sciendo
www.degruyter.com/view/j/remav

UNDERSTANDING PRICE-TO-RENT RATIOS THROUGH SIMULATION-BASED DISTRIBUTIONS AND EXPLAINABLE MACHINE LEARNING

Jonas Vogt*[1]

[1] *Finance and Data-Science, Duale Hochschule Baden Wurttemberg Mannheim, Coblitzallee 1-9, 68163 Mannheim; Germany; e-mail: jonas.a.vogt@gmail.com, ORCID: 0000-0003-0739-1705*

* *Corresponding author*

ARTICLE INFO

Keywords:

real estate finance, property valuation, price rent ratio, explainable machine learning, property investment

JEL Classification:

G10, R31

ABSTRACT

Index-level price-to-rent (PTR) ratios are a widely used metric for analyzing housing markets, employed by both real estate practitioners and policymakers. This article seeks to improve the contextualization of observed PTR values by examining the interplay between these ratios and macroeconomic and housing-market developments in a non-linear framework. We analyze historical data on housing prices, rents and macroeconomic developments from 18 advanced economies, spanning from 1870, using Boosted Regression Trees and explainable machine learning techniques. As a precursor to this analysis, we also present the empirical distribution of the price-to-rent ratio and the implied housing risk premia across all years and countries.

Citation:

Vogt, J. (2025). Understanding price-to-rent ratios through simulation-based distributions and explainable machine learning. *Real Estate Management and Valuation*, *33*(3), 36-48. https://doi.org/10.2478/remav-2025-0024

1. Introduction

The development of housing markets is highly relevant for both real-estate practitioners, life-cycle-investors and policymakers. The price-to-rent ratio can serve as a powerful tool for analyzing the current state of housing markets (Gilbukh et al., 2023; Lo et al., 2022). However, contextualizing observed values of this ratio within historical trends and understanding possible reasons behind certain values is crucial for effectively applying this tool, especially in light of the significant discrepancies in PTR ratios between the European countries depicted. In Figure 1, the PTR ratios at the index level for all countries in our dataset are shown for the latest available year (2019). The highest PTR ratios of countries like Australia or Spain exceed the lower PTR ratios by more than twofold.

There is already an extensive body of literature on the relationship between single property characteristics and single property PTR ratios, as well as on linear relationships between macro-level variables and time-series variation of PTR ratios. This article aims at complementing that body by looking into the relationship between the current and past states of national economies as well as housing-markets and the cross-country- and time-series-variation of PTR. Our analysis also aims to consider non-linear relationships. The central question of this article is therefore: What aspects of the current state as well as the development of an economy and the housing market explain variations in PTR ratios between countries and across time and should be taken into account for the contextualization of specific country level PTR ratios.

Our analysis is based on an extensive macro-dataset covering 18 countries starting from 1870, known as the Jordà-Schularick-Taylor Macrohistory (JST) dataset, as presented by Jordà, Schularick, and Taylor (2017). This dataset contains comprehensive information necessary to calculate national index-level PTR ratios and variables that describe the historical development and current state of a country's economy, financial system, and particularly the housing market. Examples for such variables are real wage growth, rental price or wage variance, the dividend rate or real house price growth. For the analysis of the single variables' relevance for the variation in index level PTR ratios across countries and over time, we do not want to limit ourselves to linear relationships and we model the PTR ratio using non-linear tree and net based models. The most promising models (boosted regression trees) are then analyzed using explainable machine learning techniques. As a precursor to this analysis, we also present the empirical distribution of the price-rent ratio across all years and countries. Our article is structured as follows: Section 2 provides a survey of related literature and elaborates on the originality of this article. Section 3 outlines the description of the analyzed data. Section 4 discusses our methodological tools and in Section 5, we discuss our results.

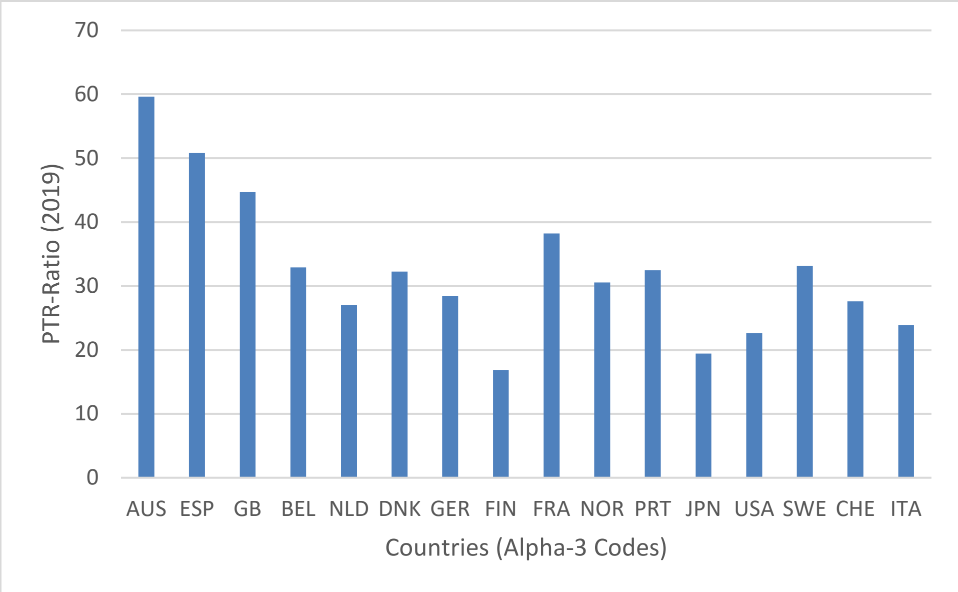


Fig. 1. PTR-Ratios for all countries included in our datasets for 2019 based on the Jordà-Schularick-Taylor Macrohistory (JST) dataset presented in Section 4. Source: own study.

2. Literature Review

The literature on housing prices and rents in general is very extensive, and existing research examining the formation of price-rent ratios can be categorized into four groups:

1. Research on actual housing returns and implied housing risk premia (e.g. (Chambers et al., 2021; Demers & Eisfeldt, 2022; Jordà et al., 2019; Mork & Trønnes, 2023))
2. Research on the impact of property, buyer and sub-area characteristics on rent-price ratios based on microdata (e.g. Bracke (2015); Clark & Lomax (2020); D'Lima & Schultz (2022); Halket et al., (2020); Hill & Syed (2016)).
3. Analyses of the intertemporal relationship among rents, prices, and macroeconomic drivers in a linear VAR- oder VECM-framework (e.g. Anundsen & Jansen (2013); Lo et al. (2023, 2021))
4. Analyses of the intertemporal relationship among rents, prices, and macroeconomic drivers based on the dynamic Gordon-Growth-Model presented for stocks by Campbell and Shiller (1988), and introduced to housing research by Campbell et al. (2009). Further research based on this approach has been presented by Huh and Kim (2021), Kishor and Morley (2015). Lo et al. (2022) and Plogmann et al. (2020).

The key findings from the various research categories are summarized below and we start by discussing the strand in the literature on housing returns. Demers and Eisfeldt (2022) analyze US micro data on family housing and estimate annual gross returns for family housing of 8.5%. Eichholtz et al. (2021) analyze micro data on housing in Paris and Amsterdam and estimate an annualized gross return of 6.89%. Jordà et al. (2019) analyze historical index-level data for several countries and estimate real annualized gross housing returns of 7.05%. Chambers, Spaenjers and Steiner (2021) analyze micro data for housing in Oxford and estimate the annualized gross housing return to be less than 5%. Bracke et al. (2018) analyze data on leasehold housing in London and estimate average real discount rates for housing to be between 2% and 3% in the long run. Piazzesi et al. (2007) analyze leasehold data and propose estimates for the real equity and housing risk premiums of approximately 3.5% each.

The second strand of the literature that presents interesting results for contextualizing property-specific price-rent ratios analyzes the impact of property, sub-area, and investor characteristics on price-rent ratios based on microdata. Clark and Lomax (2020) analyze matched sales and rental microdata for the London housing market and show that the PTR ratio is higher for flats compared to single-family houses. Bracke (2015) analyzes matched sales and rental micro data for the London housing market as well and finds that the PTR ratio tends to be higher for higher-priced and therefore bigger and more central units. Ahlfeldt, Heblich and Seidel (2023) analyze micro data from a major German real estate platform and present evidence for a positive correlation between population density and PTR ratios. Amaral et al. (2021) show, based on international long-term city-level data, that long term housing returns are lower (and the PTR ratio is therefore higher) in "main economic agglomerations". Rutherford et al. (2023) analyze the relationship between characteristics of buy and let investors and price-rent ratios based on match rental and sales micro data for the Miami-Date County. They suggest that investors who conduct more transactions per year tend to trade at higher price-to-rent ratios.

The third relevant branch of literature addresses the question of how the fluctuations in prices and rents, as components of the rent-price ratio, are intertemporally related and which macro-level variables, such as expected rent increases at the index level or inflation data, influence the ratio. Campbell et al. (2009) adapt the linearized dynamic Gordon growth model which has been introduced for stocks by Campbell & Shiller (1988) for housing. In the resulting model formula, the logarithmic PTR ratio is decomposed into expected rent increases, expected housing risk premium, and interest rates. Based on estimations of the corresponding VAR model for the logarithmic price-rent-ratio, they conclude that the variation in expected risk premium is crucial for the development of prices. Kim and Lim (2014), as well as Kishor and Morley (2015), adapt this approach and estimate the expected rental developments and the expected housing risk premium using state-space models. They confirm the high relevance of the expected housing risk premium for variations in PTR ratios. Lo et al. (2023) model changes in the PTR-ratios for different UK-sub-markets based on different macro-variables like exchange-rates, stock returns and monetary aggregates and show that these macro variables are highly relevant for the development of the price-rent-ratio in different submarkets. Lo et al. (2022) follow a similar approach to demonstrate the relationships between macro-variables and price-rent-ratios for the housing market in Hong Kong. Lo et al. (2021) estimate Vector Error Correction Models (VECMs) for price and rent aggregates in the Belfast housing market, and apply Johansen's Cointegration Test to demonstrate that prices and rents are cointegrated with the transmission channel going from prices to rents.

This article complements the existing literature on PTR ratios in two ways: Unlike other articles the author is aware of, it uses the different developments in the overall economy and the housing market to explain the cross-sectional variation of PTR ratios between countries which appears to be an important aspect to contextualize country level PTR ratios. Furthermore this article does not limit the analysis of potential relationships between PTR ratios and their drivers to linear ones. Both tree and net based machine learning methods provide methodologically interesting opportunities for such an analysis and boosted regression trees, which turned out to be the most promising modeling choice, have already been established for modeling yield variations in other asset classes (e.g. Gu et al., 2020; Kaufmann et al., 2021; Rossi & Utkus, 2024). Explainable machine learning techniques allow us to understand the resulting boosted tree model despite of its complexity.

3. Data and Methodology

Our data source is the Jordà-Schularick-Taylor Macrohistory (JST) dataset, presented by Jordà et al. (2017). This dataset includes data on rental and housing prices, inflation rates, bill rates, and other macroeconomic variables for 18 advanced countries since 1870, and contains over 2500 observations. This dataset includes annual country level information needed to calculate variables describing the current state as well as the development of the national economy and financials, as well as the housing market.

The target variable in our boosted regression tree model is the PTR ratio which has been observed once a year for all countries and years since 1870. The original dataset includes an average rental yield per year and country, which is the inverse of the PTR ratio. We include annualized real rental and housing price change over the past 5 and 20 years as a feature in our model. We do this because changes in prices and rents have a direct impact on the PTR ratio when they do not occur in parallel. Additionally, Barberis et al. (2018) provide strong evidence that past experiences with asset price developments significantly shape future expectations, which could further influence the current

PTR ratio. The original dataset includes the consumer price index $CPI_{t,i}$ and nominal housing price indices $P_{i,t}^{housing}$ for each country *i* and all years *t*. We calculate the nominal annual change in housing price $y_{i,t-1}^{housing}$ as follows:

$$y_{i,t}^{housing} = \frac{P_{i,t}^{housing} - P_{i,t-1}^{housing}}{P_{i,t-1}^{housing}} \tag{1}$$

We adjust the nominal housing price change by changes in the consumer price index and calculate the real change in the housing price index $1 + y_{i,t}^{real,housing}$

$$1 + y_{i,t}^{real,housing} = \frac{1 + y_{i,t-1}^{housing}}{CPI_{t,i}/CPI_{t-1,i}} \tag{2}$$

Real changes in rents and wages are calculated accordingly. The experienced variances of rental price changes, trend-adjusted housing prices, and average wages are also included to account for the level of uncertainty regarding these aspects. Our hypothesis is that greater uncertainty regarding wages and prices might lead to greater reluctance to invest in housing. Higher rental uncertainty, on the other hand, could have a positive effect on the PTR ratio, as purchasing property serves as insurance against rental uncertainty (Piazzesi et al., 2007).

To calculate the variance of prices, we detrend our housing price indices of trend adjusted housing price, because we want quantify uncertatinty. We therefore filter our real house price using the double exponential smoothing method according to Holt. This method accounts for a potential linear trend and produces the smoothed value $\hat{p}_{i,t}^{real,housing}$.

The variance of the trend-adjusted housing prices is then calculated as follows

$$Var\begin{pmatrix} p_{i,t}^{real,housing} - \hat{p}_{i,t}^{real,housing} \\ p_{i,t-1}^{real,housing} - \hat{p}_{i,t-1}^{real,housing} \\ \vdots \end{pmatrix} \tag{3}$$

Current dividend yields on the country index level are included as well. Even though there is no consensus in the literature on the exact nature of the transmission channel, the basic existence of a relationship between the stock market and the housing market is well documented (Okunev et al., (2000); Tsai, (2015)). The current real bill rate is chosen to approximate the risk-free rate and the interest rate development is included to account for the rate's path to its current value. Both the current dividend yield and the current yield are already included in the JST dataset.

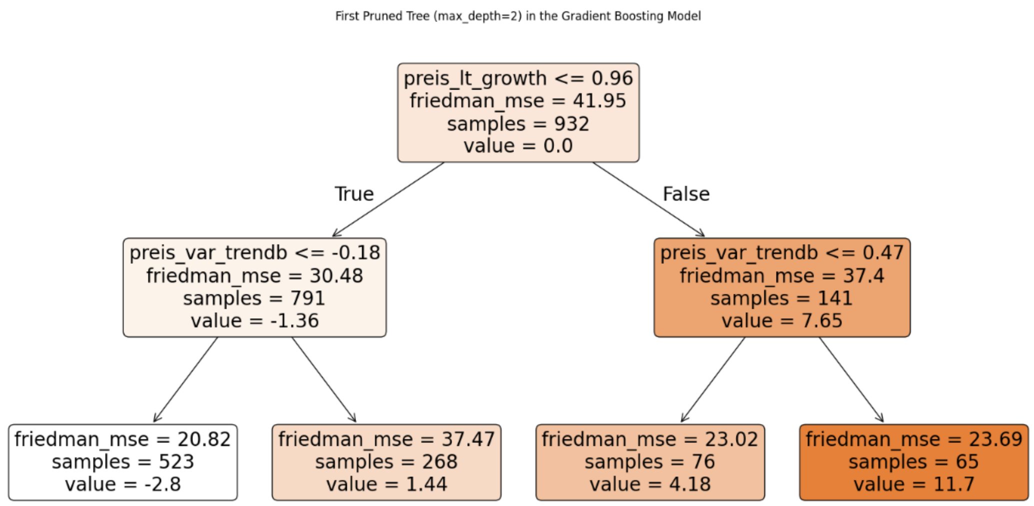


Fig. 2. Simple regression tree example explaining the standardized PTR ratio on country level based on the experienced long-term inflation adjusted term housing price growth and the experienced variance of the trend adjusted housing price. Stronger price growth seems to lead to higher PTR-ratios. Given the experienced price growth, the expected PTR ratio appears to be higher for lower experienced variances of trend and inflation adjusted housing prices. *Source*: own study.

We model the PTR ratio using boosted regression trees (specifically, the Gradient Boosting implementation in the Scikit-Learn library), as we don't want to restrict our analysis to linear relationships, while also aiming to limit the risk of overfitting. When modeling a target (in our case the PTR ratio) with regression trees, we seek a partition of the feature space (i.e., the value ranges of the macro-financial variables) such that the belonging of an observation to a specific sub-region is particularly informative with respect to the modeled variable. Since we gradually split the feature space, we are not limited to modeling linear relationships. Figure 2 shows a simple regression tree which has been calibrated based on our original data:

With large regression trees, there is a risk of overfitting (i.e. the model reflects the original data including all random noise too precisely). Boosted regression trees address this issue by calibrating several small trees instead of a single large tree, with many nodes to account for all features. In this process, the first small tree is intended to model the PTR ratio and the subsequent trees are intended to model the remaining unexplained variation in the PTR ratio. The impact of each individual tree is shrunk so that individual trees do not dominate the modeling of the target. A detailed presentation of the boosted regression trees method can be found in Friedman (2000). For our application we standardize all our features based on the mean and the standard-deviation estimated based on the training data, and eliminate values that are more than 2.5 interquartile ranges away from the 25th and 75th percentiles. We use a final test set to estimate the out-of-sample R-squared of our model and run a 5-fold cross-validation for our model selection and hyperparameter search. The test set includes randomly drawn countries which are then not included in our training set.

Essentially, it would also be possible to model non-linear relationships without specific modeling priors using neural networks (e.g., a classical neural network, convolutional networks, etc.). However, it is well-documented in research that tree-based methods outperform neural networks when it comes to modeling tabular data (Gu et al., 2020). Therefore, we chose tree-based methods. Alternatively, one could consider using random forests. Since boosted regression trees perform well in a well-known comparative study (Caruana & Niculescu-Mizil, 2006), they have been tested in the financial context (Cabrol et al., 2024; Kaufmann et al., 2021). Nevertheless we also calibrated multiplayer perceptrons and random forests and compare the results with our boosted regression tree results in the results section.

We use the Scikit-Learn implementation for all three methods. Before running the calibration algorithms on our data, we clean the data by removing observations (i.e. data for one year and one country) that deviate by more than 2.5 interquartile ranges from the median. This is to prevent our model from being strongly influenced by highly exceptional scenarios, such as war conditions. Additionally, we standardize our features to achieve greater robustness in the calibration process. The hyperparameters—such as key features of our model, like the number of trees - are determined through 5-fold cross-validation. Furthermore, about 20% of the observations are randomly selected and set aside as a final test set, ensuring that the countries included in the test set are not represented in other years in the training set.

A potential drawback of many machine learning techniques, such as boosted regression trees, is the low transparency of the models. Due to the multitude of model features, such as the "split points" in the individual trees, it is difficult to assess the impact of individual features. A discussion on the opacity of machine learning models has been presented, for instance, by Bücker et al. (2022). Due to the difficulty in directly identifying the impact of individual features, there are a variety of methods available to interpret more complex machine learning models. The following methods have been applied in our analysis: Variable importance, partial dependence plots and Shapley values. Variable importance measures, for instance, a feature's relative impact on the model's overall reduction in the modeling error, partial dependence plots depict the predicted target value when changing one feature while setting all other features at their average values (Friedman, 2000) and Shapley values are average marginal contributions of features on the actual predictions without keeping the other features constant (Rozemberczki et al., 2022).

Leading up to our actual analysis, we estimate and examine the empirical distribution of the PTR ratio as well as real price and rent growth over the last 19 years. Based on the Gordon Growth Model, we also solve for the risk premium and estimate its distribution. To that, we assume an expected rent development that corresponds to observed annualized development. The Gordon Growth Model suggests the following relationship between the PTR ratio and the fair discount rate i and the rental growth rate g:

$$PTR - Ratio_t = \frac{Price_t}{Rent_{t+1}} = \frac{1}{i-g} \quad (4)$$

For the estimating the empirical distributions, we draw 20 year blocks from our dataset 3000 times and calculate gross price-rent ratios for the last year of each block and annual real growth rates for rents and housing prices for the 19 years preceding it.

4. Results and Application

First, we look at the distribution of the PTR ratio, the real price, the rent growth rate, the implied housing risk premium. Percentiles are displayed in Table 1 and the distributions are displayed in Figure 3. For the risk premium, we consider both the gross premium (ignoring running costs) and a net running cost of 35% of our rental income or rental savings. The distributions of all metrics show a moderate right skew, which makes historical averages less informative about "typical" values. Therefore, we focus on the median as the measure of central tendency. The median PTR ratio is 20.4, the median real annual price and rental growth rates are both greater than zero with price growth rates percentiles being higher than rental price growth rates and the median of the estimated risk premium at 5.5, with the interquartile ranges of both variables being wider than that of the PTR ratio. If we take running costs into consideration and assume that 35% of our rental income or rental savings is absorbed by them, median risk premium decreases to 3.6. The empirical distribution of the PTR ratio shows that PTR values above 33.76 have historically been observed in only 10% of all years across countries and PTR values have exceeded 16.52 in 75% of all considered years. The histogram of annual price growth rates reflects stronger dispersion compared to rents. We see that PTR rations observed in 2024 for countries like Switzerland and Cyprus have been rather unusual historically but housing prices are relatively low in countries like Italy, Ireland, the UK or the Netherlands (see Figure 1). The estimated medium net risk premium of 3.66 is similar to values suggested in the literatures (see literature review). We also see that positive real rental and housing price growth has historically been very common.

Table 1

Percentiles of estimated key figures; median PTR value of 21.05 suggests that many European countries are rather highly priced compared to historical valuation levels

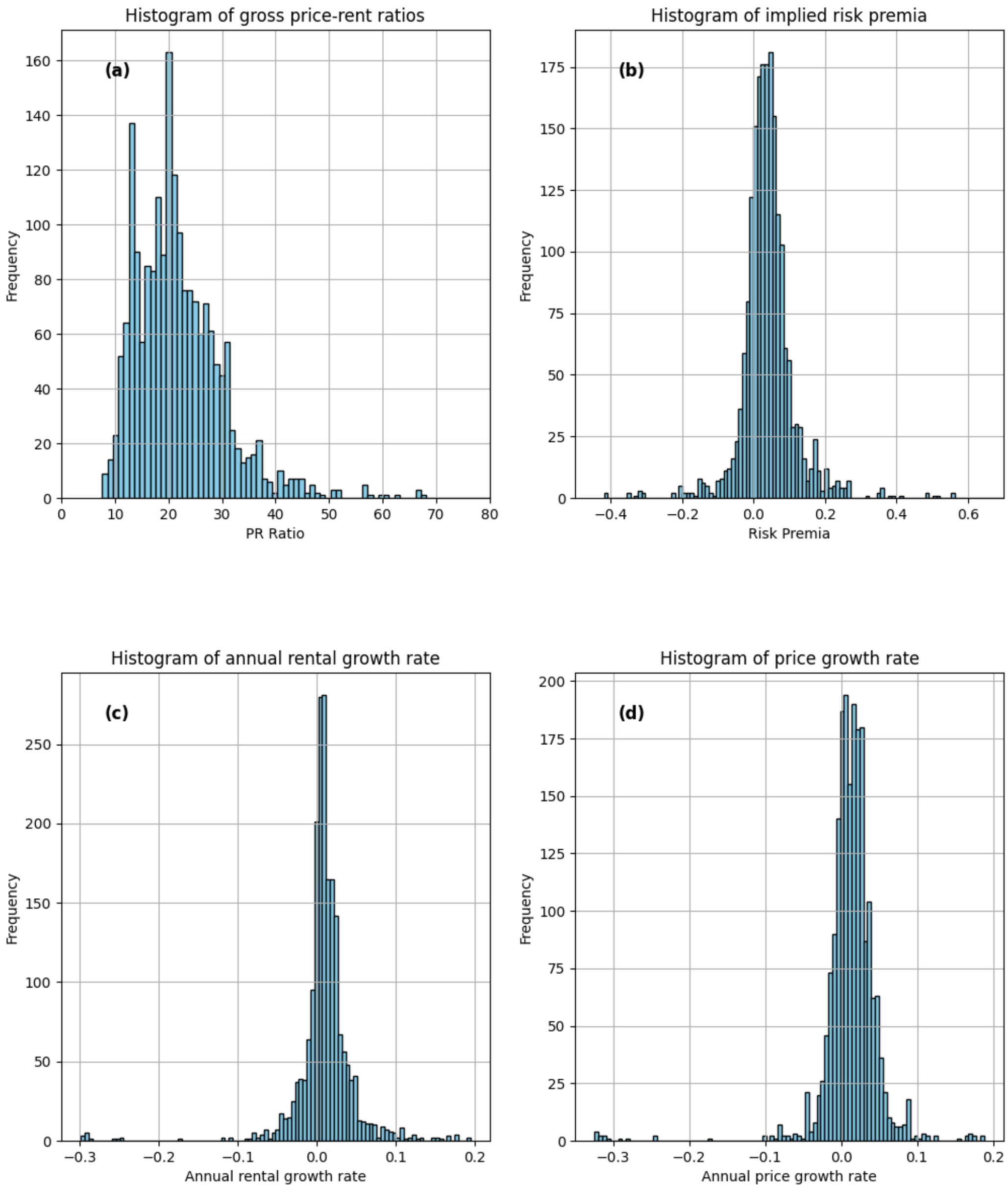
	PTR Ratio	Risk Premium	Net Risk Premium	Real price growth rate	Real rent growth rate
mean	26.68	6.17	4.42	1.37	1.05
1%	9.20	-11.12	-13.10	-5.05	-6.25
5%	11.57	-2.87	-4.42	-2.60	-3.23
10%	13.22	-0.61	-2.09	-1.54	-1.87
25%	16.52	2.46	0.72	-0.04	-0.07
50%	21.05	5.48	3.66	1.43	0.81
75%	27.31	8.51	6.97	2.70	2.12
90%	33.47	13.12	10.86	4.16	4.02
95%	43.69	18.16	16.53	5.23	5.59
99%	205.98	36.12	35.21	8.70	11.37

Source: own study.

Before analyzing our model in relation to the question of which factors explain the variation in PTR ratios between countries and, over time, within a non-linear framework, we will first discuss the accuracy and robustness of our model. Our boosted regression tree model achieves an in-sample R^2 of 74.5%. This means that 74.5% of the variation in PTR ratios can be explained for all countries included in the training sample. In this context, "training" refers to the calibration of our Random Forest model. The in-sample R^2, however, does not provide a realistic measure of the model's explanatory power on unseen data, such as new observations. To test the model's robustness, we randomly excluded some countries from the training process to create a test set. When the model calibrated using the training data is applied to the data of these previously excluded countries, R^2 drops to 45%. This substantial decrease in explanatory power highlights the model's limitations with new data. Nevertheless, the model still performs relatively well compared to financial return models in other contexts. For example, popular factor models like the Fama French three factor model explain less than 20% of the cross-sectional variation in single-stock returns (Bartholdy & Peare, 2005). Finally, cross-validation, which is used primarily for hyperparameter optimization, estimates an out-of-sample R^2 close to the in-sample value of 74%.

All our features are standardized by country, and we do not condition our expectations regarding a country's PTR value on whether a country has, on average, higher feature values to avoid overfitting. As a robustness check, we also evaluate the results without this standardization. This leads to higher in-sample R^2 values of over 90%, but lower out-of-sample R^2 values of 30%, which supports our reasoning.

Alternatively, we could standardize our target values by date, which would allow us to focus exclusively on the variation in the ratio within the cross-section. However, since our goal is to model both the variation in the cross-section and over time, we refrain from this approach.



Alt Text for Graphical Figure 3a: A histogram of the price rent ratio. The distribution is centered around 20.4 and displays a strong right skew, with more extreme values to the right than to the left. Alt Text for Graphical Figure 3b: A histogram of the housing risk premium. The distribution is centered around 3.5 and displays a moderate right skew, with some extreme values to the right of the median. Alt Text for Graphical Figure 3c: A histogram of the annualized real rental growth rate. The distribution is centered around 1.34 and displays a rather weak right skew, with few extreme values to the right of the median. Alt Text for Graphical Figure 3d: A histogram of the annualized real price growth rate. The distribution is centered around 1.34 and displays a moderate right skew, with some extreme values to the right of the median.

Fig. 3. Distributions of the PTR ratio, the implied housing risk premium as well as annual growth rate of housing and rental price changes; right skew makes arithmetic means less informative. *Source*: own study.

Before interpreting the model results, we would also like to discuss potential limitations of the model, which at the same time provide opportunities for further development of the approach. First of all structural differences (e.g., in tax structures) remain unmodeled, but naturally influence variation across the cross-section. This could be addressed by looking into possibilities to include characteristics of a country's tax system. Additional information that could be considered in our model includes data on the condition of the housing stock and fundamental differences in housing quality between properties owned by owners and those rented by tenants. Furthermore, it would be interesting to consider data that captures market sentiment well, such as news or social media data.

Regardless of the variables considered, we do not account for potential structural changes in the relationships over time but instead assume a constant relationship between the features and our PTR ratio. This could be addressed by introducing an exponential weighing scheme of observations, increasing the impact of newer observations and dampening the impact of older ones on our estimation results.

Next, we discuss the results of our hyperparameter search. Our model consists of 1,000 trees, with three tree levels. The impact of each additional trees is moreover shrunk by 1% each time. These characteristics have been set based on a cross-validation hyperparameter search, which means that we chose the characteristics that maximize the out-of-sample R^2. In addition to Boosted Regression Trees, other machine learning methods, such as Random Forests and Multilayer Perceptrons (i.e. neural nets with multiple hidden layers), were also tested. All results and the key facts for our testing procedure are presented in Table 2.

Subsequently, we look at the variable importance (Table 3), partial dependence (Figures 4-6) and Shapley values (Figure 7) of our features. Unsurprisingly, the variable importance of price growth is very high, and both our partial dependence plots and the Shapley value plot indicate that significant increases in prices lead to an increase in the price-to-rent ratio, while rental increases tend to decrease the price-rent ratio. However, based on the Partial Dependence Plot and the Shapley Values, this effect does not appear to be constant across the entire range of growth rates but is instead disproportionately strong in cases of more pronounced price increases. A stronger decoupling of prices from rents is, in other words, more likely to be observed when there have been significant price increases. The PTR ratio appears to be less sensitive to price change for lower price levels. This aligns with the findings of Barberis et al. (2018) that substantial price increases, detached from rental income, have a strong effect on expectations for future price increases.

Table 2

Five-fold cross-validation was used for a hyperparameter search and model selection and a test-set was split to evaluate the "real" out-of-sample-performance of all models.

Test and Validation Settings	
Validation	5-fold Cross-Validation
Test-Set	0.20
Precision and Hyperparameter Boosting	
OOS R2	0.48
IS R2	0.75
Number Trees	1000.00
Tree Depth	3.00
Precision Benchmark models	
OOS R2 Random Forest	0.42
OOS R2 Neural Net	0.40
OOS R2 OLS	0.04

Note: The selected model was a boosted regression model with 1000 trees and a tree depth of three. The other tested alternatives were random forests and multilayer perceptron / neural nets, but they performed slightly worse compared to boosted regression trees

Source: own study.

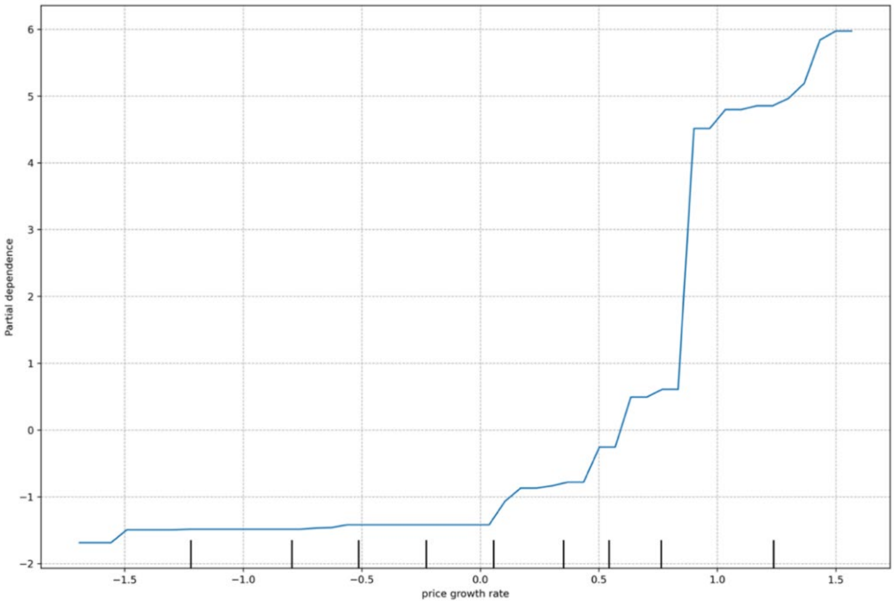
Furthermore, the variable importance for the variance of trend adjusted housing prices is relatively high. The partial dependence plots and the Shapley value plots indicate that the experience of stronger price fluctuations result in higher price-rent ratios, suggesting that investors' willingness to pay for housing increases with greater price uncertainty. This seems counter intuitive at first glance, as we would usually expect a positive correlation between the risk and yield of investments. Owner occupied homes do, however, have option-like characteristics, since they can be occupied by investors in case of decreasing prices and can be sold if prices increase significantly. The effect of the price variance on PTR ratios moreover shows a saturation for higher variance levels. The variable importance of income variance appears to be relatively high as well, and partial dependence plots and Shapley plot indicate that higher income variance leads to a lower PTR ratio. This can be explained by a lower willingness to invest larger amounts of money in cases of higher income uncertainty. Based on Shapley values and the partial dependence plot, this effect appears to be asymmetric as well, and the sensitivity of PTRs ratios to income variance is greater lower variance levels and the effect of changes in the variance seems to be moderate if the variance levels are already high. Our model also confirms a strong connection between stock and housing markets with dividend yields having a negative impact on the price-rent ratios. Finally, our model does not confirm a positive relationship between rental price uncertainty (measured by rental price variance) and price-rent ratios as posited in the literature. The effect of the local bill-rate is moreover only moderate, which is in line with the findings by Campbell et al. (2009), Huh & Kim, 2021 and Kishor & Morley, 2015.

Table 3

Variable importance for all features included in our model

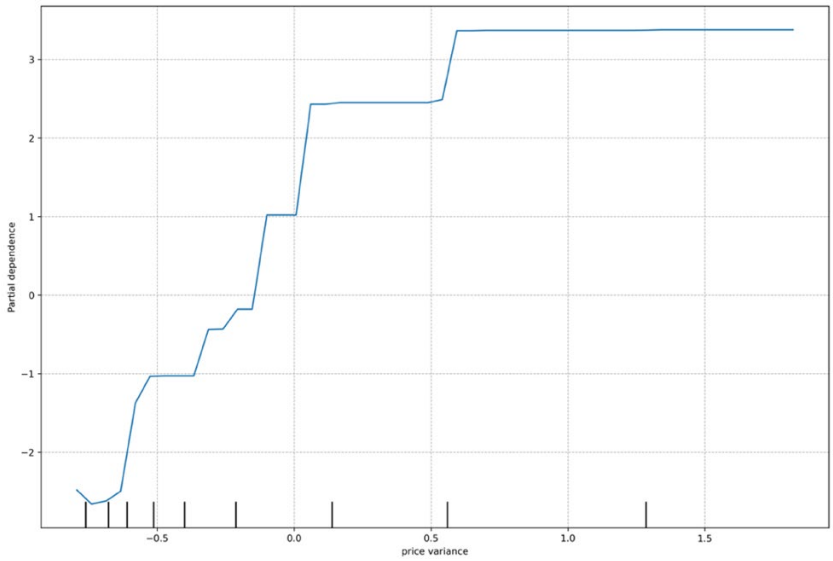
annualized price growth (last 20 years)	0.32
trend adjusted price variance	0.3
annualized rental growth (last 20 years)	0.09
wage variance	0.07
rental price variance	0.06
dividend yield local stock market	0.05
bill rate	0.05

Source: own study.



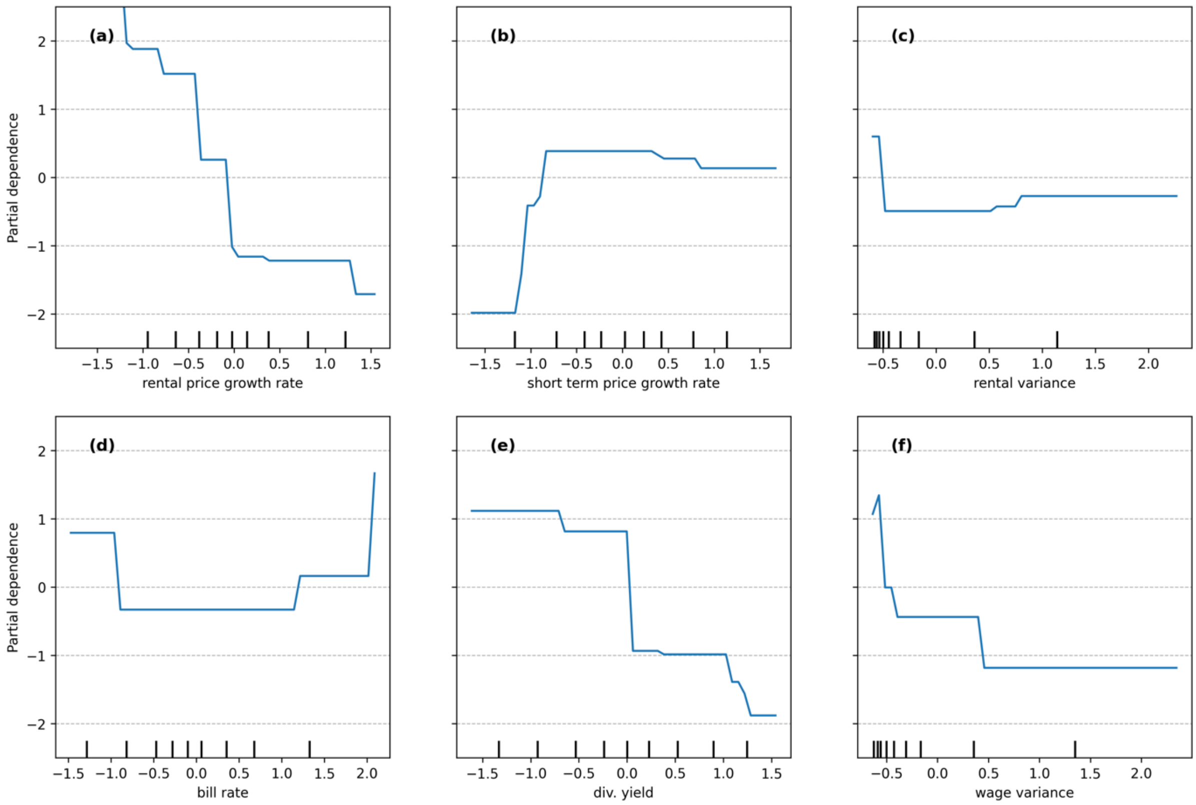
Alt Text for Graphical Figure 4: A partial dependence plot for the annual real price growth rate. The curve slopes upwards and the slope is more extreme for higher values of the annual real price growth rate.

Fig. 4. Partial dependence plot for price growth rate. *Source*: own study.



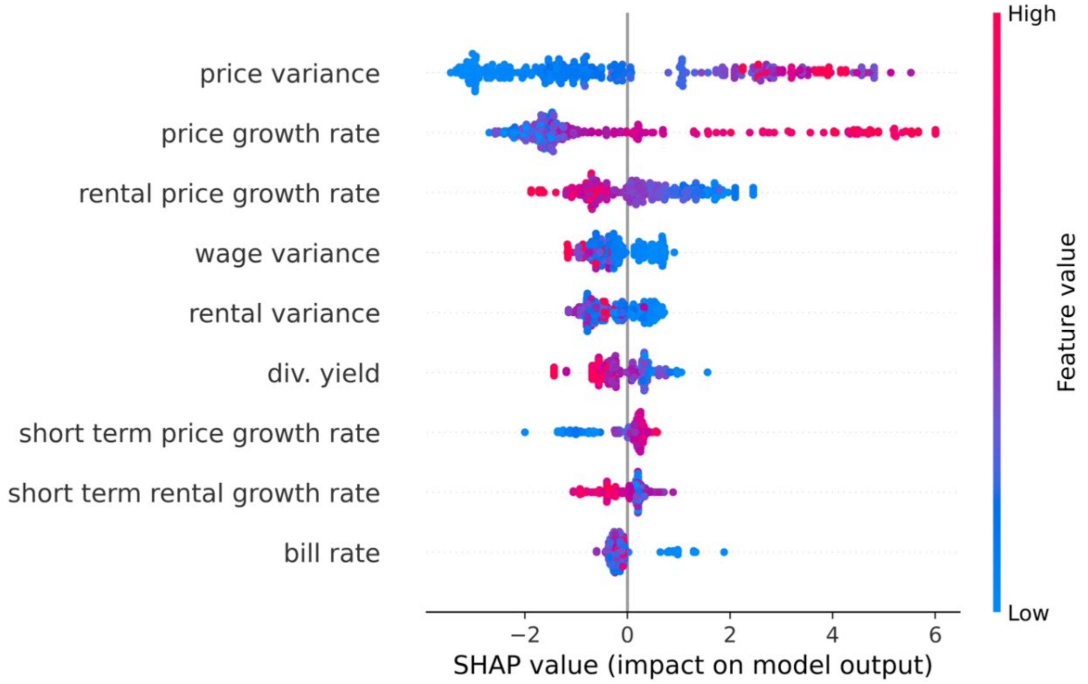
Alt Text for Graphical Figure 5: A partial dependence plot for variance of the trend and inflation adjusted housing price. The curve slopes upwards and the slope is more extreme for lower values of the price variance.

Fig. 5. Partial dependence plot for trend adjusted price variance. *Source*: own study.



Alt Text for Graphical Figure 6-a: A partial dependence plot for the annualized real rental price growth rate. The curve slopes downwards. Alt Text for Graphical Figure 6-b: A partial dependence plot for the short-term annualized real price growth rate. The curve slopes strongly upwards for low levels of the growth rate and is rather flat for higher values. Alt Text for Graphical Figure 6-c: A partial dependence plot for the variance of real rental price growth rate. The curve slopes strongly downwards for low levels of the rental variance and is rather flat for higher values. Alt Text for Graphical Figure 6-d: A partial dependence plot for the bill rate. The curve is moderately U-shaped. Alt Text for Graphical Figure 6-e: A partial dependence plot for the dividend yield. The curve slopes downwards. Alt Text for Graphical Figure 6-f: A partial dependence plot for the wage variance. The curve slopes strongly downwards for low levels of the variance and is rather flat for higher values.

Fig. 6. Partial dependence plot for all other features. *Source*: own study.



Alt Text for Graphical Figure 5: This graph shows the Shapley-values for all features in our model. The X-axis shows the average impact of a feature value on the target (i.e. the price-rent ratio) and the color of each data point reflects the respective feature value. The impact on the target is strongly positive for high levels of the price variance, low levels of the rental price growth rate, low levels of the wage variance, low levels of the dividend-yield.

Fig. 7. General Shapley values for all features. *Source*: own study.

Our study primarily explores the factors relevant to the variation in PTR ratios across countries and over time, with a focus on an applied scientific perspective. However, as previously outlined, the resulting model can also provide practical value by aiding in the interpretation of observed PTR ratios.

Specifically, we can compare the observed PTR ratios at the country level with the model's expectations and analyze whether a particular PTR ratio is lower or higher than our model would expect. If the expected value conditioned on all model features falls significantly below the observed value, this can be interpreted as an unusually high valuation given these features. Additionally, we can enhance this comparison by examining the impact of individual features on a specific prediction by adapting the concept of Shapley values to single forecasts, rather than using them to analyze the model as a whole. In Table 4, we present the actual PTR values and the expected PTR values based on the model for the last year included in the dataset (2019). Such a comparison can help us to identify unexpected valuation patterns. The results are shown in Table 3. Our results indicate that the high valuation levels in Australia, Spain, and Great Britain are rather unexpected, whereas the high valuation level in France can be well explained. At the same time, there were markets in 2019 whose valuations were lower than predicted by our model, such as Japan and the United States. In Figure 8, we see a breakdown of the forecast for the country with the largest discrepancy between the actual value and the value expected based on the model (Australia). We observe that the already relatively high forecast is strongly driven by a very smooth price development with minimal fluctuations. However, the relatively low dividend yield of the national stock market (4%) - particularly in comparison to other countries in the short term - and the not exceptionally low bill rate of 2% prevent an even higher forecast which would be closer to the actual PTR ratio of Australia in 2019. Overall the actual value of the PTR ratio lies substantially above the expectation based on the model, which can be interpreted as a sign for over-valuation.

Table 4

Actual and predicted PTR-ratios for all countries in 2019. We see that the values in Australia and Spain are particuarly difficult to explain

Country	PTR-Ratio	Predicted PTR-Ratio	Absolute Error
AUS	59.63	36.48	23.16
ESP	50.80	36.38	14.42
GB	44.70	34.38	10.32
BEL	32.91	28.32	4.59
NLD	27.06	24.11	2.95
DNK	32.28	29.62	2.66
GER	28.44	25.90	2.54
FIN	16.89	19.20	2.32
FRA	38.21	36.46	1.74
NOR	30.55	29.16	1.39
PRT	32.45	33.65	1.20
JPN	19.45	20.16	0.71
USA	22.62	23.12	0.51
SWE	33.16	32.76	0.40
CHE	27.62	27.28	0.34
ITA	23.89	24.03	0.14

Source: own study.

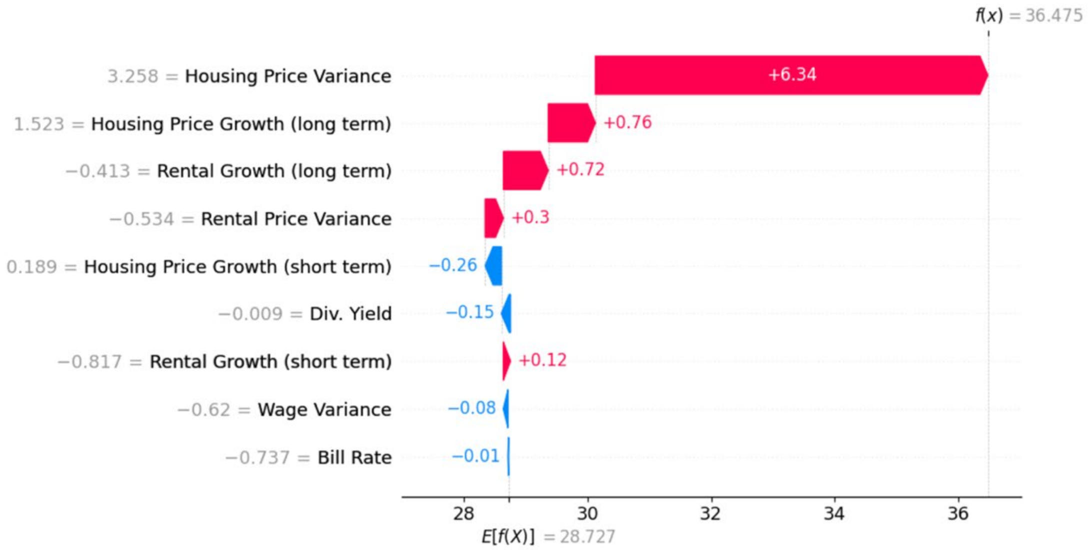


Fig. 8. Shapley values for the Australia 2019 forecast. *Source*: own study.

The difference between the expected and the actual PTR ratio could be utilized within a quantitative investment approach for REITs. For example, it would be possible to model the expected return of real estate investments in a particular market considering this difference and incorporate it into the investment decision-making process, similar to factor-based equity strategies. A more detailed examination of such an approach, however, goes beyond the scope of this article and represents a potential avenue for further research.

Another application of the described difference between the concrete PTR ratio and the expected values derived from the model could be the design of housing policies. An unexpectedly high PTR ratio could, for example, indicate that, given the current rent level, the attractiveness of property acquisition (e.g., through new construction) is too low, and a housing shortage could possibly be explained by potential rent caps like the ones implemented in Germany.

5. Conclusions

Our analysis shows that the dispersion of historically and currently observed PTR values, and the implied housing risk premium is rather wide. PTR ratios above 33.47 have only been observed in 10% of all cases and a value of 16.52 was exceeded in 75% of the observed years. This also means that current (2024) observed values for some European countries can be considered as very high in the context of our long-term distribution, while the values for some countries are relatively low. The median value of the implied housing risk premium is 5.48 if we ignore running costs and 3.66 when we assume that running costs amount to 35% of rental prices which is in line with theoretical results and results based on micro data. Boosted regression trees modeling the variation in PTR-ration across countries and across time based on the current state as well as past development of the national economy and the national housing market can explain 75% of the variation of the PTR ratio in-sample and 43% out-of-sample. Our model analysis suggests that the impact of several macro variables on the PTR ratio is nonlinear. The experienced increase in prices and a high experienced variance of the trend adjusted housing prices have a strong impact on PTR ratios. The sensitivity of the PTR ratio towards these variables is moreover higher when current values are already high. An increase in wage variance has a negative effect on PTR ratios, and their sensitivity toward changes in wage variance is higher for low variance levels. The relevance of the dividend yield supports the hypothesis of a strong relation between stock and housing markets. Finally our results support the hypothesis of a merely moderate effect of changes in interest rates.

References

Anundsen, A. K., & Jansen, E. S. (2013). Self-reinforcing effects between housing prices and credit. *Journal of Housing Economics, 22*, 192–212. https://doi.org/10.1016/j.jhe.2013.07.001

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics, 129*, 203–227. https://doi.org/10.1016/j.jfineco.2018.04.007

Bartholdy, J., & Peare, P. (2005). Estimation of expected return: CAPM vs. Fama and French. *International Review of Financial Analysis, 14*, 407–427. https://doi.org/10.1016/j.irfa.2004.10.009

Bracke, P. (2015). House prices and rents: Microevidence from a matched data set in central London. *Real Estate Economics, 43*, 403–431. https://doi.org/10.1111/1540-6229.12062

Bracke, P., Pinchbeck, E. W., & Wyatt, J. (2018). The time value of housing: Historical evidence on discount rates. *Economic Journal (London), 128*, 1820–1843. https://doi.org/10.1111/ecoj.12501

Bücker, M., Szepannek, G., Gosiewska, A., & Biecek, P. (2021). Transparency, auditability, and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society*, *73*(1), 70–90. https://doi.org/10.1080/01605682.2021.1922098

Cabrol, A., Drobetz, W., Otto, T., & Puhan, T. (2024). Predicting corporate bond illiquidity via machine learning. *Financial Analysts Journal, 80*, 1–25. https://doi.org/10.1080/0015198X.2024.2350952

Cabrol, A., Drobetz, W., Otto, T., & Puhan, T. (2024). Predicting corporate bond illiquidity via machine learning. *Financial Analysts Journal*, *80*(3), 103–127. https://doi.org/10.1080/0015198X.2024.2350952

Campbell, J. Y., & Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies, 1*, 195–228. https://doi.org/10.1093/rfs/1.3.195

Campbell, S. D., Davis, M. A., Gallin, J., & Martin, R. F. (2009). What moves housing markets: A variance decomposition of the rent-price ratio. *Journal of Urban Economics, 66*, 90–102. https://doi.org/10.1016/j.jue.2009.06.002

Caruana, R., & Niculescu-Mizil, A. 2006. An empirical comparison of supervised learning algorithms, in: Proceedings of the 23rd International Conference on Machine Learning, ICML '06. Association for Computing Machinery, New York, NY, USA, pp. 161–168. https://doi.org/10.1145/1143844.1143865

Chambers, D., Spaenjers, C., & Steiner, E. (2021). The rate of return on real estate: Long-run micro-level evidence. *Review of Financial Studies, 34*, 3572–3607. https://doi.org/10.1093/rfs/hhab028

Chen, J., Chen, Y., Hill, R. J., & Hu, P. (2022). The user cost of housing and the price-rent ratio in Shanghai. *Regional Science and Urban Economics, 92*, 103738. https://doi.org/10.1016/j.regsciurbeco.2021.103738

Clark, S., & Lomax, N. (2020). Rent/price ratio for English housing sub-markets using matched sales and rental data. *Area, 52*, 136–147. https://doi.org/10.1111/area.12555

D'Lima, W., & Schultz, P. (2022). Buy-to-rent investors and the market for single family homes. *The Journal of Real Estate Finance and Economics, 64*, 116–152. https://doi.org/10.1007/s11146-020-09790-5

Demers, A., & Eisfeldt, A. L. (2022). Total returns to single-family rentals. *Real Estate Economics, 50*, 7–32. https://doi.org/10.1111/1540-6229.12353

Friedman, J. H. (2000). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*, 1189–1232.

Gu, S., Kelly, B., Xiu, D. (2020). Why do tree-based models still outperform deep learning on tabular data? Empirical Asset Pricing via Machine Learning. *Review of Financial Studies, 33*, 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Halket, J., Nesheim, L., & Oswald, F. (2020). The housing stock, housing prices, and user costs: The roles of location, structure, and unobserved quality. *International Economic Review, 61*, 1777–1814. https://doi.org/10.1111/iere.12475

Hill, R. J., & Syed, I. A. (2016). Hedonic price–rent ratios, user cost, and departures from equilibrium in the housing market. *Regional Science and Urban Economics, 56*, 60–72. https://doi.org/10.1016/j.regsciurbeco.2015.11.001

Huh, S., & Kim, I. (2021). The housing risk premium in a production economy. *Applied Economics Letters, 28*, 213–219. https://doi.org/10.1080/13504851.2020.1740158

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., & Taylor, A. M. (2019). The rate of return on everything, 1870–2015*. *The Quarterly Journal of Economics, 134*, 1225–1298. https://doi.org/10.1093/qje/qjz012

Kaufmann, H., Messow, P., & Vogt, J. (2021). Boosting the equity momentum factor in credit. *Financial Analysts Journal, 77*, 83–103. https://doi.org/10.1080/0015198X.2021.1954377

Kim, J., & Lim, G. (2014). Understanding the Irish price–rent ratio: An unobserved component approach. *Applied Economics Letters, 21*, 836–841. https://doi.org/10.1080/13504851.2014.892191

Kishor, N. K., & Morley, J. (2015). What factors drive the price–rent ratio for the housing market? A modified present-value analysis. *Journal of Economic Dynamics & Control, 58*, 235–249. https://doi.org/10.1016/j.jedc.2015.06.006

Lo, D., McCord, M., Davis, P. T., McCord, J., & Haran, M. E. (2023). Causal relationships between the price-to-rent ratio and macroeconomic factors: A UK perspective. *Journal of Property Investment & Finance, 41*, 11–34. https://doi.org/10.1108/JPIF-08-2021-0068

Lo, D., McCord, M.J., McCord, J., Davis, P.T. & Haran, M. (2021). Rent or buy, what are the odds? Analysing the price-to-rent ratio for housing types within the Northern Ireland housing market. *International Journal of Housing Markets and Analysis, 14*(5), 1062-1091. https://doi.org/10.1108/IJHMA-08-2020-0103

Lo, D., Yau, Y., McCord, M., & Haran, M. (2022). Lead-lag relationship between the price-to-rent ratio and the macroeconomy: An empirical study of the residential market of Hong Kong. *Buildings, 12*, 1345. https://doi.org/10.3390/buildings12091345

Mork, K. A., & Trønnes, H. A. (2023). Expected long-term rates of return when short-term returns are serially correlated. *International Review of Financial Analysis, 88*, 102696. https://doi.org/10.1016/j.irfa.2023.102696

Okunev, J., Wilson, P., & Zurbruegg, R. (2000). The causal relationship between real estate and stock markets. *The Journal of Real Estate Finance and Economics, 21*, 251–261. https://doi.org/10.1023/A:1012051719424

Piazzesi, M., Schneider, M., & Tuzel, S. (2007). Housing, consumption and asset pricing. *Journal of Financial Economics, 83*, 531–569. https://doi.org/10.1016/j.jfineco.2006.01.006

Plogmann, J., Mußhoff, O., Odening, M., & Ritter, M. (2020). What Moves the German Land Market? A Decomposition of the Land Rent-Price Ratio. *Ger. J. Agric. Econ., 69*. Advance online publication. https://doi.org/10.30430/69.2020.1.1-18

Rossi, A. G., & Utkus, S. (2024). The diversification and welfare effects of robo-advising. *Journal of Financial Economics, 157*, 103869. https://doi.org/10.1016/j.jfineco.2024.103869

Rozemberczki, B., Watson, L., Bayer, P., Yang, H.-T., Kiss, O., Nilsson, S., Sarkar, R. (2022). The shapley value in machine learning. https://doi.org/10.24963/ijcai.2022/778

Tsai, I.-C. (2015). Dynamic information transfer in the United States housing and stock markets. *The North American Journal of Economics and Finance, 34*, 215–230. https://doi.org/10.1016/j.najef.2015.09.012