relationship found in McGurk (2016) (Ch. 2), Norwegian exchange rates are relatively good predictors of US housing inflation. Oil prices are also related to the US economy and

housing prices, directly. Second, due to the instability of the Russian exchange rate with the dollar (Ruble/Dollar), investors from non-EU Europe looking for less risky assets, purchase US Treasury securities.³² With the increase in demand for longer term assets, longer-term interest rates decrease (mortgage rates). Thus causing US housing prices to increase.

Similar to the results in McGurk (2016) (Ch. 2), capital inflows perform better and produce more accurate forecasts for the 4 quarter forecast horizon compared to the 2 quarter forecast horizon. This result is likely a consequence of the illiquidity of housing.

Following the “global glut savings” hypothesis, increased capital inflows into the US, caused world interest rates to decline (higher supply of loanable funds). Thus the price of housing increase given the lower cost of financing housing causing the housing bubble. Due to the illiquidity and high transaction costs of housing, it takes longer for the housing market to clear.

For example, say mortgage rates decreased causing investors or those demanding housing services to increase demand for housing. At any given point in time, housing stock is fixed. Existing homes would need to enter the market, and new homes would have to be built for the market to finally clear.³³ At the same time each home is unique and demanders of housing services may have constraints by quality and locations. Thus causing more time for markets to clear. Given this, it is expected capital inflows would be able to forecast better for longer time horizons.

3.8 Combination Methods

As is the case for much as the macroeconomic forecasting literature, there are a large number of possible predictors (N) with a relatively smaller time series (T), $N > T$. With the new estimation of financial account data produced by the BEA, and data ranges from 2003-2015, this is also the case.

For regional quarterly data, there is around 60 usable predictors and when lags are included the number ranges up to 200 and only 50 observations. Due to this, dimension reduction techniques can be used to develop forecasting models which are able to maximize signal rather than noise. In the following section, I will describe a number of dimension reduction methods and apply them to forecasting US housing inflation prediction.

3.8.1 Forecast Combination

With any forecasting exercise, the goal is to provide a more accurate forecast with relatively limited additional cost. Sifting through these results to determine which model produces the most accurate

³²See: <http://www.tradingeconomics.com/russia/currency>

³³On average it takes 8 months for a new home to be built. See: <http://eyeonhousing.org/2013/10/how-long-does-it-take-to-build-a-house/>

forecast for each measure and forecast horizon would be rather costly and would likely suffer from data mining problems (Inoue and Kilian (2005)). Given the relative closeness in forecast accuracy between models, this would likely only provide small, comparable gains in forecast accuracy. Forecast combination techniques provide a lower cost solution to this problem.

Intuitively, forecast combination is similar to asset diversification. By holding a wide range of assets, the variance of returns is smaller than just by holding one asset. Individual forecasting models may capture different forecasting information at different times. Specific models may be better during certain periods of time and worst during others. So if one model is superior before a structural break and another after this break., a combination of both models would be better than either over the entire sample.

A number of forecasting combination methods have been proposed, ranging from complex to simple. In practice, the more complex combination techniques rarely produce marginally better forecasts compared to the simple methods (Rapach and Strauss (2009)). Given this empirical observation, I only include the simplest forecast combination techniques. Explicitly, I only use the mean, trimmed mean, and median forecast combination methods.

Mean

For this method, an arithmetic mean is taken for each forecast period and then compared against the realized housing inflation measure. Assume there are K (in the case of this forecasting exercise $K = 10$) number of different forecasts

$$\hat{h}_{t+\tau} = K^{-1} \sum_{i=1}^K \hat{h}_{i,t+\tau} \quad (3.8)$$

3.8 shows the simple mean of all forecasts, for forecast $t + \tau$. Like the individual forecast models accuracy is judged by MSFE. Using the asset diversification analogy, this would be equivalent to an investor holding a "naïve" portfolio where every forecast is given equal weight.

The trimmed mean and median forecast combination are also computed. For the trimmed mean forecast combination, only the middle 50 percentile of forecasts is used, the bottom and top 25 percentile of forecasts are dropped. Finally, the median measure takes the median forecast out of the set forecasts for each forecasted period.

All of this measures assume that the true and best forecasts are centrally located in the set of forecasts. If this is not the case, then all of these methods overall would not be able to produce more accurate forecasts compared to the baseline model.

3.8.2 Diffusion Indexes

Another method used in the forecasting literature is the development of diffusion indexes. This technique extracts relevant information on latent factors by estimating a dynamic factor model. The general idea is there is an unobservable component that drives a portion of the movement of predictor set. Eq 3.9 and 3.10 detail the standard dynamic factor model set up.

$$P_t = \lambda(L)f_t + e_t \quad (3.9)$$

$$f_t = \psi(L)f_{t-1} + \nu_t \quad (3.10)$$

P_t is vector of all predictors at time t . Each predictor is then a function of these unobservable factors, f_t and $\lambda(L)$ is a lag polynomial matrix of factor loadings. In the context of capital inflows, P_t represents the set of predictors and f_t is the unobserved factors that are likely important in forecasting US housing inflation.

Stock and Watson (2011) provides an overview of the most used methods of estimating dynamic factor models and the associated diffusion index. Stock and Watson (2011) details three generations of dynamic factor model estimation.

The first generation estimates dynamic factor models using maximum likelihood estimation (MLE) and the Kalman filter. This method can only be used when the number of predictors (N) is relatively small compared to the time series (T).

This approach is estimated by first rewriting Eq 3.9 and 3.10 as a linear state space model provided in Eq 3.11 and Eq 3.12.

$$P_t = \Lambda F_t + e_t \quad (3.11)$$

$$\Psi(L)F_t = G\nu_t \quad (3.12)$$

Here $F_t = (f'_t, f'_{t-1}, \dots, f'_{t-p})'$ and $\Lambda = (\lambda_0, \lambda_1, \dots, \lambda_p)$. Ψ is a matrix comprised of 1's, 0's, and elements of ψ and G is a matrix of 1's and 0's. Finally ν_t is assumed to be $MVN(0, I_p)$ and e_t is $MVN(0, \Omega)$. A number of assumptions can be made on Ω to ease estimation.

This method has an advantage in that it can deal with predictors that are observed in a different frequency. So a diffusion index can be calculated using data that is observed for multiple frequencies. For example, if different capital inflows data was observed in either monthly, quarterly, or yearly frequency a quarterly diffusion index can still be estimated.

The second generation estimates dynamic factor models using principal component analysis. This method is not limited by the number of predictors, and can be estimated when N is relatively large

compared to T by using cross-sectional averaging. ^{34 35}

$$\min_{F_1, \dots, F_T, \Lambda} \frac{1}{NT} \sum_{t=1}^T (P_t - \Lambda F_t) \quad \text{s.t. } N^{-1} \Lambda' \Lambda = I_r \quad (3.13)$$

Λ is a matrix of eigenvectors of the sample variance of the predictor matrix, P_t . This method has the advantage of being less time consuming to estimate compared to the first generation method.

Finally, the third generation uses both the state space method of the first generation and the convenience of principal component analysis in the second generation. To estimate this the first step is to estimate factors by principal components. Next, in the second step, the estimated factors are used to compute the unknown parameters in the state space model. This has the advantages of both of the estimation methods.

Due to the relatively small number of predictors that produce more accurate forecasts, I estimate diffusion indexes similar to the first generation methods. In my estimation, Ω is assumed to have unequal variance across individual predictors errors. ³⁶ Finally, the varimax rotation is used to estimate Λ and F_t . To determine the number of factors I use an information criteria method (AIC), the number of factors is determined. With the goal in forecasting using housing prices, the number of factors is not necessarily important. ³⁷

In this study, I estimate three diffusion indexes in two unique ways. For the first, which from I call unrealistic diffusion indexes, is estimated using all observation available. This method is unrealistic because I use information which is not available when the forecast is being made. Such that the diffusion indexes are only estimated once. The second, which I call realistic diffusion indexes, are estimated first for the full in-sample period, and then re-estimated after each forecast is realized. The method can be used in actual real-time forecasting and only contains information which is available at the time the forecast is being made.

3.9 Combination Results

The results below show the combination methods. Table 3.8 and 3.9 show the national level out-of-sample results. Tables 3.9 and 3.10 show the metropolitan area out-of-sample results. As a note, predictors used in the diffusion index and forecast combination are NEE Long Term, NEE Other Investment, LA Inflow, AP Investment Asset, ME Asset, ME Direct Investment, ME Equity Share, ME Short Term, ME Long Term, and OPEC Equity share.

³⁴Cross sectional averaging is done by taking a weighted average of the idiosyncratic disturbances. When N is large enough by the weak law of large numbers the weighted average of the idiosyncratic disturbances will go to zero and only the factors remain.

³⁵See: [Stock and Watson \(2011\)](#) for further discussion

³⁶There are a number Ω specifications; I choose this one because of it has the lowest AIC

³⁷AIC results determine the prescribed number of factors as one. As a test of robustness, I include three.

3.9.1 Out-of-Sample National Level

Overall, there seems to be some gain from combining methods. Specifically, the unrealistic dynamic factor model performs the best. There is a large gain in accuracy of 55 percent for the 2 quarter forecast horizon. There is even a larger gain the 4 quarter forecast horizon, with an increase of accuracy by up to 66 percent.

The goal of this forecasting exercise is to test the theory presented in Ferrero (2015). The results from the first factor show that there is some truth to this theory. The second and third factor do not perform as well. The second factor is unable to produce a significant forecast for any measure or forecast horizon. The realistic diffusion index also does not perform as well. Where only the 4 quarter forecast horizon and Case-Shiller measure is significant. The difference between the results is most likely due to the small times series. With more observations more accurate relationships can be found.

The forecast combination methods are significant for every national measure and forecast horizon. All centrality measure performs about as well. The mean method produces the best forecast for 2 quarter forecast horizon, with an increase of between 18-28 percent. All other methods produce almost identical results. For the 4 quarter forecast horizon, the median measure performs the best, and it is able to produce 42 percent increase in forecast accuracy for the Composite. The other forecast combination methods are able to produce similar results with gains between 39 percent and 35 percent. For the Case-Shiller measure, all forecast combination methods are able to produce around 25 percent gain in forecast accuracy.

These results are interesting. The forecast combination methods are able to produce more accurate forecasts across the board, where the diffusion indexes produce significantly better forecasts but only for a few of cases. The results may be just limited to the length of the time series. The gains from the first factor of the unrealistic method seem to support this case. This diffusion index produces a larger increase in forecast accuracy compared to the forecast combination method, where the diffusion index produces more accurate forecasts by around 50 percent while the later increases by 20 percent. For the 4 quarter forecast, horizon gains differ by 54-67 percent and 27-40 percent.

These results differ from the individual results where only the model, NEE Long Term, can produce significant results for all measure and forecast horizons. The unrealistic first forecast not only is also significant for all measures and forecast horizons, but it also produces better forecasts compared to the NEE Long Term model. The forecast combination methods are so also significant for each measure but with similar forecasting gains as the NEE Long Term model.

Figure 3.16 and 3.17 shows a graph of the squared forecasting errors of both the baseline AR(2) and the unrealistic first factor and mean forecast combination method. Figure 3.16 shows the 2 quarter forecast horizon, while Figure 3.17 shows the 4 quarter forecast horizon.

For Figure 3.16, The unrealistic first factor performs rather well compared to the AR(2) model. For the Case-Shiller measure, the unrealistic first factor is able to produce better forecasts for Q2:2012-Q4:2013 and Q4:2014-q3:2015. While, the Mean forecast combination is able to produce a better forecast for all most every out-of-sample forecast period, although only slightly better. The results are similar for both national housing inflation measures.

Figure 3.17 differs from the 2 quarter forecast horizon. The forecasts here seem to be more volatile, especially for the Composite measure. The first unrealistic factor still performs well compared to the AR(2) model from Q4:2012-Q3:2014. Both the mean and first unrealistic forecast performs worse than the AR(2) model from the Q3:2013-Q2:2014 for the Composite measure.

Although in general, these forecast combination methods are able to produce more accurate forecast when housing inflation changes direction. Again like the results from the individual models. This further gives the idea that capital inflows are not only useful in forecasting US housing inflation but can produce better forecasts when housing inflation changes direction. This is similar to the evidence from McGurk (2016) (Ch. 2).

3.9.2 Out-of-sample Metropolitan Level

Tables 3.9 and 3.10 shows the metropolitan area results for the combination methods. The results are similar to the national level. The first unrealistic diffusion index performs rather well, across both forecast horizons. The gains are substantial.

Table 3.9 specifically shows the results from the metropolitan area and two-quarter forecast horizon. Like the national results, the unrealistic first factor produces significant forecasts across all metropolitan areas. A majority of the gains are around 30-40 percent. The largest gain is found for Las Vegas with a gain of over 70 percent. The smallest gain is from Boston with around 30 percent. The first unrealistic factor also produces the most accurate forecasts for every metropolitan area.

The second and third unrealistic factors perform relatively poorly, and none are able to produce significant results for any metropolitan areas. Similarly, the realistic factors also are not able to produce many significant results. Only the third factor produces significant results for New York with a gain of 13 percent.

In contrast with the factor models, the combination methods are able to produce significant results across all metropolitan areas except for Miami and every measure. The gains here are not as large as the unrealistic first factor and range from 10-35 percent. The largest gain is for San Francisco of around 35 percent while the smallest positive gains are for Las Vegas with around 10 percent. In general, it does show that there are gains that can simply be made by taking an averaging the best forecasting models.

These results do differ from the individual models which are unable to produce significant for some

of the metropolitan areas. In fact, the best model (NEE Long Term) is only able to produce a significant result for six out of the 10 metropolitan areas.

Table 3.10 shows the results for the 4 quarter forecast horizon. These results are similar to 2 quarter forecast horizon. The unrealistic first factor produces significant results for all 10 metropolitan areas. There seem to be larger gains in forecast accuracy compared to the 2 quarter forecast horizon, with most gains around 50 percent. Again the greatest gain is for Las Vegas of about 80 percent. Chicago and Boston see the smallest gains of around 30 percent. Unlike the results from 2 quarter forecast horizon, the third unrealistic factor produces significant forecast in 7 out of the 10 metropolitan areas. The largest gains being for Denver and by 43 percent.

The second unrealistic factor and all of the realistic factors perform relatively poor and only produce limited increases in forecast accuracy. The third realistic factor produces significant results for 4 metropolitan areas. The largest gains are for Los Angeles, New York, and Chicago with gains of around 30 percent.

Finally, like again the results for the 2 quarter forecast horizon, the forecast combination methods produce significantly better results for all metropolitan areas except for Miami. All three measures perform about equally as well with the only difference of about 3 percent between each. The gains here, range from around 15-35 percent. The biggest gains come from Los Angeles and San Diego both by 35 percent.

These combination methods do well compared to the individual models. The best models (NEE Long Term and ME Long Term) are only able to produce significant results for 7 and 8 metropolitan areas., while a number of the combination methods are able to produce significant results for 9 or all of the metropolitan areas.

In general, for both forecast horizon, there seems to be a gain from using the combinations methods. The results from the unrealistic factors perform very well. These factors could not have been used in actually forecasting housing inflation due to using information not contained in the information set. Due to the rather limited sample size the estimated factors do change, giving the relative difference in forecast accuracy between the unrealistic and realistic factors. This does provide evidence that going forward these factors may be useful in forecasting US housing inflation.

The forecast combination methods also perform rather well. It is likely that these centrality measures contain some hidden information.

3.10 Conclusion

With more accurate forecast, better decisions can be made. The decrease in forecast accuracy particularly after the housing price bubble and bust has lead to a whole new literature using new predictors and new methods to produce more accurate forecasts. In continuing with the work of McGurk (2015) (Ch. 2) further tests the added predictability of international capital inflows as predictors in US housing inflation forecasting. In general, these results continue with this story, a number of region specific capital inflow are able to produce more accurate forecasts of US housing inflation.

The new data from BEA allows for a further study of the effect of capital inflows on US housing inflation. My results provide additional empirical evidence that capital inflows have an important relationship with the US macroeconomy. Further these results provide evidence, only speculated in the international finance theory, that oil producing countries are likely to be important to US housing inflation, given the gains in forecast accuracy from the middle east and non- EU Europe.

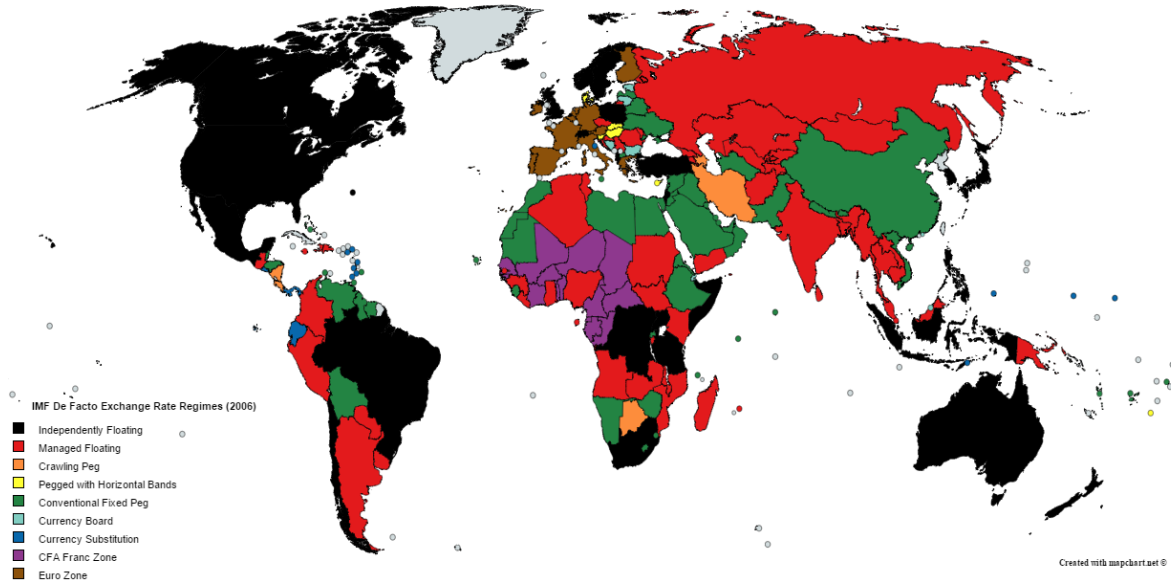
These results should help policy makers and other financial institutions in determining holdings of housing assets and policy. Specifically, due to banks owning housing assets, using capital inflows in housing inflation forecasting may be useful in predicting future market developments. In the case of the Federal Reserve, its capital movements and more accurate forecasts of housing inflation is useful for determining what current policy should be. It also may be useful for understanding the overall risk of from the mortgage-backed securities that the Federal Reserve and other large banks are holding. For the FDIC, this information is useful in preparing for future financial instability.

Although the new BEA bilateral capital flow data is useful in studying how well capital inflows predict US housing inflation, there are some are flaws. Specifically, this data is only available from 2003 onward. The true forecast ability may not necessarily be known. Another problem is that most of the data set takes place during the housing bubble and burst, Great Recession and its aftermath. This relationship may just be one in times of financial stress and recession and recover. As the time series evolves, the true relationship may be able to be determined.

Further works can run with this simple exercise and apply this to other countries. The role of capital inflows may certainly be stronger in Euro Area countries. Where for most countries there is no limit to how long a foreigner may work or live. Also due to the restructuring of the data, there are 30 or so individual country bilateral capital flow data. This country level data is not as complete as the regional level data. With a longer time series country level study may deepen the evidence that capital inflows, particularly from countries with managed exchange rate regimes, are important in forecasting US housing inflation.

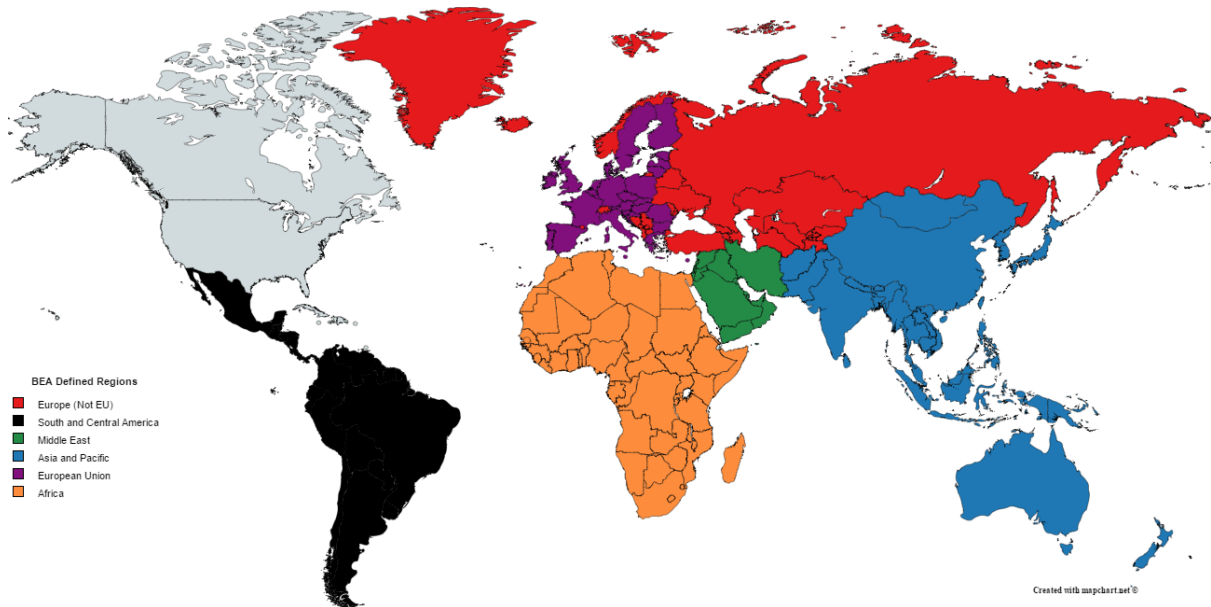
3.11 Tables and Figures

Figure 3.1: IMF De Facto Exchange Rate Regimes



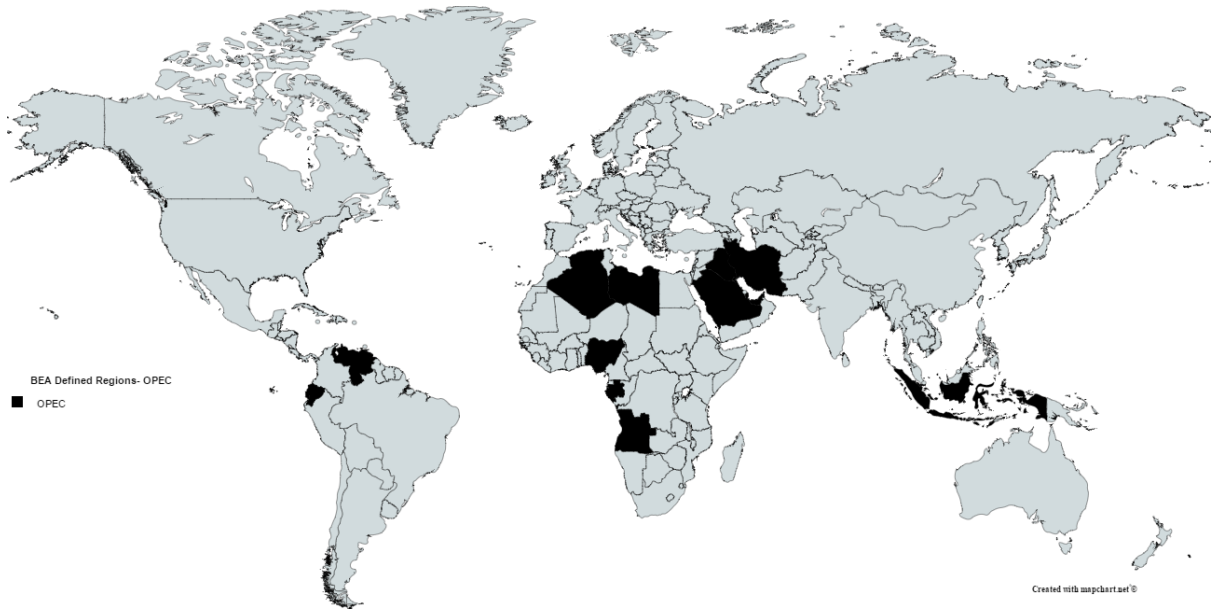
Note: Figure 3.1 shows all countries where the IMF has defined exchange rate regimes

Figure 3.2: BEA Defined Regions



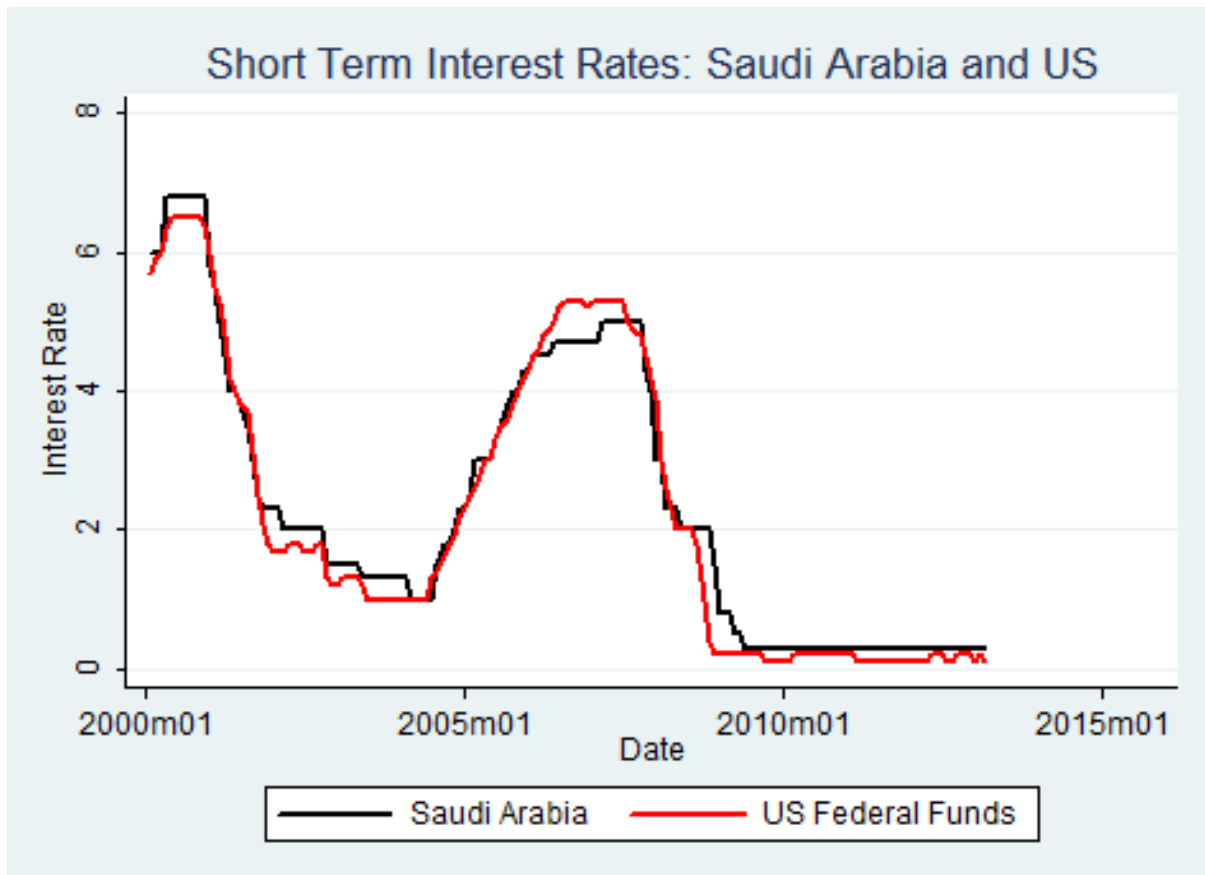
Note: Figure 3.2 shows regional areas defined by the BEA. As a note, the BEA defines regions by not only geography but historical association. As such a number of countries located in Asia are considered part of Europe.

Figure 3.3: BEA Defined Regions-OPEC



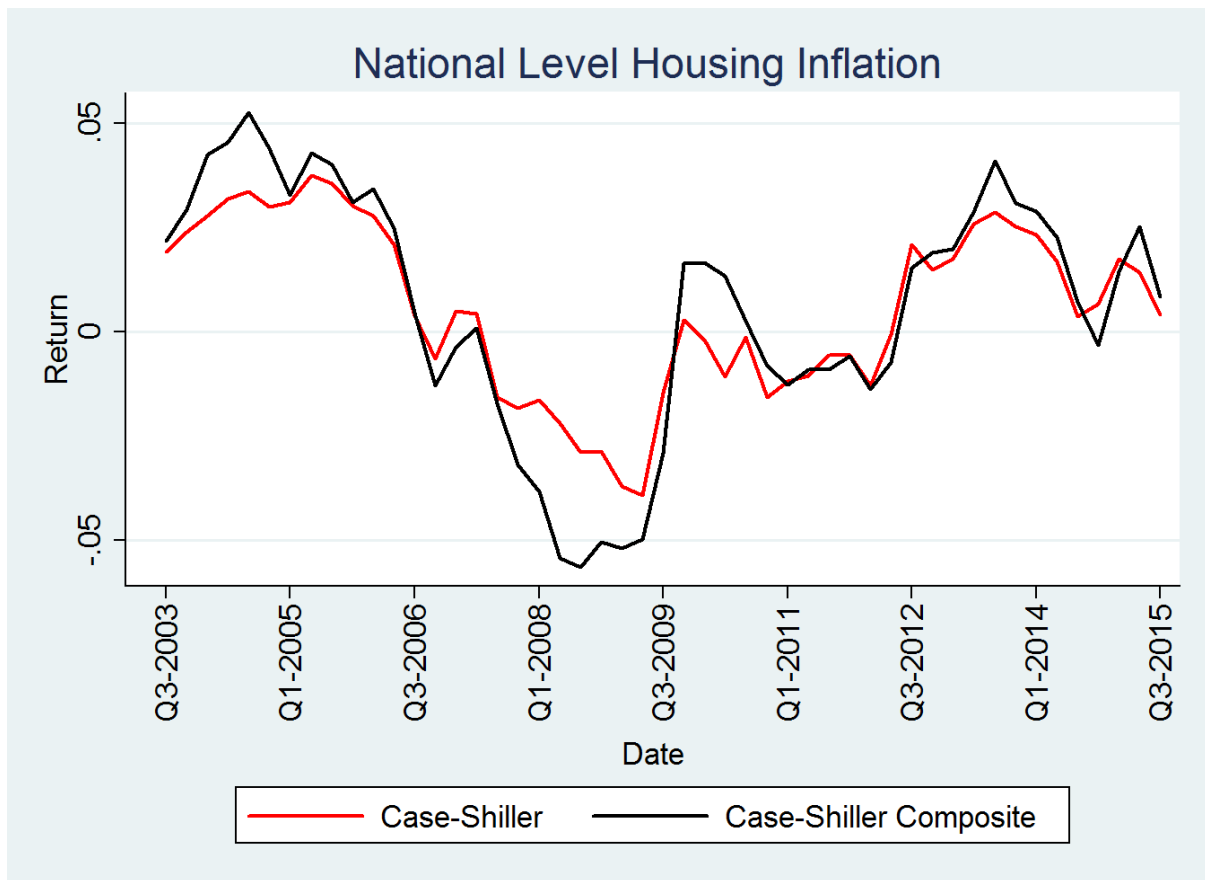
Note: Figure 3.3 shows all countries which are considered members of OPEC during the sample period

Figure 3.4: Short Term Interest Rates: Saudi Arabia and US



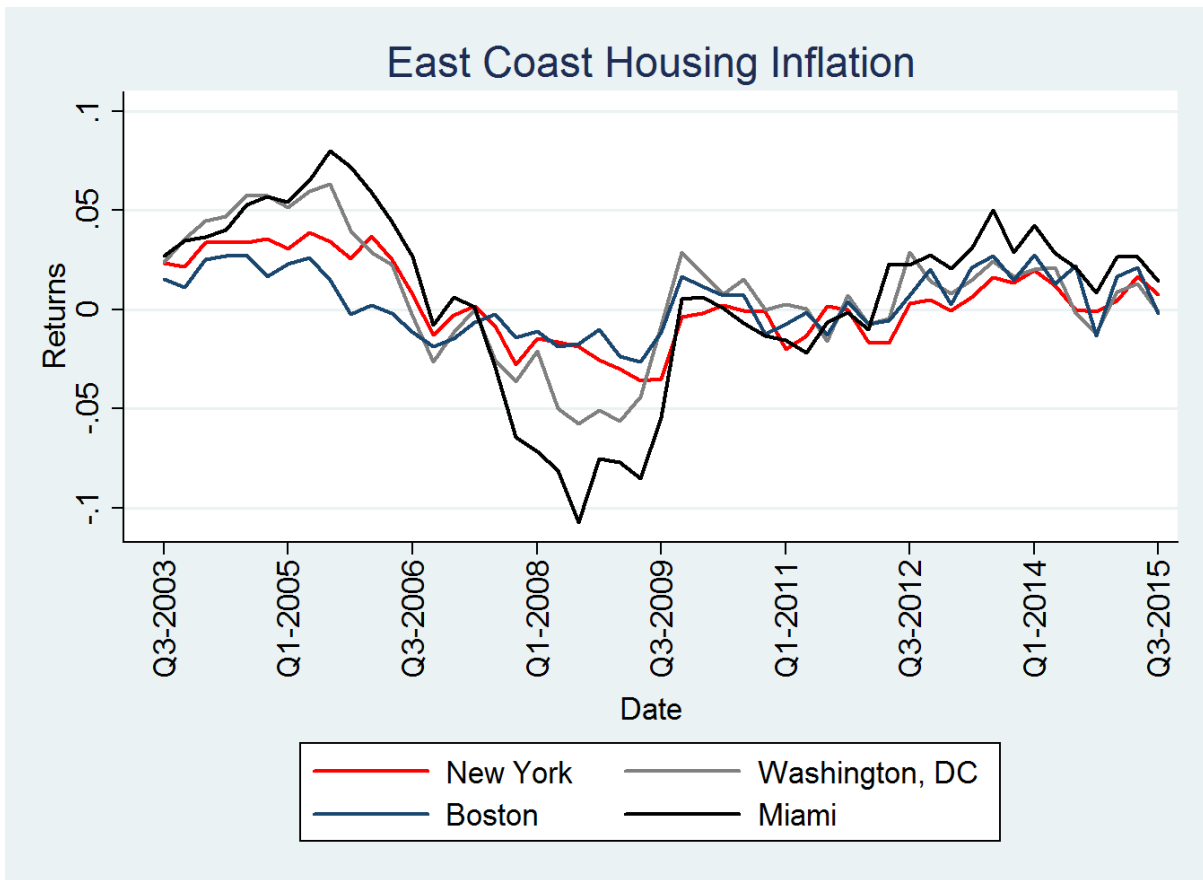
Note: Figure 3.4 shows the discount rate for Saudi Arabia and the US Federal Funds rate from M2:2000-M3:2013.

Figure 3.5: US National Level Housing Inflation



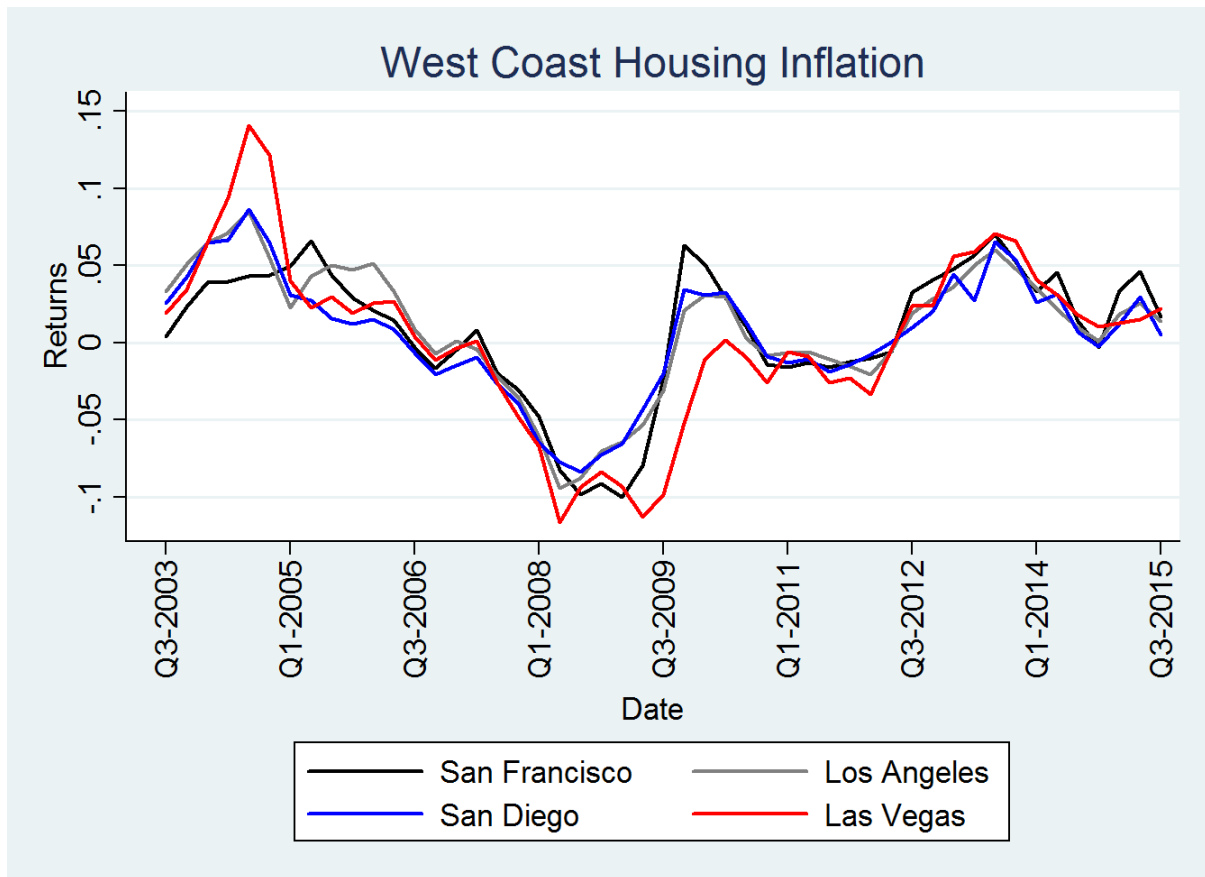
Note: Figure 3.5 shows the movement of the Case-Shiller and Case-Shiller Composite housing price inflation measure from Q2:2003-Q3:2015.

Figure 3.6: East Coast Metropolitan Area Housing Inflation



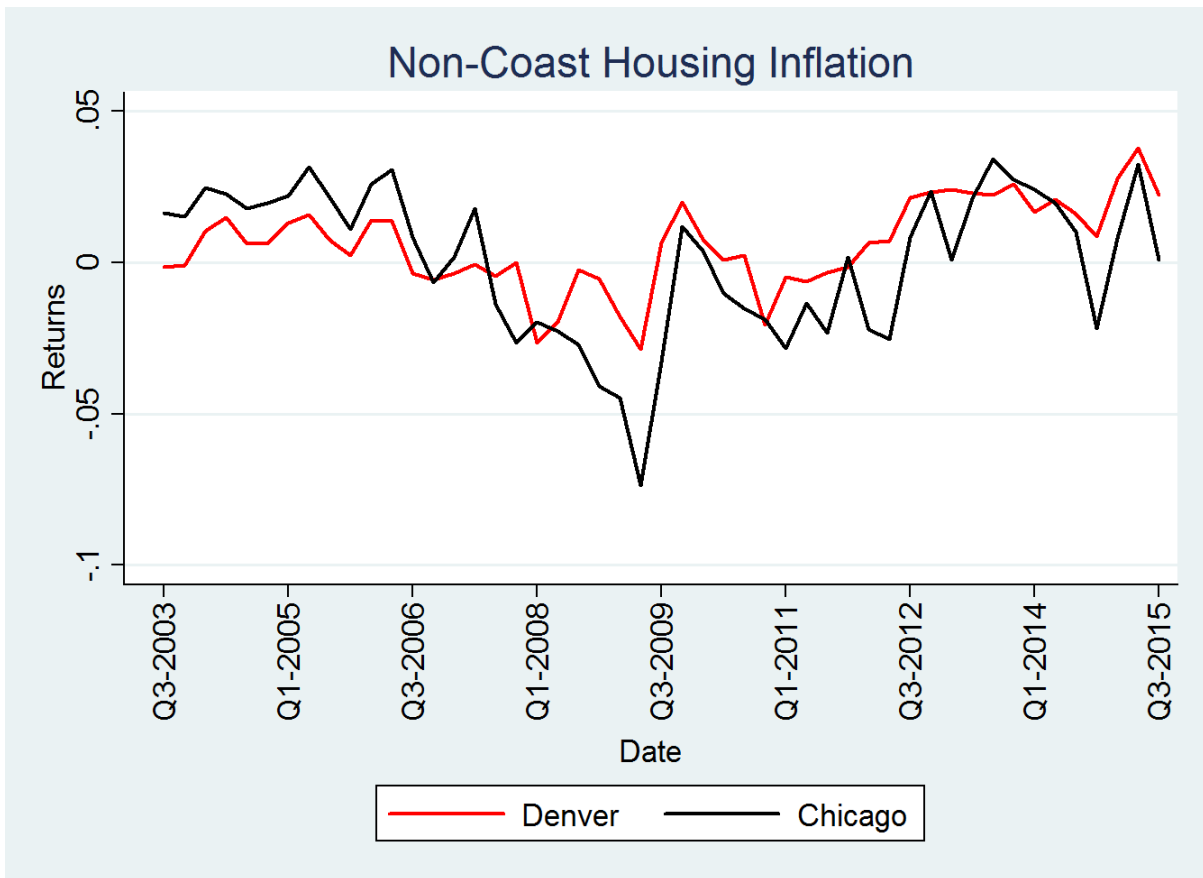
Note: Figure 3.6 shows the movement of the Case-Shiller metropolitan area housing inflation for broadly defined east coast cities. These include Miami, Washington, DC, New York and Boston. These range from Q2:2003-Q3:2015.

Figure 3.7: West Coast Metropolitan Area Housing Inflation



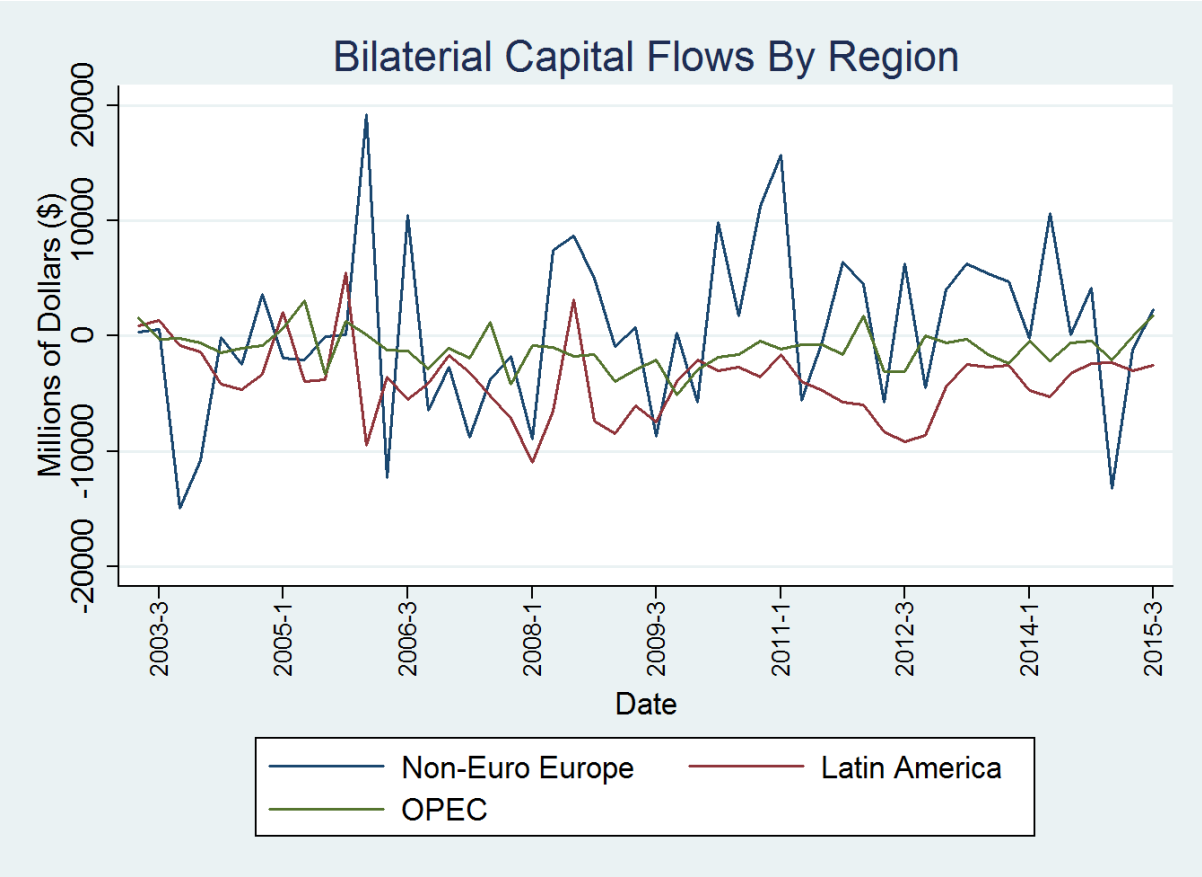
Note: Figure 3.7 shows the movement of the Case-Shiller metropolitan area housing inflation for broadly defined west coast cities. These include San Diego, Los Angeles, Las Vegas, and San Francisco. These range from Q2:2003-Q3:2015.

Figure 3.8: Non-Coast Metropolitan Area Housing Inflation



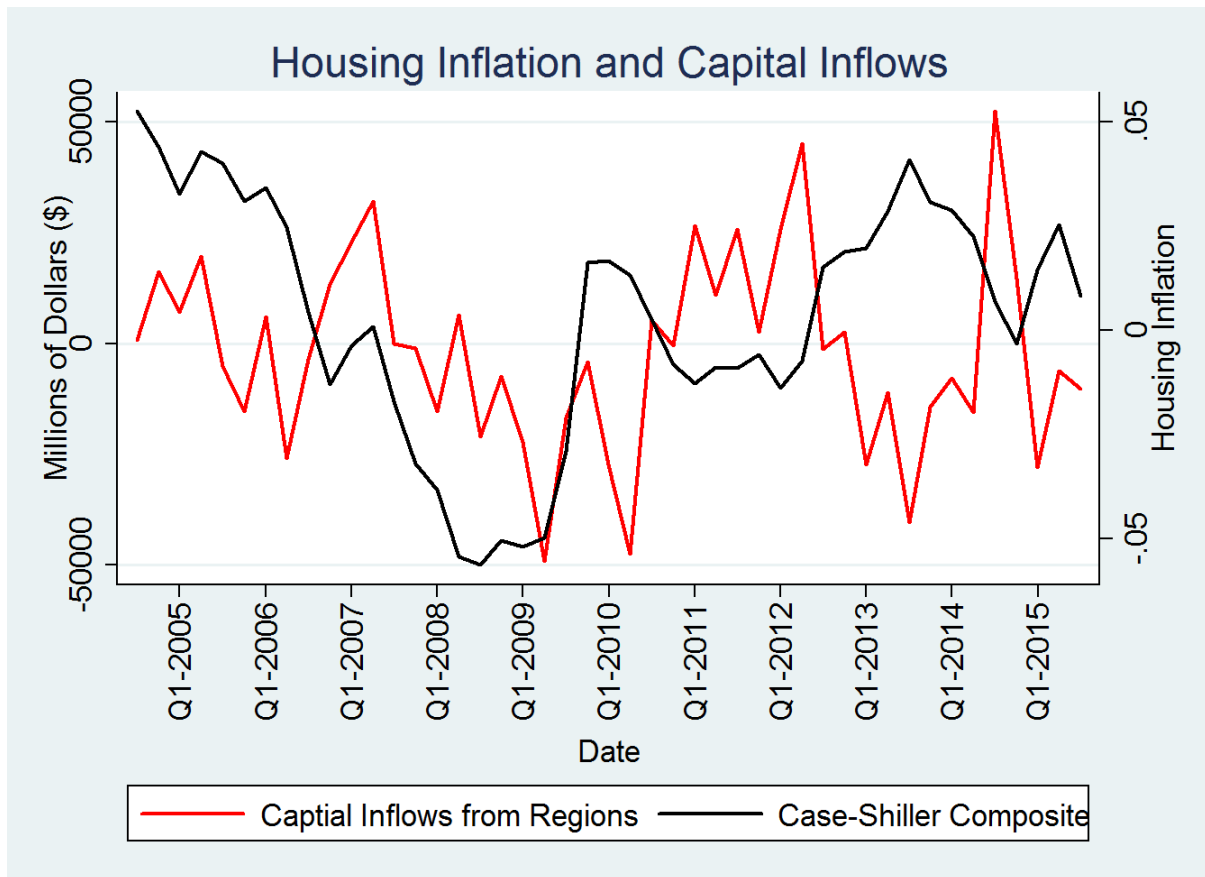
Note: Figure 3.8 shows the movement of the Case-Shiller metropolitan area housing inflation for broadly defined non coastal cities. These include Chicago and Denver. These range from Q2:2003-Q3:2015.

Figure 3.9: Regional Bilateral Capital Inflow



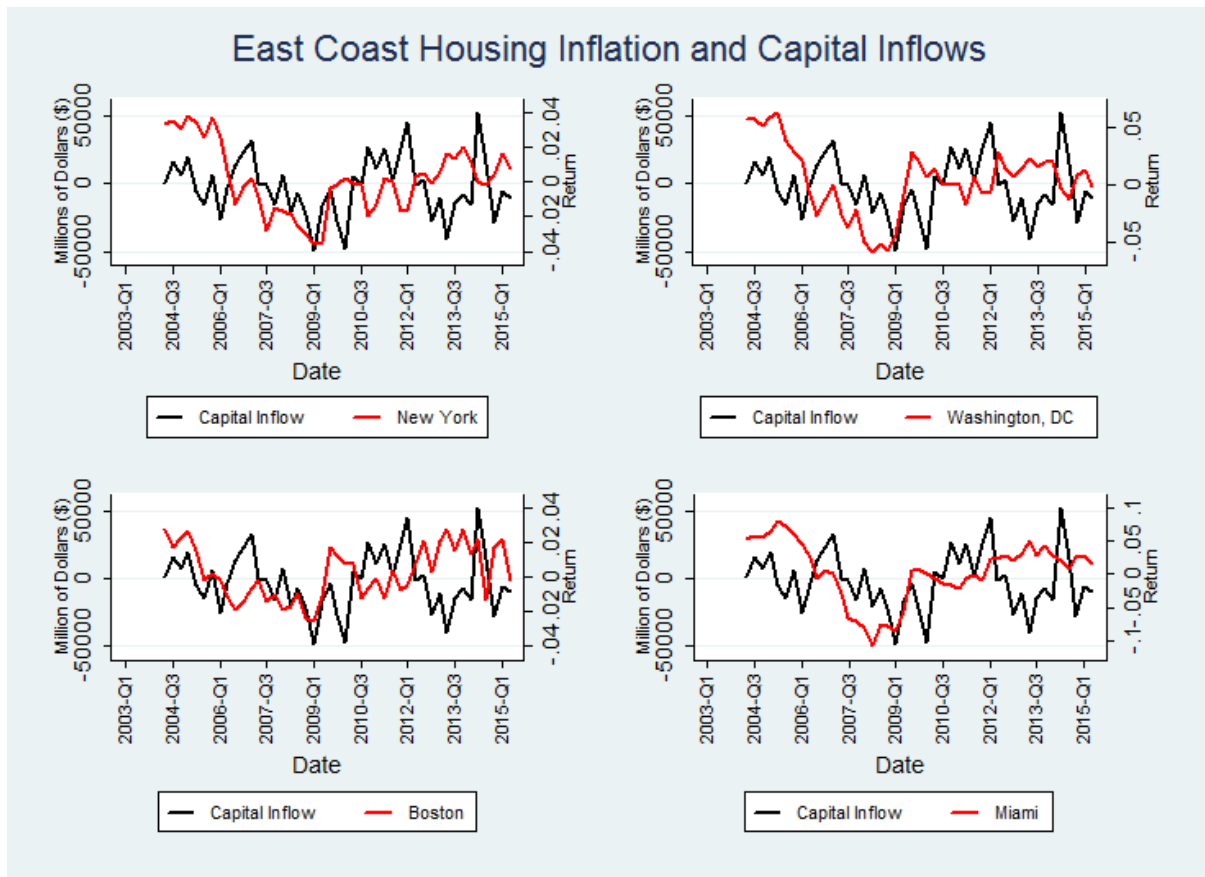
Note: Figure 3.9 shows the overall capital inflow from two regions, Non-Euro Europe and Latin America, and from the international organization, OPEC. These range from Q2:2003-Q3:2015.

Figure 3.10: National Level Housing Inflation and 4th Lag of Capital Inflows



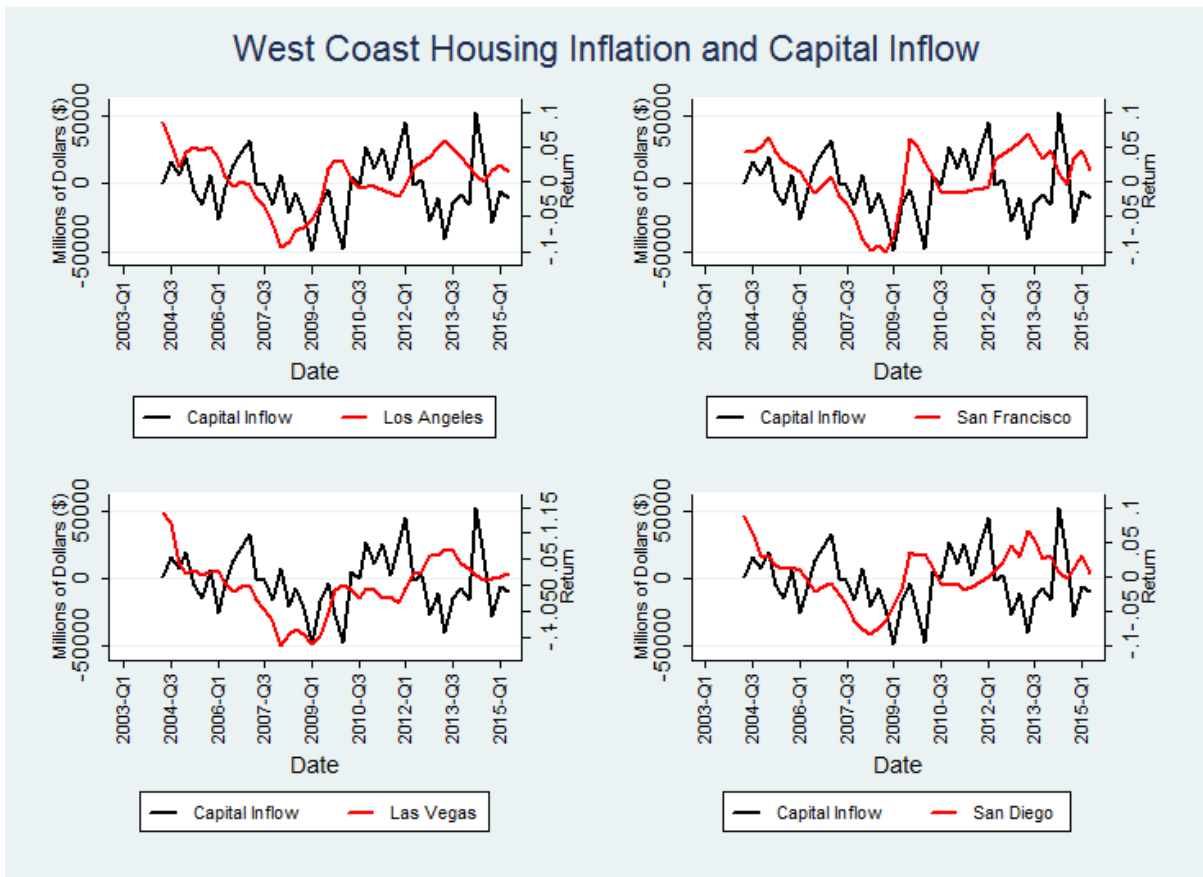
Note: Figure 3.10 shows the 4th lag of the sum of Latin America, Asia and Pacific, and Non-Euro Europe, labeled Capital Inflows from Regions and the current Case-Shiller Composite housing inflation. These range from Q2:2003-Q3:2015.

Figure 3.11: East Coast Housing Inflation and 4th Lag of Capital Inflows



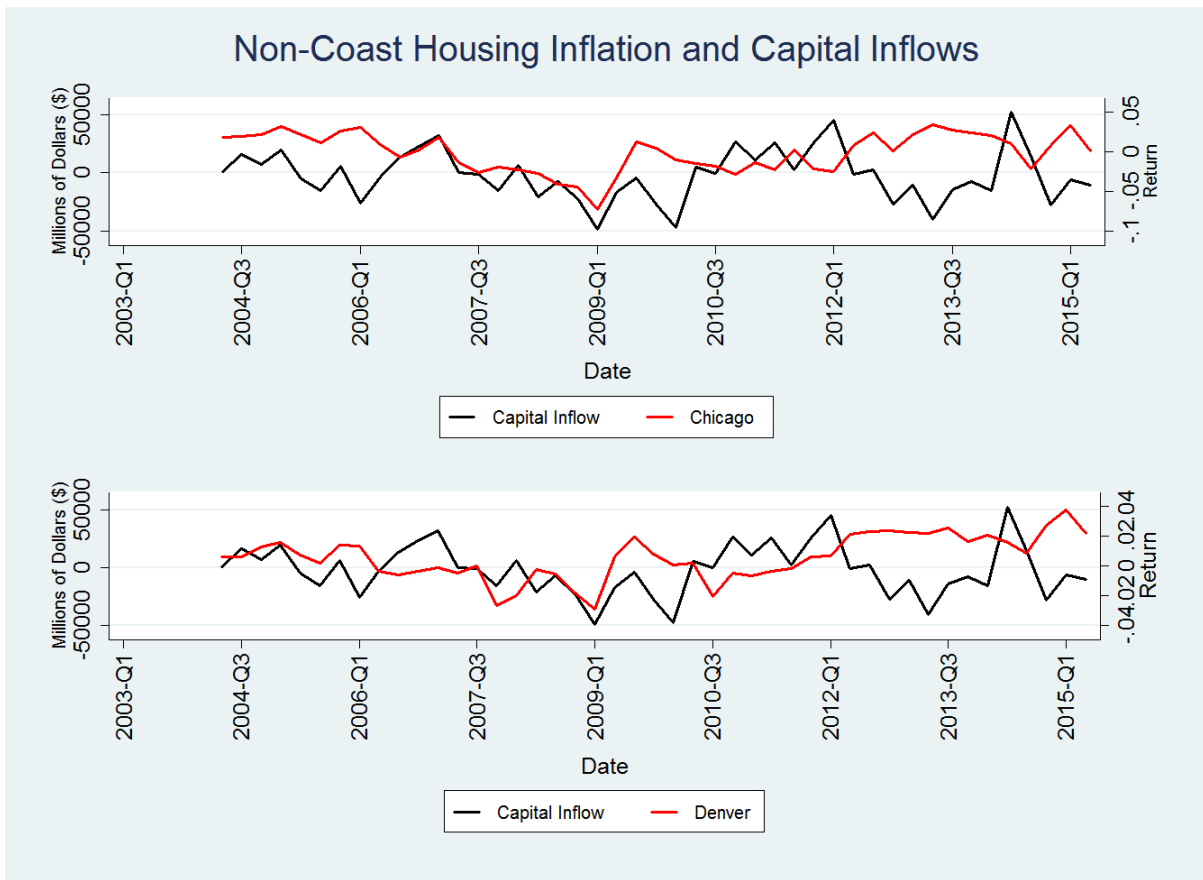
Note: Figure 3.11 shows the 4th lag of the sum of Latin America, Asia and Pacific, and Non-Euro Europe, labeled Capital Inflows from Regions and the Case-Shiller metropolitan area housing inflation for East Coast cities. These range from Q2:2003-Q3:2015.

Figure 3.12: West Coast Housing Inflation and 4th Lag of Capital Inflows



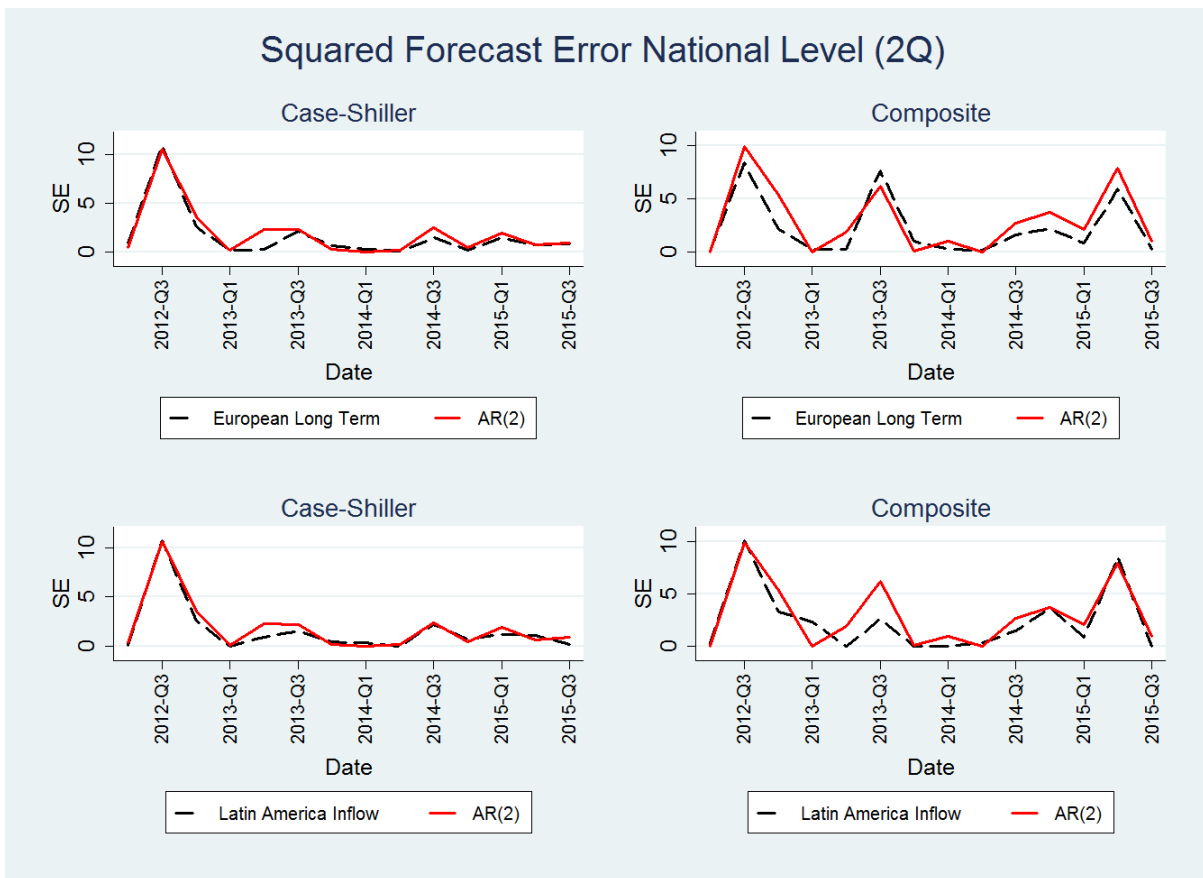
Note: Figure 3.12 shows the 4th lag of the sum of Latin America, Asia and Pacific, and Non-Euro Europe, labeled Capital Inflows from Regions and the Case-Shiller metropolitan area housing inflation for West Coast cities. These range from Q2:2003-Q3:2015.

Figure 3.13: Non-Coast Housing Inflation and 4th Lag of Capital Inflows



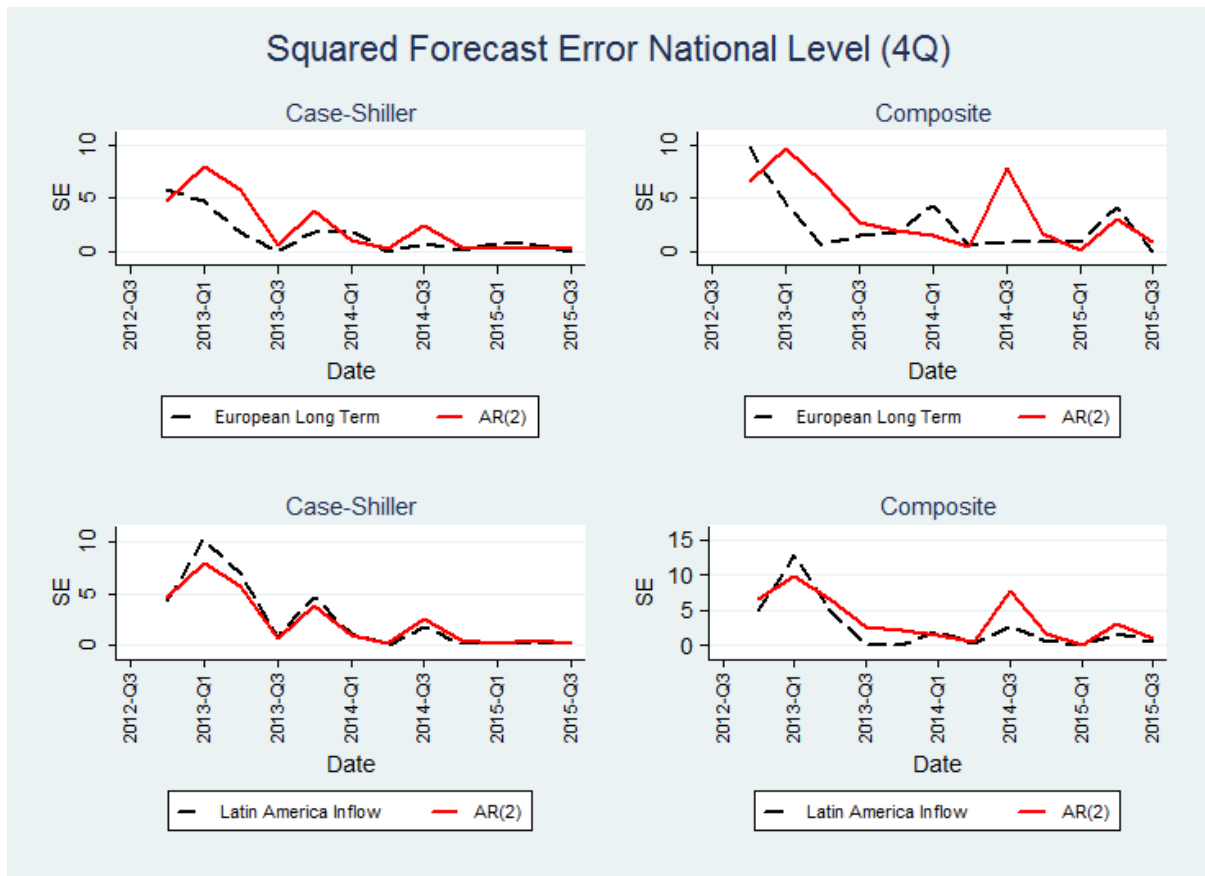
Note: Figure 3.13 shows the 4th lag of the sum of Latin America, Asia and Pacific, and Non-Euro Europe, labeled Capital Inflows from Regions and the Case-Shiller metropolitan area housing inflation for Non-Coast cities. These range from Q2:2003-Q3:2015.

Figure 3.14: Squared Forecasting Errors: AR(2) and Capital Inflows (2Q)



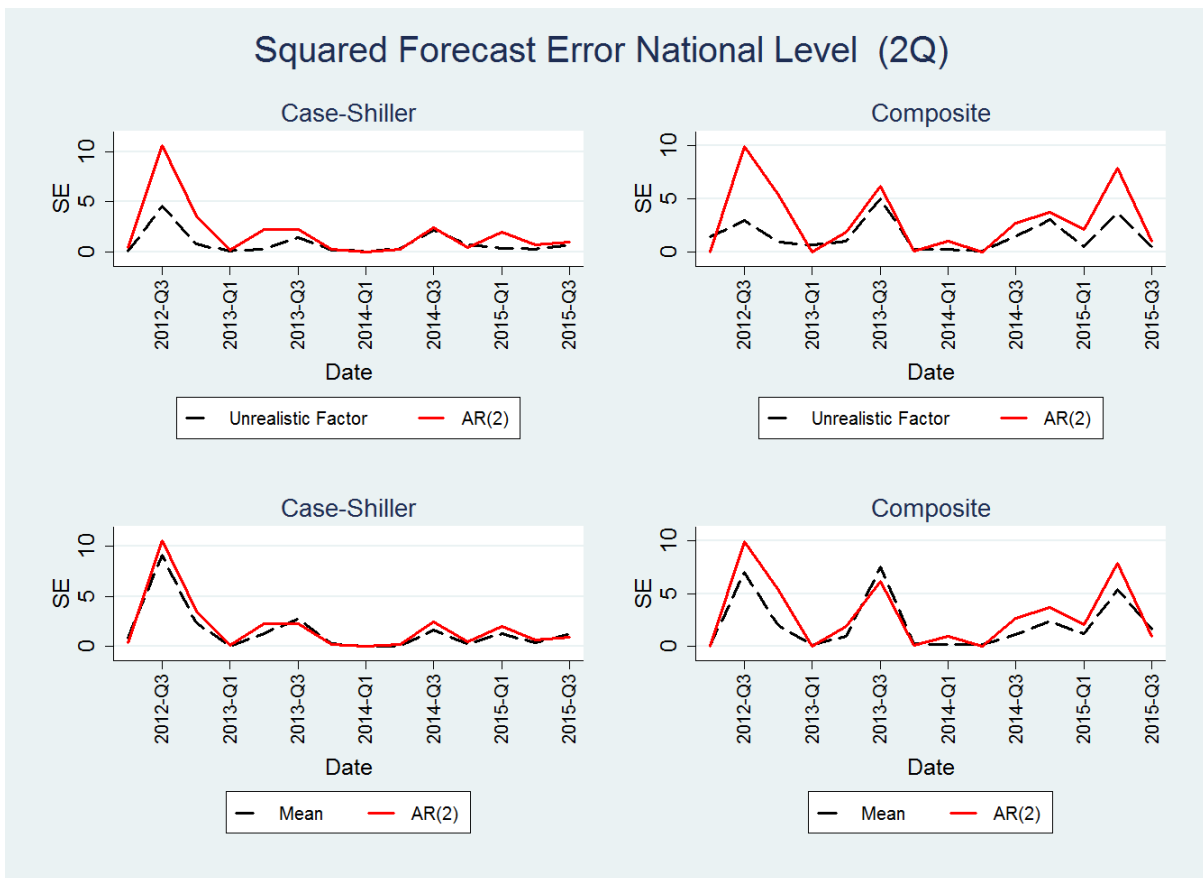
Note: Figure 3.14 shows the 2 quarter forecast horizon out-of-sample squared forecasting errors for the AR(2) model and the non-Euro Europe Long Term inflows and Overall Latin America capital inflow models. The left side shows the results for the Case-Shiller measure, while the left side shows the results for the Case-Shiller Composite. These range from Q2:2012-Q3:2015.

Figure 3.15: Squared Forecasting Errors: AR(2) and Capital Inflows (4Q)



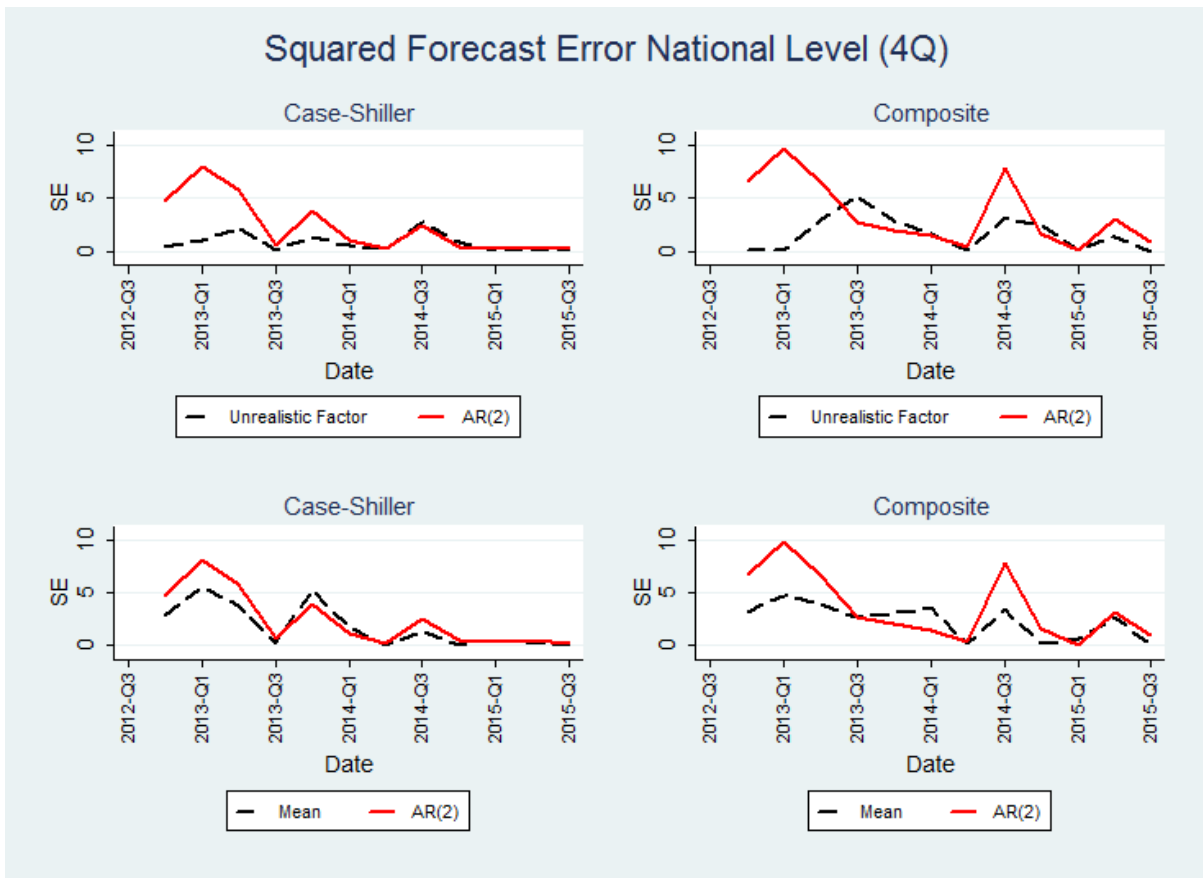
Note: Figure 3.15 shows the 4 quarter forecast horizon out-of-sample squared forecasting errors for the AR(4) model and the non-Euro Europe Long Term inflows and Overall Latin America capital inflow models. The left side shows the results for the Case-Shiller measure, while the left side shows the results for the Case-Shiller Composite. These range from Q4:2012-Q3:2015.

Figure 3.16: Squared Forecasting Errors: AR(2) and Combination Methods (2Q)



Note: Figure 3.16 shows the 2 quarter forecast horizon out-of-sample squared forecasting errors for the AR(2) model and mean forecast combination results and the first unrealistic factor model. The left side shows the results for the Case-Shiller measure, while the left side shows the results for the Case-Shiller Composite. These range from Q2:2012-Q3:2015.

Figure 3.17: Squared Forecasting Errors: AR(2) and Combination Methods (4Q)



Note: Figure 3.17 shows the 4 quarter forecast horizon out-of-sample squared forecasting errors for the AR(2) model and mean forecast combination results and the first unrealistic factor model. The left side shows the results for the Case-Shiller measure, while the left side shows the results for the Case-Shiller Composite. These range from Q4:2012-Q3:2015.

Table 3.1: Descriptive Statistics

	Mean	Std Dev	Corr:Case-Shiller	Corr:Composite
Case-Shiller	0.568	2.066	1	0.95
Composite	0.593	2.931	0.95	1
NEE Long Term	527.174	19591.122	0.241	0.293
NEE Other Investment	-319.311	15406.843	0.17	0.178
LA Inflow	-3986.369	3165.191	0.047	-0.019
AP Investment Asset	11706.022	9268.762	0.047	-0.023
ME Asset	1235.597	3916.625	-0.141	-0.201
ME Direct Investment	6415.428	10764.392	-0.031	0.005
ME Equity Shares	5384.612	8292.166	0.036	0.056
ME Short Term	4498.97	6052.519	0.155	0.218
ME Long Term	1875.428	4814.186	0.14	0.125
OPEC Equity Shares	5738.458	7868.958	-0.034	-0.012

Note: Table 3.1 shows the descriptive statistics for the Case-Shiller national level housing price inflation measures, and selected regional capital inflows. The first column shows the mean. The second column the standard deviation. The third and fourth column show their correlation of the 4th lag of the predictors with the two national level housing price inflation measures and the correlation between the two housing inflation measures.

Table 3.2: In-sample National Level Predictability Test

	2 Quarter Horizon		4 Quarter Horizon	
	Case-Shiller	Composite	Case-Shiller	Composite
NEE Long Term	2.887**	2.577*	3.685**	3.714**
NEE Other Investment	0.504	0.386	1.029	1.361
LA Inflow	1.667	1.929	0.344	0.6
AP Investment Asset	0.171	0.529	0.201	0.074
ME Asset	6.026***	8.698***	5.616***	10.674***
ME Direct Investment	2.101	2.738**	2.786**	3.001**
ME Equity Shares	2.142*	3.444**	3.432**	3.781**
ME Short Term	1.815	2.23*	3.295**	3.733**
ME Long Term	0.92	1.496	1.629	2.42*
OPEC Equity Shares	2.443*	3.627**	2.588*	3.441**

Note: Table 2.2 shows the national level in-sample results for bilateral capital inflow models and two forecast horizons. The numbers represent the F-stat. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 3.3: In-sample Metropolitan Level Predictability Test(2Q)

	SFCase	LACase	NYCase	SDCase	DenCase	LVCCase	MICase	DCCase	CHICase	BOSCase
NEE Long Term	3.669**	1.521	2.09	2.977**	0.815	2.574*	2.275*	3.133**	2.408*	1.868
NEE Other Investment	0.54	0.559	0.37	1.445	0.826	1.326	0.596	0.158	0.84	0.271
LA Inflow	1.094	1.941	1.125	1.208	2.756**	1.404	1.454	0.425	0.986	3.028**
AP Investment Asset	0.841	0.36	0.234	0.527	0.505	0.432	0.117	0.714	0.821	1.429
ME Asset	7.619***	5.701***	5.137***	7.81***	4.397***	2.326*	6.461***	7.024***	6.595***	4.648***
ME Direct Investment	2.603*	1.962	2.039	2.959**	1.526	0.765	0.899	2.41*	3.329**	4.685***
ME Equity Shares	3.882***	2.097	3.161**	2.936**	2.087	1.021	0.888	2.761**	4.319***	5.727***
ME Short Term	3.703**	1.299	2.209*	1.82	3.656**	1.22	1.493	1.887	3.222**	2.692**
ME Long Term	1.111	2.729**	0.537	2.425*	2.148*	1.267	1.665	1.121	0.65	0.682
OPEC Equity Shares	3.791**	2.136*	3.593**	2.517*	2.466*	1.227	1.245	3.653**	3.369**	5.658***

Note: Table 3.3 shows the 2 quarter forecast horizon metropolitan level in-sample results for selected bilateral capital inflow models. The numbers represent the F-stat. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 3.4: In-sample Metropolitan Level Predictability Test (4Q)

	SFCase	LACase	NYCase	SDCase	DenCase	LVCCase	MICase	DCCase	CHICase	BOSCase
NEE Long Term	4.72***	3.389**	3.085**	4.478***	1.021	10.232***	3.013**	4.175***	3.385**	2.595*
NEE Other Investment	1.769	2.132*	0.407	2.425*	0.24	4.595***	1.839	1.215	1.043	0.548
LA Inflow	0.211	0.694	0.819	0.714	0.46	1.076	0.298	0.096	0.471	1.576
AP Investment Asset	0.323	0.102	0.205	0.149	0.093	0.162	0.094	0.349	0.036	0.699
ME Asset	7.64***	7.561***	5.499***	7.222***	5.042***	3.85**	7.408***	6.808***	6.979***	2.986**
ME Direct Investment	2.449*	2.235*	1.459	2.251*	2.157*	1.065	1.055	1.878	2.012	4.643***
ME Equity Shares	4.892***	2.268*	2.095	2.413*	2.847**	0.618	1.801	2.548*	2.358*	3.682**
ME Short Term	4.537***	1.833	3.11**	1.846	5.286***	1.782	2.706**	2.167*	2.864**	2.646**
ME Long Term	1.723	3.639**	1.51	2.737**	2.97**	2.045	1.828	1.569	1.345	0.914
OPEC Equity Shares	3.254**	2.2*	1.723	1.885	2.419*	0.416	1.604	2.941**	1.163	3.026**

Note: Table 3.4 shows the 4 quarter forecast horizon metropolitan level in-sample results for selected bilateral capital inflow models. The numbers represent the F-stat. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 3.5: National level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance

	Forecast Horizon	2 Quarter		4 Quarter	
		Case-Shiller	Composite	Case-Shiller	Composite
NEE Long Term	MSFE Ratio	0.857	0.738	0.653	0.696
	MSFE F-stat	2.338***	4.972***	6.386***	5.253***
NEE Other Investment	MSFE Ratio	0.937	0.969	0.845	0.901
	MSFE F-stat	0.944	0.441	2.196**	1.322*
LA Inflow	MSFE Ratio	0.856	0.803	1.093	0.698
	MSFE F-stat	2.355***	3.44***	-1.026	5.18***
AP Investment Asset	MSFE Ratio	0.975	0.961	0.986	1.068
	MSFE F-stat	0.356	0.574	0.171	-0.761
ME Asset	MSFE Ratio	0.974	1.246	1.308	1.912
	MSFE F-stat	0.379	-2.762	-2.826	-5.725
ME Direct Investment	MSFE Ratio	0.947	0.954	0.797	0.917
	MSFE F-stat	0.784	0.681	3.052***	1.079
ME Equity Shares	MSFE Ratio	0.969	0.817	0.642	0.463
	MSFE F-stat	0.45	3.146***	6.693***	13.908***
ME Short Term	MSFE Ratio	1.233	1.333	1.279	1.523
	MSFE F-stat	-2.644	-3.499	-2.621	-4.122
ME Long Term	MSFE Ratio	1.292	1.322	1.417	1.637
	MSFE F-stat	-3.166	-3.414	-3.531	-4.671
OPEC Equity Shares	MSFE Ratio	0.973	0.87	0.743	0.573
	MSFE F-stat	0.381	2.099**	4.143***	8.949***

Note: Table 3.5 shows selected national level out-of-sample results for exchange rates and the net change in financial accounts (NFA) and the three forecast horizons. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 3.6: Metropolitan Level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance (2Q)

		SFC _{Case}	LAC _{Case}	NYC _{Case}	SDC _{Case}	Den _{Case}	LVC _{Case}	MIC _{Case}	DCC _{Case}	CHIC _{Case}	BOSC _{Case}
NEE Long Term	MSFE Ratio	0.614	0.768	1.06	0.761	1.011	0.814	1.201	1.085	0.842	0.889
	MSFE F-stat	8.817***	4.222***	-0.791	4.398***	-0.157	3.191***	-2.35	-1.095	2.625***	1.742***
NEE Other Investment	MSFE Ratio	0.965	1.04	1.057	0.8	1.031	1.019	1.13	1.111	0.993	1.038
	MSFE F-stat	0.507	-0.54	-0.749	3.509***	-0.42	-0.255	-1.57	-1.397	0.094	-0.511
LA Inflow	MSFE Ratio	0.932	0.614	0.888	0.757	0.95	0.622	1.07	1.176	1.128	0.842
	MSFE F-stat	1.029	8.795***	1.761***	4.504***	0.743	8.49***	-0.88	-2.092	-1.585	2.627***
AP Investment Asset	MSFE Ratio	0.965	1.023	0.956	1.005	0.949	1.078	0.94	0.907	0.978	0.968
	MSFE F-stat	0.513	-0.31	0.649	-0.074	0.755	-1.007	0.93	1.428**	0.315	0.467
ME Asset	MSFE Ratio	1.419	1.959	0.886	1.114	1.136	2.123	1.46	1.046	0.848	0.971
	MSFE F-stat	-4.137	-6.852	1.8***	-1.435	-1.676	-7.405	-4.42	-0.616	2.509***	0.412
ME Direct Investment	MSFE Ratio	0.856	0.939	1.384	0.835	1.15	1.199	1.29	1.211	0.703	0.977
	MSFE F-stat	2.346***	0.905	-3.882	2.762***	-1.822	-2.326	-3.17	-2.439	5.91***	0.327
ME Equity Shares	MSFE Ratio	0.783	0.975	1.181	0.785	1.036	1.103	1.34	0.995	0.723	0.877
	MSFE F-stat	3.881***	0.365	-2.145	3.829***	-0.483	-1.31	-3.58	0.074	5.374***	1.96***
ME Short Term	MSFE Ratio	1.737	1.579	1.199	1.26	0.646	1.322	1.84	1.616	1.136	1.047
	MSFE F-stat	-5.938	-5.133	-2.325	-2.89	7.66***	-3.406	-6.38	-5.335	-1.678	-0.627
ME Long Term	MSFE Ratio	1.309	1.521	1.281	1.115	1.02	1.768	1.64	1.888	1.22	1.202
	MSFE F-stat	-3.302	-4.794	-3.071	-1.445	-0.271	-6.083	-5.46	-6.584	-2.528	-2.348
OPEC Equity Shares	MSFE Ratio	0.764	1.111	1.114	0.891	1.149	1.188	1.28	0.954	0.806	0.996
	MSFE F-stat	4.331***	-1.395	-1.432	1.705***	-1.811	-2.215	-3.06	0.677	3.38***	0.055

Note: Table 3.6 shows selected 2 quarter forecast horizon metropolitan area out-of-sample results for Bilateral Capital Inflow. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 3.7: Metropolitan Level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance (4Q)

		SFC _{Case}	LAC _{Case}	NYC _{Case}	SDC _{Case}	Den _{Case}	LVC _{Case}	MIC _{Case}	DCC _{Case}	CHIC _{Case}	BOSC _{Case}
NEE Long Term	MSFE Ratio	1.01	0.734	0.671	0.683	1.011	0.789	1.025	0.783	0.822	0.933
	MSFE F-stat	-0.124	4.346***	5.884***	5.572***	-0.128	3.214***	-0.29	3.319***	2.603***	0.866
NEE Other Investment	MSFE Ratio	1.015	0.972	0.961	0.622	0.973	0.808	1.23	1.065	1.065	1.021
	MSFE F-stat	-0.178	0.349	0.481	7.296***	0.327	2.851***	-2.241	-0.727	-0.728	-0.245
LA Inflow	MSFE Ratio	1.164	0.556	0.782	0.645	1.225	0.649	1.354	0.873	1.298	0.843
	MSFE F-stat	-1.689	9.583***	3.341***	6.609***	-2.201	6.477***	-3.139	1.743**	-2.758	2.24**
AP Investment Asset	MSFE Ratio	1.24	1.201	1.007	1.121	1.012	1.224	0.966	0.982	1.03	0.958
	MSFE F-stat	-2.323	-2.005	-0.085	-1.294	-0.14	-2.195	0.418	0.225	-0.347	0.525
ME Asset	MSFE Ratio	3.603	2.838	1.519	1.638	0.932	1.944	1.809	2.447	1.54	1.269
	MSFE F-stat	-8.67	-7.771	-4.099	-4.672	0.88	-5.827	-5.365	-7.096	-4.205	-2.545
ME Direct Investment	MSFE Ratio	1.657	1.057	1.333	1.004	0.961	0.93	1.645	2.444	0.907	0.894
	MSFE F-stat	-4.758	-0.646	-3.001	-0.053	0.486	0.899	-4.704	-7.09	1.234*	1.416*
ME Equity Shares	MSFE Ratio	0.808	0.62	0.874	0.668	0.661	0.833	1.074	1.219	0.687	0.83
	MSFE F-stat	2.86***	7.36***	1.725**	5.961***	6.159***	2.411***	-0.825	-2.156	5.457***	2.46***
ME Short Term	MSFE Ratio	2.195	1.549	1.302	1.33	0.895	1.14	2.286	2.188	1.334	1.081
	MSFE F-stat	-6.534	-4.254	-2.783	-2.975	1.406*	-1.477	-6.751	-6.517	-3.005	-0.901
ME Long Term	MSFE Ratio	1.986	2.072	1.242	1.292	0.976	1.749	1.555	2.297	1.432	1.142
	MSFE F-stat	-5.957	-6.208	-2.341	-2.713	0.291	-5.141	-4.284	-6.776	-3.618	-1.489
OPEC Equity Shares	MSFE Ratio	0.779	0.747	0.998	0.8	0.776	0.912	1.091	1.206	0.877	0.922
	MSFE F-stat	3.399***	4.073***	0.025	2.994***	3.457***	1.156	-1	-2.054	1.677**	1.017

Note: Table 3.7 shows selected 4 quarter forecast horizon metropolitan area out-of-sample results for Bilateral Capital Inflow. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 3.8: National level: Out-of-Sample Combination Methods Forecasting Performance

	Forecast Horizon	2 Quarter		4 Quarter	
		Case-Shiller	Composite	Case-Shiller	Composite
F1A	MSFE Ratio	0.451	0.514	0.337	0.465
	MSFE F-stat	17.024***	13.258**	23.659***	13.826**
F2A	MSFE Ratio	1.618	1.73	2.179	2.37
	MSFE F-stat	-5.348	-5.906	-6.492	-6.936
F3A	MSFE Ratio	1.178	1.306	0.53	0.638
	MSFE F-stat	-2.112	-3.284	10.629**	6.822***
F1	MSFE Ratio	0.991	0.964	0.988	1.177
	MSFE F-stat	0.127	0.529	0.141	-1.807
F2	MSFE Ratio	1.108	1.094	1.795	2.418
	MSFE F-stat	-1.367	-1.204	-5.314	-7.037
F3	MSFE Ratio	0.988	1.024	0.762	0.948
	MSFE F-stat	0.175	-0.323	3.755***	0.657
Mean	MSFE Ratio	0.822	0.72	0.759	0.652
	MSFE F-stat	3.035***	5.434***	3.82***	6.391***
Trim-Mean	MSFE Ratio	0.83	0.748	0.759	0.608
	MSFE F-stat	2.877***	4.706***	3.814***	7.732***
Median	MSFE Ratio	0.825	0.756	0.733	0.586
	MSFE F-stat	2.977***	4.528***	4.373***	8.466***

Note: Table 3.8 shows selected national level out-of-sample results for combination methods and the two forecast horizons. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 3.9: Metropolitan Level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance (2Q)

		SFC _{Case}	LAC _{Case}	NYC _{Case}	SDC _{Case}	DenC _{Case}	LVC _{Case}	MIC _{Case}	DCC _{Case}	CHIC _{Case}	BOSC _{Case}
F1A	MSFE Ratio	0.459	0.506	0.638	0.541	0.679	0.283	0.676	0.699	0.406	0.695
	MSFE F-stat	16.515**	13.655**	7.954***	11.891*	6.608***	35.48***	6.713***	6.035***	20.453***	6.143***
F2A	MSFE Ratio	1.931	2.091	1.714	1.423	1.306	1.766	2.137	2.057	1.481	1.032
	MSFE F-stat	-6.749	-7.305	-5.834	-4.159	-3.28	-6.072	-7.448	-7.194	-4.544	-0.428
F3A	MSFE Ratio	1.44	1.216	1.899	1.178	1.243	1.158	1.134	1.151	1.3	1.33
	MSFE F-stat	-4.277	-2.487	-6.628	-2.118*	-2.738	-1.913	-1.657	-1.839	-3.231	-3.472
F1	MSFE Ratio	0.974	1.008	0.977	0.986	1.043	1.103	0.925	0.94	0.977	1.026
	MSFE F-stat	0.373	-0.105	0.331	0.195	-0.572	-1.305	1.13	0.897	0.329	-0.353
F2	MSFE Ratio	1.068	1.157	1.079	1.017	0.971	1.058	1.022	1.033	1.007	0.941
	MSFE F-stat	-0.891	-1.904	-1.031	-0.231	0.422	-0.773	-0.306	-0.442	-0.095	0.884
F3	MSFE Ratio	1.057	0.951	0.876	1.077	1.203	1.002	0.959	0.933	1.043	1.096
	MSFE F-stat	-0.754	0.727	1.989**	-0.996	-2.359**	-0.023	0.592	1.013	-0.571*	-1.23*
Mean	MSFE Ratio	0.66	0.771	0.835	0.752	0.838	0.905	1.056	0.827	0.699	0.777
	MSFE F-stat	7.212***	4.161***	2.772***	4.621***	2.716***	1.465*	-0.741	2.923***	6.017***	4.022***
Trim-Mean	MSFE Ratio	0.678	0.772	0.837	0.732	0.89	0.855	1.066	0.865	0.737	0.793
	MSFE F-stat	6.661***	4.144***	2.717***	5.127***	1.722**	2.378***	-0.866	2.19**	4.995***	3.653***
Median	MSFE Ratio	0.635	0.725	0.855	0.734	0.896	0.843	1.031	0.912	0.737	0.758
	MSFE F-stat	8.053***	5.307***	2.38***	5.079***	1.633**	2.615***	-0.426	1.355*	4.995***	4.473***

Note: Table 3.9 shows selected 2 quarter forecast horizon metropolitan area out-of-sample results for Combination Methods. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

Table 3.10: Metropolitan Level: Out-of-Sample Bilateral Capital Inflow Forecasting Performance (4Q)

		SFCCase	LACCase	NYCase	SDCase	DenCase	LVCCase	MICCase	DCCCase	CHICCase	BOSCase
F1A	MSFE Ratio	0.506	0.443	0.543	0.441	0.473	0.19	0.567	0.807	0.702	0.698
	MSFE F-stat	11.697**	15.058**	10.087**	15.24**	13.347**	51.126***	9.168***	2.864***	5.093***	5.185***
F2A	MSFE Ratio	2.539	2.3	2.1	1.732	2.215	1.7	2.189	3.223	1.925	1.227
	MSFE F-stat	-7.273	-6.783	-6.287	-5.071	-6.583	-4.942	-6.517	-8.277	-5.766	-2.221
F3A	MSFE Ratio	0.802	0.6	0.994	0.842	0.567	0.832	0.941	1.984	0.777	0.818
	MSFE F-stat	2.963***	8.012***	0.076	2.256**	9.164***	2.425***	0.753	-5.953	3.444***	2.662***
F1	MSFE Ratio	1.461	1.279	1.143	0.971	1.137	1.509	1.246	1.026	1.264	1.135
	MSFE F-stat	-3.786	-2.617	-1.5	0.36	-1.447	-4.046	-2.369	-0.302	-2.506	-1.423
F2	MSFE Ratio	2.597	2.083	1.994	1.538	1.822	1.924	2.733	3.279	1.765	1.541
	MSFE F-stat	-7.379	-6.24	-5.981	-4.196	-5.415	-5.762	-7.61	-8.34	-5.202	-4.211
F3	MSFE Ratio	1.659	0.799	0.786	1.178	1.349	0.863	1.489	1.128	0.744	1.256
	MSFE F-stat	-4.765	3.017***	3.263***	-1.817	-3.108	1.913**	-3.939	-1.358	4.14***	-2.444
Mean	MSFE Ratio	0.833	0.759	0.82	0.72	0.816	0.779	0.938	0.84	0.909	0.847
	MSFE F-stat	2.407***	3.815***	2.635***	4.676***	2.701***	3.41***	0.791	2.292***	1.202*	2.169**
Trim-Mean	MSFE Ratio	0.767	0.641	0.816	0.678	0.843	0.767	1.002	0.749	0.866	0.854
	MSFE F-stat	3.654***	6.727***	2.711***	5.708***	2.231**	3.636***	-0.021	4.012***	1.85**	2.053**
Median	MSFE Ratio	0.799	0.619	0.799	0.649	0.876	0.794	1.062	0.723	0.866	0.87
	MSFE F-stat	3.01***	7.374***	3.02***	6.49***	1.705**	3.116***	-0.702	4.609***	1.853**	1.787**

Note: Table 3.10 shows selected 4 quarter forecast horizon metropolitan area out-of-sample results for Combination Methods. Both the MSFE ratio and MSFE F-stats are shown. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level.

3.12 Appendix

Table 3.11: Data Source

Variable	Source
Housing Price Index	St.Louis FRED
Capital Inflow	Bureau of Economic Analysis (International Data)

Table 3.12: Predictor Definition

Predictor	Definition
Inflow	Overall net purchases of assets minus net liabilities holding.
Direct Investment	asset purchases (minus liabilities) which give some significant degree of control in an enterprise.
Portfolio	net purchases of financial assets like stocks. Owners play no role in decision making.
Equity Shares	a portion of portfolio investments, only concerning with specifically owning of equity assets.
Securities	Net purchases of US securities.
Short Term	Part of Securities and only contains shorter term securities.
Long Term	Part of Securities and only contains longer term securities.
Other Investment	Net purchases of other assets including currency, deposits, loans, insurance technical reserves, trade credit and advances and allocations of special drawing rights.
Currency	Part of Other Investment. Net Foreign holding of US currency.
Loans	Part of Other Investment. Foreign loans given to Domestic residents.
Asset	Same as Inflow but liability data are unavailable
Direct Asset	Same as Direct Investment but liability data are unavailable.
Other Asset	Same as Other Investment but liability data are unavailable.
Loan Asset	Same as Loan but liability data are unavailable.

Table 3.13: In-sample National Level Predictability Test

	2 Quarter Horizon		4 Quarter Horizon	
	Case-Shiller	Composite	Case-Shiller	Composite
NEE Asset	1.156	1.184	0.878	0.765
NEE Inflow	0.239	0.587	0.819	1.293
NEE Direct Investment	1.567	1.156	1.938	1.78
NEE Portfolio	0.439	0.342	2.54*	1.598
NEE Equity Shares	1.31	0.93	0.894	0.911
NEE Securities	0.349	0.479	0.501	0.388
NEE Short Term	2.172*	1.722	1.772	1.079
NEE Currency	2.05	1.642	1.925	1.463
NEE Loans	4.496***	3.112**	3.591**	2.796**
LA Asset	0.386	0.927	0.686	0.552
LA Direct Investment	0.925	1.471	0.536	1.613
LA Portfolio	3.656**	2.755**	2.943**	2.263*
LA Equity Shares	1.512	2.327*	1.623	3.254**
LA Securities	0.294	0.171	0.841	0.9
LA Short Term	1.638	1.984	1.739	3.905**
LA Long Term	0.667	0.607	0.197	0.245
LA Other Investment	0.303	0.443	0.97	1.375
LA Currency	1.186	1.773	0.715	0.639
LA Loans	1.423	1.489	0.963	1.194
AP Asset	0.272	0.16	0.392	0.166
AP Inflow	1.056	2.277*	1.381	2.515*
AP Direct Investment	0.228	0.317	0.585	0.239
AP Portfolio	0.531	1.083	0.372	1.171
AP Equity Shares	0.693	0.908	1.056	0.85
AP Securities	2.338*	2.502*	1.223	2.355*
AP Short Term	4.124***	5.619***	1.562	2.675**
AP Long Term	1.159	1.351	0.688	1.518
Ap Other Investment	2.5*	4.137***	1.209	3.636**
AP Currency	0.377	0.542	0.225	0.156
AP Loans	0.996	1.887	0.71	1.177
ME Direct Asset	1.057	0.96	0.611	1.013
ME Portfolio	0.737	0.588	1.677	1.146
ME Equity Shares	2.142*	3.444**	3.432**	3.781**
ME Securities	2.54*	2.534*	0.955	1.298
ME Other Asset	4.067***	7.449***	3.376**	6.477***
ME Loan Assets	4.879***	5.105***	2.714**	3.799**
ME Loans	0.396	0.386	2.003	2.133*
OPEC Asset	2.615*	4.346***	2.046	4.709***
OPEC Inflow	0.234	0.203	0.287	0.208
OPEC Direct Investment	2.198*	2.82**	2.606*	3.176**
OPEC Portfolio	0.721	0.657	2.3*	1.591
OPEC Securities	4.14***	3.465**	1.118	1.405
OPEC Short Term	1.963	2.52*	3.659**	3.672**
OPEC Long Term	0.964	2.029	2.46*	4.096***
OPEC Other Investment	0.442	0.908	1.009	1.846
OPEC Currency	0.664	1.037	0.276	0.514
OPEC Loans	0.73	0.673	1.693	2.072

Table 3.14: In-sample Metropolitan Level Predictability Test(2Q)

	SFCase	LACase	NYCase	SDCase	DenCase	LVCase	MICase	DCCase	CHICase	BOSCase
NEE Asset	0.864	0.896	0.335	1.602	0.213	1.082	1.447	2.575*	0.243	0.863
NEE Inflow	0.865	1.186	0.14	2.515*	1.683	1.661	0.886	0.177	0.422	0.649
NEE Direct Investment	2.205*	0.181	1.097	1.797	0.641	0.519	1.088	2.717**	2.071	1.229
NEE Portfolio	0.867	0.272	0.412	1.786	0.766	0.24	0.38	0.414	1.253	0.427
NEE Equity Shares	1.592	0.067	1.207	0.805	0.494	0.316	0.864	2.822**	1.979	1.707
NEE Securities	0.483	0.677	0.503	0.459	1.169	0.285	0.382	0.035	0.757	0.511
NEE Short Term	2.723**	0.319	2.169*	0.876	0.873	0.45	1.182	2.821**	3.264**	1.892
NEE Currency	2.483*	0.778	1.561	1.801	0.479	0.656	1.103	4.197***	1.119	1.669
NEE Loans	3.558**	2.052	2.82**	2.287*	5.597***	2.865**	2.967**	3.065**	3.731**	5.018***
LA Asset	1.27	0.619	0.129	1.255	0.522	0.374	1.69	2.122*	0.586	1.256
LA Direct Investment	2.851**	0.923	0.703	1.434	2.274*	1.182	0.948	1.786	1.584	2.146*
LA Portfolio	3.264**	2.314*	2.635**	3.557**	1.603	1.745	1.359	2.285*	4.721***	4.831***
LA Equity Shares	3.936***	1.864	1.634	2.874**	2.522*	2.008	1.707	2.774**	2.526*	2.564*
LA Securities	0.182	0.276	0.318	0.094	0.148	0.359	0.134	0.185	0.017	0.469
LA Short Term	3.17**	1.64	1.515	2.169*	1.358	1.284	1.214	2.245*	1.534	2.572*
LA Long Term	1.24	0.281	0.262	0.234	0.308	0.19	0.73	1.37	0.961	1.217
LA Other Investment	0.489	1.134	0.313	1.051	1.099	1.23	0.37	0.251	0.454	0.905
LA Currency	2.037	0.922	1.201	1.425	2.096	0.766	0.803	0.554	1.902	1.332
LA Loans	1.818	1.219	1.711	0.746	0.776	1.168	1.271	1.132	1.423	1.427
AP Asset	0.581	0.165	0.249	0.742	0.475	0.688	0.453	0.113	0.515	0.441
AP Inflow	1.315	3.256**	2.063	3.512**	1.589	1.243	0.631	0.584	1.557	1.321
AP Direct Investment	0.465	0.486	0.263	0.412	1.665	1.632	0.288	0.363	0.928	0.456
AP Portfolio	0.463	1.871	0.965	1.967	0.072	1.053	0.673	0.57	1.028	0.367
AP Equity Shares	0.782	1.561	0.13	1.242	2.015	3.236**	0.424	0.32	0.635	0.795
AP Securities	3.308**	1.115	2.302*	1.44	2.623*	0.707	3.605**	2.104*	1.727	2.483*
AP Short Term	3.99***	4.974***	2.662**	4.115***	9.543***	6.51***	3.022**	2.371*	3.764**	5.292***
AP Long Term	1.297	1.617	1.04	1.563	0.366	1.815	1.781	0.708	0.833	0.718
AP Other Investment	6.344***	2.447*	1.566	3.992***	3.85**	2.275*	1.816	2.19*	4.017***	4.221***
AP Currency	0.596	1.111	0.278	0.479	1.296	2.294*	0.665	0.305	0.487	0.626
AP Loans	2.845**	1.008	1.512	1.88	2.129*	0.219	1.536	1.502	3.293**	1.491
ME Direct Asset	0.524	0.815	0.898	1.056	0.785	1.01	1.683	0.954	0.744	0.649
ME Portfolio	0.628	0.343	0.18	1.405	0.353	0.202	0.445	0.444	0.632	1.026
ME Equity Shares	3.882***	2.097	3.161**	2.936**	2.087	1.021	0.888	2.761**	4.319***	5.727***
ME Securities	2.757**	1.296	4.097***	1.437	1.02	0.53	1.008	2.933**	3.225**	4.079***
ME Other Asset	8.718***	6.302***	3.54**	7.395***	4.072***	2.271*	3.551**	6.409***	4.887***	4.174***
ME Loan Assets	5.847***	3.056**	3.822**	2.46*	4.993***	1.694	2.75**	3.6**	8.126***	4.441***
ME Loans	1.674	0.171	0.274	0.935	1.62	0.365	0.289	1.382	0.827	0.499
OPEC Asset	4.245***	3.711**	4.216***	4.005***	2.453*	1.546	3.268**	3.171**	3.413**	2.387*
OPEC Inflow	0.481	0.344	0.084	0.241	0.76	0.324	0.14	0.383	0.551	0.677
OPEC Direct Investment	2.277*	2.076	2.439*	2.844**	1.854	1.356	1.237	2.638**	2.774**	4.168***
OPEC Portfolio	0.714	0.442	0.17	2.328*	0.637	0.473	0.458	0.42	0.694	1.013
OPEC Securities	4.696***	1.312	5.463***	1.399	2.068	0.472	1.948	4.861***	5.346***	5.662***
OPEC Short Term	3.289**	1.477	2.361*	2.066	3.022**	1.945	1.524	2.676**	2.325*	2.099
OPEC Long Term	3.992***	2.014	0.325	4.735***	4.018***	0.901	0.895	1.674	2.581*	2.374*
OPEC Other Investment	0.757	1.922	0.206	1.937	1.031	0.933	1.164	0.747	0.404	0.251
OPEC Currency	2.717**	0.37	1.868	0.921	2.184*	0.243	0.236	1.168	2.866**	5.096***
OPEC Loans	2.216*	0.292	0.57	0.985	1.874	0.445	0.55	2.066	1.324	0.865

Table 3.15: In-sample Metropolitan Level Predictability Test(4Q)

	SFCase	LACase	NYCase	SDCase	DenCase	LVCCase	MICase	DCCase	CHICase	BOSCase
NEE Asset	0.488	0.573	0.929	1.107	0.217	1.293	0.366	1.951	0.263	1.358
NEE Inflow	2.301*	1.143	0.212	2.688**	3.108**	2.298*	1.53	0.605	0.658	1.183
NEE Direct Investment	1.604	0.708	1.19	1.664	1.479	0.924	1.895	2.308*	1.964	1.703
NEE Portfolio	1.603	1.605	1.263	1.804	1.055	1.874	1.218	1.401	1.792	1.093
NEE Equity Shares	0.855	0.264	0.605	0.664	1.128	0.478	1.017	1.056	1.345	2.439*
NEE Securities	0.144	0.283	0.6	0.297	0.792	0.066	0.469	0.143	0.754	0.4
NEE Short Term	0.956	0.244	1.459	0.447	2.175*	0.483	0.855	0.688	2.369*	2.259*
NEE Currency	2.027	0.692	2.948**	1.351	0.547	0.822	0.625	3.478**	1.756	2.96**
NEE Loans	1.935	1.938	3.837**	1.732	3.493**	2.701**	2.842**	1.663	2.231*	4.193***
LA Asset	0.692	0.48	0.418	0.717	0.909	0.38	0.622	0.943	0.507	0.863
LA Direct Investment	2.408*	1.113	1.319	1.79	1.079	1.668	1.223	1.147	1.829	1.127
LA Portfolio	2.071	1.631	3.647**	1.644	1.073	1.29	1.168	2.078	4.901***	3.629**
LA Equity Shares	4.477***	2.324*	2.731**	3.395**	1.745	2.99**	2.153*	2.416*	3.643**	2.1
LA Securities	0.739	1.012	0.694	1.012	1.295	0.768	0.746	0.2	0.764	1.584
LA Short Term	4.656***	3.004**	3.015**	4.242***	2.026	2.294*	2.013	2.177*	3.091**	4.366***
LA Long Term	0.358	0.409	0.258	0.18	0.582	0.265	0.608	0.291	0.25	0.759
LA Other Investment	1.198	2.189*	0.415	1.495	1.655	1.947	0.913	0.323	1.037	0.686
LA Currency	0.642	0.699	0.183	1.481	2.621*	1.524	0.273	0.047	0.34	0.343
LA Loans	1.704	0.784	1.993	1.031	1.557	0.61	0.637	0.741	1.885	1.412
AP Asset	0.707	0.067	0.134	0.523	0.569	0.374	0.686	0.269	0.379	0.307
AP Inflow	0.957	3.504**	1.044	2.712**	2.569*	0.847	0.857	0.502	0.798	1.57
AP Direct Investment	0.891	0.288	0.038	0.484	1.1	1.405	0.34	0.143	0.372	0.724
AP Portfolio	0.253	2.384*	0.327	2.077	0.557	1.426	0.424	0.609	0.379	0.511
AP Equity Shares	1.462	1.115	0.2	2.071	1.716	3.169**	0.779	0.24	0.624	1.773
AP Securities	2.334*	1.154	1.61	1.483	3.421**	0.315	1.746	0.847	1.455	1.918
AP Short Term	1.615	3.002**	2.126*	3.657**	3.096**	5.563***	1.806	0.344	1.52	3.793**
AP Long Term	0.954	1.52	1.606	1.319	0.412	1.433	1.209	1.232	0.822	0.484
Ap Other Investment	3.405**	2.113	1.066	2.364*	1.732	1.882	2.282*	1.183	1.815	1.257
AP Currency	0.222	0.413	0.212	0.055	0.142	0.533	0.114	0.195	0.157	0.449
AP Loans	1.439	0.704	0.408	1.279	2.779**	0.319	0.954	0.482	0.672	1.614
ME Direct Asset	0.492	1.3	0.559	0.842	0.506	1.786	2.021	0.573	0.221	0.693
ME Portfolio	1.393	1.168	0.652	2.065	0.84	1.524	0.662	0.874	2.108	3.22**
ME Equity Shares	4.892***	2.268*	2.095	2.413*	2.847**	0.618	1.801	2.548*	2.358*	3.682**
ME Securities	1.167	0.75	2.428*	0.771	0.449	0.174	0.876	1.349	1.109	1.621
ME Other Asset	5.533***	4.309***	6.075***	4.601***	3.425**	2.308*	3.044**	4.451***	5.296***	3.258**
ME Loan Assets	3.604**	2.058	3.591**	1.827	2.391*	1.319	2.529*	1.945	4.455***	2.261*
ME Loans	4.697***	1.409	0.416	2.976**	3.619**	1.002	1.303	1.677	1.62	1.439
OPEC Asset	3.201**	3.665**	3.81**	2.658**	2.166*	1.209	2.894**	2.972**	3.272**	1.379
OPEC Inflow	0.496	0.245	0.208	0.266	0.429	0.168	0.166	0.091	0.769	0.252
OPEC Direct Investment	2.092	2.685**	1.555	2.274*	2.157*	1.117	1.117	2	1.339	4.026***
OPEC Portfolio	1.976	1.634	0.997	2.795**	1.144	2.076	0.69	1.162	3.234**	4.156***
OPEC Securities	1.6	0.577	2.507*	0.595	0.871	0.274	0.923	1.691	0.952	1.852
OPEC Short Term	2.886**	1.951	2.87**	1.793	5.927***	1.965	3.038**	1.902	2.1	1.66
OPEC Long Term	5.673***	4.052***	1.813	5.719***	4.305***	3.198**	1.844	2.275*	2.564*	1.27
OPEC Other Investment	1.578	3.04**	0.676	2.297*	1.286	1.779	1.261	1.223	1.073	0.306
OPEC Currency	1.783	0.336	0.406	0.804	0.678	0.217	0.185	0.656	0.554	2.117*
OPEC Loans	4.693***	1.202	0.653	2.417*	3.586**	0.506	1.207	1.618	1.625	1.559

Table 3.16: Out-of-Sample National Level Predictability Test

	2 Quarter Horizon		4 Quarter Horizon	
	Case-Shiller	Composite	Case-Shiller	Composite
NEE Asset	1.017	0.94	1.15	1.298
NEE Inflow	1.174	1.245	1.161	1.369
NEE Direct Investment	1.242	1.154	1.467	1.546
NEE Portfolio	3.281	3.665	3.97	6.096
NEE Equity Shares	1.379	1.24	1.334	1.347
NEE Securities	1.017	1.086	0.899	0.983
NEE Short Term	1.14	1.036	0.967	1.05
NEE Currency	1.124	0.976	1.059	1.092
NEE Loans	1.038	1.22	1.156	1.334
LA Asset	1.389	1.348	1.349	1.77
LA Direct Investment	1.368	1.366	1.504	1.939
LA Portfolio	1.12	0.973	1.522	1.9
LA Equity Shares	1.225	1.295	1.305	1.551
LA Securities	1.143	1.291	0.98	1.254
LA Short Term	1.422	1.68	1.234	1.89
LA Long Term	1.826	1.777	1.62	2.138
LA Other Investment	1.849	1.804	2.254	2.332
LA Currency	1.325	1.526	1.078	1.406
LA Loans	1.726	1.726	1.96	2.269
AP Asset	1.153	1.38	1.216	1.378
AP Inflow	1.121	1.183	0.912	0.778
AP Direct Investment	1.946	2.276	1.649	2.92
AP Portfolio	1.024	0.997	1.005	0.947
AP Equity Shares	2.766	2.706	2.05	2.313
AP Securities	1.382	1.596	1.079	1.328
AP Short Term	0.787	0.892	0.973	0.829
AP Long Term	1.242	1.293	1.142	1.055
Ap Other Investment	1.094	1.081	1.144	1.437
AP Currency	1.502	1.777	1.376	1.594
AP Loans	1.389	1.31	1.101	1.364
ME Direct Asset	1.745	2.03	1.886	2.655
ME Portfolio	1.501	1.26	2.264	2.285
ME Equity Shares	0.969	0.817	0.642	0.463
ME Securities	1.143	1.101	1.261	1.427
ME Other Asset	0.924	0.882	1.014	1.098
ME Loan Assets	0.881	0.837	0.99	0.959
ME Loans	2.15	2.215	1.452	1.965
OPEC Asset	1.449	1.578	1.468	1.916
OPEC Inflow	1.181	1.217	1.146	1.31
OPEC Direct Investment	1.015	1.104	0.846	1.067
OPEC Portfolio	1.349	1.232	1.625	1.626
OPEC Securities	1.163	1.132	1.324	1.403
OPEC Short Term	1.114	1.109	1.183	1.333
OPEC Long Term	1.36	1.43	1.28	1.754
OPEC Other Investment	1.359	1.466	1.482	1.8
OPEC Currency	1.078	1.068	1.067	1.163
OPEC Loans	2.08	2.096	1.449	1.892

Table 3.17: Out-of-Sample (MSFE Ratio) Metropolitan Level (2Q)

	SFCCase	LACCase	NYCase	SDCase	DenCase	LVCCase	MICCase	DCCCase	CHICCase	BOSCase
NEE Asset	0.853	1.204	1.212	0.906	1.177	1.373	1.187	1.126	1.2	1.044
NEE Inflow	1.354	1.402	1.175	1.161	0.944	1.614	1.322	1.249	1.102	1.011
NEE Direct Investment	1.324	1.238	1.442	0.911	1.373	1.446	1.195	1.117	1.558	1.199
NEE Portfolio	4.208	4.249	3.04	2.73	2.108	5.918	8.156	7.243	2.399	2.827
NEE Equity Shares	1.375	1.289	1.476	1.132	1.353	1.702	1.339	1.103	1.291	1.09
NEE Securities	1.256	1.216	0.959	1.118	0.872	1.321	1.409	1.14	0.906	0.922
NEE Short Term	1.035	1.017	1.302	1.054	1.251	1.286	1.039	0.794	1.101	0.943
NEE Currency	0.901	1.064	1.263	0.873	1.124	1.134	1.12	1.264	1.154	1.033
NEE Loans	1.218	1.223	1.393	0.964	0.79	1.165	1.207	1.399	1.2	0.815
LA Asset	1.563	1.51	1.775	1.098	1.269	1.165	1.476	1.363	1.381	1.175
LA Direct Investment	1.4	1.465	1.335	1.478	1.259	1.463	1.583	1.588	1.316	1.185
LA Portfolio	1.363	1.356	1.456	0.743	1.28	1.639	1.829	1.747	0.817	0.674
LA Equity Shares	1.537	1.505	1.188	1.282	1.064	1.512	1.637	1.517	1.219	1.048
LA Securities	1.468	1.286	1.261	1.347	1.115	1.35	1.392	1.397	1.28	1.177
LA Short Term	2.023	1.921	1.427	1.564	1.376	1.913	2.083	2.066	1.473	1.156
LA Long Term	1.886	2.41	1.624	1.376	1.898	2.95	3.894	1.554	1.544	1.216
LA Other Investment	2.04	2.062	1.447	1.221	1.431	2.259	3.25	2.187	1.679	1.33
LA Currency	1.793	1.885	1.533	1.507	1.223	2.34	2.371	2.247	1.469	1.301
LA Loans	2.007	1.896	1.797	1.516	1.584	2.568	2.164	1.872	2.014	1.335
AP Asset	1.976	1.553	1.163	1.087	0.983	1.523	1.669	1.549	1.399	1.092
AP Inflow	1.32	1.451	1.04	0.801	1.082	2.059	2.105	1.579	1.182	1.159
AP Direct Investment	1.697	3.735	1.887	1.557	0.969	2.582	3.043	3.164	1.416	1.335
AP Portfolio	1.169	1.194	0.826	0.998	1.138	1.216	1.762	1.467	1.019	1.06
AP Equity Shares	2.969	3.405	3.064	1.447	1.423	2.811	3.248	5.158	1.754	1.935
AP Securities	1.933	1.526	1.469	1.272	1.149	1.358	1.481	1.748	1.557	1.331
AP Short Term	1.102	0.763	0.903	0.901	0.385	1.082	1.142	1.477	0.906	0.615
AP Long Term	1.288	1.291	1.1	1.188	1.348	1.722	1.758	1.424	1.15	1.319
AP Other Investment	1.324	1.312	1.082	1.073	1.017	1.647	1.706	1.275	1.134	1.037
AP Currency	1.733	1.618	1.658	1.629	1.937	2.13	1.596	1.616	1.512	1.883
AP Loans	1.263	1.507	1.904	1.05	1.465	1.558	1.683	1.092	1.392	1.423
ME Direct Asset	2.157	2.865	1.771	1.492	1.658	3.486	2.83	1.734	1.842	1.454
ME Portfolio	1.519	1.472	2.469	1.243	1.982	3.894	1.166	2.191	2.63	1.366
ME Equity Shares	0.783	0.975	1.181	0.785	1.036	1.103	1.344	0.995	0.723	0.877
ME Securities	1.284	1.32	1.196	1.067	1.257	1.322	1.518	1.141	1	0.947
ME Other Asset	1.012	1.144	0.726	0.865	0.832	1.412	1.312	1.081	0.704	0.816
ME Loan Assets	0.894	0.871	0.793	1.035	0.962	1.26	1.074	0.88	0.818	0.848
ME Loans	2.792	2.407	2.328	1.274	1.872	2.222	2.4	2.822	2.534	2.092
OPEC Asset	1.674	1.887	1.166	1.238	1.792	1.995	1.798	1.551	1.221	1.192
OPEC Inflow	1.454	1.133	1.299	1.178	1.213	1.293	1.24	1.442	1.312	1.238
OPEC Direct Investment	0.879	1.172	1.326	0.927	1.22	1.363	1.344	1.416	0.877	1.119
OPEC Portfolio	1.182	1.627	2.177	1.122	1.634	2.938	2.082	1.837	2.081	1.455
OPEC Securities	1.352	1.383	1.111	1.101	1.236	1.521	1.609	1.189	1.053	0.915
OPEC Short Term	1.288	1.305	1.116	1.089	0.776	1.22	1.52	1.29	1.043	1.014
OPEC Long Term	1.574	1.992	1.447	1.048	0.808	2.229	1.877	1.7	1.109	1.046
OPEC Other Investment	1.466	1.88	1.513	1.075	1.199	1.966	1.75	1.759	1.279	1.226
OPEC Currency	1.165	1.228	1.015	1.185	0.983	1.327	1.198	1.031	0.962	0.977
OPEC Loans	2.479	2.224	2.227	1.311	1.869	2.178	2.209	2.334	2.477	2.035

Table 3.18: Out-of-Sample (MSFE Ratio) Metropolitan Level(4Q)

	SFCCase	LACCase	NYCase	SDCase	DenCase	LVCCase	MICCase	DCCCase	CHICCase	BOSCase
NEE Asset	1.933	1.698	1.148	1.174	1.088	1.697	1.494	1.196	1.187	1.085
NEE Inflow	1.664	1.641	1.149	1.368	0.771	1.627	1.502	1.49	1.309	0.954
NEE Direct Investment	2.306	1.512	1.527	1.233	1.638	1.615	1.589	2.153	1.559	1.299
NEE Portfolio	9.884	5.527	5.675	5.357	3.479	5.367	13.201	18.073	5.508	4.028
NEE Equity Shares	1.857	1.488	1.638	1.416	1.469	1.958	1.379	1.426	1.297	1.018
NEE Securities	1.382	1.266	0.82	1.211	0.96	1.319	1.234	1.32	0.84	1.001
NEE Short Term	1.231	0.978	1.416	1.006	1.179	1.19	1.006	0.842	1.054	0.863
NEE Currency	1.601	1.281	1.018	1.008	1.096	1.29	1.292	1.199	1.043	1.056
NEE Loans	2.069	1.453	0.815	1.154	1.062	1.178	1.395	1.234	1.037	0.745
LA Asset	2.354	2.327	1.944	1.074	1.111	1.355	1.627	2.239	1.473	1.285
LA Direct Investment	2.462	2.167	1.571	1.617	1.249	2.057	1.742	2.003	1.672	1.148
LA Portfolio	3.321	3.016	1.765	1.871	1.582	3.068	1.574	1.412	1.562	1.039
LA Equity Shares	1.896	1.791	1.243	1.267	1.071	1.802	1.459	1.825	1.433	0.998
LA Securities	1.479	1.182	1.284	1.237	1.129	1.103	1.533	1.528	1.247	0.989
LA Short Term	2.659	2.116	1.499	1.411	1.109	1.956	1.816	2.568	1.55	0.87
LA Long Term	3.094	2.718	1.733	1.725	1.506	2.63	3.328	1.536	1.733	1.169
LA Other Investment	3.365	2.759	1.637	1.772	1.811	2.284	3.338	2.93	1.8	1.453
LA Currency	1.92	1.56	1.597	1.405	0.698	1.49	1.936	2.064	1.295	1.272
LA Loans	3.138	2.795	2.17	1.955	1.302	3.24	2.215	2.47	2.438	1.531
AP Asset	1.712	1.66	1.205	1.04	1.176	1.533	1.579	1.62	1.351	1.153
AP Inflow	1.201	0.877	0.848	0.685	0.696	1.812	1.822	1.367	1.188	0.947
AP Direct Investment	2.604	3.89	2.703	2.077	1.111	2.624	2.448	4.277	1.923	1.416
AP Portfolio	1.567	0.948	1.148	0.855	0.93	1.18	1.426	1.293	1.311	1.094
AP Equity Shares	3.435	2.278	4.034	1.113	1.26	1.88	1.577	4.603	2.646	1.623
AP Securities	1.77	1.408	1.5	1.118	0.779	1.275	1.503	1.479	1.201	1.062
AP Short Term	1.521	0.885	0.746	0.6	0.827	0.696	0.63	0.982	1.028	0.585
AP Long Term	1.832	1.426	0.842	1.414	1.114	1.731	1.594	1.06	1.166	1.185
AP Other Investment	2.406	1.848	1.133	1.337	1.138	1.699	1.925	1.181	1.045	1.112
AP Currency	2.497	1.781	1.079	1.821	1.506	1.982	1.632	1.093	1.329	1.273
AP Loans	1.622	1.732	1.527	1.538	0.798	1.541	1.565	1.016	1.279	0.893
ME Direct Asset	3.967	3.415	2.181	2.34	1.575	3.513	3.276	2.977	1.821	1.319
ME Portfolio	3.336	2.437	2.155	1.718	2.305	1.663	2.414	6.068	2.288	1.174
ME Equity Shares	0.808	0.62	0.874	0.668	0.661	0.833	1.074	1.219	0.687	0.83
ME Securities	2.038	1.564	1.528	1.129	1.304	1.546	1.134	1.816	1.335	1.312
ME Other Asset	1.917	1.366	1.022	1.148	0.708	1.342	1.131	1.502	1.238	1.033
ME Loan Assets	1.351	0.98	0.989	1.134	0.951	1.083	1.061	1.334	0.988	0.98
ME Loans	3.685	2.433	2.938	1.111	0.948	2.007	2.132	2.223	1.943	1.384
OPEC Asset	2.764	2.047	1.802	1.822	1.13	1.907	1.747	2.58	1.833	1.425
OPEC Inflow	1.679	1.464	1.289	1.498	1.235	1.458	1.899	1.203	1.294	1.153
OPEC Direct Investment	1.611	1.163	1.134	0.996	0.989	0.926	1.642	3.047	0.978	0.956
OPEC Portfolio	2.399	1.861	1.645	1.356	2.081	1.16	2.488	4.287	1.595	1.053
OPEC Securities	2.001	1.615	1.64	1.204	1.262	1.692	1.406	1.722	1.424	1.229
OPEC Short Term	1.631	1.352	1.301	1.304	0.916	1.095	2.368	1.861	1.29	1.158
OPEC Long Term	2.193	2.328	1.454	1.34	0.864	1.846	2.088	2.861	1.73	1.419
OPEC Other Investment	2.099	2.194	1.443	1.346	1.162	1.845	1.931	2.343	1.551	1.244
OPEC Currency	1.47	1.34	1.13	1.181	0.946	1.348	1.306	1.307	1.127	1.038
OPEC Loans	3.731	2.47	2.873	1.235	1.012	2.13	1.91	1.883	1.896	1.368

Bibliography

- Joshua Aizenman and Yothin Jinjarak. Current account patterns and national real estate markets. *Journal of Urban Economics*, 66(2):75–89, 2009.
- Christopher Avery and Judith Chevalier. Identifying investor sentiment from price paths: The case of football betting (digest summary). *Journal of Business*, 72(4):493–520, 1999.
- Malcolm Baker and Jeffrey Wurgler. Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680, 2006.
- Malcolm Baker and Jeffrey Wurgler. Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2):129–152, 2007.
- Malcolm Baker, Jeffrey Wurgler, and Yu Yuan. Global, local, and contagious investor sentiment. *Journal of Financial Economics*, 104(2):272–287, 2012.
- Mehmet Balcilar, Rangan Gupta, and Stephen M Miller. The out-of-sample forecasting performance of nonlinear models of regional housing prices in the us. *Applied Economics*, 2015.
- Eli Bartov, Lucile Faurel, and Partha S Mohanram. Can twitter help predict firm-level earnings and stock returns? *Available at SSRN 2631421*, 2015.
- Ben Bernanke. The global savings glut and the u.s. current account deficit. *Remarks at the Homer Jones Lecture, St. Louis, MO.*, 2005.
- John Y Campbell and Robert J Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228, 1988.
- John Y Campbell and Tuomo Vuolteenaho. Bad beta, good beta. *American Economic Review*, pages 1249–1275, 2004.
- Sean D Campbell, Morris A Davis, Joshua Gallin, and Robert F Martin. What moves housing markets: A variance decomposition of the rent–price ratio. *Journal of Urban Economics*, 66(2):90–102, 2009.

- Karl Case and Robert Shiller. The efficiency of the market for single-family homes. *American Economic Review*, 79(1):125–37, 1989.
- Karl E Case and Robert J Shiller. Prices of single family homes since 1970: New indexes for four cities, 1987.
- Ambrogio Cesa-Bianchi, Luis Felipe Céspedes, and Alessandro Rebucci. Global liquidity, house prices, and the macroeconomy: Evidence from advanced and emerging economies. *Journal of Money, Credit and Banking*, 47(S1):301–335, 2015.
- Hailiang Chen, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, 27(5):1367–1403, 2014.
- Menzie D Chinn and Yi Zhang. Uncovered interest parity and monetary policy near and far from the zero lower bound. Technical report, National Bureau of Economic Research, 2015.
- San-Lin Chung, Chi-Hsiou Hung, and Chung-Ying Yeh. When does investor sentiment predict stock returns? *Journal of Empirical Finance*, 19(2):217–240, 2012.
- Todd Clark, Michael McCracken, et al. Advances in forecast evaluation. *Handbook of Economic Forecasting*, 2:1107–1201, 2013.
- Todd E Clark. Can out-of-sample forecast comparisons help prevent overfitting? *Journal of Forecasting*, 23(2):115–139, 2004.
- Todd E Clark and Michael W McCracken. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1):85–110, 2001.
- Todd E Clark and Michael W McCracken. Evaluating direct multistep forecasts. *Econometric Reviews*, 24(4):369–404, 2005a.
- Todd E Clark and Michael W McCracken. The power of tests of predictive ability in the presence of structural breaks. *Journal of Econometrics*, 124(1):1–31, 2005b.
- Zhi Da, Joseph Engelberg, and Pengjie Gao. In search of attention. *The Journal of Finance*, 66(5):1461–1499, 2011.
- Zhi Da, Joseph Engelberg, and Pengjie Gao. The sum of all fears investor sentiment and asset prices. *Review of Financial Studies*, 28(1):1–32, 2015.
- Antonello D’Agostino and Paolo Surico. Does global liquidity help to forecast us inflation? *Journal of Money, Credit and Banking*, 41(2-3):479–489, 2009.

- Francis X Diebold and Roberto S Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995.
- Rudiger Dornbusch. Expectations and exchange rate dynamics. *The Journal of Political Economy*, pages 1161–1176, 1976.
- Casey Dougal, Joseph Engelberg, Diego Garcia, and Christopher A Parsons. Journalists and the stock market. *Review of Financial Studies*, page hhr133, 2012.
- Jack Favilukis, Sydney C Ludvigson, and Stijn Van Nieuwerburgh. The macroeconomic effects of housing wealth, housing finance, and limited risk-sharing in general equilibrium. Technical report, National Bureau of Economic Research, 2010.
- Jack Favilukis, David Kohn, Sydney C Ludvigson, and Stijn Van Nieuwerburgh. International capital flows and house prices: Theory and evidence. In *Housing and the Financial Crisis*, pages 235–299. University of Chicago Press, 2012.
- Jack Favilukis, Sydney C Ludvigson, and Stijn Van Nieuwerburgh. Foreign ownership of us safe assets: good or bad? Technical report, National Bureau of Economic Research, 2014.
- Andrea Ferrero. House price booms, current account deficits, and low interest rates. *Journal of Money, Credit and Banking*, 47(S1):261–293, 2015.
- Marcel Fratzscher, Luciana Juvenal, and Lucio Sarno. Asset prices, exchange rates and the current account. *European Economic Review*, 54(5):643–658, 2010.
- Atish Ghosh, Anne-Marie Gulde, and Holger Wolf. Exchange rate regimes: Classification and consequences. *Center for Economic Performance*, 2002.
- Eric Ghysels, Alberto Plazzi, Walter N Torous, and Rossen I Valkanov. Forecasting real estate prices. *Handbook of Economic Forecasting*, 2, 2012.
- Anthony Y Gu. The predictability of house prices. *Journal of Real Estate Research*, 24(3):213–234, 2002.
- Atsushi Inoue and Lutz Kilian. In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, 23(4):371–402, 2005.
- Alejandro Justiniano, Giorgio E Primiceri, and Andrea Tambalotti. Credit supply and the housing boom. Technical report, National Bureau of Economic Research, 2015.
- Roy Kouwenberg and Remco Zwinkels. Forecasting the us housing market. *International Journal of Forecasting*, 30(3):415–425, 2014.

- Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011.
- Tim Loughran and Bill McDonald. Measuring readability in financial disclosures. *The Journal of Finance*, 69(4):1643–1671, 2014.
- Michael W McCracken. Asymptotics for out of sample tests of granger causality. *Journal of Econometrics*, 140(2):719–752, 2007.
- Tanushree Mitra and Eric Gilbert. The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 49–61. ACM, 2014.
- Vasilios Plakandaras, Rangan Gupta, Periklis Gogas, and Theophilos Papadimitriou. Forecasting the us real house price index. *Economic Modelling*, 45:259–267, 2015.
- David E Rapach and Jack K Strauss. Differences in housing price forecastability across us states. *International Journal of Forecasting*, 25(2):351–372, 2009.
- David E Rapach, Jack K Strauss, et al. Forecasting real housing price growth in the eighth district states. *Federal Reserve Bank of St. Louis. Regional Economic Development*, 3(2):33–42, 2007.
- Barbara Rossi and Atsushi Inoue. Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, 30(3):432–453, 2012.
- Filipa Sá and Tomasz Wieladek. Capital inflows and the us housing boom. *Journal of Money, Credit and Banking*, 47(S1):221–256, 2015.
- Maik Schmeling. Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, 16(3):394–408, 2009.
- James H Stock and Mark W Watson. Forecasting output and inflation: the role of asset prices. Technical report, National Bureau of Economic Research, 2001.
- James H Stock and Mark W Watson. Dynamic factor models. *Oxford Handbook of Economic Forecasting*, 1:35–59, 2011.
- Matt Taddy. Measuring political sentiment on twitter: Factor optimal design for multinomial inverse regression. *Technometrics*, 55(4):415–425, 2013a.
- Matt Taddy. Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108(503):755–770, 2013b.

Paul C Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168, 2007.

Kenneth D West. Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, pages 1067–1084, 1996.