Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA
DESDE 1911

MESTRADO

ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO

DISSERTAÇÃO

PREDICTING AGGREGATE RETURNS USING VALUATION RATIOS OUT-OF-SAMPLE

ANA CARLA NATAL DA SILVA SEQUEIRA

SETEMBRO - 2012

Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA
DESDE 1911

MESTRADO EM ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO

DISSERTAÇÃO

PREDICTING AGGREGATE RETURNS USING VALUATION RATIOS OUT-OF-SAMPLE

ANA CARLA NATAL DA SILVA SEQUEIRA

ORIENTAÇÃO:

JOÃO VALLE E AZEVEDO, BANCO DE PORTUGAL

SETEMBRO - 2012

ABSTRACT

It is well established that valuation ratios provide, in-sample, relevant signals regarding future returns on assets. This pattern of predictability is pervasive across financial markets. In this dissertation we assess the ability of valuation ratios to predict out-of-sample aggregate returns for the stock and the housing markets in the U.S.. We apply linear models and multivariate filters to produce the forecasts and employ powerful out-of-sample tests for inference. We find that there is statistical evidence supporting the extension of the in-sample results to an out-of-sample framework. The dividend-price ratio and the rent-price ratio display a significant ability for predicting stock and housing returns, respectively. Nevertheless, we note that these findings may be sample dependent. Especially for the stock market, the end of the sample, including the recent financial crisis, may be responsible for the good results.

RESUMO

É amplamente reconhecido que os *valuation ratios* fornecem, *in-sample*, indicações relevantes sobre os retornos futuros de ativos. Este padrão de previsibilidade é comum a uma larga maioria de mercados. Nesta dissertação, avaliamos a capacidade de certos *valuation ratios* para prever, *out-of-sample*, os retornos agregados para o mercado de ações e para o mercado imobiliário, nos E.U.A.. Aplicamos modelos lineares e filtros multivariados para gerar as previsões e utilizamos "poderosos" testes *out-of-sample* para fazer inferência estatística. Verificamos que existe evidência estatística que suporta a passagem dos resultados *in-sample* para um contexto *out-of-sample*. O rácio dividendo-preço e o rácio renda-preço apresentam uma capacidade significativa para prever os retornos de ações e imóveis, respetivamente. Notamos, contudo, que estes resultados podem depender da amostra. Sobretudo para o mercado de ações, o final da amostra (que inclui a recente crise financeira) pode ser o responsável pelos bons resultados.

ACKNOWLEDGMENTS

I would like to thank my advisor João Valle e Azevedo for his entire availability, ideas and relevant observations, as well as for the constant support and motivation. I am also grateful to Ana Pereira for providing me her valuable *Mathematica* programming codes and clarifying some questions, Martín Saldías for his important suggestions and Mário Centeno for his comprehension. I owe an appreciation to the Bank of Portugal, particularly to the Research Department, that provided all the necessary conditions to successfully develop this dissertation.

Special thanks go to Hugo for his patience and continuous support, ideas and fruitful discussions and for his reviews and constructive comments. I would also like to thank my master colleagues, particularly Luís, for their vital encouragement and helpful remarks.

Finally, I wish to express my deep gratitude to my family for their unconditional support.

Thank you all.

Ana Carla Sequeira

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Predicting returns is one of the most discussed topics in the academic financial world. Cochrane (2011) summarizes a pattern of predictability that is pervasive across markets. For a wide set of markets (stocks, bonds, houses, credit spreads, foreign exchange and sovereign debt), he concludes (in-sample) that a yield or a valuation ratio predicts excess returns, instead of cashflow or price change.[1] For the stock market, Cochrane (2011) argues that the dividend yields predict returns and do not predict dividend growth. More than that, low dividend-yield ratios mean low future returns and high dividend-yield ratios mean high future returns. For the housing market, the argument is similar: high prices, relative to rents, imply low returns, and do not signal the permanent increase of rents or prices.

From an asset pricing perspective, we can explain this phenomenon using the fundamental present value relation. That is, the price of a financial asset should equal the present value of its future cashflows or, briefly, asset prices should equal expected discounted cashflows. In the case of the housing market, this means that the price of a house should equal the present value of its future rents (the analogy to the stock market is straightforward). This relation then implies that observed fluctuations in financial asset prices should reflect variation in future cashflow, in future discount rates, or in both.

[1] See, Fama and French (1988, 1989) for stocks; Fama and Bliss (1987), Campbell and Shiller (1991) and Piazzesi and Swanson (2008) for Treasuries; Fama (1986) for Bonds; Hansen and Hodrick (1980) and Fama (1984) for foreign exchange; Gourinchas and Rey (2007) for foreign debt.

In this paper, we intend to verify whether this pervasive phenomenon holds out-of-sample, i.e., whether a forecaster would be able to predict excess returns systematically, if he stood at the forecast moment without further information. Since there are relatively few studies about predicting the housing returns, we decided to focus on the housing market. The stock market analysis appears as an important reference. We use linear models and multivariate filters to produce the forecasts for the two aforementioned markets and employ equal accuracy tests and forecast encompassing tests for inference.

Our results show that there is statistical evidence supporting the extension of the in-sample results to an out-of-sample framework. Especially for the housing market, we conclude that the rent-price ratio has a huge ability for predicting returns (performing the equal accuracy test, we note that all the values are statistically different from 1 at the 1% significance level).

Given the lack of out-of-sample studies for the housing market, we consider that our findings are a considerable contribution to the literature. Using a diverse set of models to produce the forecasts of returns, we also apply relatively powerful test statistics and a bootstrap approach (because our models are nested) to conduct robust inference. We obtain all the results for the housing market using two different data sources (the Case-Shiller-Weiss (CSW) index and the Office of Federal Housing Enterprise Oversight (OFHEO) price index).

As Rapach and Wohar (2006), our purpose is testing for the existence of return predictability in population. As for this paper, we are not interested in exploring "whether a practitioner in real time could have constructed a portfolio that earns extra-normal returns".

The remainder of this paper is organized as follows. In Section 2 we briefly review the relevant literature and in Section 3 we provide a theoretical distinction between the in-sample and out-of-sample concepts. Section 4 describes the data used to obtain the empirical results, while Section 5 reports the in-sample results. In Section 6 we expose the econometric methodology. Section 7 discusses our main findings and the last two sections present ideas for future research and the conclusions.

2. A BRIEF REVIEW OF THE RELEVANT LITERATURE

As mentioned before, there are about a handful of papers examining the predictability of housing returns. Case and Shiller (1990) investigate the prices and excess returns (in-sample) predictability in the housing market based on a set of independent variables including the rent-price ratio. For this variable, the estimated coefficient in the ordinary least squares (OLS) regression is positive and statistically significant.

Using quarterly data and based on a long-horizon regression, Gallin (2008) shows that changes in real rents tend to be larger than usual and changes in real prices tend to be smaller than usual, when house prices are high relative to rents.

With a different focus, Campbell *et al.* (2009) apply the dynamic Gordon growth model to the housing market and find that changes in expected future housing *premia* are an important source of volatility in rent-price ratios.

More recently, Plazzi $et\ al.$ (2010) conclude (in-sample) that the rent-price ratio predicts expected returns for apartments, retail properties and industrial properties (but does not predict expected returns of office buildings).

For the stock market, the literature is voluminous. Several authors have already examined the ability of the most common financial variables to be good predictors for the aggregate returns or the equity premium.

Goyal and Welch (2003) assess the performance of the dividend-price ratio when used to predict the CRSP (Center for Research in Security Prices) value-weighted annual excess returns. Contrary to the in-sample results, they find that the out-of-sample forecasts produced through a model with the dividend-price ratio have a worse performance than those created by a model of constant returns (that is, a model that includes only the constant term).

Along the same line, Goyal and Welch (2008) explore the existence of gains when one uses the financial variables with a reasonable in-sample performance to forecast (out-of-sample) the equity premium. They conclude that almost all models produce poor results out-of-sample, which suggest "that most models are unstable or even spurious".

Against this background, Rapach and Wohar (2006), using annual data over the $1927-1999$ period, conclude that several financial variables have a good in-sample and out-of-sample ability to forecast stock returns. As justification for these results, they emphasize the fact that the tests employed are robust for inference (specifically, they use the tests presented in Clark and McCracken (2001) and McCracken (2007)).

Following a slightly different approach, Rapach *et al.* (2010) use forecast combining methods to produce out-of-sample forecasts and find that this approach provides significant out-of-sample gains when compared to the historical mean.

3. IN-SAMPLE AND OUT-OF-SAMPLE

Although we aim at exploring out-of-sample forecasts, we consider important to understand the differences between in-sample predictability and out-of-sample predictability. In this section, we will distinguish these concepts.

For sake of simplicity, let us consider the following regression model:

(1) $$y_{t+h} = \alpha + \beta x_t + u_{t+h}, \qquad t = 1, \dots, T$$

where y_{t+h} is the return from holding the financial asset from t until $t + h$, $h > 0$ is the forecast horizon, x_t is the financial variable used to predict y_{t+h} and u_{t+h} is a disturbance term.

An in-sample analysis consists of estimating the equation (1) using the available $T - h$ observations and then, examine the $t - statistic$ associated to the OLS estimate of β and the goodness-of-fit measure R^2 to assess the predictive ability of x_t.[2] When the null hypothesis is rejected and the R^2 is high, we can conclude that x_t has predictive power over y_{t+h}.

There are some potential problems related to this perspective, specifically the small-sample bias (x_t is not an exogenous regressor in equation (1); see Stambaugh

[2] The null hypothesis ($H_0\colon \beta = 0$) reflects the lack of ability of x_t to forecast y_{t+h}.

(1986, 1999)) and the dependence between the observations for the regressand in (1) (these observations are overlapping when the forecast horizon is greater than 1; see Richardson and Stock (1989)). The serial correlation induced in the disturbance term should be taken into consideration when conducting inference. The Newey and West (1987) standard errors robust to the autocorrelation and the heteroskedasticity are a usual solution.

An out-of-sample analysis implies the generation of the forecasts for y_{t+h}. Typically, the researcher chooses one of the three most common schemes (fixed, recursive or rolling) that allow producing the predictions in-real time, as if the forecaster stood in the moment when the prediction is made (i.e. using the data available up to that time).

Here, we describe the concept of out-of-sample predictability only based in the recursive scheme for this is the scheme we use in our empirical applications.

We should start by determining the sample-split parameter (R), that is, the period of the first prediction (we discuss this issue in more detail in Section 6.2.). Once we obtain predictions for different forecast horizons (h), determining R is not the same as determining the period of the last observation used in estimating the model (which will be $t = R - h$). Fixing R, we ensure that the first forecast obtained refers to the same period, for each h.

Next, we split the total sample (T observations) into an in-sample portion (includes the first $R - h$ observations) and an out-of-sample portion (composed of the observations from $t = R$ until $t = T$). The first sub-sample is used to estimate the model (equation (1), for example). The other allows evaluating the performance of the obtained forecasts through the analysis of the forecast errors.

Using the OLS estimates of the coefficients in equation (1), we construct the forecast for the period R given the information until $R-h$, that is:

$$\hat{y}_{R|R-h} = \hat{\alpha}_{R-h} + \hat{\beta}_{R-h} x_{R-h}.$$

And then we compute the forecast error ($e_{R|R-h}$) for $t = R$:

$$e_{R|R-h} = y_R - \hat{y}_{R|R-h},$$

where y_R is the observed value of the dependent variable at $t = R$.

The remaining predictions are obtained by repeating this procedure for $r = R + 1, R + 2, \dots, T + h$, that is:

$$\hat{y}_{r|r-h} = \hat{\alpha}_{r-h} + \hat{\beta}_{r-h} x_{r-h}, \qquad r = R + 1, \dots, T + h.$$

In the end, we have $T - (R-1) + h$ forecasts but only $T - (R-1)$ forecast errors. We can then determine the Mean Squared Forecast Error ($MSFE$):

$$MSFE_h = \frac{1}{T-(R-1)} \sum_{t=R}^{T} \left(y_t - \hat{y}_{t|t-h}\right)^2 = \frac{1}{T-(R-1)} \sum_{t=R}^{T} \left(e_{t|t-h}\right)^2$$

and compare the forecasts obtained through the different models (which are described in Section 6.1.).

4. DATA

In this section, we describe and characterize the data (available at John Cochrane's website) used in the models estimation.

Stock Market:

As Lettau and Ludvigson (2001), we use quarterly data for the U.S. stock market. Our sample covers the period $1947{:}Q1 - 2010{:}Q2$ ($T = 254$) and our dependent variable is the equity premium from holding stocks from period t to $t + h$.

As usual, we define equity premium as the return on the stock market minus the return on a short-term (risk-free) interest rate. In our case, we use the CRSP value-weighted return less the 3-month Treasury bill return (the 3-month Treasury bill is a proxy for the risk-free rate). Formalizing, we can write the variable to forecast h periods ahead (y_{t+h}^{S}) as:

$$y_{t+h}^{S} = \prod_{j=1}^{h} R_{t+j}^{*} - \prod_{j=1}^{h} RTb_{t+j}$$

where $R_t^* = 1 + r_t^*$, r_t^* is the return (including dividends) on the Value-Weighted Index, $RTb_t = 1 + rTb_t$ and rTb is the 3-month Treasury bill return.

The dividend-price ratio (dp_t) is the financial variable which potentially predicts the equity premium:

$$dp_t = \frac{Dividend_t}{Price_t} = \frac{R_t^*}{R_t} - 1.$$

We intend to verify whether the pervasive phenomenon identified by Cochrane (2011) holds out-of-sample. Our variables were therefore constructed following the definitions presented in Cochrane (2011) – we use the simple returns instead of log returns. At all events, Goyal and Welch (2003) tried both specifications and found similar conclusions.

Housing Market:

In our applications for the housing market, we use quarterly data from $1960{:}Q1$ to $2010{:}Q1$ $(T = 201)$. There are two different available samples with similar information. One comes from the Case-Shiller-Weiss (CSW) price data, the other consists in the houses prices and rents from the Office of Federal Housing Enterprise Oversight (OFHEO) “purchase-only” price index.

Our dependent variable (y_{t+h}^{H}) is the log return from holding the house from t until $t + h$ and the predictor is the respective rent-price ratio (rp_t). That is:

$$y_{t+h}^{H} = \sum_{j=1}^{h} lreturn_{t+j} = \sum_{j=1}^{h} ln\left(\frac{price_{t+j} + rent_{t+j}}{price_{t+j-1}}\right)$$

$$rp_t = \frac{rent_t}{price_t}.$$

As mentioned before, for both markets, we construct the variables based on the definitions and methods presented in Cochrane (2011) (for more details, see Appendix

C).[3] Table I in Appendix B contains the usual descriptive statistics for all the analyzed series.

5. IN-SAMPLE FIT

As mentioned in Goyal and Welch (2008), the out-of-sample performance is only interesting when the model has a good in-sample performance. Hence, in this section, we discuss the results obtained through the in-sample regressions and present some motivations to the out-of-sample exercise.

Table II in Appendix B provides the results of regressing the returns from holding the financial asset from t to $t+h$ (y_{t+h}) on the corresponding valuation ratio (x_t). Specifically, for each market in question, we estimate:

(2) $$y_{t+h} = \alpha + \beta x_t + u_{t+h}$$

where y_{t+h}, x_t and u_{t+h} have the meanings introduced in Section 3; h, as before, is the forecast horizon in quarters.

The equation (2) is estimated by OLS and the Newey and West (1987) standard errors, which are robust to heteroskedasticity and serial correlation, are used to compute the $t-statistics$. Following Rapach and Wohar (2006), we use the Bartlett kernel and a lag truncation parameter equal to $\lfloor 1.5 \cdot h \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part, for $h > 1$; and zero for $h = 1$ to calculate these standard errors.

[3] The stock market data and the housing market data are available at http://faculty.chicagobooth.edu/john.cochrane/research/index.htm. The housing market data are also available at http://www.lincolninst.edu/subcenters/land-values/rent-price-ratio.asp.

As mentioned before, we assess the in-sample predictive performance based on values of the $t - statistic$ and R^2. We can also interpret the OLS estimate of β as an indicator of the x_t significance to forecast y_{t+h}.

Analyzing the results shown in Table II (Appendix B), we can detect a set of characteristics that are common across the two markets. The estimate of β and the R^2 are higher for longer forecast horizons, and the observed $t - statistics$ always reject the null hypothesis of no predictability. In addition, the signal of the estimates is positive, which confirms the conclusions presented in Cochrane (2011): higher valuation ratios indicate higher returns. Or, more specifically, high prices, relative to dividends (or rents, for the housing market) can be a sign of low returns.

Hereupon, and since this in-sample predictability may mean nothing out-of-sample, it is of all the interest to examine the predictability of these variables out-of-sample.

6. ECONOMETRIC PROCEDURE

In this section, we discuss the regression models used to produce the out-of-sample forecasts, the methods employed to compare them and, lastly, the equal accuracy tests and the forecast encompassing tests applied to statistically analyze the results. We also describe the bootstrap procedure used to generate the critical values.

6.1. Predictive regression models

Apart from assessing the out-of-sample performance of the valuation ratios to predict aggregate returns, we also aim at identifying which model(s) provides the best forecasts compared to the historical mean. Therefore, we select several methods to generate different sets of predictions for the same variable – aggregate returns. All of them are estimated using OLS.

In what follows, $\hat{y}_{t+h|t}$ denotes the forecast of y_{t+h} (the return from holding the financial asset from t to $t+h$), given the information up to period t, and x_t is the valuation ratio that might have predictive power for y_{t+h}.

A direct method requires that only information available up to t is used to obtain the forecast for $y_{t+h}, h > 1$. By contrast, the iterated method generates the prediction for $y_{t+h}, h > 1$, using one-step ahead forecasts. For $h = 1$, the direct and the iterated models produce the same forecasts. The direct approach is computacionally simpler.

• ($Model\ 0$) Historical mean:

$$\hat{y}_{t+h|t} = \frac{1}{t}\sum_{s=1}^{t} y_s \quad , \qquad t = R, \dots, T.$$

As Goyal and Welch (2003, 2008) and Campbell and Thompson (2008), we use the historical mean as a benchmark forecasting model, since it represents the hypothesis of no predictability, consistent with the most common interpretation of the efficient markets hypothesis.

• ($Model$ 1) Direct autoregressive (AR) with fixed lag order (p):

$$\hat{y}_{t+h|t} = \hat{\alpha} + \sum_{j=0}^{p-1} \hat{\beta}_j y_{t-j}\,, \qquad t = R, \dots, T,$$

where $\hat{\alpha}$ and $\hat{\beta}_j$ $(j = 1, \dots, p-1)$ are the OLS estimates. In our empirical applications, we fix $p = 2$.

• ($Model$ 2) Direct AR using the Akaike Information Criterion (AIC) (Akaike, 1974) to determine the lag order (p^*):

$$\hat{y}_{t+h|t} = \hat{\alpha} + \sum_{j=0}^{p^*-1} \hat{\beta}_j y_{t-j}\,, \qquad t = R, \dots, T \;\; and\; p^* is\; an\; integer\; \epsilon\; [1\,, 4].$$

In this method, we only define the maximum lag order ($p_{max} = 4$). After that, whenever a forecast is generated, we apply the AIC to determine the optimal number of lags (p^*), given all the past information. Thus, for each period t, the p^* employed to produce the prediction $\hat{y}_{t+h|t}$ can be different.

• ($Model$ 3) Direct augmented AR using the AIC to determine the lag order (p^*):

$$\hat{y}_{t+h|t} = \hat{\alpha} + \sum_{j=0}^{p_1^*-1} \hat{\beta}_j y_{t-j} + \sum_{j=0}^{p_2^*-1} \hat{\delta}_j x_{t-j}\,,$$

$$t = R, \dots, T\; and\; p_1^*, p_2^*\; are\; integers\; \epsilon\; [1\,, 4].$$

$\hat{\alpha}, \hat{\beta}_j (j = 1, \dots, p_1^* - 1)$ and $\hat{\delta}_j (j = 1, \dots, p_2^* - 1)$ are the OLS estimates. The expression "augmented" denotes the introduction of valuation ratios as explanatory variables in the regression. Again, the lag order is determined by the AIC (the above comment applies).

• ($Model$ 4) Direct regression with or without lags:

$$\hat{y}_{t+h|t} = \hat{\alpha} + \sum_{i=0}^{p-1} \hat{\beta}_i x_{t-i}\,, \qquad t = S, \ldots, T\,, p = 1,..,4.$$

In this method, the autoregressive part is not taken into consideration.

• ($Model$ 5) Univariate and multivariate filters:

Following the argument presented in Valle e Azevedo and Pereira (2012), when we choose this method to generate our forecasts, we assume that we are interested in predicting the low frequencies of y_t (say, $w_t = B(L)y_t$, where $B(L) = \sum_{j=-\infty}^{\infty} B_j L^j$ is a band-pass filter eliminating the fluctuations with period smaller than a specified cut-off) and using these predictions as forecasts of y_t itself. Explicitly, we will consider predictions of the low frequencies of aggregate returns as forecasts of aggregate returns itself. The weights of the ideal filter ($B(L)$) are given by:

$$B_0 = \frac{\omega_h}{\pi}, \qquad B_j = \frac{sin[\omega_h j]}{\pi j}, |j| \geq 1, \quad \omega_h = \frac{2\pi}{cut-off}.$$

Nevertheless, since $B(L)$ is an infinite (absolutely summable and stationary) polynomial in lag operator L and we only have available a finite sample ($\{y_t\}_{t=1}^{T}$), we approximate the low frequencies of y_t (that is, w_t) through a weighted sum of elements of y_t ($\widehat{w}_t$), which will be considered a forecast for y_t. That is:

$$\widehat{w}_t = \sum_{j=-f}^{p} \hat{B}_j^{p,f} y_{t-j}\,.$$

p and f denote the number of observations in the past and in the future, respectively, that are considered.

We obtain the multivariate filter when we include, in that weighted sum, elements of c series of covariance-stationary covariates $z_1, \dots, z_c$, where $z_j = (z_{j,1}, \dots, z_{j,T})'$ $(j = 1, \dots, c)$. Namely:

$$\widehat{w}_t = \sum_{j=-f}^{p} \widehat{B}_j^{p,f} y_{t-j} + \sum_{s=1}^{c} \sum_{j=-f}^{p} \widehat{R}_{s,j}^{p,f} z_{s,t-j}.$$

Solving the problem:

$$(3) \qquad \underset{\{\widehat{B}_j^{p,f}, \widehat{R}_{1,j}^{p,f}, \dots, \widehat{R}_{c,j}^{p,f}\}_{j=-f,\dots,p}}{Min} E[(y_t - \hat{y}_t)^2],$$

where the information set is implicitly restricted by p and f, we determine $\widehat{w}_t$ (the weights of the filter are found solving a linear system with $(p + f + 1) \times (c + 1)$ equations and unknowns). The solution to problem (3) is discussed in Valle e Azevedo (2011).

To extract the signal $w_{t+h} = B(L)y_{t+h}$ for $h > 0$, we should set $f = -h$ in the solution. As a result, only the available information up to period t is employed.[4]

After choosing the model, we apply the recursive scheme described in Section 3 to generate the forecasts.

6.2. Estimation period

As mentioned in Section 3, the first step of an out-of-sample analysis is to determine the in-sample period and the out-of-sample period. Specifically, we should fix a value to the sample-split parameter (R). Nevertheless, there is no criterion that defines how

[4] More detailed explanations about the multivariate filter can be found in Valle e Azevedo (2011).

to choose R. We should make a compromise between the number of observations used to estimate the coefficients of the model and the number of available observations to assess the forecasts performance.

It is natural to compare predictions at different horizons referring to the same period of time (regardless the forecast horizon). To make this possible, the forecast for period t must be generated based in the information until period $t - h$, which implies that, for longer horizons, less observations are available to estimate the coefficients.

Additionally, since we are forecasting the aggregate returns (the returns from holding the financial asset from t until $t + h$), we lose the first h observations of the series of interest. Again, longer forecast horizons imply losing more initial observations.

Taking this information into consideration, we fix $R = 153$, which corresponds to the first quarter of 1985 in the stock market data (we consider the predictions for the period $1985{:}\,Q1 - 2010{:}\,Q2$) and to the first quarter of 1998 in the housing market data (we consider the predictions for the period $1998{:}\,Q1 - 2010{:}\,Q1$).

6.3. Forecast evaluation

We choose the (out-of-sample) Mean Squared Forecast Error ($MSFE$) ratio as evaluation metric to compare the sets of forecasts obtained through the models described in Section 6.1.

Given a set of h-step ahead forecasts generated by the model k $\left(\left\{\hat{y}_{t|t-h}^{k}\right\}_{t=R}^{T+h}, \ k = 0,1,\ldots,5\right)$, we calculate the forecast errors as:

$$e_{t|t-h}^{k} = y_t - \hat{y}_{t|t-h}^{k}, \qquad t = R, \ldots, T,$$

where y_t is the observed value of the dependent variable at t and $\hat{y}_{t|t-h}^{k}$ is the forecast of y_t, generated by model k, given the information up to $t-h$.

Consequently, the $MSFE$ for model k is equal to:

$$MSFE_h^k = \frac{1}{T-(R-1)}\sum_{t=R}^{T}\left(y_t - \hat{y}_{t|t-h}^{k}\right)^2 = \frac{1}{T-(R-1)}\sum_{t=R}^{T}\left(e_{t|t-h}^{k}\right)^2.$$

Denoting the $MSFE$ of the benchmark model (the historical mean) by $MSFE_h^0$ $(k = 0)$ and the $MSFE$ of the competing model by $MSFE_h^k$ $(k = 1,\ldots,5)$, the $MSFE$ ratio is given by:

$$MSFE\ ratio = \frac{MSFE_h^k}{MSFE_h^0}.$$

When the $MSFE\ ratio$ is less than 1, the competing model predicts better than the benchmark model, suggesting that there are out-of-sample forecasting gains. Otherwise, the historical mean (which signals constant expected returns) is the best possible forecast.

We also use a graphical analysis to examine the relative performance of the forecasting models. As proposed in Goyal and Welch (2003), we construct charts with the difference of the cumulative squared forecast errors of the benchmark model

($SSE_{h,S}^{bench}$) and the cumulative squared forecast errors of the competing model ($SSE_{h,S}^{comp}$). Formalizing:

$$Net\ SSE_{h,S} = SSE_{h,S}^{0} - SSE_{h,S}^{k} = \sum_{t=R}^{S} \left(e_{t|t-h}^{0}\right)^2 - \sum_{t=R}^{S} \left(e_{t|t-h}^{k}\right)^2,$$

$$S = R, \dots, T; \quad k = 1, \dots, 5.$$

When this difference is positive, the competing model outperforms the benchmark model (the sum of the squared forecast errors from R through S (i.e., the date in the x-axis) is greater for the benchmark model than for the competing model).

6.4. Out-of-Sample tests

We assess the statistical significance of the obtained results considering equal accuracy tests (under the null hypothesis, the $MSFE$ from two distinct models are statistically equal) and forecast encompassing tests (we test whether a given set of forecasts generated by a simpler model embody all the useful predictive information contained in another set of forecasts).

Before describing these tests in more detail, it is important to make a distinction between nested and non-nested models and, above all, note that, excluding the multivariate filter (model 5), all of our models are nested.

We say that two models are nested when there is a set of regressors that is common between them. In our studies, whenever we compare the competing model k ($k = 1, .., 4$) with the benchmark model (model 0), we are comparing nested models due to the constant term in each regression (setting the coefficients $\hat{\beta}_j$ and $\hat{\delta}_j$ equal to zero in models 1,2,3 and 4, we obtain the model 0). Briefly, the benchmark model (which includes only the constant term) is a restricted version of the model of interest.

This clarification is relevant because, as stressed in Clark and McCracken (2005), when we have nested models, the population errors of the analyzed models are exactly the same, under the null hypothesis that the restrictions imposed in the benchmark model are true. This implies that the asymptotic difference between the $MSFE$ of two models is exactly zero with zero variance and, consequently, the standard distributions are asymptotically invalid.

Because our models are nested and different forecast horizons ($h = 1, \dots, 24$, in quarters) are explored, we use, as recommended in literature, a bootstrap procedure for inference.

6.4.1. Equal Accuracy test

Using the $MSFE$ as the evaluation metric, the equal accuracy test allows testing whether the $MSFE$ ratio is statistically equal to 1, against the alternative that the forecasts produced by the competing model are better (have a lower $MSFE$). Tantamount, we can write:

$$H_0: \frac{MSFE_h^k}{MSFE_h^0} = 1 \quad vs \quad H_1: \frac{MSFE_h^k}{MSFE_h^0} < 1,$$

where $MSFE_h^k$ is the $MSFE$ of the competing model k and $MSFE_h^0$ is the $MSFE$ of the benchmark model. This test is a one-sided to the left.

Using the set of $N = T - (R - 1)$ h-steps ahead forecast errors from model k, we can also express the null hypothesis as follows:

$$H_0: E\left[\left(e_{t|t-h}^k\right)^2\right] = E\left[\left(e_{t|t-h}^0\right)^2\right] \Leftrightarrow E\left(d_i^k\right) = 0, \;\; i \equiv t - (R - 1),$$

$$t = R, \dots, T \;\; or \;\; i = 1, \dots, N, \qquad k = 1, \dots, 5,$$

where $d_i^k = \left(e_{t|t-h}^k\right)^2 - \left(e_{t|t-h}^0\right)^2, \; i = 1, \dots, N.$

First proposed by Diebold and Mariano (1995), the $MSFE - t$ test statistic can be written as:

$$MSFE - t = \sqrt{N}\frac{\bar{d}}{\sqrt{\widehat{V_{LR}}(d)}}$$

where:

$\bar{d} = N^{-1}\sum_{i=1}^{N} d_i^k; \; \widehat{V_{LR}}(d) = \hat{\gamma}_0^d + 2\sum_{j=1}^{M} K(j, M).\hat{\gamma}_j^d;$

$\hat{\gamma}_j^d$ $(j = 0, \dots, M)$ is the estimated jth autocovariance of d:

$$\hat{\gamma}_j^d = N^{-1}\sum_{i=j+1}^{N}\left(d_i^k - \bar{d}\right)\left(d_{i-j}^k - \bar{d}\right), \quad j = 0, \dots, M, \qquad \hat{\gamma}_j^d = \hat{\gamma}_{-j}^d \, ;$$

M is the truncation parameter and $K(j, M) = \left(1 - \frac{j}{M+1}\right)$ is the Bartlett Kernel.

Following Clark and McCracken (2005), we fix $M = 0$ for $h = 1$ and $M = \lfloor 1.5 \cdot h \rfloor$ for $h > 1$. This test statistic has an asymptotic standard normal distribution when used to compare non nested models forecasts.

Notwithstanding, there is evidence that the $MSFE - t$ test could be over-sized for $h > 1$ in small and moderate samples. Thus, Harvey $et\ al.$ (1997) proposed a small-sample correction that resulted in the following test statistic:

(4) $$modified\ MSFE - t = \left[\frac{N + 1 - 2h + N^{-1}h(h-1)}{N}\right]^{1/2} MSE - t.$$

These authors recommend comparing the values of the modified statistic with critical values from the Student's t distribution with $N - 1$ degrees of freedom, when comparing forecasts from non nested models.

Due to the emergence of other problems (namely, the degeneracy of the long-run variance of d_i^k), McCracken (2007) develops the $MSFE - F$ test statistic:

(5) $$MSFE - F = \sqrt{N}\frac{\bar{d}}{MSFE_k},$$

where $MSFE_k$ is the $MSFE$ of the competing model $(k = 1, \ldots, 4)$.

In our case, a significant $MSFE - F$ statistic means that the forecasts from the competing model have statistically more predictive power than those from the historical mean model.

For practical purposes, we will use the $modified\ MSFE - t$ and the $MSFE - F$ test statistics which have non standard distributions with nested models. [5,6] Clark and McCracken (2001) and McCracken (2007) provide tables with asymptotic critical values

[5] Clark and McCracken (2001) find that $MSFE - F$ has higher power than $MSFE - t$.

[6] Each of them can be written as functions of stochastic integrals of Brownian motion.

for $h = 1$. Since we are interested in a multi-step analysis, we based our inference in a bootstrap approach (except for the multivariate filter model).

6.4.2. Forecast Encompassing tests

According to Clements and Harvey (2009), a set of forecasts encompasses a rival set if the latter does not contribute to a statistically significant reduction in $MSFE$ when used in combination with the original set of forecasts. Applying this concept to our study, if the historical mean forecast encompasses the forecast produced by the model with the valuation ratio, then the financial variable does not contain useful additional information for predicting the aggregate returns.

We will present three alternative definitions for forecast encompassing. The way how the test is applied depends on the chosen setting.

The most general formulation, proposed by Fair and Shiller (1989), considers that $\hat{y}^0_{t|t-h}$ encompasses $\hat{y}^k_{t|t-h}$ if the value of θ_2 is zero in equation:

$$y_t = \mu + \theta_1 \hat{y}^0_{t|t-h} + \theta_2 \hat{y}^k_{t|t-h} + \varepsilon_t, \tag{6}$$

where $\hat{y}^k_{t|t-h}$ denotes the h -steps ahead forecast of y_t produced by model k ($k = 0$ for the restricted model (historical mean) and $k = 1,2,3,4$ for the unrestricted model). As a result, we should test $H_0: \theta_2 = 0$ ($\hat{y}^0_{t|t-h}$ encompasses $\hat{y}^k_{t|t-h}$) against $H_1: \theta_2 > 0$ ($\hat{y}^0_{t|t-h}$ does not encompass $\hat{y}^k_{t|t-h}$).

Assuming that the individual forecasts are efficient, we can impose the restriction $\theta_1 + \theta_2 = 1$ in the equation (6), obtaining the regression:[7]

(7) $$y_t = \mu + (1-\theta)\hat{y}^0_{t|t-h} + \theta\hat{y}^k_{t|t-h} + \varepsilon_t.$$

However, to apply the forecast encompassing test, we consider the Andrews $et\ al.$ (1996) approach instead of equation (7).[8] Consequently, the encompassing is defined by $\theta = 0$ (we test $H_0\colon \theta = 0$ against $H_1\colon \theta > 0$) in the regression:

(8) $$e^0_{t|t-h} = \mu + \theta\left(e^0_{t|t-h} - e^k_{t|t-h}\right) + \varepsilon_t,$$

where $e^0_{t|t-h} = y_t - \hat{y}^0_{t|t-h}$ and $e^k_{t|t-h} = y_t - \hat{y}^k_{t|t-h}$.

Finally, dropping the intercept in equation (8), which means fixing $\mu = 0$, we require that the individual forecasts are unbiased and efficient. In this last specification we define encompassing by $\theta = 0$ in the equation:

(9) $$e^0_{t|t-h} = \theta\left(e^0_{t|t-h} - e^k_{t|t-h}\right) + \varepsilon_t.$$

As mentioned in Clements and Harvey (2009), if the restrictions imposed ($\mu = 0$ when the forecasts are unbiased, and $\theta_1 + \theta_2 = 1$ when the forecasts are efficient) do not hold, the exposed definitions are not equivalent, implying that we can take different conclusions using distinct classifications. Nevertheless, when the restrictions imposed are true, the tests based in the modified equations ((8) or (9)) should be more powerful.

---

[7] A forecast $\hat{y}_{t+h|t}$ is said to be Mincer-Zarnowitz efficient if $\mu = 0$ and $\lambda = 1$ in a regression $y_t = \mu + \lambda\hat{y}_{t+h|t} + \varepsilon_t$, which implies no correlation between the forecast and the forecast error.

[8] This approach results from some transformations in equation (7). Specifically,

$$y_t = \mu + (1-\theta)\hat{y}^0_{t|t-h} + \theta\hat{y}^k_{t|t-h} + \varepsilon_t \Leftrightarrow y_t - \hat{y}^0_{t|t-h} = \mu + \theta\left(\hat{y}^k_{t|t-h} - \hat{y}^0_{t|t-h} + y_t - y_t\right) + \varepsilon_t \Leftrightarrow$$
$$e^0_{t|t-h} = \mu + \theta\left[\left(y_t - \hat{y}^0_{t|t-h}\right) - \left(y_t - \hat{y}^k_{t|t-h}\right)\right] + \varepsilon_t \Leftrightarrow e^0_{t|t-h} = \mu + \theta\left(e^0_{t|t-h} - e^k_{t|t-h}\right) + \varepsilon_t.$$

Additionally, if the optimal value of θ_2 in equation (6) is zero, we should conclude that, in $MSFE$ sense, the forecast of y_t cannot be improved by adding the $\hat{y}^k_{t|t-h}$ forecast in the linear function of $\hat{y}^0_{t|t-h}$ (i.e., $\mu + \theta_1 \hat{y}^0_{t|t-h}$), which does not imply that the $\hat{y}^0_{t|t-h}$ is the optimal forecast for y_t.

In order to test for encompassing, the standard $t - statistic$ cannot be used since the regression errors may not be independent (we consider $h \geq 1$) nor normally distributed.[9,10] Therefore, Harvey $et\, al.$ (1998) proposed an approach, based in Diebold and Mariano (1995) test statistic, which consists in testing whether the series $\{d_i\}_{i=1}^{N}$ (N is the number of forecast errors from each model) has zero mean. The test statistic is similar to that presented in Section 6.4.1 for the equal accuracy test, changing just the definition of d_i, which depends on the regression ((6), (8) or (9)). The following table describes the three possible cases: [11]

Characteristics of the individual forecasts	d_i^k
Biased and Inefficient: Regression (6)	$d_i^k = \eta_{0i}\eta_{ki}$
Biased and Efficient: Regression (8)	$d_i^k = \left(e^0_{t\|t-h} - \overline{e^0}\right)\left[\left(e^0_{t\|t-h} - \overline{e^0}\right)\left(e^k_{t\|t-h} - \overline{e^k}\right)\right]$
Unbiased and Efficient: Regression (9)	$d_i^k = e^0_{t\|t-h} \cdot \left(e^0_{t\|t-h} - e^k_{t\|t-h}\right)$

where $\boldsymbol{i \equiv t - (R-1), t = R, \dots, T}$; $\boldsymbol{\eta_{0i}}$ and $\boldsymbol{\eta_{ki}}$ denote the errors from regressions of $\boldsymbol{y_t}$ and $\boldsymbol{\hat{y}^k_{t|t-h}}$, respectively, on a constant and $\boldsymbol{\hat{y}^0_{t|t-h}}$; $\boldsymbol{\overline{e^0} = \sum_{t=R}^{T} e^0_{t|t-h}}$ and $\boldsymbol{\overline{e^k} = \sum_{t=R}^{T} e^k_{t|t-h}}$.

[9] Note that the optimal forecast errors is expected to follow a moving-average process of order $h - 1$.

[10] Harvey $et\, al.$ (1998) examine this problem in the context of unbiased and efficient individual forecasts.

[11] Clements and Harvey (2009).

Formally, the test statistic is given by:

$$ENC - t = \sqrt{N}\frac{\bar{d}}{\sqrt{\widehat{V_{LR}}(d)}},$$

where $\bar{d}$ and $\widehat{V_{LR}}(d)$ have the aforementioned meaning. We can also compute the modified approach suggested by Harvey $et\ al.$ (1997).

Clark and McCracken (2001, 2005), admitting that the individual forecasts are unbiased and efficient, developed the following test statistic:

$$ENC - F = \sqrt{N}\frac{\bar{d}}{MSFE_k}, \qquad (11)$$

which is more powerful than the previous test statistics for forecast encompassing.[12]

The test statistics exposed do not have a standard distribution (and, most often, neither a pivotal asymptotic distribution) in the case of multi-step predictions. The procedure of obtaining forecasts using estimated regression models (the estimation uncertainty affects the encompassing tests) and the existence of nested models also difficult the deduction of the critical values. Therefore, it is widely suggested in the literature to use the critical values generated by bootstrap methods.

In our practical application, we employ the $modified\ ENC - t$ and the $ENC - F$ statistics to test the forecast encompassing. The critical values used are obtained by bootstrapping.

[12] Clark and McCracken (2001) provide the critical values for this test statistic, when $h = 1$ and the forecast errors are conditionally homoskedastic.

6.5. Bootstrap procedure

We follow Mark (1995) and Kilian (1999) to define a bootstrap method which allows obtaining the critical values for the test statistics described in previous sections.

As Goyal and Welch (2008), we impose the null hypothesis of no predictability assuming that the data generating process (DGP) is:

$$(12) \qquad y_t = \alpha + \varepsilon_{1,t}$$

$$(13) \qquad x_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{2,t},$$

where y_t denotes the aggregate returns and x_t is the predictor. We estimate the equations (12) and (13) by OLS, using the full sample.[13]

The next step is to generate B (we set $B = 10\ 000$) innovation sequences, of length T, by drawing randomly with replacement from fitted residuals $\hat{\varepsilon}_{1,t}$ and $\hat{\varepsilon}_{2,t}$ $(t = 1, \dots, T)$. Using the sequences $\left\{\hat{\varepsilon}_{1,t}{}^{b}\right\}_{t=1}^{T}$ and $\left\{\hat{\varepsilon}_{2,t}{}^{b}\right\}_{t=1}^{T}$ $(b = 1, \dots, B)$ and the OLS coefficient estimates obtained in first step, we produce B sequences of T observations for y_t and x_t. Specifically, for $b = 1, \dots, B$, we construct:

$$y_t{}^{b} = \hat{\alpha} + \hat{\varepsilon}_{1,t}{}^{b}$$

$$x_t{}^{b} = \hat{\beta}_0 + \hat{\beta}_1 x_{t-1}{}^{b} + \hat{\varepsilon}_{2,t}{}^{b}.$$

Since x_t follows an autoregressive process of order 1, we need an initial observation, namely, an observation that is prior to sample used to estimate the

[13] We could use the seemingly unrelated regressions (SUR) to estimate these equations. However, the SUR estimates are not necessarily more efficient than OLS estimates in finite samples (although this is true asymptotically), therefore we follow the literature and use the OLS method.

equations. Whenever necessary, this observation will be randomly selected by picking one date from the available data.

Finally, for each set of T observations, we apply the recursive scheme described in Section 3. In the end, we will have B sets of $(T-(R-1)+h)$ forecasts and B sets of the corresponding forecast errors. After calculating the values of the test statistics (we have B observed values for each statistic test), we determine the 1%, 5% and 10% critical values as the 99%, 95% and 90% percentiles of the resulting statistics, respectively.

7. EMPIRICAL RESULTS

In this section, we expose and discuss the main results obtained using the methodology described before. This analysis will be done separately for each market. We first present the findings for the stock market (specifically, the results of the out-of-sample statistical tests and the interpretation of the $Net\ SSE_h$ charts), and then we do the same for the housing market.[14]

Stock Market:

Although we have generated forecasts using different models, we only statistically analyze those obtained by the direct regression model (model $4-$ without lags) and multivariate filter (model $5-$ using the dividend-price ratio), since these are the models that produce better results. As the benchmark (the historical mean) and the

[14] The empirical results presented in this paper were obtained using the software $Mathematica$.

direct regression are nested models, we use the bootstrap critical values to perform the out-of-sample tests.

Table III in Appendix B reports the $MSFE$ ratios for each model. We conclude that only the direct regression generates forecasts that can beat the benchmark for all horizons. These ratios are statistically different from 1 at conventional significance levels when we use the $MSFE-F$ statistic (equation (5)) to perform the equal accuracy test. Both the quality and the statistical significance of the predictions increase with the forecast horizon, which suggests that the dividend–price ratio ability to predict the aggregate returns improves when we use longer horizons. These findings are consistent with the in-sample results exposed in Section 5, where we note that the in-sample predictability increases with the horizon.

The univariate filter model failed to outperform the benchmark model for all horizons, but the multivariate filter has $MSFE$ ratios less than 1 for $h=20$ and $h=24$ (despite not being statistically different from 1 when we use the $modified\ MSFE-t$ statistic (equation (4)) to apply the test).

Similar conclusions can be drawn when we analyze the forecast encompassing results presented in Table IV, Appendix B (which contains the observed values of the test statistics). In particular, when we use the $ENC-F$ (equation (11)) to perform the test, we have statistical evidence to reject the null hypothesis (the historical mean forecasts encompass those produced by direct regression model) at a 5% significance level for $h=4\ ,6,18,20,24$.

The following analysis rests on the evaluation of the $Net\,SSE_h$ charts $(h = 1, 4, 8, 12, 18, 24)$ which display the cumulative squared forecast errors of the benchmark model (from 1985:Q1 through the date in the x-axis) minus the squared forecast errors of the competing model (from 1985:Q1 trough the date in the x-axis), for each horizon. A positive value means that the competing model has outperformed the benchmark model and a positive slope indicates that the competing model had lower forecasting error than the historical mean model, in a given quarter.

For the stock market, we chose to plot merely the $Net\,SSE_h$ $(h = 1, 4, 8, 12, 18, 24)$ for the direct regression model (without lags) and the multivariate filter model (with dividend-price ratio) since these illustrate the main findings.

Considering the shorter forecast horizon (1 quarter, see figure 1), we note that the direct regression curve exhibits a volatile pattern. This competing model had a good performance in $1987{:}Q4 - 1995{:}Q4$, $2002{:}Q2 - 2003{:}Q3$ and $2008{:}Q2 - 2010{:}Q2$ and had its poorest performance from $1997{:}Q3$ to $2001{:}Q1$ (although it begins to recover (the curve has a positive slope) from $2000{:}Q1$). For $h = 1$, the multivariate filter consistently has a worse performance than the direct regression model.

Figure 1 also shows the cumulative SSE difference for $h = 8$, when both models underperformed the benchmark. For this horizon, the dividend-price ratio model had large prediction errors from $1997{:}Q3$ to $2003{:}Q3$.

For longer forecast horizons ($h = 24$, for example; see figure 1 in Appendix A), the curves are smoother and we can identify three distinct periods (which have become more apparent as the horizon increases). Namely: an initial period when the forecasts produced by the competing models are better, an intermediate period when the models had a negative performance and a final period of recovery. We note that this final period may be responsible for the good results out-of-sample, meaning that if we dropped the last observations of the sample, the direct regression model probably could not beat the benchmark. It is also worth noticing that the direct regression model curve has an extremer behavior than the multivariate filter curve, that is, it had the best performance but also has the worst in given portions of the sample.

Housing Market:

As we mentioned in Section 4, we have two data sources for the housing market. The results are quite similar.

Tables V and VI in Appendix B contain the $MSFE$ ratios between the competing and the benchmark model for horizons $h\,(in\,quarters) = 1, 4, 6, 8, 12, 18, 20, 24$. For forecast horizons shorter than 3 years (12 quarters), we find that all the competing models produce better forecasts than the benchmark model. However, and importantly, for longer horizons (over 3 years), only the models that contain the rent-price ratio exhibit $MSFE$ ratios lower than 1.

In particular, the $MSFE$ ratio between the direct regression and the benchmark model decreases as the horizon increases. Applying the relatively powerful $MSFE - F$

statistic (equation (5)) to conduct the equal accuracy test, we note that all the values are statistically different from 1 (at the 1% significance level). Comparing with the results obtained in Section 5, we verify that the predictability pattern identified in-sample holds out-of-sample for the housing market.

As regards to the multivariate filter, we observe that, although the $MSFE$ ratios are always lower than 1, we only have statistical significance for $h = 1$. Tables VII and VIII (in Appendix B) display the forecast encompassing statistics for concluding that the historical mean forecasts never encompass the forecasts generated by the direct regression model (the null hypothesis is always rejected at a significance level of 1%).

Figures 2 and 3 in Appendix A contain the charts with the $Net\ SSE_h$ $(h = 1, 4, 8, 12, 20, 24)$ for the CSW data and the OFHEO data. We only examine the curves from three models: the direct AR model (with $p_{max} = 4$), the direct augmented AR model (also with $p_{max} = 4$) and the direct regression model (without lags). Although we choose to show the charts based on the two data sources, we note that there are few differences between them (we will do a general analysis).

Examining the figures 2 and 3, for $h = 1$, we conclude that the direct regression model had mild underperformance from $1998{:}Q1$ to $2006{:}Q4$, conversely it had a superior performance in the rest of the sample. The other two models exhibit a really good performance from $2006{:}Q1$ to $2010{:}Q1$ (before that, the $Net\ SSE$ is almost zero for both models).

When we consider the forecast horizons of 8 and 12 quarters, the competing models only beat the historical mean model from $2008{:}Q1$ (approximately) and the models that include the rent-price ratio start to exhibit a better performance than the model with only the autoregressive component. This pattern is obvious when we analyze the figures 2 and 3, for $h = 18$ and $h = 24$, where the cumulative SSE difference between the AR and the benchmark model is constantly negative. From $2008{:}\,Q1$, the direct regression curve grows almost exponentially, evidencing the predictive power of the rent-price ratio.

8. FUTURE RESEARCH

In this section, we propose ideas for future research, some of which are improvements to our paper.

An obvious gap in our study is the lack of robust critical values to statistically assess the quality of forecasts produced by the autoregressive models (which are also nested models). To solve this problem, it should be defined a bootstrap procedure that generate this critical values under the null hypothesis of no predictability.

Additionally, it would be interesting to extend this research to other markets, namely the bonds, the treasuries, the sovereign debt or the foreign debt markets. There are relatively few papers about predicting returns of these markets, out-of sample. Another suggestion would be to reproduce this study using data for the European markets instead of to the U.S. markets.

From a financial perspective, and since our results reveal that the valuation ratios can be used to successfully predict the aggregate returns for stock and housing markets, an academic researcher could also explore the existence of profitable investment strategies.

9. CONCLUSION

In this dissertation, we found evidence that the known in-sample pattern of return predictability across markets holds out-of-sample. Considering the stock and the housing market, we verify that there are gains when we use valuation ratios to predict aggregate returns. The relatively powerful out-of-sample tests applied corroborate these results. In particular, for the stock market, we found that the direct regression model beats the benchmark, for all horizons. Additionally, we note that the dividend–price ratio's ability to predict the aggregate returns improves at longer horizons. For the housing market, only the models that contain the rent-price ratio consistently exhibit $MSFE$ ratios lower than 1, for all horizons.

The sample dependence identified through the analysis of $Net\ SSE$ charts, for both markets, deserves further attention. It will be interesting to investigate this issue in detail, notably by examining the stability of the forecast function while linking it to specific events affecting these markets or, more generally, the U.S. economy.

REFERENCES

[1] Andrews, M. J., Minford, A. P. L. and Riley, J. (1996). "On comparing macroeconomic forecasts using forecast encompassing tests". *Oxford Bulletin of Economics and Statistics*, 58, pp. 279 – 305.

[2] Akaike, H. (1974). "A new look at the statistical model identification''. *IEEE Transactions on Automatic Control*, 19, pp. 716 – 722.

[3] Calhoun, C. (1996). "OFHEO house price indices: HPI technical description". From: http://www.ofheo.gov (Retrieved September 17, 2012)

[4] Campbell, J. Y. and Shiller, R. J. (1991). "Yield spreads and interest rate movements: A bird's eye view". *Review of Economic Studies*, 58, pp. 495– 514.

[5] Campbell, J. Y. and Thompson, S. B. (2008). "Predicting excess stock returns out of sample: Can anything beat the historical average?". *Review of Financial Studies*, 21, pp. 1509— 1531.

[6] Campbell, S. D., Davis, M. A., Gallin, J. and Martin, R. F. (2009). "What Moves Housing Markets: A Trend and Variance Decomposition of the Rent-price Ratio". *Journal of Urban Economics*, 66, pp. 90– 102.

[7] Case, K. E. and Shiller, R. J. (1990). "Forecasting Prices and Excess Returns in the Housing Market". *Real Estate Economics*, 18, pp. 253– 273.

[8] Clark, T. E. and McCracken, M. W. (2001). "Tests of Equal Forecast Accuracy and Encompassing for Nested Models". *Journal of Econometrics,* 105, pp. 85– 110.

[9] Clark, T. E., and McCracken, M. W. (2005). "Evaluating direct multi-step forecasts". *Econometric Reviews*, 24, pp. 369 – 404.

[10] Cochrane, J. (2011). “Discount Rates”. *Journal of Finance*, 66, pp. 1047 − 1108.

[11] Clements, M. P. and Harvey, D. I. (2009). “Forecast Combination and Encompassing”. In: Mills, T.C. and Patterson, K., (Eds.) Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics Palgrave Macmillan, pp. 169 − 198.

[12] Diebold, F. X. and Mariano, R. S. (1995). “Comparing predictive accuracy”. *Journal of Business and Economic Statistics*, 13, pp. 253 − 263.

[13] Fair, R. C. and Shiller, R. J. (1989). “The informational content of ex ante forecasts”. *Review of Economics and Statistics*, 71, pp. 325 − 331.

[14] Fama, E. F. (1984). “Forward and spot exchange rates”. *Journal of Monetary Economics*, 14, pp. 319 − 338.

[15] Fama, E. F. (1986). “Term premiums and default premiums in money markets”. *Journal of Financial Economics*, 17, pp. 175– 196.

[16] Fama, E. F. and Bliss, R. R. (1987). “The information in long-maturity forward rates”. *American Economic Review*, 77, pp. 680– 692.

[17] Fama, E. F. and French, K. R. (1988). “Dividend yields and expected stock returns”. *Journal of Financial Economics*, 22, pp. 3– 25.

[18] Fama, E. F. and French, K. R. (1989). “Business conditions and expected returns on stocks and bonds”. *Journal of Financial Economics*, 25, pp. 23– 49.

[19] Gallin, J. (2008). “The long-run relationship between house prices and rents”. *Real Estate Economics*, 36 (4), pp. 635 − 658.

[20] Gourinchas, P. and Rey, H. (2007). “International financial adjustment”. *Journal of Political Economy*, 115, pp. 665– 703.

[21] Goyal, A. and Welch, I. (2003). “Predicting the equity premium with dividend ratios”. *Management Science*, 49, pp. 639– 654.

[22] Goyal, A. and Welch, I. (2008). “A Comprehensive Look at the Empirical Performance of Equity Premium Predictions”. *Review of Financial Studies*, 21, pp. 1455– 508.

[23] Hansen, L. P. and Hodrick, R. J. (1980). “Forward exchange rates as optimal predictors of future spot rates: An econometric analysis”. *Journal of Political Economy*, 88, pp. 829– 853.

[24] Harvey, D. I., Leybourne, S. and Newbold, P. (1997). “Testing the equality of prediction mean squared errors”. *International Journal of Forecasting*, 13, pp. 281 − 291.

[25] Harvey, D. I., Leybourne, S. and Newbold, P. (1998). “Tests for forecast encompassing”. *Journal of Business and Economic Statistics*, 16, pp. 254 − 259.

[26] Kilian, L. (1999). “Exchange rates and monetary fundamentals: what do we learn from long–horizon regressions?”. *Journal of Applied Econometrics*, 14(5), pp. 491– 510.

[27] Lettau, M. and Ludvigson, S. C. (2001). “Consumption, Aggregate Wealth, and Expected Stock Returns”. *Journal of Finance*, 56, pp. 815– 49.

[28] Mark, N. C. (1995). “Exchange rates and fundamentals: evidence on long–horizon predictability”. *American Economic Review*, 85(1), pp. 201– 18.

[29] McCracken, M. W. (2007). “Asymptotics for Out of Sample Tests of Granger Causality”. *Journal of* Econometrics, 140, pp. 719– 52.

[30] Newey, W. K. and West, K. D. (1987). “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”. *Econometrica*, 55, pp. 703– 708.

[31] Piazzesi, M. and Swanson, E. (2008). "Futures prices as risk-adjusted forecasts of monetary policy". *Journal of Monetary Economics*, 55, pp. 677– 691.

[32] Plazzi, A., Torous, W. and Valkanov, R. (2010). "Expected Returns and the Expected Growth in Rents of Commercial Real Estate". *Review of Financial Studies*, 23 (9), pp. 3469 – 3519.

[33] Rapach, D. and Wohar, M. (2006). "In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining". *Journal of Empirical Finance*, 13, pp. 231– 247.

[34] Rapach, D., Strauss, J. and Zhou, G. (2010). "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *Review of Financial Studies*, 23(2), pp. 821 – 862.

[35] Richardson, M. and Stock, J. H. (1989). "Drawing inferences from statistics based on multiyear asset returns". *Journal of Financial Economics*, 25, pp. 323– 348.

[36] Stambaugh, R. F. (1986). "Biases in regressions with lagged stochastic regressors". Graduate School of Business, University of Chicago, Working Paper 156.

[37] Stambaugh, R. F. (1999). "Predictive regressions". *Journal of Financial Economics*, 54, pp. 375– 421.

[38] Valle e Azevedo, J. (2011). "A Multivariate Band-Pass filter for Economic Time Series", *Journal of the Royal Statistical Society (C)*, 60(1), pp. 1 – 30.

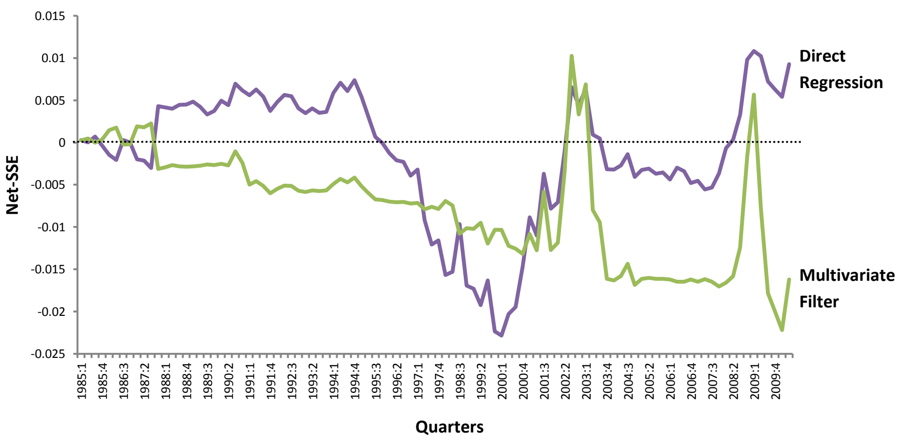
[39] Valle e Azevedo, J. and Pereira, A. (2012). "Forecasting Inflation (and the Business Cycle?) with Monetary Aggregates", Bank of Portugal, Working Paper 24 – 2010.

APPENDIX A – FIGURES

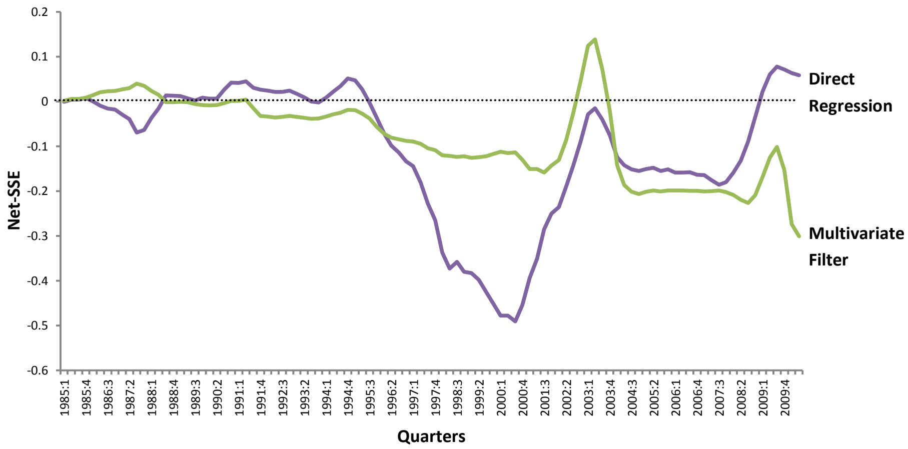
Figure 1. Cumulative SSE Difference charts for the Stock Market (horizons of 1, 4, 8, 12, 18 and 24 quarters).

Notes: This figure plots the $Net\ SSE_h$ for $h = 1, 4, 8, 12, 20, 24$, that is, the cumulative squared forecast errors of the benchmark model (the historical mean) minus the squared forecast errors of the competing model, for each horizon. We consider two competing models: the direct regression model (purple curve) and the multivariate filter model (green curve). A positive value means that the competing model has outperformed the benchmark model. A positive slope indicates that the competing model had lower forecasting error than the historical mean model, in a given quarter.

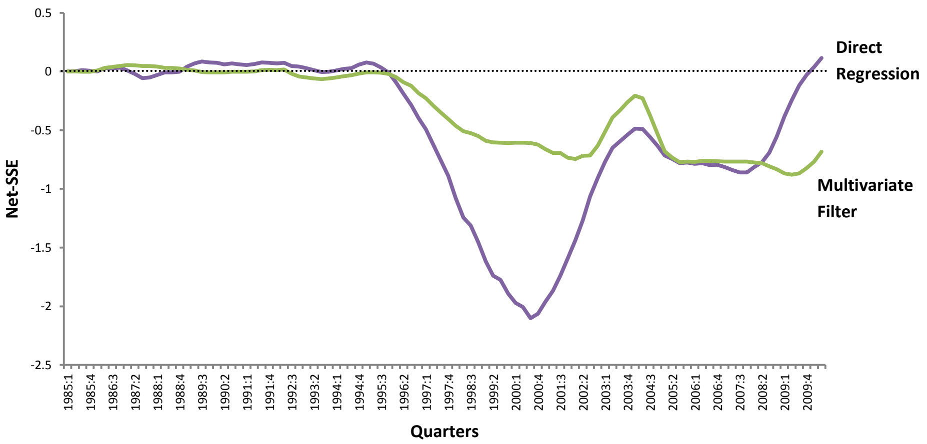
Cumulative SSE Difference for $h = 1$



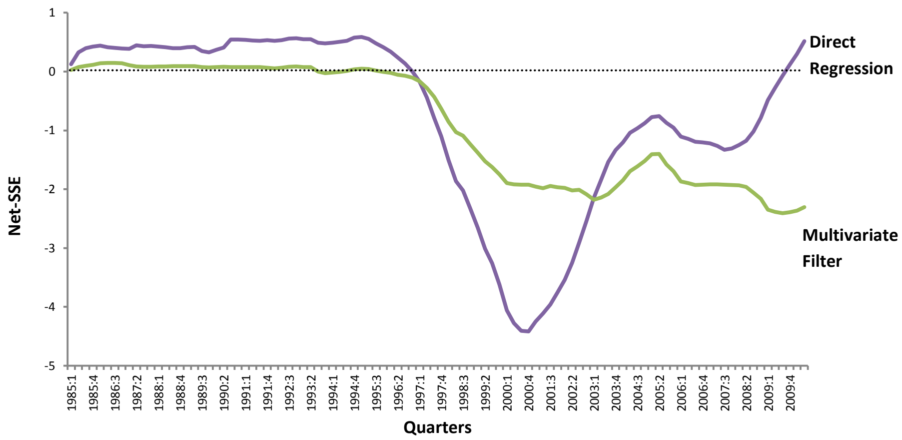
Cumulative SSE Difference for $h = 4$



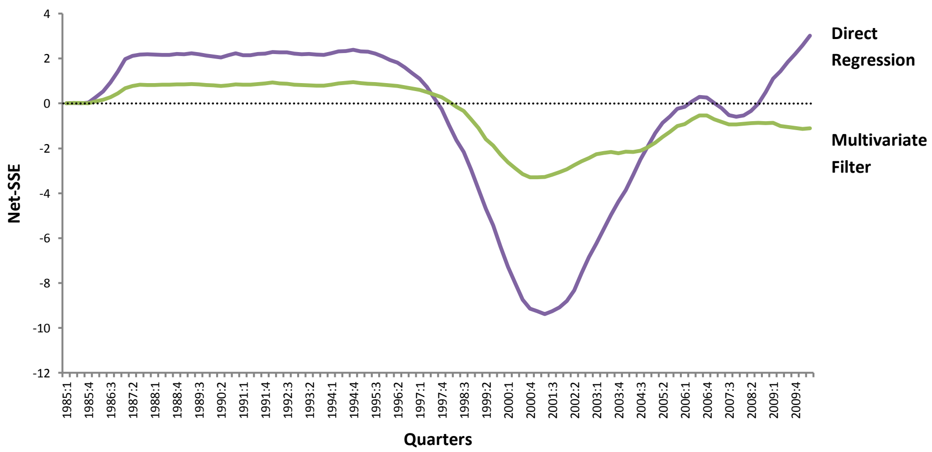
Cumulative SSE Difference for $h = 8$



Cumulative SSE Difference for $h = 12$.



Cumulative SSE Difference for $h = 18$



Cumulative SSE Difference for $h = 24$

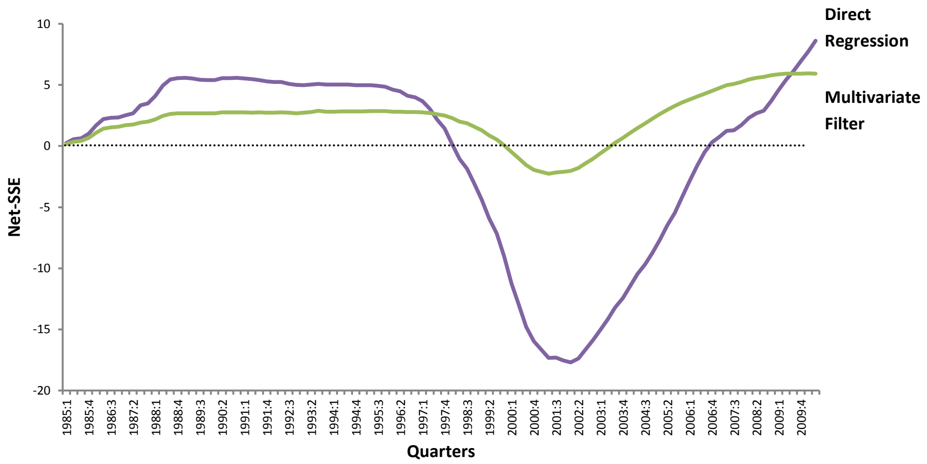
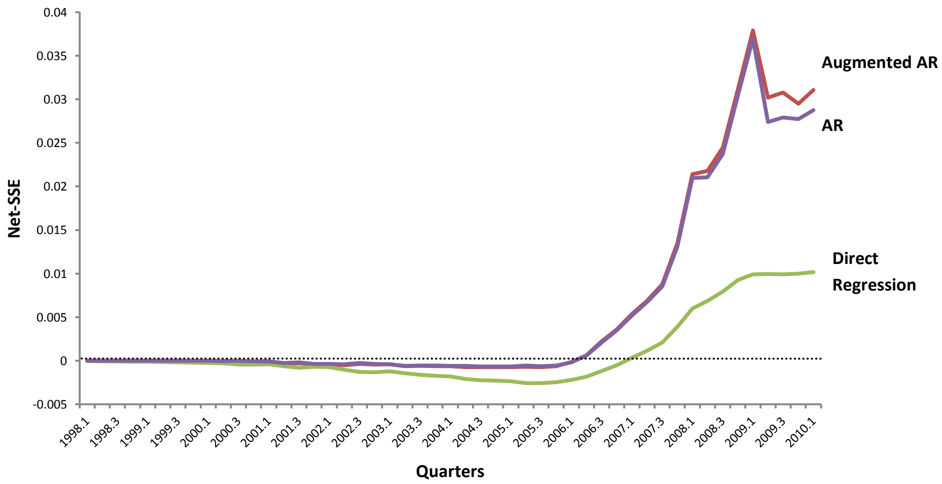


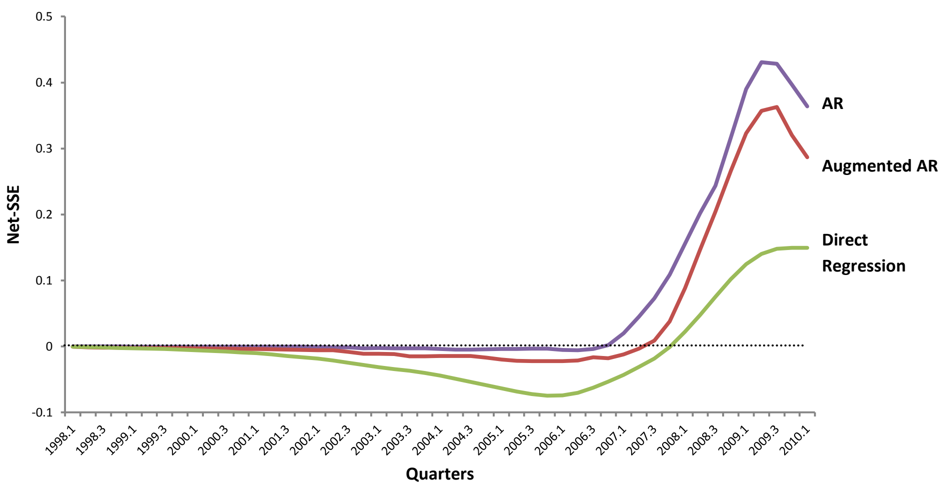
Figure 2. Cumulative SSE Difference charts for the Housing Market using CSW data (horizons of 1, 4, 8, 12, 18 and 24 quarters).

Notes: This figure plots the $Net\ SSE_h$ for $h = 1, 4, 8, 12, 20, 24$, that is, the cumulative squared forecast errors of the benchmark model (the historical mean) minus the squared forecast errors of the competing model, for each horizon. We consider three competing models: the direct regression model without lags (green curve), the direct autoregressive (AR) model (purple curve) and the direct augmented AR model (red curve). The lag order of the models with autoregressive component is determined by the Akaike Information Criterion ($p_{max} = 4$). A positive value means that the competing model has outperformed the benchmark model. A positive slope indicates that the competing model had lower forecasting error than the historical mean model, in a given quarter.

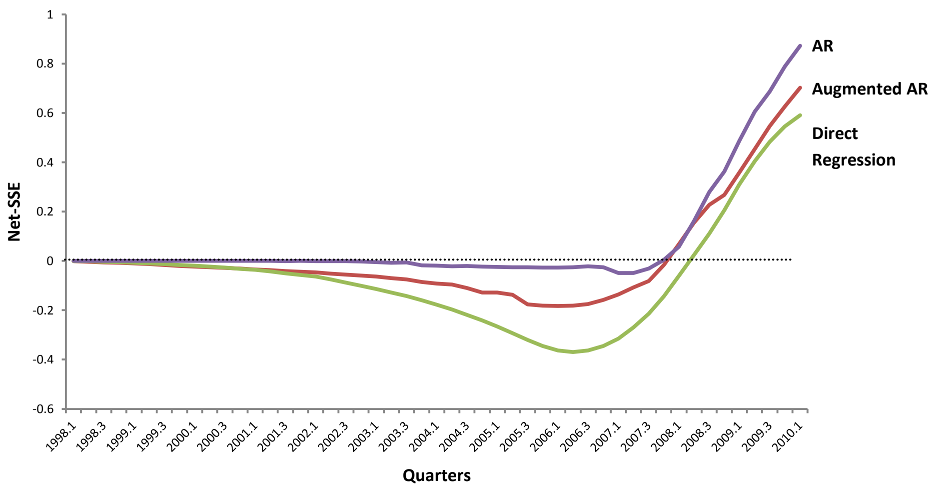
Cumulative SSE Difference for $h = 1$



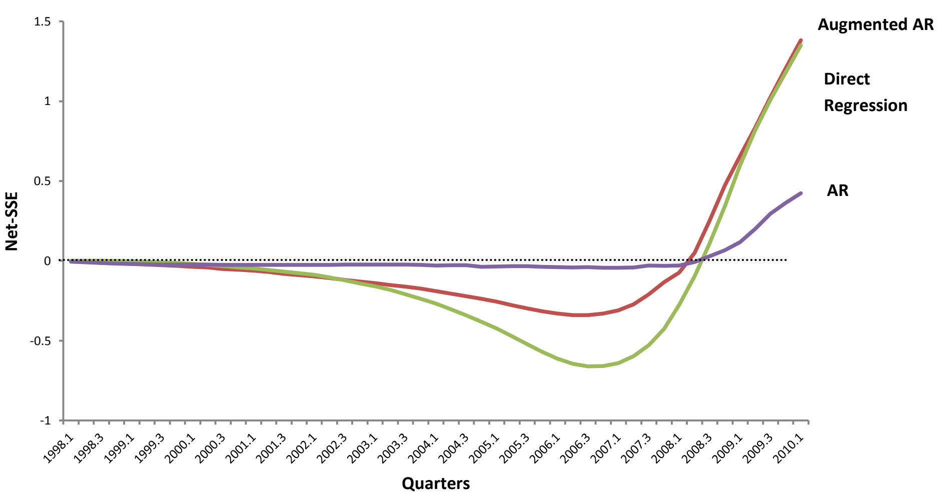
Cumulative SSE Difference for $h = 4$

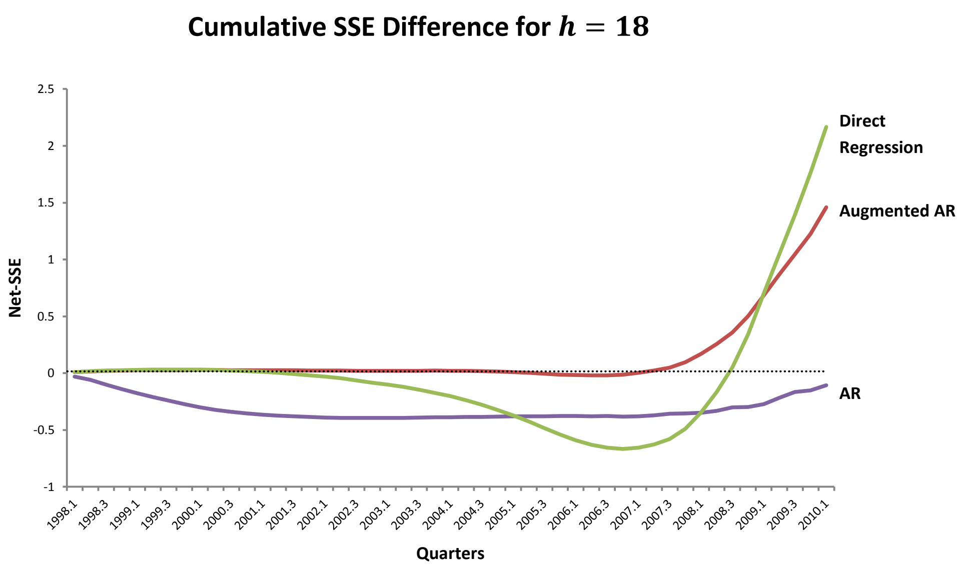


Cumulative SSE Difference for $h = 8$



Cumulative SSE Difference for $h = 12$





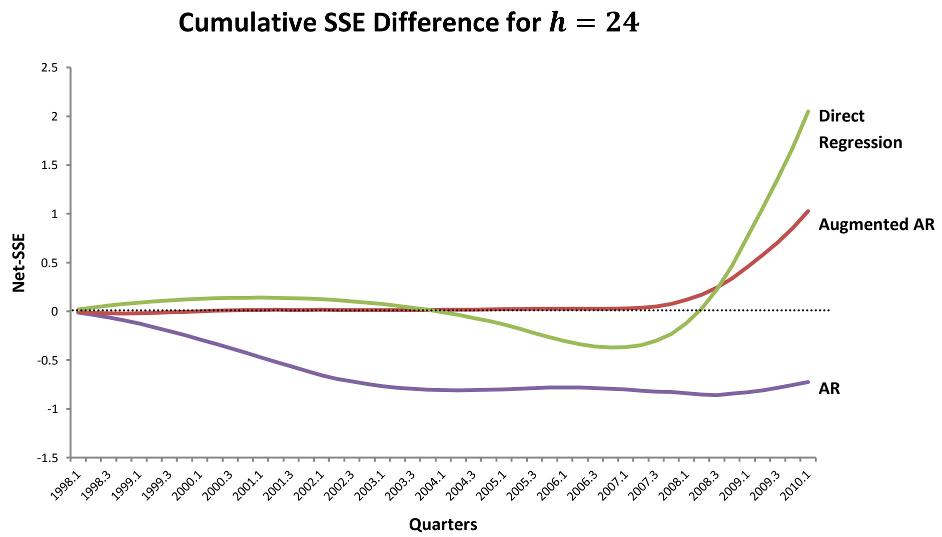
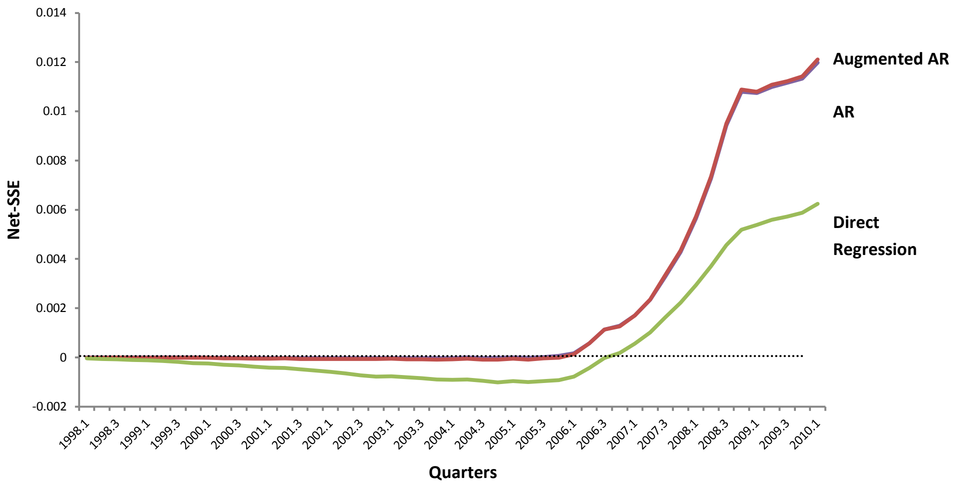


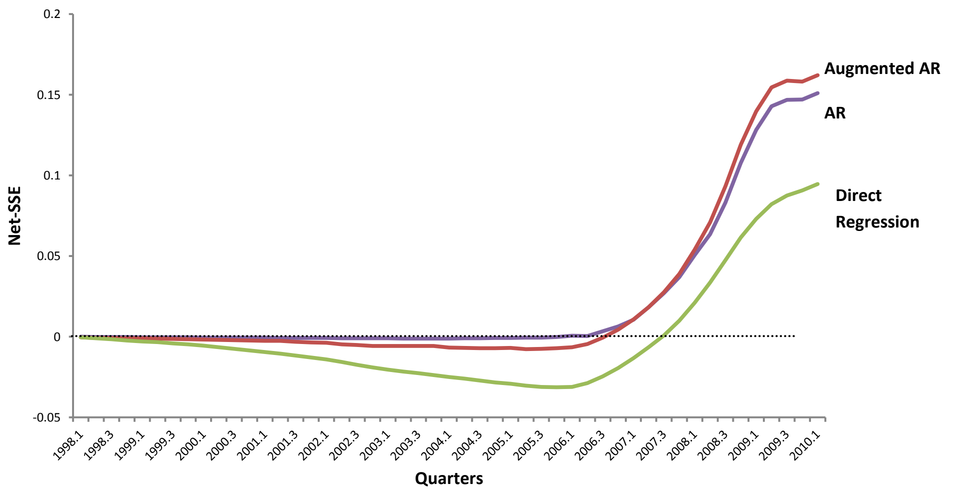
Figure 3. Cumulative SSE Difference charts for the Housing Market using OFHEO data (horizons of 1, 4, 8, 12, 18 and 24 quarters).

Notes: This figure plots the $Net\ SSE_h$ for $h = 1, 4, 8, 12, 20, 24$, that is, the cumulative squared forecast errors of the benchmark model (the historical mean) minus the squared forecast errors of the competing model, for each horizon. We consider three competing models: the direct regression model without lags (green curve), the direct autoregressive (AR) model (purple curve) and the direct augmented AR model (red curve). The lag order of the models with autoregressive component is determined by the Akaike Information Criterion ($p_{max} = 4$). A positive value means that the competing model has outperformed the benchmark model. A positive slope indicates that the competing model had lower forecasting error than the historical mean model, in a given quarter.

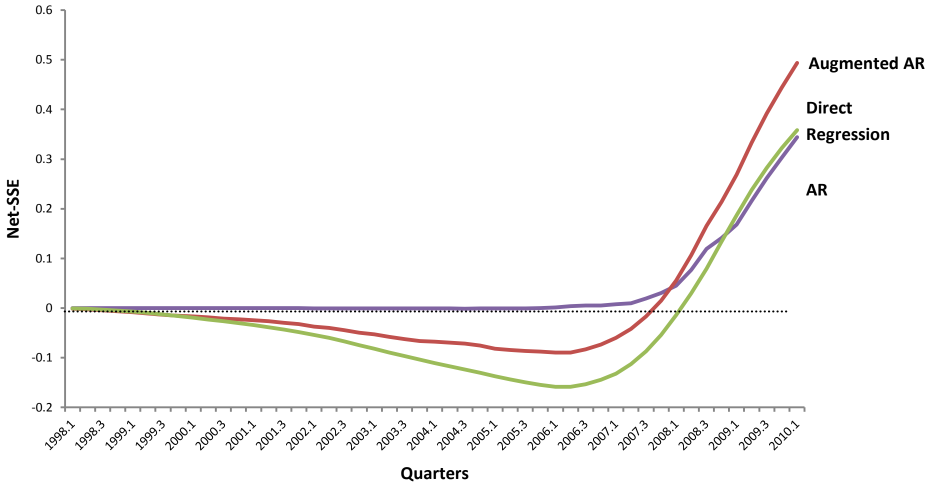
Cumulative SSE Difference for $h = 1$.



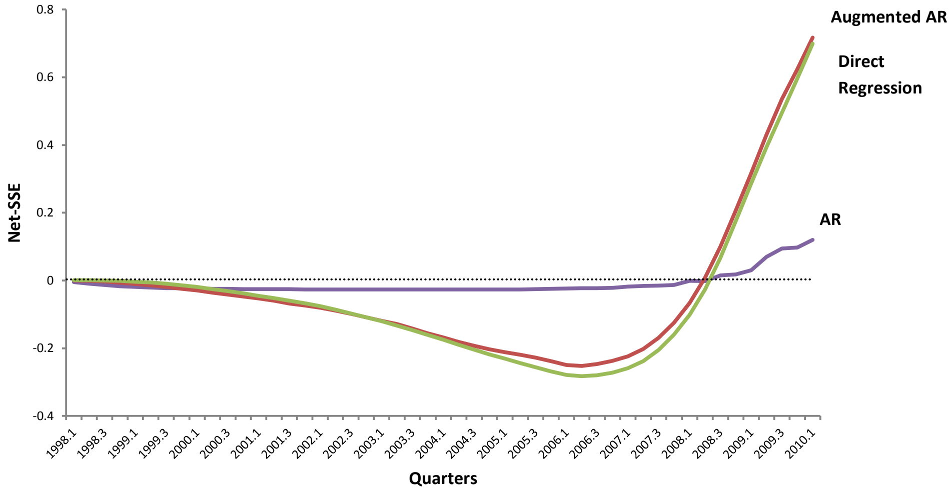
Cumulative SSE Difference for $h = 4$.



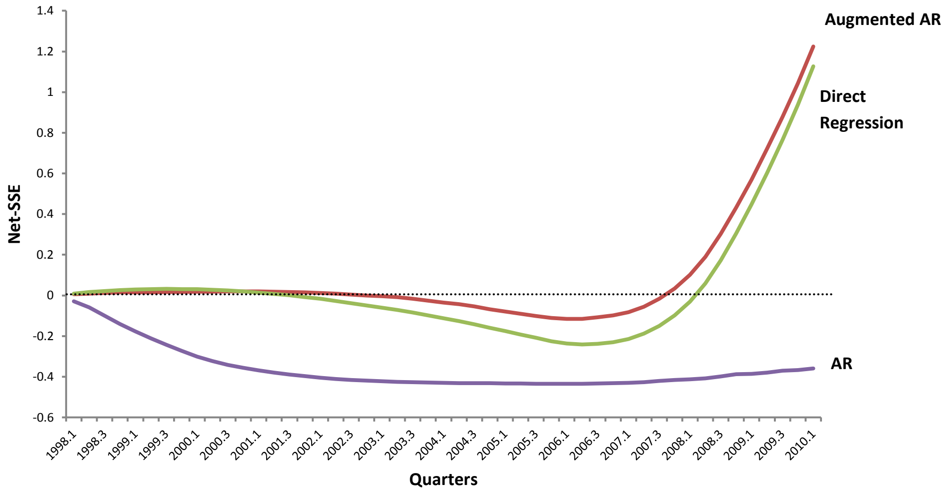
Cumulative SSE Difference for $h = 8$.



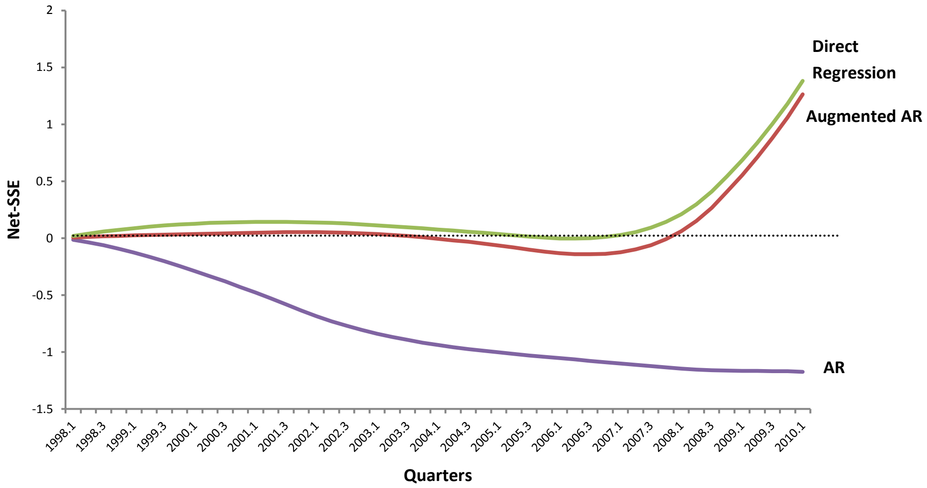
Cumulative SSE Difference for $h = 12$.



Cumulative SSE Difference for $h = 18$.



Cumulative SSE Difference for $h = 24$.



APPENDIX B – TABLES

Table I Descriptive Statistics

	Stock Market data Sample period 1947Q1-2010Q2		Housing Market data Sample period 1960Q1-2010Q1			
	Equity Premium	DP Ratio	Case-Shiller data		OFHEO data	
			Returns	RP Ratio	Returns	RP Ratio
Mean	2,751	0,872	5,316	4,994	5,390	5,027
Sdev.	8,213	0,365	1,750	0,649	1,179	0,559
Median	3,574	0,831	5,695	5,000	5,643	4,995
Min	-24,650	0,258	-3,483	3,098	0,749	3,633
Max	24,376	2,492	7,789	6,083	7,538	6,083

Notes: The data are described in Section 4. All variables are in percentage.

Table II In-Sample Regressions (horizons of 1, 4, 6, 8, 12, 18, 20 and 24 quarters).

	Stock Market data Sample period 1947:Q1-2010:Q2				Housing Market data Sample period 1960:Q1-2010:Q1							
Horizon (quarters)					Case-Shiller data				OFHEO data			
	$\hat{\beta}$	*t-stat*	R^2 %	Adj. R^2 %	$\hat{\beta}$	*t-stat*	R^2%	Adj. R^2%	$\hat{\beta}$	*t-stat*	R^2%	Adj. R^2%
1	3,80	(2,89)	2,85	2,46	1,27	(5,24)	22,30	21,91	1,21	(8,47)	32,95	32,61
4	16,57	(3,14)	11,23	10,88	5,90	(2,88)	38,80	38,48	5,39	(4,73)	45,47	45,19
6	24,42	(3,09)	15,60	15,26	9,44	(3,09)	47,53	47,26	8,44	(5,02)	52,94	52,70
8	32,08	(3,38)	19,97	19,64	12,86	(3,49)	54,85	54,61	11,41	(5,52)	59,06	58,84
12	46,35	(3,97)	25,38	25,07	18,84	(4,62)	64,31	64,12	16,67	(6,81)	67,00	66,82
18	74,17	(5,17)	33,05	32,76	25,27	(5,73)	67,05	66,87	22,73	(7,98)	69,76	69,60
20	90,46	(5,69)	36,82	36,54	26,87	(5,72)	65,56	65,37	24,37	(7,97)	69,15	68,98
24	121,28	(6,52)	44,47	44,23	29,68	(5,44)	61,46	61,24	27,18	(7,63)	66,90	66,71

Notes: The regression equation is $\boldsymbol{y_{t+h} = \alpha + \beta x_t + u_{t+h}}$, where $\boldsymbol{y_{t+h}}$ and $\boldsymbol{x_t}$ are the equity premium and the dividend-price ratio, respectively, for the Stock Market; and the log returns and the rent-price ratio for the housing market. The data are described in Section 4. *t-stat* denotes the Newey-West adjusted $\boldsymbol{t - statistic}$.

Table III MSFE ratios and Equal Accuracy test results for the stock market (horizons of 1, 4, 6, 8, 12, 18, 20 and 24 quarters).

Horizon (quarters)	1	4	6	8	12	18	20	24
Direct regression without lags								
modified MSFE-t (bootstrap critical values)	0,988	0,984	0,976	0,987	0,969	0,909	0,883	0,883
MSFE-F (bootstrap critical values)	0,988*	0,984**	0,976**	0,987*	0,969***	0,909***	0,883***	0,883***
Direct regression (p=2)	0,991	1,001	0,999	1,008	0,996	0,946	0,934	0,934
Direct regression (pmax=4))	1,001	1,005	1,006	1,011	1,009	0,985	0,977	0,977
Multivariate filter (p=100, cut-off=32, M=50)								
modified MSFE-t (t(n-1) critical values)								
without indicators	1,028	1,122	1,140	1,168	1,230	1,152	1,099	1,099
with dividend-price ratio	1,020	1,082	1,070	1,080	1,140	1,034	0,983	0,983

Notes: This table reports the $MSFE$ for each model, considering $h = 1, 4, 6, 8, 12, 10, 20, 24$, and the Equal Accuracy test results for the direct regression model (without lags) and for the multivariate filter model (H_0: $MSFE = 1$, that is, under the null hypothesis the benchmark model (historical mean) predicts better). The critical values for the direct regression model are generated using a bootstrap procedure. For the other competing model (which is non-nested), critical values from the Student's t distribution with $(N - 1)$ degrees of freedom are used (N is the number of forecast errors). Predictions were generated for the period $1985{:}Q1 - 2010{:}Q2$. Significance levels at 10%, 5%, and 1% are denoted by one, two, and three stars, respectively. The data are described in detail in Section 4.

Table IV Forecast Encompassing test results for the stock market (horizons of 1, 4, 6, 8, 12, 18, 20 and 24 quarters).

Horizon (quarters)	1	4	6	8	12	18	20	24
Direct regression (without lags)								
modified ENC-t (bootstrap critical values)	1,361*	1,184	0,619	0,118	-0,062	0,507	0,946	1,845*
ENC-F (bootstrap critical values)	0,813	2,224**	1,870**	0,471	-0,361	2,615***	4,375***	5,892***
Multivariate filter (p=100, cut-off=32, M=50)								
modified MSFE-t (t(n-1) critical values)								
without indicators	0,522	0,074	-0,283	-1,014	-2,064	-2,699	-3,207	-3,275
with dividend-price ratio	0,665	0,294	-0,085	-1,419	-2,699	-1,713	-1,275	-0,223

Notes: This table reports the observed values of the test statistics employed to conduct the Forecast Encompassing test for the direct regression model and the multivariate filter model. Under the null hypothesis, the benchmark model (the historical mean model) encompasses the competing model. The test was applied assuming that the forecasts are biased and inefficient. The horizons of $h = 1, 4, 6, 8, 12, 10, 20, 24$ quarters are considered. The critical values for the direct regression model are generated using a bootstrap procedure. For the other competing model (which is non-nested), critical values from the Student's t distribution with $(N - 1)$ degrees of freedom are used (N is the number of forecast errors). Predictions were generated for the period $1985{:}Q1 - 2010{:}Q2$. Significance levels at 10%, 5%, and 1% are denoted by one, two, and three stars, respectively. The data are described in detail in Section 4.

Table V MSFE ratios and Equal Accuracy test results for the housing market using CSW data (horizons of 1, 4, 6, 8, 12, 18, 20 and 24 quarters).

Horizon (quarters)	1	4	6	8	12	18	20	24
Direct autoregression (p=2)	0,453	0,358	0,448	0,572	0,864	1,024	1,085	1,085
Direct autoregression (pmax=4)	0,391	0,362	0,462	0,554	0,868	1,029	1,085	1,085
Direct augmented AR (pmax=4)	0,342	0,315	0,344	0,414	0,561	0,401	0,373	0,373
Direct regression (without lags)								
modified MSFE-t (bootstrap critical values)	0,785***	0,738*	0,724	0,697	0,579	0,417	0,401	0,401
MSFE-F (bootstrap critical values)	0,785***	0,738***	0,724***	0,697***	0,579***	0,417***	0,401***	0,401***
Direct regression (p=2)	0,466	0,465	0,397	0,488	0,559	0,409	0,386	0,386
Mult. filter (p=90, cut-off=32, M=40)								
modified MSFE-t (t(n-1) c.v.)								
without indicators	0,554**	0,602	0,684	0,794	0,945	1,025	1,038	1,038
with rent-price ratio	0,541**	0,553	0,627	0,716	0,824	0,829	0,827	0,827

Notes: This table reports the $MSFE$ for each model, considering $h = 1, 4, 6, 8, 12, 10, 20, 24$, and the Equal Accuracy test results for the direct regression model (without lags) and for the multivariate filter model (H_0: $MSFE = 1$, that is, under the null hypothesis the benchmark model (historical mean) predicts better). The critical values for the direct regression model are generated using a bootstrap procedure. For the other competing model (which is non-nested), critical values from the Student's t distribution with $(N - 1)$ degrees of freedom are used (N is the number of forecast errors). Predictions were generated for the period $1998{:}Q1 - 2010{:}Q1$. Significance levels at 10%, 5%, and 1% are denoted by one, two, and three stars, respectively. The data are described in detail in Section 4.

Table VI MSFE ratios and Equal Accuracy test results for the housing market using OFHEO data (horizons of 1, 4, 6, 8, 12, 18, 20 and 24 quarters).

Horizon (quarters)	**1**	**4**	**6**	**8**	**12**	**18**	**20**	**24**
Direct autoregression (p=2)	0,110	0,230	0,375	0,603	0,930	1,215	1,386	1,660
Direct autoregression (pmax=4)	0,118	0,229	0,355	0,500	0,897	1,227	1,386	1,647
Direct augmented AR (pmax=4)	0,108	0,172	0,204	0,283	0,388	0,227	0,239	0,302
Direct regression (without lags)								
modified MSFE-t (bootstrap critical values)	0,541***	0,516*	0,502*	0,479*	0,403	0,288	0,264*	0,238*
MSFE-F (bootstrap critical values)	0,541***	0,516***	0,502***	0,479***	0,403***	0,288***	0,264***	0,238***
Direct regression (p=2)	0,162***	0,214*	0,271	0,347	0,400	0,291	0,260	0,224
Mult. filter (p=90, cut-off=32, M=40)								
modified MSFE-t (t(n-1) c.v.)								
without indicators	0,297***	0,472*	0,613	0,760	0,970	1,131	1,162	1,17
with rent-price ratio	0,255***	0,351	0,433	0,518	0,630	0,722	0,760	0,829

Notes: See Table VI. Significance levels at 10%, 5%, and 1% are denoted by one, two, and three stars, respectively.

Table VII Forecast Encompassing test results for the housing market using CSW data (horizons of 1, 4, 6, 8, 12, 18, 20 and 24 quarters).

Horizon (quarters)	1	4	6	8	12	18	20	24
Direct regression (without lags)								
modified ENC-t (bootstrap critical values)	3,308***	1,499	1,447	1,351	1,652	1,253	1,211	1,156
ENC-F (bootstrap critical values)	3,248***	3,055***	2,492***	3,620***	11,272***	28,392***	27,639***	20,581***
Mult. filter (p=90, cut-off=32, M=40)								
modified MSFE-t (t(n-1) c.v.)								
without indicators	2,995***	2,167**	1,764**	1,639*	1,342*	1,059	1,088	1,322*
with rent-price ratio	3,715***	2,053**	1,696**	1,607*	1,441*	1,212	1,160	1,101

Notes: This table reports the observed values of the test statistics employed to conduct the Forecast Encompassing test for the direct regression model and the multivariate filter model. Under the null hypothesis, the benchmark model (the historical mean model) encompasses the competing model. The test was applied assuming that the forecasts are biased and inefficient. The horizons of $h = 1, 4, 6, 8, 12, 10, 20, 24$ quarters are considered. The critical values for the direct regression model are generated using a bootstrap procedure. For the other competing model (which is non-nested), critical values from the Student's t distribution with $(N-1)$ degrees of freedom are used (N is the number of forecast errors). Predictions were generated for the period $1998{:}Q1 - 2010{:}Q1$. Significance levels at 10%, 5%, and 1% are denoted by one, two, and three stars, respectively. The data are described in detail in Section 4.

Table VIII Forecast Encompassing test results for the housing market using OFHEO data (horizons of 1, 4, 6, 8, 12, 18, 20 and 24 quarters).

Horizon (quarters)	1	4	6	8	12	18	20	24
Direct regression (without lags)								
modified ENC-t (bootstrap critical values)	5,343***	2,451**	1,949*	1,404	3,763***	1,787	1,606	1,421
ENC-F (bootstrap critical values)	4,225***	5,591***	5,368***	4,848***	11,585***	32,739***	33,88***	27,802***
Mult. filter (p=90, cut-off=32, M=40)								
modified MSFE-t (t(n-1) c.v.)								
without indicators	3,565***	1,692**	1,247	1,477*	1,706**	1,139	1,321*	1,581*
with rent-price ratio	4,076***	1,989**	1,634*	1,688**	3,575***	1,444*	1,348*	1,291

Notes: See Table VII. Significance levels at 10%, 5%, and 1% are denoted by one, two, and three stars, respectively.

APPENDIX C – DATA DESCRIPTION

This Appendix provides additional information about the data used in the estimation of models (Section 4 contains the main information about the data).

Stock Market:

For the stock market, our dependent variable is the $CRSP$ value-weighted return less the 3-month Treasury bill return. Specifically, we use the following series from the $CRSP$ database:

Name of Series	Description
Vwretd	Return on the Value-Weighted Index – contains the returns, including all distributions, on a value-weighted market portfolio (excluding ADRs).
Vwretx	Return on the Value-Weighted Index – contains returns, excluding all dividends, on a value-weighted market portfolio (excluding ADRs).
T90ret	3-month Treasury bill return.

Denoting the return including dividends (held from the beginning of $t-1$ to the beginning of t) by r_t^* and the return excluding dividends (held from the beginning of $t-1$ to the beginning of t) by r_t, we can formally write:

$$r_t^* = \frac{(P_t + D_t) - P_{t-1}}{P_{t-1}}$$

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}},$$

where P_{t-1} denotes the price of a stock portfolio at the beginning of period $t-1$, P_t denotes the price of a stock portfolio at the beginning of period t and D_t denotes the total dividends paid on the portfolio during period t.

Housing market:

As mentioned in Section 4, for the housing market, we use two different data sources. The Case-Shiller-Weiss price data corresponds to a national home price index that is calculated from data on repeat sales of single-family homes. The OFHEO index is a national house price index for single-family detached properties that considers data on conventional conforming mortgage transactions (these data are obtained from the Federal Home Loan Mortgage Corporation (Freddie Mac) and the Federal National Mortgage Association (Fannie Mae)).[15]

Although these indexes apply the same repeat-valuations approach, there are some differences between them. In particular, they consider a different geographic coverage (only the OFHEO index is calculated using data from all states); and a distinct weighting method (the CSW index is value-weighted (the price trends for more expensive homes have greater influence on estimated price changes than other homes) while the OFHEO index weights price trends equally for all properties).

[15] Calhoun (1996).