PREDICTING AGGREGATE RETURNS USING VALUATION RATIOS OUT-OF-SAMPLE*

Ana Sequeira**

ABSTRACT

It is well established that valuation ratios (indicators of the financial market situation) provide, in-sample, relevant signals regarding future returns on assets. Specifically, periods of high prices, relative to dividends, are proceeded by years of low returns; and periods of low prices, relative to dividends, precede years of high returns. This pattern of predictability is pervasive across financial markets. In this paper, 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 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 in real-time stock and housing returns, respectively. Nevertheless, we note that these findings may be sample dependent. Especially for the stock market, the sample's ending data, including the recent financial crisis, may be responsible for the good results.

1. Introduction

Predicting returns is one of the most discussed topics in the academic financial world. Cochrane (2011) summarizes the evidence that shows the existence of a pattern of predictability that is pervasive across markets (stocks, bonds, houses, foreign exchange and sovereign debt) and concludes (in-sample) that valuation ratios predict excess returns, instead of future cashflows. In this context, for the stock market, one concludes that the dividend-price ratio predicts returns and does not predict dividend growth. Moreover, low dividend-price ratios signal low future returns; and high dividend-price ratios signal high future returns. For the housing market, the argument is similar: high prices, relative to rents, imply low returns; do not signal the permanent increase of rents or prices.

In this paper, we intend to verify whether this pervasive phenomenon holds out-of-sample, that is, whether a forecaster would be able to predict excess returns systematically, if he stood at the forecast moment without further information (simulating the production of forecasts in real-time). We focus on the housing market (there are relatively few studies about predicting the housing returns) and use the stock market analysis as an important reference.

Among the studies that examine the predictability of housing returns, Case and Shiller (1990) show

---

* This paper is a summary of my master thesis (which contains all the details), developed in Economics and Research Department (Banco de Portugal). I would like to thank my advisor João Valle e Azevedo for his permanent availability, ideas and relevant observations. I am also grateful to Ana Pereira for providing me her Mathematica programming codes and clarifying some questions, and Martín Saldías for his important suggestions. The opinions expressed in the article are those of the author and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. Any errors and omissions are his sole responsibility.

** Banco de Portugal, Economics and Research Department.

that the rent-price ratio has a good performance when used to predict (in-sample) the housing excess returns. Lately, Plazzi et al. (2010) conclude (also, 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. Goyal and Welch (2003, 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 (stock returns over the return of a short-term risk-free interest rate) and conclude that almost all models produce poor results out-of-sample. On the other hand, Rapach and Wohar (2006) find that several financial variables have a good in-sample and out-of-sample ability to predict stock returns. Rapach et al. (2010) also find significant out-of-sample gains using forecast combining methods.

Our results show that there is statistical evidence supporting the extension of the in-sample results to an out-of-sample framework in both markets. Especially for the housing market, we conclude that the rent-price ratio has a huge ability for predicting returns. Given the lack of out-of-sample studies for the housing market, these findings are a contribution to the literature.

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 describe the data used to obtain the empirical results. Section 3 reports the in-sample results while Section 4 exposes the econometric methodology. In Section 5 we discuss our main findings and in Section 6 we present the conclusions and ideas for future research.

2. Data

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 (sample with \( T = 254 \) observations) and our dependent variable is the equity premium from holding stocks (represented in an index) 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 Center for Research in Security Prices (CRSP) value-weighted return and the 3-month Treasury bill (as a proxy for the risk-free rate). The dividend-price ratio is the financial variable which potentially predicts the equity premium.

Housing Market:

In our applications for the housing market, we use quarterly data from 1960:Q1 to 2010:Q1 (sample with \( T = 201 \) observations). We consider two house price indexes: the price index computed by Case, Shiller and Weiss and the “purchase-only” price index released by the Office of Federal Housing Enterprise Oversight (OFHED). Since the results were very similar, we chose to present only the conclusions for the data from Case, Shiller and Weiss. Our dependent variable is the log return from holding the house from \( t \) until \( t + h \) and the predictor is the respective rent-price ratio.

2 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 is also available at http://www.lincolninst.edu/subcenters/land-values/rent-price-ratio.asp.

3 See Calhoun (1996) for detailed information.
3. In-Sample Fit

In this section, we discuss the results obtained through the in-sample regressions which motivate the out-of-sample exercise.

Let us consider the following regression model:

\[ y_{t+h} = \alpha + \beta x_t + u_{t+h}, \quad t = 1, \ldots, T \]  

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

To assess the predictive ability of \( x_t \) in-sample, we can estimate the equation (1) using the available \( T - h \) observations and then, examine the \( t \)-statistic associated to the OLS estimate of \( \beta \) and the goodness-of-fit measure (\( R^2 \)).

When there is evidence to reject the null hypothesis of \( \beta = 0 \) and the \( R^2 \) is high, we can conclude that \( x_t \) has predictive power over \( y_{t+h} \).

There are some problems related to this perspective, specifically the small-sample bias (see Stambaugh 1986, 1999) and the dependence between the observations for the regressand in (1) (see Richardson and Stock, 1989). The serial correlation induced in the disturbance term should be taken into consideration when conducting inference.

For each market in question, we estimate equation (1) by OLS and use the Newey and West (1987) standard errors to compute the usual \( t \)-statistic. Table 1 provides the results.

Analyzing the results, we detect a common pattern across the two markets. The estimate of \( \beta \) 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): high prices relative to dividends (or rents, for the housing market) can be a sign of low returns.

We must therefore investigate whether the in-sample predictability is kept when subject to a simulation exercise of forecasts in real time (out-of-sample).

Table 1

<table>
<thead>
<tr>
<th>Horizon (quarters)</th>
<th>Stock Market</th>
<th>Housing Market</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample period 1947:Q1-2010:Q2</td>
<td>Sample period 1960:Q1-2010:Q1</td>
</tr>
<tr>
<td></td>
<td>Case-Shiller-Weiss</td>
<td>OFHEO data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Horizon (quarters)</th>
<th>( \hat{\beta} )</th>
<th>t-stat</th>
<th>R(^2 )%</th>
<th>Adj. R(^2 )%</th>
<th>( \hat{\beta} )</th>
<th>t-stat</th>
<th>R(^2 )%</th>
<th>Adj. R(^2 )%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,80 (2,89)</td>
<td>2,85</td>
<td>2,46</td>
<td>1,27</td>
<td>(5,24)</td>
<td>22,30</td>
<td>21,91</td>
<td>1,21 (8,47)</td>
</tr>
<tr>
<td>4</td>
<td>16,57 (3,14)</td>
<td>11,23</td>
<td>10,88</td>
<td>5,90</td>
<td>(2,88)</td>
<td>38,80</td>
<td>38,48</td>
<td>5,39 (4,73)</td>
</tr>
<tr>
<td>8</td>
<td>32,08 (3,38)</td>
<td>19,97</td>
<td>19,64</td>
<td>12,86</td>
<td>(3,49)</td>
<td>54,85</td>
<td>54,61</td>
<td>11,41 (5,52)</td>
</tr>
<tr>
<td>12</td>
<td>46,35 (3,97)</td>
<td>25,38</td>
<td>25,07</td>
<td>18,84</td>
<td>(4,62)</td>
<td>64,31</td>
<td>64,12</td>
<td>16,67 (6,81)</td>
</tr>
<tr>
<td>18</td>
<td>74,17 (5,17)</td>
<td>33,05</td>
<td>32,76</td>
<td>25,27</td>
<td>(5,73)</td>
<td>67,05</td>
<td>66,87</td>
<td>22,73 (7,98)</td>
</tr>
<tr>
<td>24</td>
<td>121,28 (6,52)</td>
<td>44,47</td>
<td>44,23</td>
<td>29,68</td>
<td>(5,44)</td>
<td>61,46</td>
<td>61,24</td>
<td>27,18 (7,63)</td>
</tr>
</tbody>
</table>

Source: Author’s calculations.

Notes: The regression equation is \( y_{t+h} = \alpha + \beta x_t + u_{t+h} \), where \( y_{t+h} \) and \( x_t \) 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. t-stat denotes the Newey and West (1987) adjusted \( t \)-statistic.

4 The null hypothesis (\( H_0: \beta = 0 \)) reflects the lack of ability of \( x_t \) to forecast \( y_{t+h} \).
4. Econometric Procedure

In this section, we discuss the regression models used to produce the out-of-sample forecasts and the statistic tests applied to analyze the results.

4.1 Out-of-sample analysis

An out-of-sample analysis implies the simulation of forecasts for \( y_{t+h} \), in period \( t \). Given the sample in study, we should determine a value for the sample-split parameter (\( R \)), which will match with the period of the first prediction (there is no criterion that defines how to choose \( R \); one should make a compromise between the number of observations used to estimate the coefficients of the models and the number of available observations to assess the forecasts performance). We use the sample that includes the first \( R - k \) observations to estimate the model and produce the predictions for \( t = R \). Then, the next observation (which corresponds to the period \( t = R + 1 \)), is added to the sample, the model is re-estimated and a new forecast (for period \( t = R + 1 \)) is generated. This process is repeated until the end of the sample.

4.2 Predictive regression models

We select several methods to generate the set of predictions. In what follows, \( \hat{y}_{t+h} \) 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 \( y_{t+h} \).

- **Historical mean**: \( \hat{y}_{t+h} = \frac{1}{T} \sum_{s=1}^{T} y_s \), \( t = R, \ldots, T \). 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.

- **Direct autoregressive model**: \( \hat{y}_{t+h} = \hat{\alpha} + \sum_{j=1}^{p} \hat{\beta}_j y_{t-j} \). In our empirical applications, we consider a version with a fix lag order (\( p = 2 \)) and a version with the lag order determined using the AIC Criterion (\( p \) integer \( \in [1,4] \)). The coefficients are estimated by OLS.

- **Direct augmented autoregressive model**: \( \hat{y}_{t+h} = \hat{\alpha} + \sum_{j=1}^{p_1} \hat{\beta}_j y_{t-j} + \sum_{j=1}^{p_2} \hat{\delta}_j x_{t-j} \), \( p_1, p_2 \) integers \( \in [1,4] \). The lag orders (\( p_1^* \) and \( p_2^* \)) are determined using the AIC Criterion and coefficients are estimated by OLS.

- **Direct regression model with or without lags**: \( \hat{y}_{t+h} = \hat{\alpha} + \sum_{j=1}^{p} \hat{\beta}_j y_{t-j} \). Here, we also consider a fix lag order (\( p = 2 \)) and a lag order determined using the AIC Criterion (\( p \) integer \( \in [1,4] \)). The coefficients are estimated by OLS.

- **Univariate and multivariate low-pass filters**: Following the argument presented in Valle e Azevedo and Pereira (2012), we use this method to generate our forecasts when it is useful to predict only the low frequencies of \( y_t \) (\( w_t = B(L)y_t \), where \( B(L) \) is a band-pass filter eliminating the fluctuations with period smaller than 32 quarters, as is usual in business cycle studies). Thus, we consider the predictions of the low frequencies of aggregate returns as forecasts of aggregate returns itself. Since the available sample is finite (\( \{y_t\}_{t=R+1} \)) and supposing that there are \( c \) series of covariance-stationary covariates (\( z_t, \ldots, z_c \)) available, we approximate the low frequencies of \( y_t \) (that is, we approximate \( w_t \)) through a weighted sum of elements of \( y_t \) and \( z_t, \ldots, z_c \):

\[
\hat{w}_t = \sum_{j=-f}^{p} \hat{B}_j y_{t-j} + \sum_{s=1}^{c} \sum_{j=-f}^{p} \hat{R}_{s,j} z_{s,t-j}
\]
The number of observations in the past and in the future, respectively, that are considered. The coefficients are estimated solving a minimization problem. To extract the signal \( w_{t+h} = B(L)y_{t+h} \) for \( h > 0 \), we should set \( f = -h \) in the solution of the mentioned problem (as a result, only the available information up to period \( t \) is employed). We obtain the univariate filter when we drop the covariates \( z_1, \ldots, z_c \) from equation (2).

### 4.3 Forecast evaluation

As an evaluation metric to compare the sets of forecasts obtained through the models described before, we chose the ratio between the mean squared forecast error of the competing model and those of the benchmark model (MSFE ratio). When the MSFE ratio is lower than 1, the competing model generates better predictions (according to the mentioned criteria) than the benchmark model (historical mean).

We also use a graphical analysis to examine the relative performance of the forecasting models over the sample. As proposed in Goyal and Welch (2003), we construct charts with the difference between the cumulative squared forecast errors of the benchmark model and the cumulative squared forecast errors of the competing model (hereafter, we will refer to this difference as Net – SSE). When this difference is positive, the competing model outperforms the benchmark model in the sample between the first prediction and the date in the \( x \)-axis.

### 4.4 Out-of-sample tests

We assess the statistical significance of the obtained results considering equal accuracy tests and forecast encompassing tests.

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). We apply the test statistics modified MSFE – \( t \) and MSFE – \( F \) to test this null hypothesis (see Diebold and Mariano, 1995; Harvey et al. 1997, 1998 and McCracken, 2007). Excluding the multivariate filter, all our models are nested, therefore the critical values for these statistics were generated by simulation. According to Harvey et al. (1997), 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. So, with the forecast encompassing test, we assess whether a given set of forecasts generated by a simpler model embody all the useful predictive information contained in another 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, the financial variable does not contain useful additional information for predicting the aggregate returns. The test statistics employed, modified ENC – \( t \) and ENC – \( F \), result from an adaptation to this problem of the Diebold and Mariano (1995) test statistic. Again, since the test statistics considered do not have a standard distribution, the critical values were generated using a bootstrap procedure.

---

6 More detailed explanations about the multivariate filter can be found in Valle e Azevedo (2011) and Valle e Azevedo and Pereira (2012).

7 Two models are nested when there is a set of regressors that is common between them (see Clark and McCracken, 2005). In our studies, we have nested models due to the constant term in most models.

8 We rely on methods provided by Kilian (1999) and Mark (1995) for developing the bootstrap procedure used to generate the critical values.
5. Empirical Results

In this section, we discuss the main results obtained using the methodology described before.

Stock Market:

We conclude that only the direct regression generates forecasts that can beat the benchmark for all horizons (see Table 2). The MSFE ratios are statistically lower than 1 at conventional significance levels, which means that the forecasts from the competing model have more predictive power than those from the historical mean model. The MSFE ratios decrease 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 3, 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 lower than 1 for $h=20$ and $h=24$ (despite not being statistically lower than 1 when we use the modified MSFE – $t$ statistic to apply the test).

Similar conclusions can be drawn when we analyze the forecast encompassing results (Table 3). In particular, when we use the ENC – $F$ 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.

The following analysis rests on the evaluation of the Net – SSE charts which display the cumulative squared forecast errors of the benchmark model (from 1985:Q1 through the date in the $x$ – axis) minus the cumulative squared forecast errors of the competing model (from 1985:Q1 through 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 a lower forecasting error than the historical mean model, in a given quarter.

In Chart 1 we plot the mentioned curves for $h=1$ and $h=24$, considering as competing models the direct regression model (without lags) and the multivariate filter (which uses the dividend-price ratio). Considering the shorter forecast horizon (1 quarter), 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.

For longer forecast horizons ($h=24$; see figure 1), the curves are smoother and we can identify three distinct periods (which have become more apparent as the horizon increases). Namely: an initial period

### Table 2

<table>
<thead>
<tr>
<th>Horizon (quarters)</th>
<th>1</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
<th>18</th>
<th>20</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct regression model ($p=0$)</td>
<td>0.988*</td>
<td>0.984**</td>
<td>0.976**</td>
<td>0.987*</td>
<td>0.969***</td>
<td>0.883***</td>
<td>0.883***</td>
<td></td>
</tr>
<tr>
<td>Direct regression model ($p=2$)</td>
<td>0.991</td>
<td>1.001</td>
<td>0.999</td>
<td>1.008</td>
<td>0.996</td>
<td>0.946</td>
<td>0.934</td>
<td>0.934</td>
</tr>
<tr>
<td>Direct regression model ($p_{max}=4$)</td>
<td>1.001</td>
<td>1.005</td>
<td>1.006</td>
<td>1.011</td>
<td>1.009</td>
<td>0.985</td>
<td>0.977</td>
<td>0.977</td>
</tr>
<tr>
<td>Multivariate filter without indicators</td>
<td>1.028</td>
<td>1.122</td>
<td>1.140</td>
<td>1.168</td>
<td>1.230</td>
<td>1.152</td>
<td>1.099</td>
<td>1.099</td>
</tr>
<tr>
<td>with dividend-price ratio</td>
<td>1.020</td>
<td>1.082</td>
<td>1.070</td>
<td>1.080</td>
<td>1.140</td>
<td>1.034</td>
<td>0.983</td>
<td>0.983</td>
</tr>
</tbody>
</table>

Source: Author’s calculations.

Notes: For the direct regression (nested model), we use the test statistic MSFE-$F$ and critical values generated using a bootstrap procedure; for the multivariate filter (non-nested model), we consider the test statistic modified MSFE-$t$ and critical values from the Student’s $t$ distribution with $(N-1)$ degrees of freedom ($N$ is the number of forecast errors; see Clark and McCracken, 2001 and McCracken, 2007). 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.
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.

Additionally, we decide to plot, in the same chart, the Cumulative $SSE$ Difference curve (the competing model selected only contains the dividend-price ratio as regressor – direct regression model without lags) and a price index curve (for the stock market, we chose the SP$500$ Index). With this exercise, we intend to note how the curves are related and discuss the reasons for this relationship.

The key point to emphasize is that the curves exhibit a symmetrical behaviour: the peaks in SP$500$ Index correspond to the troughs in the $Net - SSE$ curve (see Chart 2). This means that the model with the dividend price-ratio produces less accurate predictions for the period in which the stock price is increasing, while its good performance is associated with a period in which there is a fall in prices. We note these phenomena when, for example, we analyze the exuberant period associated with the Dot-com, in the late nineties. During the pre-crash period, the SP$500$ Index rises and the $Net - SSE$ decreases (the historical average is a better predictor than direct regression model over this period). Nevertheless, after the fall in prices, we identify the reverse performance: the direct regression forecasts are closer to the observed value (the $Net - SSE$ curve has a positive slope). How can we explain this relation? Resuming the introductory discussion – low dividend-price ratios signal low future returns – we can deduce that when prices increase, relative to dividends, it can be expected a reduction of returns in subsequent periods. Thus, we understand that when the price increases (and the dividends remain stable), our model that is predicting a fall in returns exhibits a worse performance. This translates into a negative

<table>
<thead>
<tr>
<th>Horizon (quarters)</th>
<th>1</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
<th>18</th>
<th>20</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct regression model ($p=0$)</td>
<td>ENC-F</td>
<td>0.813</td>
<td>2.224**</td>
<td>1.870**</td>
<td>0.471</td>
<td>-0.361</td>
<td>2.615***</td>
<td>4.375***</td>
</tr>
</tbody>
</table>

**Source:** Author’s calculations.

**Notes:** The test was applied assuming that the forecasts are biased and inefficient (more general case). The critical values are generated using a bootstrap procedure. 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.

**Chart 1**

**CUMULATIVE SSE DIFFERENCE CONSIDERING TWO COMPETING MODELS, FOR H=1 AND H=24.**

![Chart 1](chart1.png)

**Source:** Author’s calculations.
inclination for the $Net - SSE$ curve and, simultaneously, a positive inclination for the $SP500$ curve. In the post-crash period, the returns descend steeply. At this time, the competing model produces good predictions (relatively to the historical mean forecasts), and therefore the $Net - SSE$ slope is positive (whereas it is negative for the $SP500$).\(^9\)

**Housing Market:**

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 (see Table 4). In particular, the $MSFE$ ratio between the direct regression and the benchmark model decreases as the horizon increases (all the values are statistically lower than 1, at 1% significance level). Comparing with the results obtained in Section 3, we verify that the predictability pattern identified in-sample holds out-of-sample for the housing market.

Table 5 displays the forecast encompassing statistics which allow the conclusion 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%).

Figure 3 contains the charts with the $Net - SSE$ (for $h=1, 12, 18$ and $24$), considering three competing models and the Case, Shiller and Weiss data. 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 (considering $h=1$). 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 horizon of 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 horizons of 18 and 24 quarters, where the cumulative $SSE$ difference between the autoregressive model and the benchmark model is constantly negative. From 2008:Q1, the

---

\(^9\) We did the same exercise for the housing market and the conclusions are similar.
Table 4
MSFE RATIOS AND EQUAL ACCURACY TEST RESULTS FOR THE HOUSING MARKET

<table>
<thead>
<tr>
<th>Horizon (quarters)</th>
<th>1</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
<th>18</th>
<th>20</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct autoregressive model ((p=2))</td>
<td>0.453</td>
<td>0.358</td>
<td>0.448</td>
<td>0.572</td>
<td>0.864</td>
<td>1.024</td>
<td>1.085</td>
<td>1.085</td>
</tr>
<tr>
<td>Direct augmented AR ((p_{max}=4))</td>
<td>0.342</td>
<td>0.315</td>
<td>0.344</td>
<td>0.414</td>
<td>0.561</td>
<td>0.401</td>
<td>0.373</td>
<td>0.373</td>
</tr>
<tr>
<td>Direct regression model ((p=0))</td>
<td>0.785***</td>
<td>0.738***</td>
<td>0.724***</td>
<td>0.697***</td>
<td>0.579***</td>
<td>0.417***</td>
<td>0.401***</td>
<td>0.401***</td>
</tr>
<tr>
<td>Direct regression model ((p=2))</td>
<td>0.466</td>
<td>0.465</td>
<td>0.397</td>
<td>0.488</td>
<td>0.559</td>
<td>0.409</td>
<td>0.386</td>
<td>0.386</td>
</tr>
<tr>
<td>Multivariate filter without indicators</td>
<td>0.554**</td>
<td>0.602</td>
<td>0.684</td>
<td>0.794</td>
<td>0.945</td>
<td>1.025</td>
<td>1.038</td>
<td>1.038</td>
</tr>
<tr>
<td>with rent-price ratio</td>
<td>0.541**</td>
<td>0.553</td>
<td>0.627</td>
<td>0.716</td>
<td>0.824</td>
<td>0.829</td>
<td>0.827</td>
<td>0.827</td>
</tr>
</tbody>
</table>

Source: Author’s calculations.
Notes: See Table 2. 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.

Table 5
FORECAST ENCOMPASSING TEST RESULTS FOR THE HOUSING MARKET

<table>
<thead>
<tr>
<th>Horizon (quarters)</th>
<th>1</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
<th>18</th>
<th>20</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate filter with rent-price ratio</td>
<td>3.715***</td>
<td>2.053**</td>
<td>1.696**</td>
<td>1.607*</td>
<td>1.441*</td>
<td>1.212</td>
<td>1.160</td>
<td>1.101</td>
</tr>
</tbody>
</table>

Source: Author’s calculations.
Notes: See Table 3. 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.

direct regression curve grows almost exponentially, evidencing the predictive power of the rent-price ratio. Again, it is important to underline the importance of the observations corresponding to the end of the sample to the good results out-of-sample.

6. Conclusion

In this paper, we found evidence that the known in-sample pattern of return predictability (using valuation ratios) holds out-of-sample for the stock market and, especially, for the housing market. Considering the stock market, we show that a simple regression model that includes a valuation ratio outperforms the benchmark (which represents the hypothesis of no predictability of returns), for all horizons. Additionally, we note that the ability of dividend–price ratio to predict the aggregate returns improves at longer horizons. For the housing market, all the models that contain the rent-price ratio consistently exhibit MSFE ratios lower than 1, for all horizons.

The sample dependence (significance of the last observations for the good results out-of-sample) identified 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.

Additionally, it would be interesting to extend this research to other markets, namely bonds and treasuries since there are relatively few studies about predicting returns on these markets, out-of-sample. Another suggestion would be to reproduce this study using data for European markets, aiming the development of overvaluation indicators.
Chart 3

CUMULATIVE SSE DIFFERENCE CONSIDERING THREE COMPETING MODELS (HORIZONS OF 1, 12, 18 AND 24 QUARTERS).

Source: Author’s calculations.
References


