International Journal of Forecasting 41 (2025) 208–228

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

# Predicting the equity premium around the globe: Comprehensive evidence from a large sample[☆]

Fabian Hollstein[a,*], Marcel Prokopczuk[b,c], Björn Tharann[b], Chardin Wese Simen[d]

[a] Saarland University, School of Human and Business Sciences, Campus C3 1, 66123 Saarbruecken, Germany
[b] Leibniz University Hannover, School of Economics and Management, Koenigsworther Platz 1, 30167 Hannover, Germany
[c] University of Reading, Henley Business School, ICMA Centre, Reading, RG6 6BA, UK
[d] University of Liverpool, Management School, L69 7ZH, Liverpool, UK

ARTICLE INFO

*Keywords:*
International equity premium
Return predictability
Market efficiency
Market development
Cross-market return predictability

ABSTRACT

Examining 81 countries over a period of up to 145 years and using various predictor variables and forecasting specifications, we provide a detailed analysis of equity premium predictability. We find that excess returns are more predictable in emerging and frontier markets than in developed markets. For all groups, forecast combinations perform very well out of sample. Analyzing the cross-section of countries, we find that market inefficiency is an important driver of return predictability. We also document significant cross-market return predictability. Finally, domestic inflation-adjusted returns are significantly more predictable than USD returns.



## 1. Introduction

Whether or not the equity premium is predictable is an important question that has been analyzed for at least a century. One of the first attempts to predict stock returns dates back to Dow (1920). Several early studies find that stock returns are predictable by macroeconomic variables (e.g., Fama and French, 1988; Campbell, 1991; Cochrane, 1992). Conversely, Goyal and Welch (2008) argue that most predictor variables have limited out-of-sample predictive power. Following this seminal study, several articles show that, with specification adjustments, U.S. returns are predictable to a significant degree (e.g., Campbell and Thompson, 2008; Rapach et al., 2010; Johnson, 2018; Rapach and Zhou, 2020). Less is known about the predictability of the equity premium around the world.

Studying return predictability is important from both a theoretical and a practical perspective. Investors could improve their portfolio allocations across markets. Knowledge of country-specific market return predictability could enable investors to develop profitable market timing strategies. For theorists, understanding the sources and drivers of return predictability is important because we need to distinguish between rational predictability (e.g., Campbell and Cochrane, 1999; Bansal and Yaron, 2004) and irrational or friction-based predictability, which may be more concentrated in less efficient markets. Finally, it is important to ascertain whether predictability is "real" and not a spurious result of data mining.

In this context, the main objective of this paper is to analyze whether the equity premium is predictable across the globe. In doing so, we make two contributions to the existing literature. First, to the best of our

☆ We thank George Kapetanios (the editor), an anonymous associate editor, an anonymous referee, and Fabian Bätje for helpful comments and suggestions.

* Corresponding author.
*E-mail addresses:* fabian.hollstein@uni-saarland.de (F. Hollstein), prokopczuk@fmt.uni-hannover.de (M. Prokopczuk), tharann@fmt.uni-hannover.de (B. Tharann), c.wesesimen@liverpool.ac.uk (C. Wese Simen).

https://doi.org/10.1016/j.ijforecast.2024.05.002

knowledge, we provide the most comprehensive analysis of the predictability of market returns in terms of sample length and cross-section. Indeed, our sample period extends from January 1871 to December 2015 and includes 81 countries with more than 54,000 country–month observations.[1] Most of the literature has focused either on the U.S. stock market or on a small number of developed countries. This narrow focus may raise concerns about data snooping (Lo and MacKinlay, 1990). With our data, we are able to directly address these concerns and provide *out-of-sample* evidence by extending the analysis to many countries around the world. In particular, we test whether the specification adjustments that work well for the U.S. also improve predictability in international markets. Importantly, compared to existing studies, we also include less developed countries whose stock markets are only weakly correlated with those of developed countries. Thus, we add a significant number of independent observations.

Second, the large number of countries allows us to gain insights into the determinants of the predictability of market excess returns. The countries in our sample are heterogeneous along several dimensions, for example, the size of the capital market and market openness. By exploiting this heterogeneity in the cross-section of countries, we can gain insights into the economic sources of international return predictability.

Our focus is on analyzing the predictive power of various economic variables for the one-month USD denominated market excess return. Consistent with Goyal and Welch (2008), we find that international market excess returns are largely unpredictable out-of-sample using standard variables and techniques. While there is some in-sample return predictability for most variables, standard ordinary least squares (OLS) out-of-sample predictions mostly underperform the simple historical mean benchmark. Only technical indicators have both positive and significant in-sample and out-of-sample OLS $R^2$s. In particular, the use of simple forecast combinations yields substantial and significant out-of-sample $R^2$s for countries at all development stages around the world.

An important finding is that in sample and out of sample, return predictability is overall stronger in emerging and frontier markets than in developed markets. This is also true for economic utility gains, which are substantial on average and more often positive than the out-of-sample $R^2$s. Thus, international investors who rely on return predictability for their portfolio allocation strategies can benefit substantially.

Having found predictability around the world, an important question is what drives this predictability: market efficiency with rationally time-varying expected returns or mainly financial frictions? To the best of our knowledge, we are the first to analyze this in an international setting. The generally stronger predictability of returns in countries with less developed capital markets and the substantial predictive power of technical signals mentioned above already point to a market inefficiency explanation. Pursuing our analysis, we sort developed, emerging, and frontier market countries by proxies for market efficiency. Within each group, we generally find stronger predictability among the countries with lower market efficiency, proxied by market capitalization and GDP per capita.

We also analyze predictability in a panel setting by pooling together the observations from all countries for the parameter estimation. This approach works particularly well for out-of-sample forecasts of countries within the same stage of development. We also use lagged average returns of developed, emerging, and frontier markets to predict the market excess returns of each country. We find that this approach also works well. In particular, lagged average returns of similarly or more developed countries are strong and significant predictors.

When we analyze the predictability of domestic inflation-adjusted market excess returns (instead of USD returns as in our main analysis), we find even stronger evidence of in-sample and out-of-sample return predictability. We thus conclude that the exchange rate movements reduce return predictability. This is consistent with a large body of literature showing that exchange rates are difficult to predict (e.g., Meese & Rogoff, 1983; Rossi, 2013). Furthermore, the economic sign restrictions of Campbell and Thompson (2008) improve the return predictability.

Naturally, there are limitations to our study. It is possible that different sample periods and definitions of variables across countries affect the return predictability. In addition, data quality may vary across countries, depending on their political and economic conditions. We address these issues by using the variables from the same database and by accounting for potential time variation in the predictive relationship, e.g., by using a rolling rather than an expanding window in the out-of-sample analysis. In addition, we test the robustness of our results for a much shorter post-1990 period which, at the cost of reduced power of the statistical tests, substantially reduces the heterogeneity of the dataset in most dimensions. Overall, the results for this reduced sample period are very similar to those for the full sample period. Further robustness tests also show that the results are qualitatively similar for alternative rolling and expanding window specifications for the out-of-sample analysis.

This paper is related to the literature on the predictability of U.S. market excess returns, which mainly uses aggregate valuation ratios as predictors. Variables that have been extensively studied in the existing literature include the dividend–price ratio (Rozeff, 1984; Fama and French, 1988; Hodrick, 1992), short-term interest rates (Campbell, 1987; Hodrick, 1992; Ang and Bekaert, 2007), and the consumption–wealth ratio (Lettau and Ludvigson, 2001). Campbell and Shiller (1988, 1998) show that the price–earnings ratio in particular predicts long-term stock returns.

In recent years, several studies have examined a wide range of accounting-based valuation ratios (e.g., Rapach and Wohar, 2006; Rapach and Zhou, 2013). Goyal and Welch (2008) conduct a comprehensive analysis of predictability, and document that many existing methodologies produce unstable or spurious results due to serious

[1] The exact sample length for different countries is, of course, heterogeneous, depending on data availability.

econometric problems. They conclude that the previously documented predictability of the equity premium is not robust and that the historical mean is the best predictor of the equity premium. However, many subsequent studies document predictability when implementing economically motivated sign restrictions (e.g., Campbell & Thompson, 2008), forecast combinations (e.g., Rapach & Zhou, 2013), weighted least squares regressions (e.g., Johnson, 2018), and machine learning methods (e.g., Rapach & Zhou, 2020), among others. We extend this analysis to a large international sample and provide out-of-sample evidence.

Most studies of international stock return predictability focus on a limited number of European countries. Ang and Bekaert (2007) analyze Germany, the U.K., and the U.S. They document the short-term predictive power of dividend yields in combination with short-term interest rates. Golez and Koudijs (2018) analyze return predictability over four centuries by combining Dutch, British, and U.S. data. They provide evidence of a strong annual and multiyear predictability, and show that expected returns are higher in recessions. Henkel et al. (2011) examine the G7 countries and find short-term predictability of macroeconomic variables only in recessions. Rangvid et al. (2014) provide evidence on the predictability of dividend growth in an international setting. In addition, Jordan et al. (2014) consider 14 European countries and provide evidence for return predictability of the short-term interest rate and historical stock return variance. Charles et al. (2016) analyze Asian and European countries, and document the weak predictive power of financial ratios and moderate short-term predictability of several macroeconomic variables. Rapach et al. (2013) show that lagged U.S. aggregate returns have predictive power for those of 10 non-U.S. developed markets.

The paper most closely related to ours is Hjalmarsson (2010), which examines a set of 40 countries and four predictor variables. The paper's main finding is that interest-related variables are generally better predictors than dividend- or earnings-related variables. Our paper differs from Hjalmarsson (2010) in several important ways. First, we examine a much larger cross-section of more than twice as many countries and more than 54,000 country–month observations compared to about 20,000. Second, we expand the set of predictor variables to include volatility, inflation, the unemployment rate, and technical signals. This allows us to compare the information content of various variables. Third, we exploit the heterogeneity of international countries to analyze the economic sources of return predictability. Finally, we conduct a rigorous out-of-sample test, including model selection approaches and tests of the economic significance of the results.

The rest of this paper is organized as follows. Section 2 presents the data and methodology. In Section 3, we present the main empirical results. In Section 4, we repeat the analysis for the post-1990 sample. Section 5 examines the economic sources of return predictability. In Section 6, we conduct further analyses and several robustness tests. In Section 7, we draw conclusions.

## 2. Data and methodology

### 2.1. Data

We obtain the monthly time series of equity market indices for 81 countries from the Global Financial Database (GFD).[2] Our sample period is from January 1871 to December 2015. All time series are directly available in both domestic currency and USD. For the U.S. three-month Treasury-bill rate, we use the extended dataset of Goyal and Welch (2008).[3] Table 1 provides an overview of the countries studied and the number of observations for each market index. In order to structure our analysis, we group the countries according to their development status provided by Morgan Stanley Capital International (MSCI), as well as their geographical regions.

We obtain several measures that characterize the economic strength and the investment climate of each country. We obtain the GDP per capita (in USD and base year 2010) from the World Bank.[4] In addition, we use data on stock market capitalizations (in USD), adjusted by the GDP implicit price deflator.[5] Finally, we use the Chinn and Ito (2006) index, which the authors compute as the first principal component of several indicator variables measuring capital controls, as a measure of market openness. We use the average of the standardized Chinn–Ito index as the classification criterion.[6] Table A1 in the Online Appendix provides the tickers for all time series used in this paper.

### 2.2. Variables

*Market excess returns.* We calculate the market excess return as the difference between the log-return on the market index and the risk-free rate for the corresponding period:

$$ER_{t+1} = \log\left(\frac{I_{t+1}}{I_t}\right) - rf_{t+1}, \tag{1}$$

where $ER_{t+1}$ is the monthly excess return of the specific market index at the end of month $t+1$. $I_{t+1}$ and $I_t$ denote the (total return) index value at the end of months $t+1$ and $t$, respectively. To ensure that the results are comparable across countries, we use total return indices denominated in USD. $rf_{t+1}$ refers to the log risk-free rate in month $t+1$. Following Goyal and Welch (2008), we use the three-month U.S. Treasury-bill rate as a proxy for the risk-free rate.

[2] Due to a lack of data availability in the GFD, we obtain the market indices for Ecuador (in USD) and Russia (in domestic currency) from Datastream. Due to a lack of data availability in Datastream, we use the index of Ecuador in USD only.

[3] The dataset is available at http://www.hec.unil.ch/agoyal/.

[4] The dataset is available at http://data.worldbank.org/indicator/NY.GDP.PCAP.KD?locations=CL.

[5] We obtain stock market capitalizations from the GFD and the GDP implicit price deflator from the Federal Reserve Bank of St. Louis (FRED). The dataset is available at https://fred.stlouisfed.org/series/GDPDEF.

[6] The standardized Chinn–Ito index is defined between 0 (no market openness) and 1 (complete market openness).

**Table 1**
Summary statistics – Market excess returns.

| | Mean | Median | SD | Skew | Kurt | AR(1) | Nobs | First |
|---|---|---|---|---|---|---|---|---|
| **Developed markets** | | | | | | | | |
| Average | 0.018 | 0.036 | 0.217 | −0.783 | 11.23 | 0.133 | 28,169 | |
| *Asia Pacific* | | | | | | | | |
| Australia | 0.078 | 0.096 | 0.169 | −1.598 | 15.77 | 0.069 | 1,599 | Oct-1882 |
| Hong Kong | 0.051 | 0.080 | 0.319 | −0.585 | 7.995 | 0.077 | 617 | Aug-1964 |
| Japan | 0.023 | 0.034 | 0.233 | −0.299 | 12.39 | 0.121 | 1,206 | Aug-1914 |
| New Zealand | −0.004 | 0.031 | 0.194 | −1.534 | 14.41 | 0.106 | 1,020 | Jan-1931 |
| Singapore | 0.028 | 0.047 | 0.248 | −0.438 | 6.134 | 0.138 | 605 | Aug-1965 |
| *Europe* | | | | | | | | |
| Austria | 0.004 | −0.004 | 0.260 | −0.876 | 16.21 | 0.226 | 1,071 | Feb-1922 |
| Belgium | 0.001 | 0.012 | 0.197 | −0.326 | 6.074 | 0.081 | 1,357 | Feb-1897 |
| Denmark | 0.009 | 0.009 | 0.168 | −0.089 | 16.33 | 0.174 | 1,475 | Feb-1893 |
| Finland | 0.005 | 0.011 | 0.249 | −0.699 | 9.496 | 0.173 | 1,238 | Nov-1912 |
| France | −0.004 | 0.013 | 0.228 | −1.279 | 12.45 | 0.078 | 1,400 | Jan-1898 |
| Germany | −0.020 | 0.058 | 0.305 | −2.154 | 18.83 | 0.279 | 1,740 | Jan-1871 |
| Ireland | 0.013 | 0.037 | 0.182 | −0.954 | 10.23 | 0.175 | 983 | Feb-1934 |
| Israel | 0.010 | 0.027 | 0.259 | −1.833 | 13.37 | 0.047 | 803 | Feb-1949 |
| Italy | −0.010 | −0.025 | 0.275 | −0.115 | 10.42 | 0.172 | 1,317 | Oct-1905 |
| Netherlands | 0.010 | 0.043 | 0.184 | −0.817 | 6.939 | 0.123 | 1,142 | Feb-1919 |
| Norway | 0.002 | −0.009 | 0.203 | −1.037 | 9.684 | 0.159 | 1,211 | Feb-1915 |
| Portugal | 0.017 | 0.002 | 0.252 | 1.013 | 13.03 | 0.176 | 949 | Jan-1934 |
| Spain | −0.016 | 0.022 | 0.206 | −0.977 | 8.604 | 0.151 | 1,168 | Jan-1915 |
| Sweden | 0.054 | 0.081 | 0.186 | −0.721 | 7.475 | 0.124 | 1,370 | Nov-1901 |
| Switzerland | 0.019 | 0.054 | 0.170 | −0.697 | 7.748 | 0.080 | 1,207 | Jan-1914 |
| United Kingdom | 0.043 | 0.058 | 0.159 | −0.411 | 13.74 | 0.073 | 1,740 | Jan-1871 |
| *North America* | | | | | | | | |
| Canada | 0.042 | 0.069 | 0.181 | −1.140 | 9.062 | 0.147 | 1,211 | Feb-1915 |
| United States | 0.062 | 0.090 | 0.165 | −0.446 | 11.79 | 0.109 | 1,740 | Jan-1871 |
| **Emerging markets** | | | | | | | | |
| Average | 0.023 | 0.030 | 0.336 | −0.342 | 7.356 | 0.139 | 14,711 | |
| *Africa* | | | | | | | | |
| South Africa | 0.012 | 0.028 | 0.209 | −0.945 | 7.992 | 0.123 | 1,271 | Feb-1910 |
| *Asia Pacific* | | | | | | | | |
| China | 0.093 | 0.133 | 0.291 | −0.017 | 4.065 | 0.053 | 252 | Jan-1995 |
| India | −0.004 | −0.005 | 0.214 | −0.448 | 8.898 | 0.100 | 1,110 | Jul-1922 |
| Indonesia | −0.018 | −0.003 | 0.376 | −0.422 | 8.507 | 0.226 | 396 | Jan-1983 |
| Malaysia | 0.024 | 0.084 | 0.302 | −0.591 | 6.839 | 0.180 | 517 | Dec-1972 |
| Philippines | −0.062 | −0.069 | 0.314 | 0.043 | 6.077 | 0.189 | 756 | Jan-1953 |
| Republic of Korea | 0.097 | 0.089 | 0.372 | −0.114 | 9.011 | 0.008 | 647 | Feb-1962 |
| Taiwan | 0.046 | 0.028 | 0.335 | −0.218 | 6.626 | 0.107 | 587 | Feb-1967 |
| Thailand | 0.043 | 0.052 | 0.323 | −0.577 | 6.276 | 0.112 | 488 | May-1975 |
| *Europe* | | | | | | | | |
| Czech Republic | 0.029 | 0.123 | 0.316 | −0.081 | 6.687 | 0.256 | 267 | Oct-1993 |
| Greece | −0.006 | −0.074 | 0.298 | 0.280 | 6.790 | 0.181 | 744 | Jan-1954 |
| Hungary | 0.041 | 0.063 | 0.345 | −0.689 | 7.456 | 0.092 | 299 | Feb-1991 |
| Poland | −0.049 | −0.020 | 0.480 | −0.253 | 6.357 | 0.193 | 517 | Feb-1921 |
| Russian Federation | 0.100 | 0.219 | 0.472 | −0.527 | 4.879 | 0.196 | 243 | Oct-1995 |
| *South America* | | | | | | | | |
| Argentina | −0.001 | −0.000 | 0.518 | 0.063 | 5.735 | 0.118 | 588 | Jan-1967 |
| Brazil | 0.037 | −0.015 | 0.464 | −0.247 | 5.103 | 0.061 | 731 | Feb-1955 |
| Chile | −0.096 | −0.110 | 0.368 | −0.588 | 8.751 | 0.206 | 661 | Feb-1960 |
| Colombia | −0.040 | −0.056 | 0.230 | 0.248 | 11.99 | 0.219 | 1,067 | Feb-1927 |
| Mexico | 0.028 | −0.000 | 0.235 | −1.532 | 13.23 | 0.135 | 803 | Feb-1938 |
| Peru | −0.004 | −0.027 | 0.330 | 0.446 | 10.81 | 0.037 | 996 | Jan-1933 |
| **Frontier markets** | | | | | | | | |
| Average | 0.018 | 0.019 | 0.264 | −0.188 | 9.015 | 0.208 | 11,454 | |
| *Africa* | | | | | | | | |
| Botswana | 0.084 | 0.081 | 0.168 | 0.637 | 6.652 | 0.285 | 319 | Jun-1989 |
| Ghana | 0.001 | −0.041 | 0.278 | 0.890 | 9.075 | 0.370 | 236 | Dec-1990 |
| Kenya | −0.030 | −0.026 | 0.213 | 0.413 | 10.18 | 0.218 | 623 | Feb-1964 |
| Mauritius | 0.042 | −0.002 | 0.191 | −0.458 | 6.956 | 0.230 | 317 | Aug-1989 |
| Morocco | 0.045 | 0.049 | 0.194 | −2.145 | 22.17 | 0.061 | 336 | Jan-1988 |
| Namibia | −0.002 | 0.102 | 0.298 | −1.054 | 7.216 | 0.160 | 274 | Mar-1993 |

(*continued on next page*)

**Table 1** (*continued*).

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Nigeria | 0.038 | 0.083 | 0.307 | −1.062 | 12.14 | 0.058 | 336 | Jan-1988 |
| Tunisia | 0.038 | −0.010 | 0.148 | −0.081 | 5.431 | 0.049 | 216 | Jan-1998 |
| *Asia Pacific* | | | | | | | | |
| Bangladesh | 0.024 | −0.015 | 0.310 | 0.647 | 8.794 | 0.141 | 408 | Jul-1980 |
| Sri Lanka | 0.004 | −0.036 | 0.226 | 0.516 | 5.393 | 0.178 | 516 | Jan-1963 |
| Vietnam | 0.027 | −0.033 | 0.371 | −0.097 | 4.076 | 0.339 | 180 | Jan-2001 |
| *Europe* | | | | | | | | |
| Bulgaria | −0.155 | −0.074 | 0.405 | −1.579 | 8.729 | 0.337 | 267 | Oct-1993 |
| Croatia | −0.005 | −0.006 | 0.305 | −0.943 | 9.048 | 0.056 | 227 | Feb-1997 |
| Cyprus | 0.014 | −0.015 | 0.131 | 0.402 | 7.087 | −0.045 | 384 | Jan-1984 |
| Estonia | 0.126 | 0.126 | 0.357 | −0.724 | 7.702 | 0.216 | 245 | Aug-1995 |
| Iceland | 0.035 | 0.154 | 0.266 | −2.149 | 13.38 | 0.415 | 276 | Jan-1993 |
| Latvia | 0.071 | 0.114 | 0.336 | 0.114 | 9.565 | 0.296 | 236 | May-1996 |
| Lithuania | 0.072 | 0.099 | 0.324 | 0.658 | 11.65 | 0.217 | 228 | Jan-1996 |
| Luxembourg | 0.036 | 0.047 | 0.193 | −1.025 | 10.03 | 0.177 | 744 | Jan-1954 |
| Malta | 0.046 | −0.004 | 0.191 | 0.316 | 4.514 | 0.224 | 240 | Jan-1996 |
| Romania | −0.030 | 0.034 | 0.507 | −0.253 | 6.770 | 0.194 | 410 | Jan-1931 |
| Slovakia Republic | 0.016 | 0.055 | 0.298 | 0.941 | 12.57 | 0.232 | 267 | Oct-1993 |
| Slovenia | −0.000 | −0.006 | 0.269 | 0.477 | 7.464 | 0.265 | 275 | Feb-1993 |
| Ukraine | −0.074 | −0.011 | 0.461 | −0.674 | 5.825 | 0.351 | 215 | Feb-1998 |
| *Middle East* | | | | | | | | |
| Bahrain | −0.002 | 0.004 | 0.128 | −0.199 | 3.898 | 0.307 | 306 | Jul-1990 |
| Jordan | −0.016 | −0.053 | 0.245 | −0.468 | 7.189 | 0.008 | 455 | Feb-1978 |
| Kuwait | 0.034 | 0.014 | 0.222 | −0.278 | 13.97 | 0.253 | 431 | Feb-1973 |
| Lebanon | −0.008 | −0.075 | 0.251 | 1.112 | 8.734 | 0.148 | 239 | Feb-1996 |
| Oman | 0.040 | 0.061 | 0.198 | −0.556 | 6.930 | 0.219 | 277 | Dec-1992 |
| *South America* | | | | | | | | |
| Ecuador | −0.000 | 0.011 | 0.209 | 0.734 | 12.89 | 0.132 | 264 | Jan-1994 |
| Jamaica | −0.009 | −0.069 | 0.289 | −0.510 | 11.13 | 0.246 | 558 | Jul-1969 |
| Trinidad and Tobago | 0.073 | 0.060 | 0.116 | 0.277 | 8.845 | 0.402 | 201 | Apr-1999 |
| Venezuela | 0.062 | −0.000 | 0.303 | −0.070 | 11.50 | 0.135 | 948 | Jan-1937 |

This table presents summary statistics on the individual market excess returns. We sample all data at a monthly frequency. Countries are assigned to the different panels according to their MSCI market development status and geographical region. The time series of the market indices are denominated in USD. "Mean", "Median", "SD", "Skew", "Kurt", and "AR(1)" denote the (annualized) mean, (annualized) median, (annualized) standard deviation, skewness, kurtosis, and AR(1) coefficient of the monthly log market excess returns, respectively. "Nobs" denotes the number of monthly observations and "First" indicates the first month and year for which data are available for a given country. The "Average" rows show the average of the summary statistics across countries as well as the total number of country–month observations within a market segment.

*Predictor variables.* Rational return predictability requires time variation in expected returns. Expected returns are typically assumed to be high at the troughs of the business cycle and low at the peaks. Accordingly, Cochrane (1999) argues that prices are driven down when future cash flows are discounted at a higher rate. Low prices indicate high expected returns, and vice versa. Thus, in principle, all variables that have some correlation with the business cycle are potential return predictors. In particular, price-related variables are natural candidates for predictive variables.[7] Both high dividend-to-price and earnings-to-price ratios imply that assets have high expected returns. Campbell and Cochrane (1999) and Cochrane (2007) argue that return predictability based on these variables could, for example, be generated by habits that react slowly to changes in consumption.

We base our analysis on 11 main predictor variables, including both fundamental predictors and technical indicators. The fundamental variables consist of three stock market variables, three interest rate variables, and two macroeconomic variables. The stock market variables include the dividend yield (*DY*; e.g., Cochrane, 2008, 2011), the price–earnings ratio (*PE*; e.g., Campbell and Shiller, 1988), and the stock excess return volatility (*RVOL*; estimated from a 12-month rolling window using the estimator of Mele, 2007). Following, for example, Fama and French (1989), the interest rate variables include the current local three-month government bond yield (*RREL*; stochastically detrended by subtracting the previous 12-month average), the term spread (*TMS*; long-term government bond yield minus three-month government bond yield), and the default yield spread (*DFY*; long-term corporate bond yield minus long-term government bond yield).

The macroeconomic variables include the inflation rate (*INFL*; in USD; e.g., Chen et al., 1986; Ferson and Harvey, 1991) and the unemployment rate (*UE*; e.g., Boyd et al., 2005; Rapach et al., 2005). To account for publication delays, we lag both variables by one month. Finally, following Neely et al. (2014) and Rapach and Zhou (2020), we include the following technical indicators: a dummy that is one when the current (USD) total return index value exceeds its 12-month moving average ($MA_{1,12}$), another that is one when the three-month moving average exceeds the 12-month moving average ($MA_{3,12}$), and another dummy that is equal to one when the current

[7] From a statistical perspective, this raises two potential problems. First, there may be measurement error in the predictor variables. We address this in a robustness check by taking the mean of these variables. Second, the results may be affected by the Stambaugh (1999) bias. We use a robust inference technique to address this concern.

(total return) index value exceeds that of six months ago ($MOM_6$).[8]

We obtain the monthly time series for the predictor variables from the GFD.[9] All variables are country-specific.[10] Table A1 in the Online Appendix shows the tickers for each of the time series.[11]

### 2.3. Predictive regressions

We estimate the following single regression model, regressing the one-month-ahead excess return on a constant and the predictor variable:

$$ER_{t+1} = \alpha + \beta X_t + \epsilon_{t+1}, \tag{2}$$

where $ER_{t+1}$ is the market excess return from month $t$ to $t+1$, and $\beta$ is the slope parameter. $X_t$ is a predictor variable observed at the end of month $t$, and $\epsilon_{t+1}$ represents the regression error term.

Since single regression models are typically unstable over time, it may be fruitful to combine information from different sources. To this end, we extend our analysis using methods from the statistical learning literature and introduce four approaches that deal with parameter shrinkage and variable selection. In doing so, we analyze whether forecast combinations further improve the predictive power.

*Mean forecast combination (COMB).* Rapach et al. (2010) find that the use of forecast combinations yields substantial improvements in the out-of-sample predictability in the U.S., relative to single-variable forecasts and to a "kitchen sink" approach that includes multiple predictors in the same model. The authors argue that different variables capture complementary information about the state of the economy. Forecast combinations provide more stable estimates than simple multiple regression, thereby reducing the forecast volatility. To compute the combined out-of-sample forecast, we simply take the equally weighted average of all available out-of-sample forecasts of the different predictors:

$$\widehat{ER}_{t+1}^{COMB} = \frac{1}{B}\sum_{b=1}^{B} \widehat{ER}_{t+1}^{b,oos}, \tag{3}$$

where $\widehat{ER}_{t+1}^{COMB}$ is the combination forecast. $\widehat{ER}_{t+1}^{b,oos}$ is the single-regression out-of-sample forecast of predictor $b$, and $B$ is the number of available forecasts at time $t$.[12]

*Discount mean squared prediction error (DMSPE) combination.* Following Rapach et al. (2010), we also consider the DMSPE optimization approach of Stock and Watson (2004). The DMSPE gives more weight to forecasts that are expected to perform better:

$$\widehat{ER}_{t+1}^{DMSPE} = \sum_{b=1}^{B} \frac{\phi_{b,t}^{-1}}{\sum_{j=1}^{B}\phi_{j,t}^{-1}} \widehat{ER}_{t+1}^{b,oos},$$
$$\text{with } \phi_{b,t} = \sum_{s=t_0}^{t-1} \theta^{t-1-s}\left(ER_{s+1} - \widehat{ER}_{s+1}^{b,oos}\right)^2. \tag{4}$$

$\theta$ is a discount factor to give more weight to recent observations. We follow Rapach et al. (2010) and set $\theta = 0.9$. $t_0$ indicates the beginning of the training window.

*Elastic net (ENET).* Following Rapach et al. (2013), we use the elastic net estimation technique which, similar to OLS, minimizes the sum of squared residuals but subject to two penalty terms. We use the multiple predictive regression model:

$$ER_{t+1} = \alpha + \boldsymbol{\beta}'\boldsymbol{X}_t + \epsilon_{t+1}, \tag{5}$$

where $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_M)$ is the slope parameter vector. $m = 1, \ldots, M$ is the index of the elements in the slope parameter vector. $\boldsymbol{X}_t$ is the vector of $M$ predictor variables observed at the end of month $t$. The elastic net aims to address the problem that highly parameterized forecast combinations of multiple predictors are typically overfitted in sample and thus perform very poorly out of sample. The penalty terms in the optimization aim to produce a regularization that ensures a sparse model.[13] The elastic net involves minimizing the following objective function:

$$\min_{\beta}\left[\sum_{t=0}^{T-1}(ER_{t+1} - \alpha - \beta' X_t)^2 + \lambda\left(0.5\,(1-\delta)\sum_{m=1}^{M}|\beta_m| + \delta\sum_{m=1}^{M}\beta_m^2\right)\right], \tag{6}$$

where $\lambda$ is the regularization parameter for the lasso and ridge penalty terms. Following Rapach and Zhou (2020), we set $\delta = 0.5$. Based on the results of Flynn et al. (2013), we select $\lambda$ using the corrected AIC of Hurvich and Tsai (1989) rather than cross-validation.

*Combination elastic net (C-ENET).* Finally, we use the combination elastic net of Rapach and Zhou (2020). In this approach, we first obtain the single-variable out-of-sample forecasts. In a second step, an elastic net regression (see the previous paragraph) is used to regress realized returns on the out-of-sample forecasts for a so-called holdout out-of-sample period of similar length to the in-sample period:

$$ER_{t+1} = \alpha + \beta'\widehat{ER}_{t+1}^{oos} + \epsilon_{t+1}. \tag{7}$$

Finally, all variables that yield a strictly positive slope coefficient in the elastic net estimation of Eq. (7) are selected

8 Other variables used in previous studies include the consumption–wealth–income ratio, the dividend–payout ratio, the default return spread, the investment-to-capital ratio, and net equity expansion. We do not include these because of the lack of availability of international data.

9 For the unemployment rate of Ecuador, we obtain the time series from Datastream.

10 Most of the variables are ratios. Therefore, they are not currency-dependent. For others, we use USD-denominated indices. For example, to calculate the inflation rate, we use the country's consumer price index (CPI) in USD.

11 Some predictor variables are only available for a subset of the sample length of the respective countries listed in Table 1.

12 Rapach et al. (2010) also consider the median and truncated mean as two other simple combination approaches. They show that the mean forecast combination approach performs better than these alternatives. Therefore, we focus on this approach.

13 We also try a simple multiple predictor regression without the penalty terms of Eq. (6). This approach clearly lags behind the elastic net in terms of out-of-sample performance.

for a simple mean forecast combination as in Eq. (3). Note that the main difference from the mean forecast combination is that the C-ENET tries to preselect in order to focus only on the best forecasts.

## 3. International return predictability

### 3.1. Summary statistics

Before discussing our main findings, it is instructive to look at the summary statistics reported in Table 1. The countries are first classified by their MSCI market development status and, within each group, by their geographic region. There is little difference between the averages for developed, emerging, and frontier markets. The average (median) annualized USD market excess return is 1.8% (3.6%) for developed markets, 2.3% (3.0%) for emerging markets, and 1.8% (1.9%) for frontier markets. However, there is substantial heterogeneity in the average market excess returns across countries. For some countries, the average market excess returns are negative. For developed markets, this is mainly due to high negative USD returns in the early sample period, especially in the period between World War I and World War II. Most of the emerging and frontier markets with negative average market excess returns experienced weak economic performance, often with a weakening of their domestic currencies against the U.S. dollar, for much of the respective sample periods. For the U.S., we find an average (annualized) market excess return (standard deviation) of 6.2% (16.5%). These numbers are similar to those reported by, for example, Goyal and Welch (2008).

Returns do not only display cross-sectional but also time-series variation. In Fig. 1, we show the time series of market excess returns aggregated across developed, emerging, and frontier markets. As can be seen, we have long time series for countries in all three stages of development. Furthermore, high positive and negative average returns in developed and emerging markets tend to coincide with U.S. recessions in many cases. The dynamics appear to differ somewhat for frontier markets, underlining the importance of studying return predictability out of sample.

### 3.2. In-sample analysis

We begin our main analysis by examining the in-sample predictability of market excess returns. To draw inferences, we test the null hypothesis that future excess returns are not predictable using the variable $X_t$. In the case of no predictability, we expect $\beta = 0$. In this case, we would conclude that the best predictor of future market excess returns is a constant, i.e., the recursive mean.[14] On the other hand, if the slope loading is statistically significant, there is evidence of predictability. To assess the strength of predictability across countries, we report the average $R^2$s in our main tables.

We base our statistical inference on a bootstrapped distribution, as suggested by Rapach and Wohar (2006). This approach preserves the serial correlation of the predictor variables and avoids a small-sample bias (Stambaugh, 1999). First, we set up the following null hypothesis: $ER_{t+1} = a_0 + \epsilon_{1,t+1}$ and $X_{t+1} = b_0 + b_1 X_t + \epsilon_{2,t+1}$, where $a_0$, $b_0$ and $b_1$ are the regression coefficients, and $\epsilon_{1,t+1}$ and $\epsilon_{2,t+1}$ are the error terms, respectively. We then estimate the process under the null hypothesis of no predictability via OLS, and bias-adjust the $b_1$ coefficient following Shaman and Stine (1988). Second, we use the series of error terms and set up our pseudo-sample by drawing from the residuals in tandem (with replacement). For the pseudo-sample, we compute both the in-sample and out-of-sample statistics (described in the following section). We repeat this procedure (starting from the second step) 1,000 times. This approach controls for the Stambaugh (1999) bias because the residuals are drawn in tandem, preserving their contemporaneous correlation structure.

Table 2 visualizes the results and Table A2 in the Online Appendix provides detailed regression results for each country. In discussing our results, we focus on the denser presentation of Table 3, where we aggregate the regression results separately for developed, emerging, and frontier markets. We examine each potential predictor separately.[15]

First, we can confirm the finding of Hjalmarsson (2010) that interest rate variables on aggregate perform somewhat better than valuation ratios in predicting international returns. For example, the dividend yield predicts developed market returns with an average in-sample $R^2$ of 0.34%. The $R^2$ is significantly positive for 18% of the developed markets. On the other hand, the short-term interest rate (*RREL*) yields an average in-sample $R^2$ of 0.60%, which is significantly positive for 43% of developed markets. The performance of the price–earnings ratio is similar to that of the dividend yield.

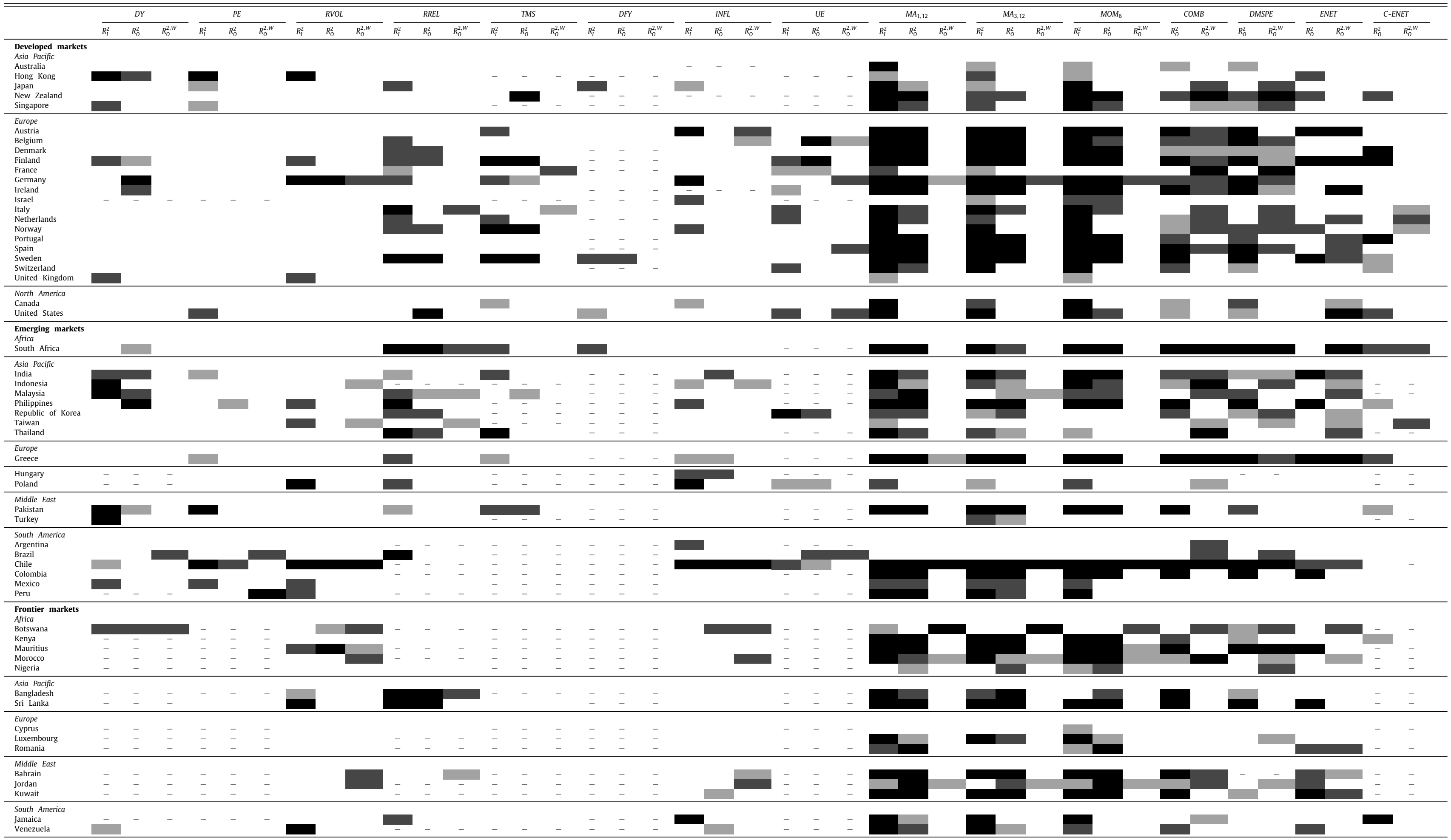
Overall, the performance of all valuation ratios, interest rate variables, and macroeconomic predictors for developed markets is modest. The highest average in-sample $R^2$ among them is 0.60% and the maximum proportion of significant observations is 43% (both for *RREL*). The highest average in-sample $R^2$s in developed markets occur for the technical signals. For example, $MA_{1,12}$ yields an average in-sample $R^2$ of 1.45%, which is statistically significant for almost all developed markets. $MA_{3,12}$ and $MOM_6$ also yield in-sample $R^2$s greater than 1%.

For emerging markets, the in-sample predictability is somewhat stronger. For nearly all fundamental variables, the average in-sample $R^2$s of emerging markets exceed those of developed markets. The average in-sample $R^2$s are higher than 1% for *RREL*, *TMS*, $MA_{1,12}$, $MA_{3,12}$, and

[14] It is not clear that the recursive mean is the best benchmark in an international setting. Therefore, we also consider an $AR(p)$ benchmark model as an alternative. It turns out that the recursive mean is indeed a better predictor of market excess returns than the $AR(p)$ model and thus a more stringent benchmark.

[15] We impose two conditions to be able to draw reliable inferences from our results. First, there must be at least 20 years of observations to include a variable for the in-sample analysis. Second, there must be at least 30 out-of-sample observations to consider the out-of-sample performance of a variable. Note that tables reporting aggregated results show equally weighted averages.

**Table 2**
Return predictability heatmap.




This table presents a heatmap summarizing information about the in-sample and out-of-sample $R^2$s of all predictor variables and model selection approaches. Countries are assigned to the different panels according to their MSCI market development status and geographical region. We sample the data at a monthly frequency and predict the future one-month USD excess returns. We present the results for the in-sample $R^2$s ($R_I^2$), the out-of-sample $R^2$s ($R_O^2$), and the out-of-sample $R^2$s from WLS forecasts ($R_O^{2,W}$). The out-of-sample results are based on 240-month rolling windows. ■, ■, and ■ denote statistical significance at the 10%, 5%, and 1% levels, respectively. White space indicates that a variable does not yield a statistically significant $R^2$, and "–" means that there are not enough data available. For all individual variables, statistical significance is determined relative to a bootstrapped distribution, while for the model selection approaches, we use the MSPE-adjusted test statistic of Clark and West (2007). Definitions of the variables can be found in Section 2.

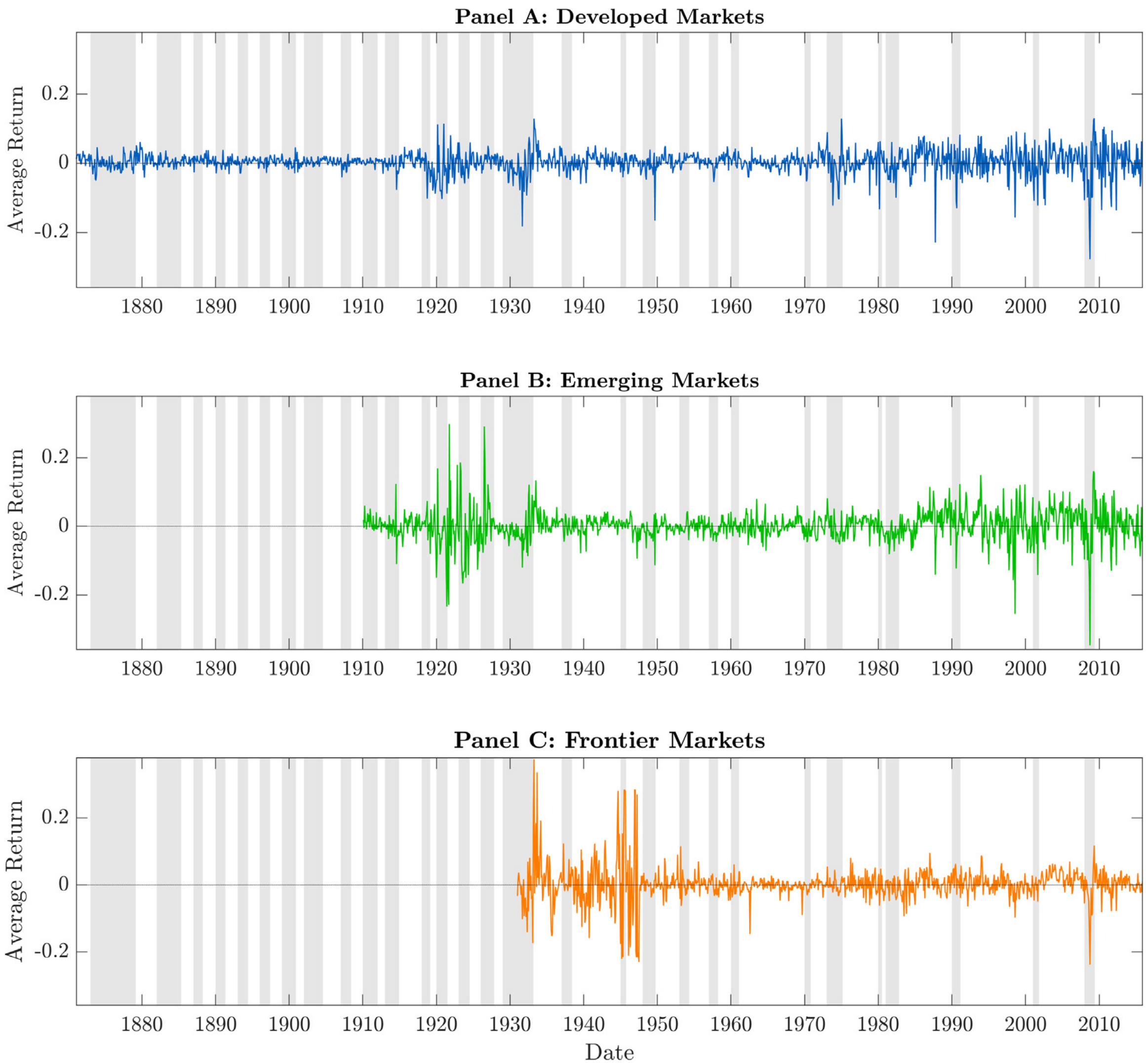



**Fig. 1. Return time series.**
This figure shows the time series of the average market excess returns for developed (Panel A; blue line), emerging (Panel B; green line), and frontier (Panel C; orange line) markets. Among all countries in a category for which we observe an excess return, we calculate the equally weighted average. The shaded areas indicate the periods identified by the National Bureau of Economic Research (NBER) as business-cycle contractions in the U.S. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$MOM_6$. Thus, interest rate variables also outperform valuation ratios in emerging markets, and technical indicators also perform well.

For frontier markets, in-sample predictability is overall stronger than for emerging markets. The average in-sample $R^2$ exceeds 1% for *DY*, *RREL*, $MA_{1,12}$, $MA_{3,12}$, and $MOM_6$. For $MA_{1,12}$, it is even 2.5% on average.

There are strong theoretical and empirical priors on the signs of the slope associated with most predictors. For all variables except *PE*, *RREL*, and *INFL*, the signs should be positive. For ease of exposition, we multiply these three predictors by $-1$ to have uniformly positive expected slope coefficients. In an untabulated analysis, we find that for developed markets, 75% of the significant $R^2$s are associated with slope coefficients of the "correct" sign for the dividend yield. For emerging and frontier markets, all significant slope estimates have the correct sign. For *PE*, *RREL*, *TMS*, *UE*, and the technical indicators, almost all significant $R^2$s are associated with slope coefficients whose signs are consistent with theory. For *INFL*, the slope coefficients have the correct sign for most markets. On the other hand, for *RVOL* and *DFY* the signs are quite mixed. For all market types, the technical indicators have the largest average $t$-statistics, indicating that they provide the strongest in-sample predictability.

Overall, we find substantial and consistent in-sample predictability across markets. Two facts point to inefficiency as an important source of return predictability: (i) it appears that the strength of the in-sample predictability is negatively related to the level of market development, and (ii) technical signals are the best in-sample predictors on average for developed, emerging, and frontier markets. The next section focuses on an out-of-sample analysis.

**Table 3**
Return predictability – Summary results.

| | Developed markets | | | | | | Emerging markets | | | | | | Frontier markets | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $R^2_{IS}$ | $(Shr)$ | $R^2_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ | $R^2_{IS}$ | $(Shr)$ | $R^2_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ | $R^2_{IS}$ | $(Shr)$ | $R^2_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ |
| **Fundamental predictors** | | | | | | | | | | | | | | | | | | |
| *DY* | 0.335 | (0.18) | −0.953 | (0.18) | 0.139 | (0.00) | 0.897 | (0.44) | −0.790 | (0.31) | 0.324 | (0.06) | 1.385 | (1.00) | 1.658 | (0.50) | 1.383 | (0.50) |
| *PE* | 0.246 | (0.18) | −0.901 | (0.00) | 0.109 | (0.00) | 0.567 | (0.26) | −0.499 | (0.11) | 1.223 | (0.11) | 0.230 | (0.00) | −0.726 | (0.00) | −0.290 | (0.00) |
| *RVOL* | 0.496 | (0.17) | −1.003 | (0.04) | 0.211 | (0.04) | 0.435 | (0.32) | −0.988 | (0.05) | 0.654 | (0.16) | 0.587 | (0.27) | −0.931 | (0.13) | 1.161 | (0.27) |
| *RREL* | 0.596 | (0.43) | −0.042 | (0.22) | 0.118 | (0.04) | 1.028 | (0.71) | −0.464 | (0.29) | −0.115 | (0.21) | 1.105 | (0.43) | 0.531 | (0.29) | 0.762 | (0.29) |
| *TMS* | 0.541 | (0.35) | −0.035 | (0.25) | 0.226 | (0.10) | 1.018 | (0.83) | 0.719 | (0.33) | −0.142 | (0.00) | 0.221 | (0.00) | −2.667 | (0.00) | 0.064 | (0.00) |
| *DFY* | 0.234 | (0.27) | −1.065 | (0.09) | 0.034 | (0.00) | 0.818 | (1.00) | −0.533 | (0.00) | 0.643 | (0.00) | | | | | | |
| *INFL* | 0.350 | (0.30) | −12.87 | (0.00) | −8.356 | (0.10) | 0.598 | (0.37) | −1.494 | (0.21) | 0.672 | (0.11) | 0.283 | (0.07) | −0.133 | (0.20) | 1.196 | (0.27) |
| *UE* | 0.258 | (0.37) | −1.012 | (0.16) | 0.285 | (0.21) | 0.797 | (0.50) | 0.529 | (0.67) | 0.627 | (0.17) | 0.027 | (0.00) | −3.607 | (0.00) | −0.199 | (0.00) |
| **Technical signals** | | | | | | | | | | | | | | | | | | |
| $MA_{1,12}$ | 1.448 | (0.96) | 0.655 | (0.65) | 0.219 | (0.04) | 1.468 | (0.79) | 0.594 | (0.68) | 0.659 | (0.11) | 2.487 | (0.87) | 2.375 | (0.87) | 1.101 | (0.20) |
| $MA_{3,12}$ | 1.039 | (0.96) | 0.294 | (0.48) | 0.217 | (0.04) | 1.010 | (0.74) | 0.262 | (0.74) | 0.644 | (0.11) | 1.830 | (0.67) | 2.060 | (0.73) | 1.107 | (0.20) |
| $MOM_6$ | 1.228 | (0.96) | 0.409 | (0.61) | 0.224 | (0.04) | 1.217 | (0.68) | 0.402 | (0.47) | 0.598 | (0.05) | 2.264 | (0.87) | 3.038 | (0.73) | 1.165 | (0.27) |
| **Forecast combinations** | | | | | | | | | | | | | | | | | | |
| *COMB* | | | 0.912 | (0.70) | 0.303 | (0.61) | | | 0.722 | (0.47) | 0.654 | (0.58) | | | 2.667 | (0.60) | 1.156 | (0.33) |
| *DMSPE* | | | 1.053 | (0.70) | 0.230 | (0.57) | | | 0.761 | (0.50) | 0.793 | (0.44) | | | 4.727 | (0.43) | 3.999 | (0.43) |
| *ENET* | | | −4.965 | (0.26) | −6.420 | (0.39) | | | −2.612 | (0.26) | 0.352 | (0.47) | | | 1.953 | (0.47) | 0.652 | (0.33) |
| *C-ENET* | | | −0.619 | (0.30) | −7.455 | (0.13) | | | −0.334 | (0.31) | 0.087 | (0.17) | | | 0.230 | (0.50) | −0.114 | (0.00) |

This table summarizes the in-sample and out-of-sample return predictability of developed, emerging, and frontier markets. We sample the data at a monthly frequency and predict the future one-month USD excess returns. $R^2_{IS}$, $R^2_{OOS}$, and $R^{2,WLS}_{OOS}$ denote the average in-sample $R^2$, out-of-sample $R^2$, and out-of-sample $R^2$s from WLS forecasts, respectively. All $R^2$s are quoted in percentage points. The out-of-sample results are based on 240-month rolling windows. In parentheses, we report the share of countries for which the respective $R^2$s are significantly positive at the 10% level. For all individual variables, statistical significance is determined relative to a bootstrapped distribution, while for the model selection approaches, we use the MSPE-adjusted test statistic of Clark and West (2007). Definitions of the variables can be found in Section 2.

### 3.3. Out-of-sample analysis

We continue the analysis by examining the out-of-sample return predictability. Following Goyal and Welch (2008), we use an initial training window of 20 years.[16] To obtain the first parameter estimates, we use only the information available in the estimation window to estimate the forecasting model presented in Eq. (2). Equipped with these parameter estimates and using the most recent observation of the predictor variable, we generate the first excess return forecast. We then re-estimate the forecasting model by moving the training window forward by one month. Thus, with the new parameter estimates, we again forecast the market excess return for the next month. We base our out-of-sample analysis on a 20-year rolling window to capture potential time variation in the coefficients of the predictive regression.

Following Johnson (2018), we also consider out-of-sample forecasts estimated with a weighted least squares (WLS) regression, where the weights are the inverse of the conditional volatility. Johnson (2018) shows that this approach yields better return forecasts than the standard OLS approach for the U.S. For the WLS approach, both the left- and right-hand-side variables are first divided by an estimate of the volatility of index excess returns. For each month, we use the monthly absolute index excess return as the volatility measure. Using the scaled variables, we run the regression of Eq. (2) and proceed analogously as described in the previous paragraph.

We use the out-of-sample $R^2$ ($R^2_{oos}$) to evaluate the performance of different models:

$$R^2_{oos} = 1 - \frac{MSE_u}{MSE_r}, \tag{8}$$

where $MSE_u$ and $MSE_r$ are the mean squared error estimates of the unrestricted and restricted models, respectively. The unrestricted model is based on Eq. (2) or one of our model selection approaches. In the case of the restricted model, we impose the null hypothesis that excess returns are unpredictable, i.e., $\beta = 0$. Thus, based on the $R^2_{oos}$, we can answer the question: What predictive power above the historical mean can be achieved by using the variable $X_t$? A variable has noteworthy predictive power if it has a positive and significant $R^2_{oos}$, indicating a significant improvement over the historical mean.

To assess whether the predictability is significantly stronger than that of the historical mean, we compute the $MSE-F$ statistic suggested by McCracken (2007):

$$MSE-F = (N-k+1) \times \left(\frac{MSE_r - MSE_u}{MSE_u}\right), \tag{9}$$

where $N$ is the number of out-of-sample predictions, and $k$ is the forecast horizon (in months). All other variables have the same definitions as above. The null hypothesis is that the restricted model performs as well as or better than the unrestricted model, i.e., $MSE_r \leq MSE_u$. The alternative is that the unrestricted model produces smaller forecast errors than the restricted model. We determine statistical significance using the bootstrap procedure described in the previous section.[17]

It is worth pointing out that out-of-sample tests are somewhat less powerful than in-sample tests of return predictability (Inoue and Kilian, 2005; Cochrane, 2008). In out-of-sample tests, the sample used to estimate the parameters is only a subset of that used for in-sample estimation. A larger sample naturally improves the precision of the estimates and increases the power of the statistical tests. As a result, we may detect a somewhat lower degree of out-of-sample predictability. On the other hand, the out-of-sample tests can account for time variation in the predictive relationship, which may negatively affect the in-sample predictability.

[16] As discussed in Section 6, using an initial training window of 10 years, as suggested by Rapach and Wohar (2006), the results are qualitatively similar.

[17] We reflect any specification of the forecasting technique in the bootstrap design. That is, for example, when assessing the WLS out-of-sample $R^2$s, we use WLS, rather than OLS, to estimate the process under the null hypothesis.

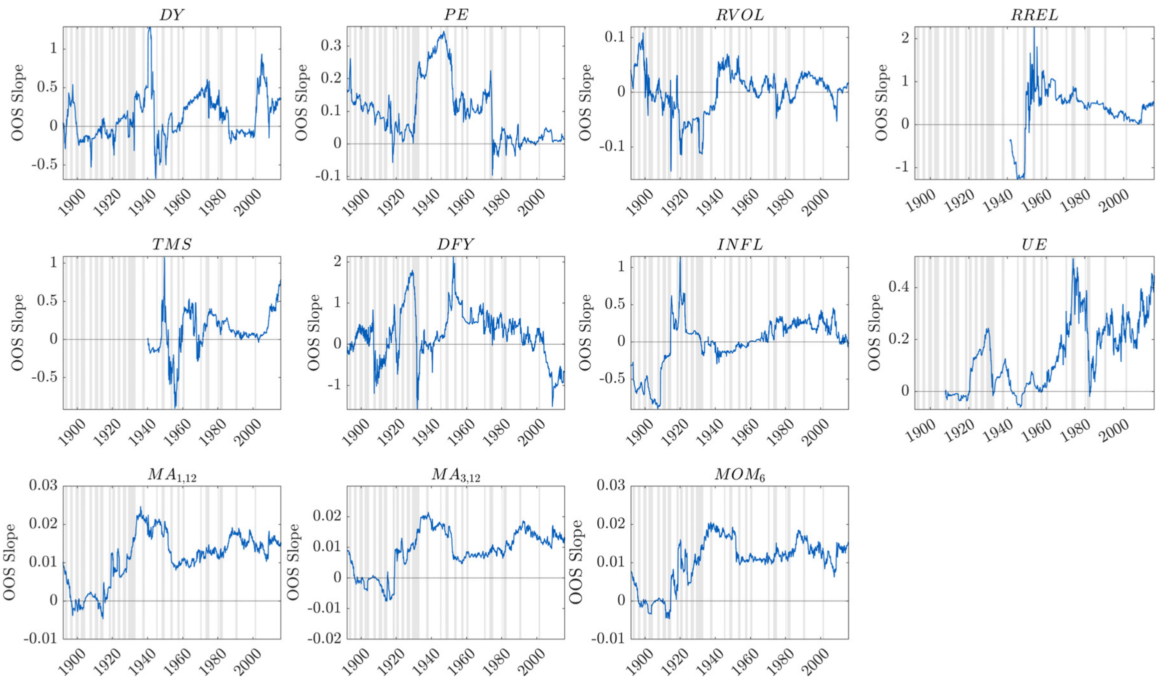



**Fig. 2. Predictive slope coefficient time series.**
This figure shows the time series of the median slope coefficients for the different predictor variables. First, we sign each predictor variable so that the sign of the slope coefficient is positive based on theory. Then, we run predictive regressions using a 240-month rolling window to compute out-of-sample slope coefficients for each country. Finally, we report the median slope coefficients across all countries for which they are available, obtained at the end of the 240-month rolling windows. The shaded areas indicate the periods identified by the NBER as business-cycle contractions in the U.S.

Similar to the in-sample analysis, Table 2 visualizes the results and Table A2 in the Online Appendix provides detailed regression results. We focus the discussion on the aggregated results in Table 3. For developed markets, we find that all average out-of-sample $R^2$s are negative, except for those of the technical signals, where the proportion of countries with significantly positive out-of-sample $R^2$s varies between 48% and 65%. For all other predictors, the proportion of countries with significantly positive out-of-sample $R^2$s is at most 25%. Thus, when using the standard methods, the out-of-sample predictability in developed markets seems to be moderate at best.

Using WLS instead of OLS turns the average out-of-sample $R^2$s positive for nearly all predictors. This extends the results of Johnson (2018) to international data. However, the WLS out-of-sample $R^2$s are only significantly positive for up to 21% of developed markets. Thus, overall, the single-variable out-of-sample predictability is not very pronounced in developed markets.

For emerging markets, the out-of-sample predictability is somewhat stronger than for developed markets. The out-of-sample $R^2$s are generally larger on average and more often significantly positive. For frontier markets, the out-of-sample predictability is even stronger than for developed and emerging markets. In particular, the technical signals perform well. They yield significantly positive out-of-sample $R^2$s for the majority of the countries.

We also examine the time variability of the predictive relationships. To do this, Fig. 2 plots the median slope coefficients at the end of each 240-month window of each predictor variable over time. We find that there is ample time variation in the slope estimates, which further strengthens our interest in the out-of-sample analysis. We also find that the median slope coefficients change sign over time for all predictors. The technical signals produce the most stable predictive relationships.

Finally, we turn to the model selection approaches that combine the information contained in different predictor variables. Given the rather unstable predictive relationships, these approaches are likely to perform better than the individual predictor variables. For convenience for the model selection approaches, we base the analysis of statistical significance on the MSPE-adjusted test statistic of Clark and West (2007).

We find that the simple mean forecast combination approach (*COMB*) works very well for out-of-sample forecasting. It yields positive average out-of-sample $R^2$s for developed, emerging, and frontier markets. For developed markets, the average OLS out-of-sample $R^2$ is 0.91%, and it is statistically significant in 70% of the countries. This average out-of-sample $R^2$ is larger than that of any single predictor variable. The average out-of-sample WLS $R^2$ is 0.30% and is significantly positive in 61% of the countries. The *DMSPE* approach performs slightly better, as evidenced by the slightly higher average out-of-sample $R^2$s. In comparison, *ENET* and *C-ENET* perform less well.

### 3.4. Economic utility gains

While the previous analysis is statistical in nature, it is very important for investors to know if, and how, predictability can be translated into economic gains from portfolio allocation strategies. However, the relationship between out-of-sample $R^2$s and economic utility gains from such a portfolio allocation strategy has proven to be non-trivial and complex (Rapach & Zhou, 2013). Therefore, in this section, we also investigate whether it is possible to obtain economic utility gains.

We assume that an investor either has mean–variance preferences or that mean–variance preferences provide a reasonable second-order Taylor approximation to the investor's true utility function (Fleming et al., 2001). The investor decides to allocate a fraction $\omega_t$ of her wealth to the risky market portfolio and the remainder, i.e., $1-\omega_t$, to the risk-free asset. Her objective function is

$$\max_{\omega_t} E_t\left(r_{p,t+1} - \frac{\gamma}{2}\sigma^2_{p,t+1}\right), \tag{10}$$

where $E_t(\cdot)$ is the conditional expectation operator, $\gamma$ is the coefficient of relative risk aversion, and $\sigma^2_{p,t+1}$ is the conditional variance of the portfolio from $t$ to $t+1$. $r_{p,t+1}$ is the simple return of the investor's portfolio between $t$ and $t+1$. Since our previous analysis is based on log rather than simple returns, for this analysis, we estimate the expected simple excess returns ($er_{t+1}$) based on all predictor variables. For the conditional return variances, we use the current estimate of *RVOL*.

Optimizing Eq. (10), one can obtain the optimal weight invested in the risky asset as

$$\omega_t = \frac{E_t(er_{t+1})}{\gamma E_t(\sigma^2_{t+1})}. \tag{11}$$

Thus, the optimal weight depends positively on the expected future excess return, while it is reduced for higher realized variance and higher levels of relative risk aversion.

For each month in our out-of-sample analysis, we compute the weight $\omega_t$ and the realized return over the next month of the portfolio. To avoid short selling and excessive leverage, we follow Campbell and Thompson (2008) and impose the constraint that $\omega_t$ must be between 0 and 1.5. The certainty equivalent return (*CER*) is

$$CER = \bar{r}_p - \frac{\gamma}{2}\sigma^2_p, \tag{12}$$

where $\bar{r}_p$ is the average simple return of the portfolio, and $\sigma^2_p$ is the variance of the portfolio returns. The utility gain ($\Delta CER$) of using a predictor is the difference between the *CER* of a strategy using that predictor and the *CER* using a strategy based on the historical mean benchmark return.

Table 4 shows the results for different $\gamma$ coefficients. We find sizable economic utility gains, especially for the technical signals and the (elastic net) model selection approaches. For $\gamma = 3$, the highest average utility gains are 0.64 percentage points per year for $MA_{1,12}$ in developed markets, 0.47 percentage points per year for $MA_{1,12}$ in emerging markets, and 1.86 percentage points per year for *UE* in frontier markets. Overall, frontier markets offer the highest utility gains.

Interestingly, the economic utility gains for the OLS forecasts clearly exceed those for the WLS forecasts. This may be because the WLS forecasts are typically more muted, with coefficient estimates that are empirically relatively small (untabulated). This is good when trying to minimize an MSE. However, it may not be economically optimal. The minimum and maximum weight restrictions described above reduce the impact of extreme forecasts, which are more common in the OLS approach. Untabulated results show that without this restriction, the OLS utility gains for most predictors are substantially smaller on average. Similarly, we find that the utility gains from using *COMB* are mostly positive, but smaller on average than those for many individual predictors. This is likely also because the averaging produces more muted forecasts that do not exhibit high statistical errors, but at the cost of reducing the economic value of the forecast.

Table A3 in the Online Appendix reports the utility gains when transaction costs are taken into account. We follow Balduzzi and Lynch (1999) and assume costs of 50 basis points per transaction, proportional to the traded asset size $|\omega_{t+1} - \omega_{t^+}|$, where $\omega_{t^+}$ is the portfolio weight before rebalancing at $t+1$. Note that there are transaction costs for both strategies: the one based on a predictor variable, and the one based on the historical mean.

## 4. Reducing heterogeneity

Due to data availability, the length of the time series is naturally heterogeneous across countries. In particular, developed markets tend to have much longer time series (some starting in 1871) than frontier markets (where the earliest series starts in 1937). Thus, some of the observed differences in the predictability of country returns may be due to differences in the time periods examined. Therefore, in this section we focus only on the post-1990 period. Examining a shorter time period significantly reduces the heterogeneity in our dataset in terms of data quality and time-series length. Thus, the main purpose of this analysis is to test the robustness of our main results to imbalances in the dataset.

We believe that 1990 is a reasonable cutoff point because it broadly coincides with the structural changes that occurred in many countries at that time, such as the collapse of the Soviet Union. In the post-1990 period, many countries are characterized by a market economy as opposed to a planned economy, and by deregulation, that has strengthened capital markets. A prominent example is the "Big Bang" deregulation promoted by Margaret Thatcher in the U.K. in 1986, but similar legislation has been introduced in many other countries. Consequently, there may be a structural break around these dates, implying changes in the predictability of aggregate excess returns. Timmermann and Granger (2004) and Chordia et al. (2008) argue that markets are currently behaving efficiently. Pesaran and Timmermann (2002) and Lettau and Van Nieuwerburgh (2008) argue that predictability is more difficult to uncover since the 1990s, due to parameter instability and other structural breaks. Thus, it is also possible that we do not observe any return predictability for this most recent subperiod.

**Table 4**
Economic utility gains.

| | Developed markets | | | | Emerging markets | | | | Frontier markets | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\Delta CER$ | $(Shr)$ | $\Delta CER^{WLS}$ | $(Shr)$ | $\Delta CER$ | $(Shr)$ | $\Delta CER^{WLS}$ | $(Shr)$ | $\Delta CER$ | $(Shr)$ | $\Delta CER^{WLS}$ | $(Shr)$ |
| **Relative risk aversion: $\gamma = 1.5$** | | | | | | | | | | | | |
| *Fundamental predictors* | | | | | | | | | | | | |
| *DY* | 0.135 | (0.59) | −0.863 | (0.14) | 0.172 | (0.50) | 0.167 | (0.56) | 1.100 | (1.00) | −0.503 | (0.50) |
| *PE* | −0.014 | (0.55) | −0.707 | (0.14) | 0.332 | (0.74) | 0.669 | (0.63) | −0.262 | (0.00) | −0.598 | (0.50) |
| *RVOL* | 0.025 | (0.57) | −0.726 | (0.17) | 0.121 | (0.53) | −0.114 | (0.42) | 0.421 | (0.53) | 0.457 | (0.47) |
| *RREL* | 0.127 | (0.70) | −0.674 | (0.17) | 0.180 | (0.43) | −0.319 | (0.43) | 1.349 | (0.43) | 0.103 | (0.43) |
| *TMS* | 0.145 | (0.60) | −0.678 | (0.10) | 0.566 | (0.50) | −0.544 | (0.50) | −0.282 | (0.50) | −0.409 | (0.50) |
| *DFY* | −0.046 | (0.36) | −0.951 | (0.00) | −0.252 | (0.00) | 0.180 | (1.00) | | | | |
| *INFL* | 0.106 | (0.60) | −0.705 | (0.20) | 0.051 | (0.53) | −0.132 | (0.37) | −0.122 | (0.33) | 0.585 | (0.53) |
| *UE* | 0.231 | (0.74) | −0.685 | (0.05) | 0.368 | (0.67) | 0.102 | (0.50) | 3.716 | (1.00) | −1.113 | (0.00) |
| *Technical signals* | | | | | | | | | | | | |
| $MA_{1,12}$ | 1.222 | (1.00) | −0.750 | (0.13) | 0.920 | (0.84) | −0.118 | (0.47) | 2.180 | (0.87) | 0.443 | (0.47) |
| $MA_{3,12}$ | 0.884 | (1.00) | −0.749 | (0.13) | 0.640 | (0.79) | −0.120 | (0.42) | 1.424 | (0.80) | 0.440 | (0.47) |
| $MOM_6$ | 1.001 | (0.96) | −0.746 | (0.13) | 0.631 | (0.79) | −0.123 | (0.42) | 2.269 | (0.87) | 0.453 | (0.47) |
| *Forecast combinations* | | | | | | | | | | | | |
| *COMB* | 0.086 | (0.87) | −0.095 | (0.13) | 0.112 | (0.84) | −0.092 | (0.16) | 0.208 | (0.93) | −0.030 | (0.40) |
| *DMSPE* | 0.087 | (0.96) | −0.095 | (0.13) | 0.117 | (0.89) | −0.112 | (0.22) | 0.215 | (0.86) | 0.177 | (0.50) |
| *ENET* | 0.913 | (0.83) | −0.138 | (0.39) | 0.832 | (0.79) | 0.094 | (0.47) | 1.759 | (0.87) | 0.047 | (0.53) |
| *C-ENET* | 0.599 | (0.78) | −0.238 | (0.09) | 0.534 | (0.77) | −0.368 | (0.15) | 0.562 | (0.50) | −0.774 | (0.00) |
| **Relative risk aversion: $\gamma = 3$** | | | | | | | | | | | | |
| *Fundamental predictors* | | | | | | | | | | | | |
| *DY* | 0.077 | (0.59) | −0.446 | (0.14) | 0.122 | (0.50) | 0.097 | (0.56) | 1.108 | (1.00) | −0.091 | (0.50) |
| *PE* | −0.012 | (0.50) | −0.357 | (0.14) | 0.176 | (0.74) | 0.344 | (0.63) | −0.131 | (0.00) | −0.299 | (0.50) |
| *RVOL* | 0.001 | (0.57) | −0.351 | (0.22) | 0.063 | (0.58) | −0.026 | (0.42) | 0.200 | (0.53) | 0.260 | (0.47) |
| *RREL* | 0.058 | (0.65) | −0.339 | (0.17) | 0.086 | (0.43) | −0.123 | (0.43) | 0.677 | (0.43) | 0.062 | (0.43) |
| *TMS* | 0.070 | (0.60) | −0.338 | (0.10) | 0.273 | (0.50) | −0.272 | (0.50) | −0.141 | (0.50) | −0.204 | (0.50) |
| *DFY* | −0.036 | (0.27) | −0.480 | (0.00) | −0.126 | (0.00) | 0.090 | (1.00) | | | | |
| *INFL* | 0.055 | (0.65) | −0.325 | (0.25) | 0.014 | (0.53) | −0.038 | (0.37) | −0.063 | (0.33) | 0.324 | (0.53) |
| *UE* | 0.128 | (0.74) | −0.341 | (0.05) | 0.184 | (0.67) | 0.051 | (0.50) | 1.858 | (1.00) | −0.556 | (0.00) |
| *Technical signals* | | | | | | | | | | | | |
| $MA_{1,12}$ | 0.643 | (0.96) | −0.363 | (0.17) | 0.471 | (0.79) | −0.027 | (0.47) | 1.119 | (0.87) | 0.252 | (0.47) |
| $MA_{3,12}$ | 0.465 | (1.00) | −0.363 | (0.17) | 0.336 | (0.74) | −0.028 | (0.42) | 0.738 | (0.80) | 0.251 | (0.47) |
| $MOM_6$ | 0.522 | (0.96) | −0.362 | (0.17) | 0.315 | (0.79) | −0.030 | (0.42) | 1.154 | (0.87) | 0.257 | (0.47) |
| *Forecast combinations* | | | | | | | | | | | | |
| *COMB* | 0.043 | (0.87) | −0.047 | (0.13) | 0.056 | (0.84) | −0.046 | (0.16) | 0.104 | (0.93) | −0.015 | (0.40) |
| *DMSPE* | 0.044 | (0.96) | −0.048 | (0.13) | 0.059 | (0.89) | −0.056 | (0.22) | 0.108 | (0.86) | 0.088 | (0.50) |
| *ENET* | 0.466 | (0.78) | −0.081 | (0.35) | 0.425 | (0.79) | 0.059 | (0.53) | 0.969 | (0.87) | 0.028 | (0.53) |
| *C-ENET* | 0.320 | (0.83) | −0.095 | (0.09) | 0.280 | (0.77) | −0.184 | (0.15) | 0.229 | (0.50) | −0.387 | (0.00) |
| **Relative risk aversion: $\gamma = 5$** | | | | | | | | | | | | |
| *Fundamental predictors* | | | | | | | | | | | | |
| *DY* | 0.051 | (0.59) | −0.271 | (0.14) | 0.073 | (0.50) | 0.058 | (0.56) | 0.665 | (1.00) | −0.054 | (0.50) |
| *PE* | −0.007 | (0.50) | −0.214 | (0.14) | 0.106 | (0.74) | 0.207 | (0.63) | −0.079 | (0.00) | −0.179 | (0.50) |
| *RVOL* | 0.001 | (0.57) | −0.183 | (0.26) | 0.018 | (0.58) | −0.009 | (0.42) | 0.118 | (0.53) | 0.156 | (0.47) |
| *RREL* | 0.035 | (0.65) | −0.204 | (0.17) | 0.052 | (0.43) | −0.065 | (0.43) | 0.401 | (0.43) | 0.041 | (0.43) |
| *TMS* | 0.041 | (0.60) | −0.202 | (0.10) | 0.161 | (0.50) | −0.163 | (0.50) | −0.084 | (0.50) | −0.123 | (0.50) |
| *DFY* | −0.027 | (0.27) | −0.288 | (0.00) | −0.076 | (0.00) | 0.054 | (1.00) | | | | |
| *INFL* | 0.033 | (0.65) | −0.159 | (0.30) | 0.009 | (0.53) | −0.016 | (0.37) | −0.040 | (0.33) | 0.194 | (0.53) |
| *UE* | 0.082 | (0.74) | −0.201 | (0.11) | 0.110 | (0.67) | 0.031 | (0.50) | 1.115 | (1.00) | −0.334 | (0.00) |
| *Technical signals* | | | | | | | | | | | | |
| $MA_{1,12}$ | 0.390 | (0.96) | −0.190 | (0.22) | 0.289 | (0.79) | −0.010 | (0.47) | 0.657 | (0.80) | 0.151 | (0.47) |
| $MA_{3,12}$ | 0.282 | (1.00) | −0.190 | (0.22) | 0.203 | (0.74) | −0.011 | (0.42) | 0.434 | (0.80) | 0.150 | (0.47) |
| $MOM_6$ | 0.316 | (0.96) | −0.189 | (0.22) | 0.189 | (0.79) | −0.012 | (0.42) | 0.684 | (0.87) | 0.154 | (0.47) |
| *Forecast combinations* | | | | | | | | | | | | |
| *COMB* | 0.026 | (0.87) | −0.028 | (0.13) | 0.034 | (0.84) | −0.028 | (0.16) | 0.063 | (0.93) | −0.009 | (0.40) |
| *DMSPE* | 0.026 | (0.96) | −0.029 | (0.13) | 0.035 | (0.89) | −0.034 | (0.22) | 0.065 | (0.86) | 0.053 | (0.50) |
| *ENET* | 0.285 | (0.74) | −0.051 | (0.35) | 0.255 | (0.79) | 0.035 | (0.53) | 0.576 | (0.87) | 0.017 | (0.53) |
| *C-ENET* | 0.203 | (0.83) | −0.024 | (0.09) | 0.168 | (0.77) | −0.110 | (0.15) | 0.137 | (0.50) | −0.232 | (0.00) |

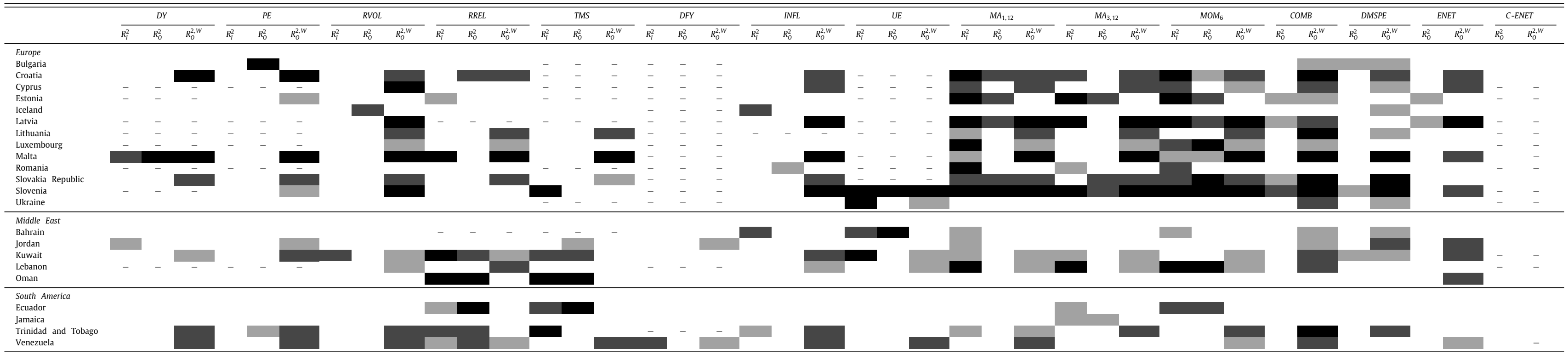
This table summarizes the economic utility gains associated with the out-of-sample return predictability of developed, emerging, and frontier markets. We sample the data at a monthly frequency and predict the future one-month USD excess returns. $\Delta CER$ and $\Delta CER^{WLS}$ denote the difference in average certainty equivalents (in annualized percentage points) of the OLS and WLS out-of-sample return forecasts, respectively, relative to the simple historical mean forecasts. The out-of-sample forecasts are based on 240-month rolling windows. In parentheses, we report the share of countries for which the corresponding $\Delta CER$s are greater than zero. Definitions of the variables can be found in Section 2.

Table 5 visualizes the detailed results, while Table 6 reports the aggregated results for in-sample and out-of-sample return predictability. In order to have enough observations in the out-of-sample period, we use a 120-month rolling window for this analysis. Overall, the results for the post-1990 period are similar to those for

**Table 5**
Return predictability heatmap (post-1990).


(continued on next page)

**Table 5** (*continued*).




This table presents a heatmap summarizing information about the in-sample and out-of-sample $R^2$s of all predictor variables and model selection approaches for the post-1990 period. Countries are assigned to the different panels according to their MSCI market development status and geographical region. We sample the data at a monthly frequency and predict the future one-month USD excess returns. We present the results for the in-sample $R^2$s ($R_I^2$), the out-of-sample $R^2$s ($R_O^2$), and the out-of-sample $R^2$s from WLS forecasts ($R_O^{2,W}$). The out-of-sample results are based on 120-month rolling windows. , , and denote statistical significance at the 10%, 5%, and 1% levels, respectively. White space indicates that a variable does not yield a statistically significant $R^2$, and "–" means that there are not enough data available. For all individual variables, statistical significance is determined relative to a bootstrapped distribution, while for the model selection approaches, we use the MSPE-adjusted test statistic of Clark and West (2007). Definitions of the variables can be found in Section 2.

**Table 6**
Return predictability – Summary results (post-1990).

| | Developed markets | | | | | | Emerging markets | | | | | | Frontier markets | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $R^2_{IS}$ | $(Shr)$ | $R^2_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ | $R^2_{IS}$ | $(Shr)$ | $R^2_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ | $R^2_{IS}$ | $(Shr)$ | $R^2_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ |
| **Fundamental predictors** | | | | | | | | | | | | | | | | | | |
| *DY* | 0.485 | (0.13) | −4.474 | (0.04) | 0.945 | (0.22) | 0.866 | (0.32) | −0.825 | (0.27) | 2.049 | (0.50) | 0.802 | (0.18) | −1.998 | (0.09) | 0.621 | (0.18) |
| *PE* | 0.180 | (0.00) | −2.555 | (0.00) | 0.862 | (0.22) | 0.795 | (0.29) | −2.339 | (0.25) | 1.911 | (0.50) | 1.383 | (0.24) | −2.322 | (0.18) | 0.892 | (0.18) |
| *RVOL* | 0.155 | (0.00) | −2.856 | (0.04) | 0.907 | (0.22) | 0.277 | (0.08) | −1.313 | (0.12) | 1.760 | (0.48) | 1.102 | (0.30) | −4.102 | (0.15) | 0.547 | (0.45) |
| *RREL* | 0.649 | (0.27) | −0.488 | (0.36) | 0.892 | (0.23) | 1.174 | (0.41) | −0.734 | (0.36) | 1.391 | (0.23) | 1.983 | (0.37) | −2.808 | (0.30) | 0.822 | (0.52) |
| *TMS* | 1.078 | (0.38) | 0.209 | (0.43) | 0.940 | (0.14) | 1.434 | (0.57) | −4.142 | (0.14) | 2.137 | (0.36) | 1.233 | (0.25) | −3.357 | (0.17) | 0.729 | (0.25) |
| *DFY* | 0.651 | (0.17) | −2.170 | (0.00) | 0.807 | (0.17) | 1.132 | (0.33) | −1.358 | (0.33) | 2.230 | (0.33) | | | | | | |
| *INFL* | 0.360 | (0.10) | −1.457 | (0.05) | 1.045 | (0.19) | 0.763 | (0.32) | −3.509 | (0.04) | 1.624 | (0.52) | 0.940 | (0.16) | −2.761 | (0.13) | 0.575 | (0.39) |
| *UE* | 1.141 | (0.38) | −1.969 | (0.33) | 0.515 | (0.10) | 1.681 | (0.43) | 0.462 | (0.36) | 2.595 | (0.43) | 1.505 | (0.46) | −3.926 | (0.23) | 1.569 | (0.38) |
| **Technical signals** | | | | | | | | | | | | | | | | | | |
| $MA_{1,12}$ | 0.782 | (0.35) | −0.815 | (0.17) | 0.907 | (0.22) | 1.332 | (0.44) | −0.687 | (0.36) | 1.731 | (0.52) | 3.581 | (0.82) | 1.119 | (0.58) | 0.824 | (0.45) |
| $MA_{3,12}$ | 0.599 | (0.26) | −0.638 | (0.26) | 0.883 | (0.13) | 1.016 | (0.40) | −1.025 | (0.28) | 1.686 | (0.40) | 2.678 | (0.64) | −0.441 | (0.48) | 0.765 | (0.52) |
| $MOM_6$ | 0.696 | (0.22) | −0.837 | (0.22) | 0.917 | (0.22) | 1.196 | (0.40) | −0.125 | (0.40) | 1.709 | (0.40) | 3.404 | (0.70) | 1.732 | (0.58) | 0.815 | (0.48) |
| **Forecast combinations** | | | | | | | | | | | | | | | | | | |
| *COMB* | | | 0.104 | (0.13) | 0.982 | (0.57) | | | 0.652 | (0.32) | 1.763 | (0.80) | | | 2.026 | (0.58) | 0.765 | (0.67) |
| *DMSPE* | | | 0.099 | (0.04) | 0.733 | (0.26) | | | 0.407 | (0.28) | 1.750 | (0.68) | | | 2.304 | (0.38) | 1.352 | (0.44) |
| *ENET* | | | −7.471 | (0.00) | −0.019 | (0.30) | | | −5.342 | (0.24) | 0.816 | (0.32) | | | −5.157 | (0.24) | 0.877 | (0.30) |
| *C-ENET* | | | −3.092 | (0.00) | 0.089 | (0.00) | | | 0.893 | (0.25) | −0.019 | (0.07) | | | 2.543 | (0.38) | −0.563 | (0.10) |

This table summarizes the in-sample and out-of-sample return predictability of developed, emerging, and frontier markets for the post-1990 period. We sample the data at a monthly frequency and predict the future one-month USD excess returns. $R^2_{IS}$, $R^2_{OOS}$, and $R^{2,WLS}_{OOS}$ denote the average in-sample $R^2$, out-of-sample $R^2$, and out-of-sample $R^2$s from WLS forecasts, respectively. All $R^2$s are quoted in percentage points. The out-of-sample results are based on 120-month rolling windows. In parentheses, we report the share of countries for which the respective $R^2$s are significantly positive at the 10% level. For all individual variables, statistical significance is determined relative to a bootstrapped distribution, while for the model selection approaches, we use the MSPE-adjusted test statistic of Clark and West (2007). Definitions of the variables can be found in Section 2.

the full sample period. For most variables, the average in-sample $R^2$s are higher than for the full sample period. Interestingly, the performance of the technical predictors is generally weaker. Importantly, the *COMB* and *DMSPE* approaches also work well for the post-1990 period.

Overall, the analysis for the post-1990 period confirms our main findings. Table A4 in the Online Appendix also reports the economic utility gains for the post-1990 period. These results are very similar to those for our entire sample period.

## 5. What drives market return predictability?

A long-standing question related to return predictability is whether it is due to rationally time-varying expected returns (e.g., Fama and French, 1989), or due to market inefficiency caused by financial frictions or possibly by irrational deviations of prices from their fundamental values (e.g., Shiller et al., 1984; Summers, 1986), or a combination of these. Using our large cross-section of countries, we provide evidence on this issue in this section. To set the stage, we first develop a hypothesis, which we then test.

Rösch et al. (2017) argue that financial frictions, such as limited capital or transaction costs, severely reduce market efficiency. This reduced market efficiency may cause slow information diffusion and delayed price reactions to new information, and may lead to predictability of returns in the time series. This leads us to the following hypothesis:

**Hypothesis 1.** Returns are more predictable in countries with less efficient capital markets.

Under this hypothesis, we expect returns to be more predictable in countries with small and restricted capital markets. We use three different proxies for market (in)efficiency: (i) the size of a capital market (larger capital markets proxy for more efficiency), (ii) capital controls/market openness (degree of capital controls proxy for inefficiency), and (iii) GDP per capita (higher GDP per capita proxies for more efficient markets, Jordan et al., 2014). The correlation between average market capitalization and GDP per capita is 0.39. Thus, the two variables contain, at least in part, different information.

Within the developed, emerging, and frontier market groups, we divide countries into terciles for each characteristic. We use the countries' average market capitalization, market openness, and GDP per capita and base our analysis on the out-of-sample predictability of the WLS mean forecast combination approach for the post-1990 period. The WLS average forecast combination approach provides both high predictability and good data availability across regions. In addition, it summarizes the information contained in different predictor variables.[18] We focus on the post-1990 period to limit the heterogeneity of the dataset in terms of the length of the time series of market excess returns and the predictor variables. Table 7 reports the median out-of-sample $R^2$s for each subgroup as well as the share of countries for which the $R^2$s are significantly positive at the 10% level.

*Market capitalization.* First, we analyze whether the size of the stock market, proxied by the aggregate USD market capitalization, systematically affects the predictability of market excess returns. We sort the countries by their USD stock market capitalization, adjusted by the GDP implicit price deflator.

Our results suggest stronger predictability in countries with lower market capitalization. For developed markets, there is little difference between low and high market capitalization countries. The median out-of-sample $R^2$s are 1.1% for both. However, within the groups of emerging

[18] Based on Cochrane (2008), one might argue that it would be better to study in-sample predictability because in-sample statistics are more powerful than out-of-sample statistics. We believe that lack of power is not a concern for the WLS average forecast combination, since it yields significant results for a majority of countries. More importantly, there is no in-sample predictor that completely summarizes the information of different variables. Nevertheless, if we perform the following analysis based on the average in-sample $R^2$s of all predictors for each country, we obtain somewhat weaker but qualitatively similar results.

**Table 7**
Economic sources of return predictability.

| | Developed markets | | Emerging markets | | Frontier markets | |
|---|---|---|---|---|---|---|
| | $R^{2,WLS}_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ |
| *Market capitalization* | | | | | | |
| Low | 1.106 | (0.50) | 2.357 | (0.83) | 4.445 | (0.80) |
| Medium | 0.658 | (0.43) | 2.486 | (1.00) | 2.244 | (0.80) |
| High | 1.080 | (0.75) | 1.544 | (0.83) | 1.218 | (0.80) |
| *Market openness* | | | | | | |
| Low | 0.658 | (0.29) | 1.657 | (0.86) | 1.526 | (0.60) |
| Medium | 1.077 | (0.63) | 2.486 | (1.00) | 2.160 | (0.70) |
| High | 1.182 | (0.71) | 1.121 | (0.43) | 2.294 | (0.70) |
| *GDP per capita* | | | | | | |
| Low | 1.008 | (0.63) | 2.280 | (0.88) | 1.536 | (0.55) |
| Medium | 0.886 | (0.57) | 1.726 | (1.00) | 2.431 | (0.73) |
| High | 0.929 | (0.50) | 1.241 | (0.50) | 1.277 | (0.73) |

This table examines the out-of-sample return predictability of developed, emerging, and frontier markets with different characteristics. For each characteristic, we sort the countries into three groups (low, medium, and high) within their market development status, breaking at the terciles of the respective sorting variable. We sample the data at a monthly frequency and predict the future one-month excess returns using the WLS mean forecasts combination (COMB). $R^{2,WLS}_{OOS}$ denotes the average out-of-sample $R^2$ (in percentage points). The results are based on a 120-month rolling window for the post-1990 period. In parentheses, we report the share of countries for which the respective $R^2$s are significantly positive at the 10% level. Statistical significance is determined using the MSPE-adjusted test statistic of Clark and West (2007).

and frontier markets, where the predictability is generally better, the out-of-sample predictability is stronger for countries with low and medium market capitalization than for countries with high market capitalization. For emerging markets, the median out-of-sample $R^2$ for high market capitalization countries is 1.5%, while that for low market capitalization countries is 2.4%. For frontier markets, the difference is even larger, 1.2% vs. 4.4%. Thus, our results provide some support for Hypothesis 1: lower market capitalization seems to be associated with better predictability.

*Market openness.* For market openness, the results are mixed. Consistent with Hypothesis 1, for emerging markets, lower openness seems to be weakly associated with stronger predictability. For developed and frontier markets, however, the relationship tends to be the opposite. On the other hand, the market openness index is probably our weakest proxy for market efficiency, as foreign capital is not necessarily needed to arbitrage away inefficiencies.

*GDP per capita.* We sort countries by their average GDP per capita. We find that aggregate excess returns are generally more predictable in countries with lower GDP per capita than in countries with higher GDP per capita. While this relationship is weak for developed markets, for both emerging and frontier markets, we find that the median out-of-sample $R^2$s of countries with medium and low GDP per capita are substantially larger than those of countries with high GDP per capita. Thus, the evidence with this proxy seems to further support Hypothesis 1: aggregate excess returns are least predictable in high GDP per capita countries.

Taken together, our results suggest that market inefficiency is an important source of the observed return predictability. First, we find that out-of-sample $R^2$s are generally larger the less developed a country's capital market is. Emerging and frontier markets are characterized by both lower market efficiency and generally higher business-cycle variability. Second, we find strong in-sample and out-of-sample predictability for the simple technical indicator variables. Predictability by technical signals challenges weak-form market efficiency as defined by Fama (1970). Third, within groups of similarly developed countries, we find that those with lower market capitalization and GDP per capita are generally more predictable, confirming that market inefficiency is likely to play a large role in return predictability. However, not all proxies for market inefficiency point uniformly in the same direction, and most individual variables fail to consistently predict out-of-sample market excess returns. Thus, rationally time-varying expected returns are also likely to play a role in return predictability.

## 6. Further analyses and robustness tests

### 6.1. Panel model return predictability

Given a large set of countries and predictor variables, a natural idea is to pool these data for joint estimation of the parameters. If the predictive relationships are similar across countries, pooling them allows us to estimate the parameters with higher precision and less noise. Thus, for each predictor variable, we pool all observations and analyze the predictability jointly for all countries. We use a country fixed effects model to account for heterogeneous levels of equity premia and jointly determine the predictive slopes. Finally, we compute the in-sample and out-of-sample $R^2$s jointly. We use the MSPE-adjusted test statistic of Clark and West (2007) to determine statistical significance. For these tests, we use the double-clustered standard errors of Cameron et al. (2011) clustered by country and month.

We present the results in Table 8. In addition to the usual categories of developed, emerging, and frontier markets, we also pool together all countries. For most predictors, the in-sample $R^2$s are somewhat smaller with the

**Table 8**
Panel model return predictability.

| | All countries | | | Developed markets | | | Emerging markets | | | Frontier markets | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $R^2_{IS}$ | $R^2_{OOS}$ | $R^{2,WLS}_{OOS}$ | $R^2_{IS}$ | $R^2_{OOS}$ | $R^{2,WLS}_{OOS}$ | $R^2_{IS}$ | $R^2_{OOS}$ | $R^{2,WLS}_{OOS}$ | $R^2_{IS}$ | $R^2_{OOS}$ | $R^{2,WLS}_{OOS}$ |
| **Fundamental predictors** | | | | | | | | | | | | |
| *DY* | 0.030 | 0.387 | −0.542 | 0.002 | 0.633* | −0.574 | 0.076** | 0.276* | −0.310 | 0.028 | 0.503 | −2.245 |
| *PE* | 0.018* | 0.101 | −0.536 | 0.010 | 0.252* | −0.701 | 0.032** | 0.057 | −0.094 | 0.020 | 0.012*** | −2.860 |
| *RVOL* | 0.081 | 0.874 | −0.520 | 0.407 | 2.082 | −0.577 | 0.003 | 0.081** | −0.346 | 0.012 | −0.191 | −0.938 |
| *RREL* | 0.022 | 0.454 | −0.567 | 0.356*** | 0.353*** | −0.668 | 0.094 | 1.526 | −0.331 | 0.568 | 0.736*** | −0.674 |
| *TMS* | 0.064 | 0.060** | −0.596 | 0.109** | 0.280* | −0.692 | 0.815*** | 1.141*** | −0.265 | 0.409 | 3.078 | −0.312 |
| *DFY* | 0.013 | 0.314 | −0.688 | 0.014 | 0.303* | −0.759 | 1.629 | 1.532*** | −0.194 | | | |
| *INFL* | 0.167 | 0.854 | −0.359 | 0.466 | 1.959 | −0.215 | 0.301* | 0.160*** | −0.336 | 0.659 | 0.258 | −0.758 |
| *UE* | 0.108*** | 0.371*** | −0.668 | 0.099** | 0.285*** | −0.660 | 0.529*** | 1.242*** | −0.630 | 0.083 | −0.476 | −1.158 |
| **Technical signals** | | | | | | | | | | | | |
| $MA_{1,12}$ | 1.362*** | 1.126*** | −0.515 | 1.255*** | 1.192*** | −0.569 | 1.144*** | 1.287*** | −0.344 | 2.266*** | 1.270*** | −0.928 |
| $MA_{3,12}$ | 0.969*** | 0.874*** | −0.516 | 0.879*** | 0.862*** | −0.569 | 0.828*** | 0.996*** | −0.348 | 1.616*** | 1.255*** | −0.934 |
| $MOM_6$ | 1.159*** | 0.986*** | −0.516 | 1.041*** | 0.981*** | −0.567 | 0.938*** | 1.079*** | −0.347 | 2.084*** | 1.369*** | −0.934 |
| **Forecast combinations** | | | | | | | | | | | | |
| *COMB* | | 1.499*** | −0.489 | | 2.029** | −0.505 | | 1.289*** | −0.337 | | 1.538*** | −0.934 |

This table summarizes the in-sample and out-of-sample return predictability of developed, emerging, and frontier markets by lagged monthly returns of the different markets. We sample the data at a monthly frequency and use a pooled panel design to predict future one-month USD excess returns. $R^2_{IS}$, $R^2_{OOS}$, and $R^{2,WLS}_{OOS}$ denote the average in-sample $R^2$, out-of-sample $R^2$, and out-of-sample $R^2$ from WLS forecasts, respectively. All $R^2$s are quoted in percentage points. The out-of-sample results are based on 240-month rolling windows. Statistical significance is determined using the MSPE-adjusted test statistic of Clark and West (2007). For these tests, we use the double-clustered standard errors of Cameron et al. (2011), clustered by country and month.

* Indicates significance at the 10% level.

** Indicates significance at the 5% level.

*** Indicates significance at the 1% level.

panel approach. Thus, the exact predictive relationship seems to vary somewhat across countries, even among countries at the same stage of market development. The situation is different for the out-of-sample OLS $R^2$s. These are generally much larger for the panel approach. Thus, the inherently less powerful out-of-sample tests benefit from the more precise parameter estimates provided by the much larger sample in a panel setup. Interestingly, this is only true for the OLS approach. All out-of-sample WLS $R^2$s are negative. Thus, the WLS return forecasting approach does not seem to work in a panel setting.

In the panel setup, the same forecasting variables work well, as in the country-by-country analysis. In particular, the technical signals together with *UE*, *TMS*, and *RREL* predict future returns quite well. The simple mean forecast combination also performs very well in the panel analysis. Finally, our results suggest that the countries should not all be pooled together. The predictability results are generally better when countries are grouped only by their level of development.

### 6.2. Spillover return predictability

Next, we examine the extent of cross-market predictability. To keep the dimensionality manageable, we only consider the lagged average returns of differently developed markets as predictors. That is, each month we calculate average market excess returns for developed, emerging, and frontier markets. We then regress the market excess returns over the next month of each country separately on each of these average market excess returns. We use exactly the same in-sample and out-of-sample approach as in the main analysis.

The results are shown in Table 9. We find that there is substantial predictability of international returns from lagged market excess returns, both in sample and out of sample. In general, the lagged returns of countries at the same level of development provide the strongest predictability for all types of markets. In addition, the lagged returns of more developed markets tend to have higher $R^2$s than those of less developed markets. Lagged frontier market returns have little predictive power for developed and emerging markets, while we observe higher predictive power in the opposite direction. Forecast combinations of spillover return predictions also work well.

These results are consistent with and extend those of Rapach et al. (2013). The authors show that lagged U.S. returns have significant predictive ability for 10 other developed markets. We show that the international cross-predictive patterns are richer than this, and that even the returns of less developed markets generally have some predictive ability. Taken together, these results underscore that a substantial portion of return predictability is likely to be due to market inefficiency. Many equity markets appear to incorporate some of the information contained in other markets only with a lag.

### 6.3. Domestic return predictability

Finally, we examine the predictability of international returns in domestic currency. Since our main results are in USD, they are economically from the perspective of a U.S. investor. The results in domestic currency can be interpreted either from the perspective of local investors for each country or from the perspective of international investors who are hedged in exchange rates. We adjust

**Table 9**
Spillover return predictability.

| | Developed markets | | | | | | Emerging markets | | | | | | Frontier markets | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $R^2_{IS}$ | $(Shr)$ | $R^2_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ | $R^2_{IS}$ | $(Shr)$ | $R^2_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ | $R^2_{IS}$ | $(Shr)$ | $R^2_{OOS}$ | $(Shr)$ | $R^{2,WLS}_{OOS}$ | $(Shr)$ |
| **Lagged returns** | | | | | | | | | | | | | | | | | | |
| *DEV* | 1.888 | (0.96) | 0.671 | (0.57) | 0.244 | (0.09) | 1.563 | (0.79) | −1.014 | (0.37) | 0.670 | (0.11) | 1.690 | (0.87) | 0.164 | (0.60) | 1.155 | (0.33) |
| *EM* | 1.155 | (0.91) | 0.266 | (0.52) | 0.586 | (0.04) | 2.446 | (0.95) | 0.599 | (0.68) | 0.753 | (0.16) | 2.143 | (0.87) | 1.421 | (0.73) | 1.162 | (0.33) |
| *FRO* | 0.295 | (0.39) | −0.846 | (0.09) | 0.158 | (0.04) | 0.457 | (0.47) | −1.169 | (0.21) | 0.646 | (0.26) | 3.109 | (0.80) | 2.488 | (0.67) | 1.079 | (0.27) |
| **Forecast combinations** | | | | | | | | | | | | | | | | | | |
| *COMB* | | | 0.946 | (0.70) | 0.246 | (0.57) | | | 0.399 | (0.58) | 0.702 | (0.53) | | | 2.127 | (0.60) | 1.133 | (0.27) |
| *DMSPE* | | | 0.898 | (0.65) | 0.236 | (0.48) | | | 0.486 | (0.44) | 0.887 | (0.61) | | | 3.309 | (0.36) | 3.993 | (0.43) |
| *ENET* | | | 0.282 | (0.52) | 0.498 | (0.65) | | | −0.687 | (0.37) | 0.952 | (0.68) | | | 2.050 | (0.53) | 0.920 | (0.33) |
| *C-ENET* | | | 0.164 | (0.50) | 0.017 | (0.18) | | | 0.032 | (0.38) | −0.216 | (0.17) | | | −2.932 | (0.00) | −0.017 | (0.00) |

This table summarizes the in-sample and out-of-sample return predictability of developed, emerging, and frontier markets by lagged average monthly market excess returns of the different markets. We sample the data at a monthly frequency and predict the future one-month USD excess returns. $R^2_{IS}$, $R^2_{OOS}$, and $R^{2,WLS}_{OOS}$ denote the average in-sample $R^2$, out-of-sample $R^2$, and out-of-sample $R^2$s from WLS forecasts, respectively. All $R^2$s are quoted in percentage points. The out-of-sample results are based on 240-month rolling windows. In parentheses, we report the share of countries for which the respective $R^2$s are significantly positive at the 10% level. For all individual variables, statistical significance is determined relative to a bootstrapped distribution, while for the model selection approaches, we use the MSPE-adjusted test statistic of Clark and West (2007).

all domestic returns for inflation and use the domestic three-month Treasury-bill rate as a proxy for the risk-free rate.[19] Adjusting the excess returns and risk-free rates for ex post inflation ensures that the results are not disproportionately driven by high unexpected inflation in some countries.[20]

We visualize the results in Table A5 in the Online Appendix and present the summary in Table A6 in the Online Appendix. Overall, the results are very similar to, but generally stronger than, our main results. The in-sample return predictability is somewhat stronger for domestic returns. The out-of-sample predictability is also stronger than for USD returns. These results suggest that exchange rate changes have a negative impact on return predictability. In particular, *INFL* performs very well for both in-sample and out-of-sample prediction of local inflation-adjusted excess returns. The *ENET* and *C − ENET* approaches also perform somewhat better for local returns. The *COMB* and *DMSPE* approaches continue to perform well regardless of how returns are measured.

### 6.4. Robustness

In this section, we conduct robustness tests. First, we examine whether economic restrictions help improve the international out-of-sample results. Second, we use a shorter rolling window to make out-of-sample predictions. Third, we examine predictability using an expanding window rather than a rolling window.

We start with the economic restrictions of Campbell and Thompson (2008). That is, if the out-of-sample slope estimate has a different sign than the in-sample estimate, or if it is negative, we set the forecast equal to the historical mean. We present the results in Table A7 in the Online Appendix. These restrictions generally increase the average out-of-sample OLS $R^2$s, but not those estimated from WLS. The simple average forecast combination approach based on the unrestricted forecasts (see Table 3) still performs better than any single restricted predictor.

Second, we turn to a rolling window estimation with a 10-year window instead of a 20-year window. This reduced window length allows us to include some additional countries with shorter time series. We present the summary results in Table A8 in the Online Appendix. These results are qualitatively similar to our main findings.

Third, we examine an expanding rather than a rolling estimation window. As in our main analysis, we set the initial estimation period at 240 months and expand this period by one month for each subsequent monthly return forecast. The results are shown in Table A9 in the Online Appendix. They are qualitatively similar to our main results for a rolling window estimation.

## 7. Conclusion

We comprehensively analyze the predictability of the equity premium using a long sample period, a broad cross-section of countries, and a wide variety of predictor variables. Using the standard setup, we find little evidence of out-of-sample predictability by fundamental predictor variables. However, technical signals exhibit significantly stronger out-of-sample predictive power. Simple forecast combinations of the variables perform consistently well.

Analyzing the determinants of return predictability, we find that market inefficiency plays an important role. On average, excess returns in emerging and frontier markets are significantly more predictable than those in developed markets. Moreover, within these groups, countries with smaller market capitalizations and GDP per capita are generally more predictable. Finally, there are rich spillover effects between the market excess returns of different countries.

## CRediT authorship contribution statement

**Fabian Hollstein:** Conceptualization, Formal analysis, Methodology, Writing – review & editing. **Marcel Prokopczuk:** Conceptualization, Methodology, Writing – review & editing. **Björn Tharann:** Conceptualization, Formal analysis, Methodology, Writing – original draft. **Chardin Wese Simen:** Conceptualization, Methodology, Writing – review & editing.

[19] For part of the sample (period), a domestic three-month Treasury-bill rate is not available. In this case, we simply set it to zero.

[20] Note that for the main analysis we do not adjust returns for inflation. In the case of USD returns, a depreciating local currency offsets most of the effect of high inflation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2024.05.002.

## References


Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there? *The Review of Financial Studies*, *20*(3), 651–707.

Balduzzi, P., & Lynch, A. W. (1999). Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, *52*(1), 47–78.

Bansal, R., & Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, *59*(4), 1481–1509.

Boyd, J. H., Hu, J., & Jagannathan, R. (2005). The stock market's reaction to unemployment news: Why bad news is usually good for stocks. *The Journal of Finance*, *60*(2), 649–672.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, *18*(2), 373–399.

Campbell, J. Y. (1991). A variance decomposition for stock returns. *The Economic Journal*, *101*(405), 373–399.

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, *107*(2), 205–251.

Campbell, J. Y., & Shiller, R. J. (1988). Stock prices, earnings, and expected dividends. *The Journal of Finance*, *43*(3), 661–676.

Campbell, J. Y., & Shiller, R. J. (1998). Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management*, *24*(2), 11–26.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509–1531.

Charles, A., Darné, O., & Kim, J. H. (2016). International stock return predictability: Evidence from new statistical tests. *International Review of Financial Analysis*, *54*, 97–113.

Chen, N.-F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, *59*(3), 383–403.

Chinn, M. D., & Ito, H. (2006). What matters for financial development? Capital controls, institutions, and interactions. *Journal of Development Economics*, *81*(1), 163–192.

Chordia, T., Roll, R., & Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics*, *87*(2), 249–268.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291–311.

Cochrane, J. H. (1992). Explaining the variance of price–dividend ratios. *The Review of Financial Studies*, *5*(2), 243–280.

Cochrane, J. H. (1999). New facts in finance. *Economic Perspectives*, *23*(3), 36–58.

Cochrane, J. H. (2007). Financial markets and the real economy. In R. Mehra (Ed.), *Handbook of the equity premium* (pp. 237–325). Amsterdam: Elsevier.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *The Review of Financial Studies*, *21*(4), 1533–1575.

Cochrane, J. H. (2011). Discount rates. *The Journal of Finance*, *66*(4), 1047–1108.

Dow, C. H. (1920). *Scientific stock speculation*. Magazine of Wall Street.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, *22*(1), 3–25.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, *25*(1), 23–49.

Ferson, W. E., & Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, *99*(2), 385–415.

Fleming, J., Kirby, C., & Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, *56*(1), 329–352.

Flynn, C. J., Hurvich, C. M., & Simonoff, J. S. (2013). Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models. *Journal of the American Statistical Association*, *108*(503), 1031–1043.

Golez, B., & Koudijs, P. (2018). Four centuries of return predictability. *Journal of Financial Economics*, *127*(2), 248–263.

Goyal, A., & Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

Henkel, S. J., Martin, J. S., & Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, *99*(3), 560–580.

Hjalmarsson, E. (2010). Predicting global stock returns. *Journal of Financial and Quantitative Analysis*, *45*(1), 49–80.

Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *The Review of Financial Studies*, *5*(3), 357–386.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307.

Inoue, A., & Kilian, L. (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, *23*(4), 371–402.

Johnson, T. L. (2018). A fresh look at return predictability using a more efficient estimator. *Review of Asset Pricing Studies*, *9*(1), 1–46.

Jordan, S. J., Vivian, A. J., & Wohar, M. E. (2014). Forecasting returns: New European evidence. *Journal of Empirical Finance*, *26*, 76–95.

Lettau, M., & Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance*, *56*(3), 815–849.

Lettau, M., & Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence. *The Review of Financial Studies*, *21*(4), 1607–1652.

Lo, A. W., & MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *The Review of Financial Studies*, *3*(3), 431–467.

McCracken, M. W. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics*, *140*(2), 719–752.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1–2), 3–24.

Mele, A. (2007). Asymmetric stock market volatility and the cyclical behavior of expected returns. *Journal of Financial Economics*, *86*(2), 446–478.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management Science*, *60*(7), 1772–1791.

Pesaran, M. H., & Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance*, *9*(5), 495–510.

Rangvid, J., Schmeling, M., & Schrimpf, A. (2014). Dividend predictability around the world. *Journal of Financial and Quantitative Analysis*, *49*(5–6), 1255–1277.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, *23*(2), 821–862.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International stock return predictability: What is the role of the United States? *The Journal of Finance*, *68*(4), 1633–1662.

Rapach, D. E., & Wohar, M. E. (2006). In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance*, *13*(2), 231–247.

Rapach, D. E., Wohar, M. E., & Rangvid, J. (2005). Macro variables and international stock return predictability. *International Journal of Forecasting*, *21*(1), 137–166.

Rapach, D. E., & Zhou, G. (2013). Forecasting stock returns. In G. Elliot, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting*: *vol. 2*, (pp. 328–383). Amsterdam: Elsevier.

Rapach, D. E., & Zhou, G. (2020). Time-series and cross-sectional stock return forecasting: New machine learning methods. In E. Jurczenko (Ed.), *Machine learning in asset management: new developments and financial applications*. Hoboken, NJ: Wiley.

Rösch, D. M., Subrahmanyam, A., & van Dijk, M. A. (2017). The dynamics of market efficiency. *The Review of Financial Studies*, *30*(4), 1151–1187.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, *51*(4), 1063–1119.

Rozeff, M. S. (1984). Dividend yields are equity risk premiums. *Journal of Portfolio Management*, *11*(1), 68–75.

Shaman, P., & Stine, R. A. (1988). The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, *83*(403), 842–848.

Shiller, R. J., Fischer, S., & Friedman, B. M. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*, *1984*(2), 457–510.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, *54*(3), 375–421.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.

Summers, L. H. (1986). Does the stock market rationally reflect fundamental values? *The Journal of Finance*, *41*(3), 591–601.

Timmermann, A., & Granger, C. W. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, *20*(1), 15–27.