ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
Bachelor Thesis Econometrics and Operations Research

Robust Bond and Currency Risk Premia

A.G. (Daniëlle) Lam (575651al)

Supervisor:	Dr. M. (Maria) Grith
Second assessor:	Dr. G. (Gustavo) Freire
Date final version:	29th June 2023

Robust Bond and Currency Risk Premia

Daniëlle Lam[a]

Abstract

The OLS estimator in a predictive regression including persistent and endogenous regressor variables has a different small-sample distribution than assumed by commonly used inference which results primarily from the negative bias in the standard errors. By implementing and modifying the parametric bootstrap procedure of Bauer and Hamilton (2018), we revisit two published studies investigating the importance of macroeconomic variables on bond and currency risk premia controlling for yield information. This robust inference shows the distortion of the Wald test, for it accumulates finite-sample distortions in t-tests, which is predominantly caused by the high persistence of the regressors. Concurrently confronting the models with more recent data raises concerns about the importance of the macroeconomic information in both financial markets.

Keywords: Yield Curve, Spanning, Return Predictability, Robust Inference, Bootstrap, Carry Trade, Forward Premium Puzzle, Factor Analysis

[a] Erasmus School of Economics Bachelor student. Contact: 575651al@eur.nl

Contents

1 Introduction

When regressor variables violate the condition of strict econometric exogeneity and are highly persistent in a predictive regression, conventional t- and Wald tests could be invalid, meaning that they reject the null hypothesis too frequently. While the coefficient estimates are consistent, they are not necessarily unbiased, and while the standard error estimates are asymptotically valid, this is not the case in small-sample sizes. Previously, the dangers of forecasting with highly serially correlated right-hand side variables have been centered around the bias occurring in the estimated coefficients, the Stambaugh bias (Stambaugh, 1999). By contrast, Bauer and Hamilton (2018) emphasize that it is not the coefficient bias, but instead, the downward bias of the estimated errors that distorts the results of conventional inference. The problems become even more severe when the right-hand side variables exhibit a trend over the observed sample and in the presence of overlapping observations. Consequently, the standard errors and the regression R^2 become less reliable. To obtain robust inference, Bauer and Hamilton (2018) propose a parametric bootstrap approach generating data samples under the null hypothesis where the serial correlation of the variables is similar to the actual data. Using this approach, we study the magnitude and consequences of the standard error bias which is shown to emerge in predicting bond risk premia by Bauer and Hamilton (2018), yet we extend this to predicting currency risk premia by investigating the study of Filippou and Taylor (2017).

A central field of research in macro-finance where these problematic features arise is in testing the spanning hypothesis. This hypothesis states that the yield curve itself spans all information relevant for forecasting bond risk premia, and no other variables than the current yield curve are needed. It has long been acknowledged that the first three principal components (PCs) of yields, labeled as level, slope, and curvature, provide an outstanding summary of the entire yield as they capture almost all cross-sectional variance of observed yields (Litterman & Scheinkman, 1991). Regardless, there is a growing agreement that the spanning hypothesis can be empirically rejected. This is generally shown in predictive regressions for bond returns on various predictors, controlling for information in the yield curve. The variables that are proposed in the literature to possess supplemental predictive power in such regressions include measures of economic growth and inflation (Joslin, Priebsch & Singleton, 2014), factors inferred from a large set of macro variables (Ludvigson & Ng, 2009), long-term trends in inflation or inflation expectations (Cieslak & Povala, 2015), higher-order PCs of yields (Cochrane & Piazzesi, 2005). However, under the null hypothesis, the predictive variables are necessarily correlated with lagged forecast errors because they summarize information in the current yield curve. As a result, this violates the exogeneity condition. In addition, the PCs and the proposed variables are commonly highly persistent. Due to these problematic features, it remains questionable whether these variables can explain bond returns beyond the information contained in the PCs when conventional inference is used. When reanalyzed with a robust inference technique, the evidence against the spanning hypothesis is significantly weaker than would appear from the published results, and at some points occurs to be spurious (Bauer & Hamilton, 2018).

Return predictability is not restricted to the bond market. Similar to bond returns, the predictability of currency returns and concurrent violations of the uncovered interest rate parity (UIRP) is shown in numerous studies that originated with Hansen and Hodrick (1980), Bilson (1980) and Fama (1984).[1] The UIRP indicates that when the foreign interest rate is higher than the domestic interest rate, risk-neutral and rational investors should expect the foreign currency to depreciate against the domestic currency by the difference between the two interest rates. In fact, high-interest rate differentials seem to lead to further appreciations on average, see among others Lustig and Verdelhan (2007). This is known as the forward premium puzzle. Due to the strong autocorrelation in the forward premium, literature has cast doubts on the finite sample accuracy of these tests. Liu and Maynard (2005) conclude, however,

[1] The literature is surveyed for example in Sarno (2005).

that the forward premium is more robust than formerly presumed as the size distortions in standard tests are not sufficient to overturn UIRP rejections. Not only is the forward discount highly persistent, but it is also not strictly exogenous (Villanueva, 2007). By the covered interest rate parity, which is supported by early literature e.g. Taylor (1987), the interest rate differential between countries equalizes the respective foreign exchange (FX) forward premium. Therefore, the currency excess return, which is the interest rate differential corrected by the change in the spot rate, is simultaneously determined with the forward premium. In addition, Lustig, Roussanov and Verdelhan (2014) document the ability of the average forward discount (AFD) of developed countries to forecast individual exchange rates and returns on other currency baskets. This follows the capability of forward rates to forecast returns on bonds of other maturities which is documented by Stambaugh (1988) and Cochrane and Piazzesi (2005).

Instead of explaining individual stock returns, interest has moved to explaining the returns on portfolios constructed by sorting equities on variables known to predict returns such as size and book-to-market ratio. Likewise by categorizing currencies into portfolios, idiosyncratic risk is eliminated (Lustig & Verdelhan, 2007). There is a vast literature on the cross-sectional predictability of the payoff from various investment strategies that exploit UIRP violations. These strategies include inter alia the carry-trade (Lustig & Verdelhan, 2007; Menkhoff, Sarno, Schmeling & Schrimpf, 2012a), a dollar carry-trade (Lustig et al., 2014), a momentum strategy (Menkhoff, Sarno, Schmeling & Schrimpf, 2012b), strategies based on information in the volatility risk premium (Della Corte, Ramadorai & Sarno, 2016), strategies sorting portfolios based on other variables, for example, the output gap (Colacito, Riddiough & Sarno, 2020).

The relation between macroeconomic fundamentals and equity return forms a central macro-finance issue (Cochrane, 2017). Although this is most extensively studied for the asset market, it is also challenging to establish for the foreign exchange market. In theory, currency returns and country-level characteristics are highly correlated, yet the high variability of exchange rates is hard to forecast by economic models (Rossi, 2013). There is a broad literature that has analyzed the connections between currency risk premia and macro fundamentals which include yield curve factors (Chen & Tsang, 2013), common domestic and global factors inferred from a large data set (Filippou & Taylor, 2017), the distance between countries (Lustig & Richmond, 2017), the relative strength of the business cycle (Colacito et al., 2020), capital flows (Gabaix & Maggiori, 2015) and global foreign exchange volatility risk (Menkhoff et al., 2012a). The importance of certain variables depends also on the strategy that is employed. For example, carry trade and momentum strategies profit from disparities observed in global market conditions and especially between debtor and creditor economies (Lustig, Roussanov & Verdelhan, 2011; Corte, Riddiough & Sarno, 2016). In contrast to the dollar carry-trade which depends more on domestic economic indicators such as the year-over-year rate of industrial production growth (Lustig et al., 2014).

In this paper, we assess the importance of factors derived from macroeconomic data sets in predicting excess returns in the bond and currency market when controlling for information in the yield curve. If macroeconomic variables do not exhibit significant explanatory power over the regressor summarizing the yield, then this would greatly simplify forecasting risk premia. We investigate this in the bond market by replicating the study of Bauer and Hamilton (2018). They reinvestigate four different influential studies by more robust inference and confronting the models with more recent data. We replicate these studies but focus mostly on the importance of factors derived from a large macroeconomic data set as originally proposed by Ludvigson and Ng (2009). We extend this by confronting the model with more recent data that has appeared since the publication by Bauer and Hamilton (2018). In a similar fashion, we investigate the study of Filippou and Taylor (2017), in the currency market however, by using the bootstrap procedure of Bauer and Hamilton (2018) and by estimating the model over data that has appeared since the publication of this study. It is important to note that Filippou and Taylor (2017) use a bootstrap procedure, however, their design imposed the null of no predictability by any variable, whereas we simulate under the null of no predictability by any macro factor.

There are some differences in the setup of these predictive regressions to follow the original studies by Ludvigson and Ng (2009) in the bond market and Filippou and Taylor (2017) in the currency market. First, in the bond market, we predict the average excess return on bonds of different maturities, whereas we study the excess return on a carry-trade strategy in the currency market. We focus in particular on the carry-trade strategy using currencies from only developed countries since these are the most actively traded (Filippou & Taylor, 2017). The carry-trade strategy profits from UIRP violations for it takes a long position in high-interest rate currencies and a short position in low-interest rate currencies. Second, in the bond market application, the first three principal components of the yield data are used to capture the information in the yield curve. In the currency market, we use the average forward discount on currencies of developed countries to represent the yield.[2] Both regressor variables are not strictly exogenous and are highly persistent. The bootstrap procedure, however, is constructed in the bond market which uses the estimated yield factors to simulate the excess returns. As the currency excess return and the average forward discount do not have a similar direct relation, we modify the bootstrap procedure to be more generally applicable. Third, for both applications, we study the importance of factors derived from large macroeconomic data sets. For the currency market, we extract factors from a US macroeconomic data set, as well as from a global macroeconomic data set. In the bond market illustration, we only study the significance of the factors from the US macroeconomic data set.

The aim of this paper is to assess how severe the standard error bias is which is already shown for the bond market by Bauer and Hamilton (2018), yet we study this also in the currency market. We do this by comparing the size and power properties of conventional t- and Wald tests to the bootstrap procedure proposed by Bauer and Hamilton (2018) and by investigating if this robust inference overturns conclusions of conventional inference. Since this was introduced in testing the spanning hypothesis in the bond market illustration, we aim to make this more generally applicable, to assess whether standard error bias is as important to identify in another financial market and to get a better understanding under which circumstances this technique provides especially useful. In addition, the testing procedure is used to profit from better size properties than conventional tests to improve our knowledge of violations of the spanning hypothesis in the bond market and of the forward premium puzzle in the currency market. Lastly, we confront the models with more recent data to assess whether the importance is due to a certain sub-sample.

For the bond market illustration specifically, we find that the macro factors are less important than previously shown by Ludvigson and Ng (2009). However, compared to Bauer and Hamilton (2018) we conclude that the macro factors are jointly more important since their publication. In this later sample, we find that the factor related to the categories output and income, the labor market, and housing results in a rejection of the spanning hypothesis. To assess if the predictability varies over the sample, we also consider rolling regressions. We find that the addition of the macro factors to the model substantially deteriorates the predicting errors, however, not significantly. When estimating the model using more recent data, we find that the addition of the macro factors increases the regression R^2. However, this measure of goodness-of-fit is included in the 95% confidence interval bootstrapped under the null hypothesis that they do not contain predictive power. Therefore, this does not provide evidence in favor of the macro factors.

In the currency market application, we find the following. First, profiting from the better-sized bootstrap tests, we assess the importance of the macro factors individually and jointly on the carry-trade pay-off using only the currencies from developed countries. We find that the bootstrap procedure overturns the conclusion from the conventional Wald test of joint significance of the macro factors. However, we find that the fifth domestic macro factor is significantly important. This factor loads especially heavily on the series from the categories of money and credit quantity aggregates and employment. The latter

[2] Extracting principal components from the forward discount (or from the excess returns of portfolios similar to Lustig et al. (2011)) leads to factors that are not persistent and that have close to zero correlation with the prediction error. Therefore the bootstrap procedure is unnecessary.

relates to a large literature exploiting the predictive power of the unemployment gap on carry trade excess returns (e.g. Berg and Mark (2018)). Second, by confronting the model with more recent data we generally find that the evidence in favor of the significance of the macro factors is even weaker. The macro factors are not jointly significant in the later sample and the adjusted R^2 deteriorates due to the macro factors. However, the latter is again included in the 95% bootstrap confidence interval. Moreover, we find that two factors are statistically significant at a 10% significance level over the complete sample controlling for the AFD. These unobserved factors correspond mostly to the stock market category. This finding relates to Brunnermeier, Nagel and Pedersen (2008) who document that the VIX helps resolve the forward premium puzzle.

In conclusion, we find that it is crucial to recognize and act on the standard error bias since the Wald test has serious size distortions in both the bond and currency markets. This is particularly due to the fact that the Wald test accumulates small sample distortions of the t-tests testing the individual significance of the macro factors. In fact, the endogeneity of the average forward discount in the currency market is much less pronounced than the endogeneity of the yield factors in the bond market. Therefore we conclude that the size distortion is particularly due to the high persistence of the regressors. Profiting from the better-sized bootstrap tests, we find that the evidence in favor of the macro factors is weaker than proposed in the original studies. Especially by confronting the models with more recent data, we observe that the macro factors are less important than originally proposed.

The remainder of this paper is structured in the following way. Section 2 starts with a motivating example and discusses bootstrap inference. Section 3 gives an overview of the data used and corresponding definitions, where the focus is especially on the foreign exchange market. Section 4 shows the estimation results of the predictive regressions in the bond market and currency market. Lastly, in Section 5 we summarize and conclude.

2 Methodology

2.1 Motivating Example

To explore the effects of the persistence and endogeneity on the small-sample behavior of the t-statistic, we perform a simulation experiment with similar settings as in Bauer and Hamilton (2018).[3] We use the following data-generating process

$$y_{t+1} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_{t+1}, \tag{1}$$

where x_{1t} and x_{2t} are scalar AR(1) processes

$$x_{1,t+1} = \mu_1 + \rho x_{1t} + \varepsilon_{1t} \tag{2}$$

$$x_{2,t+1} = \mu_2 + \rho x_{2t} + \varepsilon_{2t} \tag{3}$$

and test $H_0 : \beta_2 = 0$. Moreover, we simulate 100,000 samples by estimating (1) under the restriction that $\beta_0 = \beta_1 = \beta_2 = 0$. In addition, we assume that the errors $(\varepsilon_{1t}, \varepsilon_{2t}, u_{t+1})$ are serially uncorrelated, except $Corr(\varepsilon_{1t}, u_{t+1}) = \delta$. We are especially interested in what happens when the persistence parameter ρ is close to unity and when the regressors in x_{1t} are not strictly exogenous ($\delta \neq 0$). We also include a t-test using commonly used heteroscedasticity- and autocorrelation consistent (HAC) standard errors, which Bauer and Hamilton (2018) do not.

Panel A of Table 1 shows the simulation output when none of the variables exhibit a trend and for different levels of persistence (ρ) and endogeneity (δ). If the regressors are not persistent ($\rho = 0$)

[3] This experiment relates to Section 1.3 by Bauer and Hamilton (2018).

Table 1: Simulation Study of Standard Error Bias

Persistence	Endogeneity	Coefficient bias		SE bias		Size			
ρ	δ	β_1	β_2	OLS	HAC	OLS	HAC	Asymptotic	Bootstrap
Panel A: $\mu_1 = \mu_2 = 0$ (No trend)									
0.99	0.0	0.000	0.000	-4.7	-26.7	0.049	0.161	0.047	0.048
0.00	1.0	-0.010	0.000	-0.5	-13.4	0.050	0.114	0.051	0.050
0.90	1.0	-0.052	0.000	-15.5	-26.2	0.085	0.167	0.085	0.058
0.99	0.8	-0.054	0.000	-23.0	-36.1	0.111	0.223	0.113	0.070
0.99	1.0	-0.068	0.000	-29.7	-39.6	0.151	0.253	0.151	0.080
Panel B: $\mu_1 = 0, \mu_2 = 1$ (Trend in x_{2t})									
0.99	0.0	0.000	0.000	-5.1	-31.1	0.051	0.186	0.048	0.049
0.00	1.0	-0.010	0.000	-0.3	-13.3	0.049	0.116	0.052	0.048
0.90	1.0	-0.054	0.000	-17.1	-31.7	0.089	0.193	0.088	0.057
0.99	0.8	-0.071	0.000	-42.5	-53.8	0.184	0.331	0.111	0.077
0.99	1.0	-0.088	0.000	-50.8	-58.3	0.270	0.397	0.152	0.085
Panel C: $\mu_1 = \mu_2 = 1$ (Trend in x_{1t} and x_{2t})									
0.99	0.0	0.000	0.000	-4.1	-26.6	0.051	0.165	0.046	0.048
0.00	1.0	-0.009	0.000	-0.5	-13.4	0.050	0.116	0.050	0.050
0.90	1.0	-0.037	0.017	-11.9	-27.8	0.080	0.181	0.084	0.053
0.99	0.8	-0.036	0.035	-12.0	-32.9	0.166	0.341	0.113	0.056
0.99	1.0	-0.045	0.044	-16.0	-36.0	0.240	0.436	0.154	0.057

Note: Panel A contains the simulation output when both x_{1t} and x_{2t} do not exhibit a trend, Panel B when only x_{2t} exhibits a trend and Panel C when both sets of regressors contain a trend. Coefficient and standard error bias for predictive regression (1) with x_{1t} and x_{2t} generated by equations (2) and (3) and $Corr(\varepsilon_{1t}, u_{t+1}) = \delta$. Coefficient bias reported as $\mathbb{E}(b_i) - \beta_i$. Standard errors bias as $\mathbb{E}[(\hat{\sigma}_{b_2} - \sigma_{b_2})/\sigma_{b_2}]$. The size of the tests is for a standard t-test represented by OLS, t-test with HAC standard errors, and a Bonferroni test proposed by Campbell and Yogo (2006) to deal with Stambaugh bias, providing an approximation of the finite sample distribution when the predictor is persistent, and for the bootstrap test.

or exogenous ($\delta = 0$) the sizes of all tests except the HAC-adjusted t-test are approximately equal to the nominal size of 5%. The higher the persistence or endogeneity, the larger the size distortions. For example, when δ is held constant at 1.0 but we increase ρ from 0.90 to 0.99, we observe that the size of the conventional OLS and asymptotic t-tests increases from 8.5% to 15.1%. While this is even worse for the HAC adjusted t-test, the size of the Bootstrap t-test increases only from 5.8% to 8.0%. Panel B and C of Table 1 displays the output when only x_{2t} contains a trend (Panel B) and when both regressors contain a trend (Panel C). It shows that the commonly used t-tests are even more distorted when the regressors are trending. In this case, when the regressors are highly persistent with $\rho = 0.99$ and endogenous $\delta = 1.0$, then the OLS t-test has a size of 27% and 24% when only x_{2t} exhibits a trend and when both regressors are trending respectively. If the latter holds true, it is not only the standard error bias but also the coefficient bias in the estimator of β_2 that leads to more poorly sized tests.

Figures 1 graph the distributions of the coefficient estimates in equation (1) when the regressors are highly persistent ($\rho = 0.99$) and for different levels of endogeneity ($\delta \in \{0, 1\}$). The OLS estimators of the coefficients in (1) under standard regression settings are shown by the blue lines. The coefficient bias in the estimator of β_1 is depicted by the red lines, where the left-shifted line compared to the blue lines displays the negative bias. The magenta lines show the downward bias occurring for the standard errors that result in a wider interval for the estimator of β_2. Comparing these three figures shows that the distributions are even more distinct when the regressors exhibit a trend. Overall, these figures demonstrate that conventional inference that assumes the blue lines as coefficient distributions is likely to lead to wrong conclusions and even more so when the regressors are trending.

In conclusion, the problems that arise in testing $\beta_2 = 0$ when assuming standard regression settings while the regressors are highly persistent and not strictly exogenous are caused by the standard error bias. The conventional OLS standard errors, but also the HAC standard errors, underestimate the true sampling variance of the OLS estimates. This results in small-sample problems and leads to poorly sized tests and spurious rejections. However, the bootstrap inference overcomes these problems resulting in a better size. Additionally, Appendix A contains a derivation on how the combination of the persistence and lack of exogeneity lead the Wald statistic to not converge to the desired χ^2 distribution.

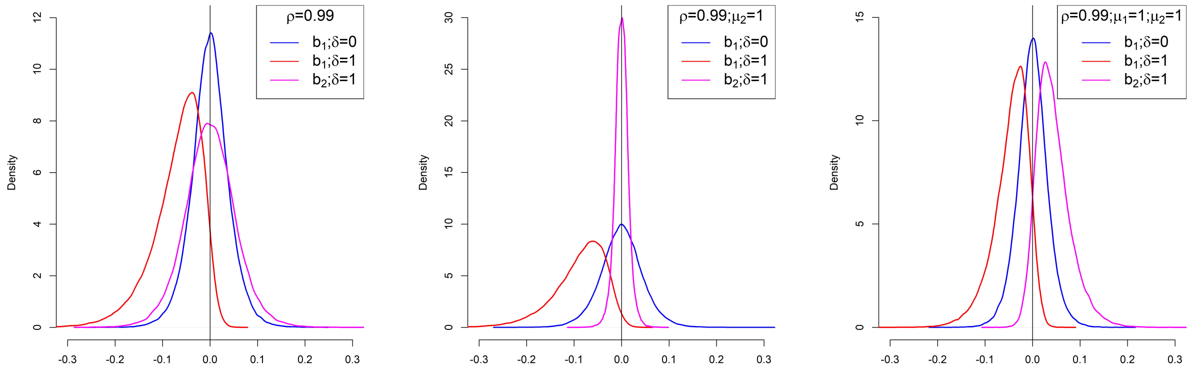

Figure 1: Simulation of the distributions of the coefficients in predictive regression (1). From left to right figures corresponds to Panel A to C of Table 1.

2.2 General Framework

Similar to many empirical studies in economics and finance, we investigate regressions of the form

$$y_{t+h} = \beta_0 + \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+h}, \tag{4}$$

where y_{t+h} reflects an h-period ahead excess return, x_{1t} is a persistent and stochastic variable related to the return at the end of period t, x_{2t} is an additional persistent regressor, and u_{t+h} is a forecast error. The null hypothesis of interest $H_0 : \beta_2 = 0$ states that the set of predictors that are contained in x_{2t} exhibit no additional explaining power over the variables in x_{1t}. Similar to the motivating example, we often observe that $Corr(x_{1t}, u_{t+h}) = \delta \neq 0$ and that the regressors x_{1t} and x_{2t} are highly persistent such that they can be simulated by a VAR(1) process with persistence parameter ρ_i close to unity. Therefore standard regression assumptions typically fail to hold resulting in poorly sized inference.

Studies of the fixed-income markets frequently include regressions as in (4), where y_{t+12} is the annual excess return on a bond, x_{1t} are the first three yield principal components ($(PC1, PC2, PC3)'$) and x_{2t} are macroeconomic variables such as factors from a large macroeconomic data set ($(F1, F2, ..., F8)'$) (Ludvigson & Ng, 2009), economic growth (GRO) and inflation (INF) (Joslin et al., 2014), trend inflation (τ) (Cieslak & Povala, 2015) and higher-order principal components of yields ($(PC4, PC5)'$) (Cochrane & Piazzesi, 2005). We focus on factors extracted from a large macroeconomic data set as explanatory variables, but Appendix E contains a discussion of the importance of the other variables when using the bootstrap inference which replicates the findings by Bauer and Hamilton (2018). The factors are estimated by the method of principal components analysis (PCA). This approach is typically used to reduce the dimensionality of the data by finding linear combinations of a variable that are uncorrelated and have maximum variance. Such a predictive regression is a special case of what is known as 'factor augmented regression' (FAR).

Equation (4) also appears in the study of the currency market, where y_{t+1} is the one-month currency excess return on a carry-trade strategy, x_{1t} is the average forward discount on currencies of developed countries (AFD) (Lustig et al., 2014) and x_{2t} are macroeconomic variables such as factors that summarize a large macroeconomic data set about the US economy ($H1, H2, ..., H9$) and factors that summarize a large macroeconomic data set about the global economy ($G1, G2, G3$) (Filippou & Taylor, 2017).

Table 2 reports the level of endogeneity (δ) and persistence (ρ) in the regressors over the sample originally examined by the study referenced in the first column. This generally shows that the variables contained in x_{1t} are not strictly exogenous, and both x_{1t} and x_{2t} are highly persistent. However, we notice that the endogeneity in the currency market illustration, shown in Panel B, is considerably smaller (0.12) than in the bond market (between 0.33 and 0.42). These problematic features are the reason to investigate these studies. We expect conventional inference to be distorted, such that it can potentially lead to invalid conclusions.

Table 2: Persistence and Endogeneity

Panel A. Bond Excess Return		δ	ρ
Ludvigson and Ng (2009)	$x_{1t} = (PC1_t, PC2_t, PC3_t)'$	0.37	0.98
	$x_{2t} = (F1_t, F2_t, .., F8_t)'$		0.91
Joslin et al. (2014)	$x_{1t} = (PC1_t, PC2_t, PC3_t)'$	0.42	0.98
	$x_{2t} = (GRO_t, INF_t)'$		0.98
Cieslak and Povala (2015)	$x_{1t} = (PC1_t, PC2_t, PC3_t)'$	0.33	0.99
	$x_{2t} = \tau_t$		0.99
Cochrane and Piazzesi (2005)	$x_{1t} = (PC1_t, PC2_t, PC3_t)'$	0.37	0.98
	$x_{2t} = (PC4_t, PC5_t)'$		0.38
Panel B. Currency Excess Return			
Filippou and Taylor (2017)	$x_{1t} = AFD_t$	0.12	0.92
	$x_{2t} = (H1_t, H2_t, ..., H9_t, G1_t, .., G3_t)'$		0.99

Note: Panel A contains an overview of the studies discussed in the bond market and Panel B of the study in the currency market, where x_{1t} and x_{2t} refer to the regressor in Equation (4). δ refers to the level of endogeneity ($Corr(x_{1t}, u_{t+h})$), and ρ to the persistence of each (vector) of regressor(s). Panel B only reports the endogeneity and persistence estimated for the currencies of developed countries, since it leads to equivalent results when estimated for all currencies. All levels of persistence and endogeneity are computed over the sample as originally studied.

2.3 Robust Inference

To obtain robust inference, the parametric bootstrap procedure proposed by Bauer and Hamilton (2018) is used. This procedure calculates 5,000 artificial data samples under the null hypothesis of x_{2t} containing no additional explanatory power and where the serial correlation of the regressors is similar to that of the actual data. The bootstrap test is then carried out by calculating the bootstrap p-value as the fraction of the sample in which the corresponding t-statistic exceeds the test statistic in the data. Appendix B displays the steps taken in the bootstrapping procedure as proposed by Bauer and Hamilton (2018). Additionally, it shows how we modify the procedure to be more generally applicable to other financial markets. This is necessary as the bootstrap procedure introduced in the bond market uses the estimated principal components ($PC1, PC2, PC3$) to estimate the yield variable. The excess return is then generated from this yield variable only. The excess return on a carry-trade strategy in the currency market and the average forward discount do not have a similar direct relation. Therefore, we modify the bootstrap procedure to bootstrap the excess return using the estimated coefficients of Equation (4) in the data by imposing the null hypothesis $\beta_2 = 0$.

Further, we investigate predictors by confronting them with new data that has come available since the publication by Bauer and Hamilton (2018) in the bond market illustration and by Filippou and Taylor (2017) in the currency market illustration. In this way, we reestimate the models over a sample period with new data and further evaluate the true out-of-sample forecasting performances of each proposed model.

3 Data

This Section provides an overview of the dataset used to replicate and expand upon previous research. We extend the data used in the bond market illustration to December 2022. We obtain Fama Bliss data on bonds from two to five years of maturity via WRDS, and we use the FRED-MD database as the large macroeconomic data set from which we extract eight common factors via PCA. For more details on the data used in the bond market illustration, we refer to the paper by Bauer and Hamilton (2018). In this Section, we focus on data in the currency market from the perspective of a US investor. Table 8 in Appendix C summarizes the data sources in the bond and currency market studies.

3.1 Currency Excess Returns

To define the currency excess returns, also referred to as currency risk premium, we denote the log of the spot exchange rate as s and the log of the forward exchange rate as f. Both are expressed in units of foreign currency per US dollar, such that a rise in s implies an appreciation of the US dollar. The log excess return (rx_{t+1}) on buying a foreign currency in the forward market and then selling it in the spot market after one month is defined as

$$rx_{t+1} = i_t^* - i_t - (s_{t+1} - s_t) = f_t - s_t - (s_{t+1} - s_t) = f_t - s_{t+1}. \tag{5}$$

The excess return consists of two parts: the forward discount ($f_t - s_t$) and the change in the spot rate ($s_{t+1} - s_t$). In addition, under the covered interest-rate parity (CIP) condition, the forward discount must be equal to the interest rate differential: $f_t - s_t \approx i_t^* - i_t$, where i_t^* and i denote the foreign and domestic nominal risk-free rates over the maturity of the forward contract. Thus, under the assumption that the CIP holds, excess returns are equal to the interest-rate differential corrected for the depreciation rate. Akram, Rime and Sarno (2008) study the deviations from the CIP and show that any deviations arising are only short-lived. Yet, this condition was frequently violated for some currencies during the financial crisis in 2008 and even after the onset of the crisis (Du, Tepper & Verdelhan, 2018). However, Lustig et al. (2014) find that including or excluding these observations does not have a major effect on their results.

In addition, the bid and ask quotes are used to adjust the long and short positions to be closer to realized excess return. The net log excess return for an investor who goes long in foreign currency is $rx_{t+1}^l = f_t^b - s_{t+1}^a$. The investor buys the foreign currency at the bid price (f_t^b) at time t and sells the foreign currency in the spot market for the ask price (s_{t+1}^a) at time $t+1$. The net log currency excess return for an investor who is short in the foreign currency is computed as $rx_{t+1}^s = -f_t^a + s_{t+1}^b$. In the regression-based analysis, we consider only net currency excess returns following Filippou and Taylor (2017).

3.2 Currency Data

The data to establish the currency risk premia comprises daily exchange rates quoted in US dollars, encompassing both spot and one-month forward markets. The data span from July 1985 to February 2023 and were sourced from Datastream by WM/Refinitiv and Barclays Bank International (BBI). To compute the logarithmic excess return for each currency, we construct monthly series by selecting end-of-the-month rates.[4] We construct two currency baskets ($j \in \{Dev, All\}$). The first consists of the currencies of developed countries (Dev), which includes a subset of 15 currencies. From the introduction of the euro in January 1999, the euro area currencies are excluded from the sample, resulting in a more

[4] Filippou and Taylor (2017) reports that the results for logarithmic returns are very close to those they presented for raw returns. However, it is common in the literature to use logarithm returns.

limited selection of the G10 currencies (Lustig et al., 2011; Filippou & Taylor, 2017). The second basket contains all currencies in the sample ($j = All$). Consistent with the literature, we eliminate observations that exhibit significant deviations from the CIP. Section C.1 of the Appendix contains a detailed list of the currencies in our sample. Moreover, we focus on the developed countries since certain currencies are pegged and are difficult to trade. In addition, Filippou and Taylor (2017) states that these currencies are most actively traded in the foreign exchange market.

3.3 Currency Portfolios

To analyze the payoff for a US investor investing in the foreign exchange market, we construct portfolios. At the end of each period t, we categorize the currencies in the sample into five portfolios for the developed countries and six portfolios for all countries following a vast stream of literature started by Lustig et al. (2011). The currencies in the sample are sorted based on their forward discounts, $f_t^i - s_t^i$ for all currencies i in currency basket j observed at the end of the month t. Portfolios are reconstructed at the end of every month. They are ranked from low to high forward discounts, the first portfolio with the lowest forward discount currencies and the fifth (sixth) with the highest forward discount currencies for the developed countries (all countries). Given the limited number of countries, especially at the sample's start, we do not want more portfolios. Using fewer portfolios would result in currencies with high forward discounts being mixed with others. We compute the excess return of each portfolio by averaging the excess returns of each currency within the portfolio.

Table 3: Summary Statistics of Portfolio Performance

Portfolio	1	2	3	4	5	1	2	3	4	5	6
	I. Developed Countries					II. All Countries					
	Spot change: Δs^k					Spot change: Δs^k					
Mean	-0.14	-0.08	-0.11	0.03	0.06	-0.08	-0.01	-0.10	-0.06	0.11	0.38
Std	2.55	2.60	2.58	2.69	3.30	2.04	1.85	2.03	2.27	2.45	2.86
	Forward discount: $f^k - s^k$					Forward discount: $f^k - s^k$					
Mean	-0.18	-0.03	0.05	0.16	0.50	-0.22	-0.05	0.05	0.16	0.37	1.01
Std	0.19	0.15	0.17	0.20	0.71	0.17	0.11	0.11	0.13	0.17	0.61
	Excess return: rx^k					Excess return: rx^k					
Mean	-0.04	0.04	0.17	0.13	0.44	-0.14	-0.04	0.15	0.23	0.26	0.62
Std	2.55	2.62	2.59	2.70	3.40	2.05	1.86	2.03	2.27	2.45	2.85
SR	-0.02	0.02	0.06	0.05	0.13	-0.07	-0.02	0.07	0.10	0.10	0.22
	Net excess return: rx_{net}^k					Net excess return: rx_{net}^k					
Mean	0.03	-0.04	0.09	0.03	0.32	-0.03	-0.13	0.04	0.09	0.07	0.35
Std	2.55	2.61	2.59	2.70	3.39	2.05	1.86	2.03	2.27	2.45	2.79
SR	0.01	-0.01	0.03	0.01	0.09	-0.01	-0.07	0.02	0.04	0.03	0.13
	High-minus-Low: $rx^k - rx^1$					High-minus-Low: $rx^k - rx^1$					
Mean		0.09	0.21	0.17	0.49		0.10	0.29	0.37	0.40	0.76
Std		1.86	1.99	2.49	3.29		1.28	1.44	1.79	1.90	2.77
SR		0.05	0.11	0.07	0.15		0.08	0.20	0.20	0.21	0.28
	High-minus-Low: $rx_{net}^k - rx_{net}^1$					High-minus-Low: $rx_{net}^k - rx_{net}^1$					
Mean		-0.07	0.05	-0.00	0.28		-0.10	0.07	0.12	0.10	0.38
Std		1.86	1.99	2.49	3.29		1.27	1.42	1.76	1.89	2.70
SR		-0.04	0.03	-0.00	0.09		-0.08	0.05	0.07	0.05	0.14

Note: The table presents summary statistics (in percentage points) of monthly portfolio performance for different portfolios in both developed countries (I. Developed Countries) and all countries (II. All Countries) over the period July 1985 - February 2023. The portfolios are evaluated based on spot change (Δs^k), forward discount ($f^k - s^k$), excess return (rx^k), net excess return (rx_{net}^k), the payoff of the high-minus-low strategy ($rx^k - rx^1$) and net of transaction cost ($rx_{net}^k - rx_{net}^1$). Mean, standard deviation (Std), and Sharpe ratio (SR) for a US investor are provided for each portfolio.

Table 3 provides an overview of the properties of the five and six currency portfolios from a US investor's perspective for the developed and all countries currency basket respectively. We display log returns because these are the sum of the forward discount and the change in spot rates. For each portfolio k, where $k = 1, .., 5$ for $j = Dev$ representing the currency basket on the developed countries and $k = 1, .., 6$ for $j = All$ representing the currency basket on all the countries, we report the average change in spot rate (Δs^k) and the average forward discount ($f^k - s^k$). We also report log currency excess returns on carry trades or high-minus-low investment strategies that go long in each portfolio except the first, and short in the first portfolio: $rx^k - rx^1$. Besides, we report returns net of transaction costs. All returns are documented in US dollars.

Following the CIP, forward discounts must equal the interest rate differential. In addition, based on the UIRP, rational investors would expect a foreign currency to depreciate or appreciate if the interest rate differential and thus the forward discount changes. This leads to the standard UIRP condition in which there is no risk premium: $E_T[\Delta s^k] = E_T[f^k - s^k]$ (Lustig et al., 2011). However, Table 3 shows that this does not hold. As one would expect from the empirical literature on UIRP, US investors earn on average negative excess returns on low forward rate currencies of minus 4 basis points for the currency basket of developed countries and large, positive returns on high forward discount rate currencies of approximately 44 basis points for the same currency basket.[5] The magnitude of these returns is particularly caused by the forward discount, which is 50 basis points for the fifth portfolio. These returns are even large when taking into account transaction costs and when measured per unit of risk. The Sharpe Ratio, defined as the ratio of average excess return to its standard deviation, on the fifth portfolio is 0.13 and the largest among its competitor portfolios. The last panel report returns on zero-cost strategies that go long in the high-interest rate portfolio and short in the low-interest rate portfolio. The UIRP hypothesized that the carry gain due to the interest rate differential is canceled by a depreciation of the high-interest-rate currency, empirically Table 3 shows that the reverse holds true. To exemplify, the spread between the net returns on the first and last portfolio is 49 basis points. We also report standard errors on the average returns. The average returns are statistically significantly different from zero. Next, we consider the second panel in Table 3 including all currencies. Likewise, the ranked portfolios lead to a higher average excess return, even when accounted for transaction costs. In addition, similar to considering only the developed countries the high-minus-low strategy benefits from the spread between high and low interest-rate currencies resulting in high Sharpe ratios.

Overall, high-yield currencies tend to appreciate, which implies predictable currency excess returns and potential trading profitability. Figure 4 in Appendix C.5 shows the cumulative return on the carry-trade strategy compared to an equally-weighted portfolio in the currency market and bond market. Although the presence of both upward and downward trends indicates that the strategy carries risks, betting against the UIRP is, on average, more profitable than an equally-weighted portfolio in both the currency and bond market. Following Filippou and Taylor (2017), we focus in particular on predicting the pay-off from this strategy. Therefore the dependent variable in the currency market application taking into account transaction costs is defined as:

$$y_{t+1} = rx^{K_j}_{net,t+1} - rx^1_{net,t+1}, \tag{6}$$

where $K_j = 5$ for $j = Dev$ and $K_j = 6$ for $j = All$. This is a long position in the highest-interest-rate currencies ($rx^{K_j}_{net}$), the investment currencies, and a short position in lowest-interest-rate currencies (rx^1_{net}), the funding currencies.

[5]This is for example shown by Lustig et al. (2011) specifically in Table 1.

3.4 Average Forward Discount

The average forward discount (AFD) on the foreign currency basket $j \in \{Dev, All\}$ against the US dollar is $AFD_t^j = \bar{f}_t^j - \bar{s}_t^j = (1/N_t^j)\sum_{i=1}^{N_t^j} f_t^i - s_t^i$, where N_t^j denotes the number of currencies in basket j at time t. By construction, the AFDs are negatively correlated with the US short-term interest rate. Figure 2 graphs the AFDs for the different currency baskets over the sample period from July 1985 to February 2023. The AFDs are almost duplicated in the first decennium but separate dramatically as the interest rates of non-developed countries shoot up. This is especially evident after the start of the financial crisis in 2008. This discrepancy indicates that these series do not contain the same predictive power. This is also shown by Lustig et al. (2014) but holds even more so after the sample period they study.[6] In addition, the AFD of the developed countries is more persistent with a first-order autocorrelation coefficient of 0.92, whereas the persistence of the AFD of all countries is only 0.80. Similar to Filippou and Taylor (2017), we follow Lustig et al. (2014) by using the AFD on the currencies of developed countries.

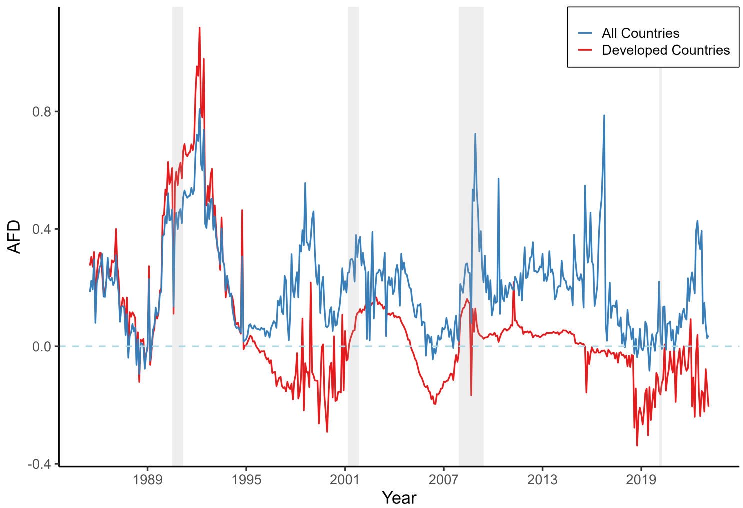

Figure 2: Average forward discounts on two currency baskets (All Countries and Developed Countries). The shaded areas are US recessions according to NBER. The sample period is 1985:7-2023:2.

3.5 Macroeconomic Data

Following Filippou and Taylor (2017), we extract factors by PCA from a domestic data set of macroeconomic and financial series for the US economy and a global data set that compromises mainly the G10 countries. The data is sourced from Datastream and spans the period from July 1985 to February 2023. Therefore, we extend the data set used by Filippou and Taylor (2017) which use data until March 2012. The domestic data set contains a panel of 125 monthly series and the global data set consists of 96 series.[7] Appendices C.2 and C.3 offer a detailed description of the data, which we transform and standardize accordingly. In particular, similar to Filippou and Taylor (2017) we use nine factors of the domestic data ($H1, H2, ..., H9$) and three factors on the global data ($G1, G2, G3$). Appendix C.4 shows that the first nine domestic factors capture 73% of the total variation in the US data, while the three global factors capture 25% of the variation in the global data.

[6] Lustig et al. (2014) study the sample period 1983:11-2010:6. They use the average 12-month forward discount and add a currency basket containing currencies of emerging countries. Therefore our figure deviates from Figure 1 by Lustig et al. (2014), however, conclusions remain the same.

[7] Filippou and Taylor (2017) use a panel of 127 US variables and 97 global variables, however, for both panels we were unable to find all series on Datastream.

4 Results

4.1 Bond Risk Premia

Table 4 and 5 replicate the regression output by Bauer and Hamilton (2018) and we extend the later sample shown in panel B to end in December 2022 instead of December 2016.[8] The model specification is very similar to Ludvigson and Ng (2009), however, following Bauer and Hamilton (2018) the yield curve is captured by the first three PCs instead of the CP factor constructed by Cochrane and Piazzesi (2005). In addition, Ludvigson and Ng (2009) search over different lag orders to obtain the combination of factors that minimize the BIC-criterion in- and out-of-sample. We focus, instead, only on one specification including all factors, without lags, following the specification of Bauer and Hamilton (2018). Also, the dependent variable is the excess return averaged over the bond with different maturities, instead of investigating each maturity separately. Since we estimate annual returns using monthly data, we have overlapping observations. The common approach to address the resulting serial correlation in the residuals is to use standard errors and the test statistic proposed by Newey and West, referred to as HAC. In regressions for annual returns with monthly data researchers typically use 18 lags (Cochrane & Piazzesi, 2005; Ludvigson & Ng, 2009).

Table 4 displays the estimation output, including the coefficient estimates, the p-values, size, and power of the HAC-based t- and Wald test and the bootstrap equivalents. The estimation output is similar to the results by Bauer and Hamilton (2018). Likewise, we find that mainly the Wald test is distorted with a size of 32.2%, whereas the bootstrap Wald test obtains a size of 5.1% which is considerably closer to the nominal size of 5%. Profiting from the better-sized test, we conclude that the macro factors are indeed jointly significant, as well as the eighth macro factor. We attempt to provide some economic intuition behind the latent factor. Since the factors are unobserved, we cannot link them directly to any macroeconomic variables, but some factors load heavily on particular macroeconomic or financial variables. To help us identify the factors, Appendix D graphs the marginal R^2 from regressing each of the economic series onto the factor.[9] Figure 5 shows that the eighth factor ($F8$) loads heavily on the series from the stock market category. Therefore we suggest that in particular variables on the stock market lead to a violation of the spanning hypothesis over the sample period from 1964 to 2008.

We are especially interested in what happens when we extend the data set used by Bauer and Hamilton (2018) shown in Panel B of Table 4. We find that it is important to use the bootstrap inference since the macro factors are jointly significant at a 1% level by the conventional Wald test, whereas they are only jointly significant at a 10% level by the bootstrap Wald test. However, these p-values are considerably lower than Bauer and Hamilton (2018) documents (0.004 and 0.258 resp.) when using the later sample of 1985 until 2016. Therefore we conclude that the macro factors are especially important in the last six years. In addition, in this later sample, we find that only the first macro factor ($F1$) is statistically significant at a 5% level by the bootstrap t-test. Figure 6 in Appendix E shows that the first factor ($F1$) loads heavily on series that measure output and income, the labor market, and housing. Due to the significance of this macro factor in predicting bond risk premia, we conclude that especially these macroeconomic categories are important determinants of risk premia from 1985 until 2023.

Table 5 displays the goodness-of-fit, for a model including only a constant and the yield factors R_1^2, a model including all regressors (R_2^2), and the difference in fit due to the addition of the macro factors ($R_2^2 - R_1^2$). In addition to these adjusted R^2 as observed in the data, we include the average adjusted R^2 and the 95% confidence interval under the bootstrap procedure where it is assumed that the macro factors do not contain explanatory power. Although we find that the adjusted R^2 increases due to the macro factors, this increase does not provide convincing evidence as they are included in the 95% confidence

[8] They relate to Table 5 of Bauer and Hamilton (2018).

[9] Ludvigson and Ng (2009) follow the same procedure. For the numbering of the variables, we refer to the FRED-MD database.

Table 4: Bond Return Predictability: Factors of Large Macro Data Sets

	PC1	PC2	PC3	F1	F2	F3	F4	F5	F6	F7	F8	Wald
A. Early sample, 1964:1-2007:12												
Coefficient	0.136	2.052	-5.010	0.742	0.147	0.072	-0.528	-0.321	0.576	0.401	0.551	
p-value												
HAC	0.121	**0.010**	**0.007**	0.064	0.704	0.544	0.056	0.192	**0.027**	**0.019**	**0.003**	**0.000**
Bootstrap				0.152	0.753	0.588	0.143	0.284	0.089	0.065	**0.009**	**0.006**
Size (HAC)				0.130	0.121	0.091	0.132	0.119	0.135	0.119	0.090	0.323
Bootstrap				0.056	0.052	0.050	0.056	0.057	0.056	0.065	0.046	0.051
Power (HAC)				0.612	0.144	0.121	0.577	0.334	0.550	0.646	0.849	0.997
Bootstrap				0.430	0.068	0.071	0.411	0.217	0.375	0.504	0.782	0.952
B. Later sample, 1985:1-2022:12												
Coefficient	0.264	0.457	-1.335	0.442	-0.479	-1.122	0.261	0.089	0.146	-0.054	-0.051	
p-value												
HAC	**0.000**	0.594	0.661	**0.003**	**0.006**	**0.004**	0.262	0.295	0.680	0.821	0.736	**0.000**
Bootstrap				**0.030**	0.065	0.067	0.386	0.373	0.759	0.860	0.780	0.078

Note: Predictive regression for annual bond excess returns (y_{t+12}), averaged over two- to five-year bond maturities. $y_{t+12} = \beta_0 + \beta_1^{'} x_{1t} + \beta_2^{'} x_{2t} + u_{t+12}$, where $x_{1t} = (PC1_t, PC2_t, PC3_t)^{'}$ and $x_{2t} = (F1_t, F2_t, ..., F8_t)^{'}$. Results in panel A are for the same period as studied by Ludvigson and Ng (2009), and panel B includes the period 1985-2022. HAC statistic and p-values are calculated using Newey-West standard errors with 18 lags. The column *Wald* reports p-values for the hypothesis that $F_1, F_2, ..., F_8$ have no predictive power. p-values below 5% are in bold. Bootstrap indicates the Bootstrap procedure presented in Appendix B where we obtain bootstrap samples under $H_0 : \beta_2 = 0$.

Under *Size*, we report estimates of the size of the tests, based on simulations from the simple bootstrap under the null hypothesis. Both tests have a nominal size of 5%.

Under *Power*, we report estimates of the power of the tests, based on simulations from the simple bootstrap under the alternative hypothesis.

intervals.

It is well known that variables that are found to have predictive power in-sample do not necessarily have predictive power out-of-sample. Therefore, we evaluate the true out-of-sample performance by considering the period from January 2007 to December 2021 as an out-of-sample period. The model is reestimated every month during this period using an expanding window approach. Panel B of Table 5 shows the $RMSE$-ratio of a model including the macro factors and without. The unrestricted model reduces the $RMSE$ by 15%, yet it increases the $RMSE$ by 130% in the out-of-sample period. However, we find that this increase in the prediction errors in the out-of-sample period is not statistically significant by the Diebold-Mariano (DM) test. The left panel of Figure 3 shows the forecasts of the restricted and unrestricted model compared to the observed values. It shows a high peak in the forecast occurring shortly after the start of 2020 resulting in a substantial forecasting error which is penalized heavily by the $RMSE$. It can be concluded that the large forecast of the unrestricted model is particularly due to the spike in macroeconomic factors. This causes the unrestricted model to perform particularly badly in the out-of-sample period.

Table 5: Bond Return Predictability: Goodness-of-fit

A. In-sample performance

	R_1^2	R_2^2	$R_2^2 - R_1^2$
I. Early sample, 1964:1-2007:12			
Data	0.25	0.35	0.10
Bootstrap	0.21	0.24	0.03
	(0.06, 0.40)	(0.09, 0.43)	(0.00, 0.11)
II. Later sample, 1985:1-2022:12			
Data	0.16	0.25	0.09
Bootstrap	0.32	0.36	0.04
	(0.12, 0.52)	(0.16, 0.55)	(0.00, 0.13)

B. Out-of-sample performance

	In-sample	Out-of-sample	DM-test p-value
RMSE-ratio	0.846	2.295	0.404

Note: Adjusted R^2 of predictive regression for annual bond excess returns (y_{t+12}), average over two- to five-year bond maturities, using three PCs ($x_{1t} = (PC1_t, PC2_t, PC3_t)'$) and factors extracted from a large macro data set ($x_{2t} = (F1_t, F2_t, ..., F8_t)'$). Results in panel A.I are for the same period as studied by Ludvigson and Ng (2009), and panel A.II includes the period 1985-2022. The first row of panels A.I and A.II reports the statistics in the original data. The following reports bootstrap mean and 95% percentiles in parentheses. The bootstrap procedure assumes that the macro factors contain no incremental predictive power. The first column corresponds to the predictive regression that includes x_{1t} and the second column to the inclusion of x_{1t} and x_{2t}. Panel B assesses the predictive power by the $RMSE$. The $RMSE$-ratio is the relative $RMSE$ of a model including x_{1t} and x_{2t} and a model without x_{2t}. The in-sample period is 1964:1-2006:12, and the out-of-sample period is 2007:1-2021:12. The out-of-sample predictions are estimated using an expanding window approach. The Diebold-Mariano (DM) test for equal forecasting accuracy considering the prediction errors out-of-sample for the unrestricted and restricted model.

In conclusion, the results show that conventional measures of fit and hypothesis tests are not reliable for assessing the spanning hypothesis. However, the bootstrap tests have close to the correct size and therefore are more reliable. While this was already shown by Bauer and Hamilton (2018), we find that it distortion of the Wald test is not due to a particular subsample. The evidence that the macro factors have predictive power beyond yield factors is weaker than proposed in the original study of Ludvigson and Ng (2009). Compared to Bauer and Hamilton (2018), we document that the macro factors are more important in recent years. However, when considering a rolling regression, we find that the model including the macro factors results in a large average forecasting error resulting from a large spike in the forecast error around the start of 2020. Therefore, similarly to Bauer and Hamilton (2018), we conclude

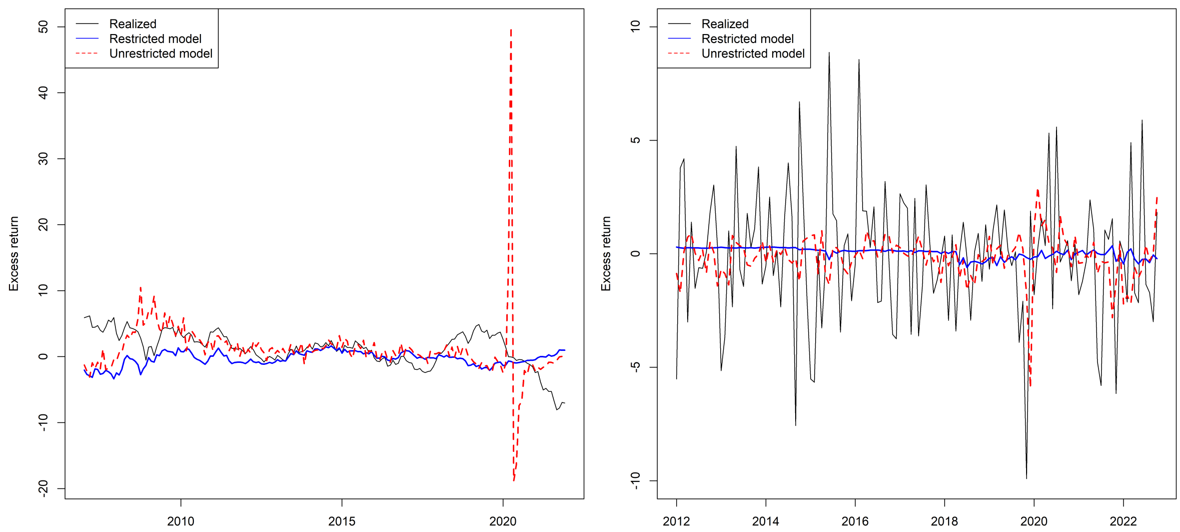

Figure 3: Out-of-sample prediction of average excess return on bonds (left) and of carry-trade excess return (right). The models are estimated each month using an expanding window. The restricted model uses only the yield information, whereas the unrestricted model extends this by including the macro factors.

that again econometric problems and subsample stability shows concerns about the importance of the macro factors beyond the yield factors as shown by Ludvigson and Ng (2009).

4.2 Currency Risk Premia

Next, we investigate the importance of macroeconomic information in predicting currency risk premia in a similar way to the bond risk premia of the previous Section. We estimate regression (4) where we regress the payoff of a carry-trade strategy defined in (6) on the AFD ($x_{1t} = AFD_t$) and the macro factors from US macroeconomic data set and from the global macroeconomic data set ($x_{2t} = (H1_t, H2_t, ..., H9_t, G1_t, .., G3_t)'$). We do this for the two currencies basket including only the currencies of the developed countries and of all countries in the panel. We discuss the estimation output for the currency basket using only the developed countries here, and Appendix F contains a detailed discussion of the estimation output for all the currencies. The model specification is very similar to Filippou and Taylor (2017). However, Filippou and Taylor (2017) search over different lag orders to obtain a model specification that minimizes the BIC and AIC-criterion. Similar to the previous bond market illustration, we focus only on one specification that includes all factors, without lags or nonlinear terms. Similar to Filippou and Taylor (2017), we use HAC standard errors with four lags following Villanueva (2007).

Table 6 contains the estimation output for currencies of the developed countries. First, we discuss the estimation results over the sample studied by Filippou and Taylor (2017) displayed in panel A. Similar to the bond market illustration, we find that the Wald test is considerably distorted with a size of 21.1% which deviates largely from the nominal size of 5%. The bootstrap Wald test, however, has a size of 5.3% that is much closer to the correct size. This is in contrast to the t-tests which test the individual significance of the macro factors. Overall, the commonly used t-tests have a size of approximately 6%, which is close to the proper size of 5%. Therefore, we conclude that the Wald test collects the small size distortions of the individual t-tests which results in a greater distortion.

Profiting from the better-sized tests, we find that the macro factors are not jointly significant. It is important to note that using the conventional Wald test would in fact lead to the conclusion that they are significant at a 5% level. Additionally, none of the macro factors are statistically significant at a 5% level by the bootstrap t-test. However, the fifth domestic factor is statistically significant at a 10% significance

Table 6: Carry-Trade Excess Return Predictability: Developed Countries

A. Early sample, 1985:7-2012:3

	AFD	H1	H2	H3	H4	H5	H6	H7	H8	H9	G1	G2	G3	Wald
Coefficient	2.557	0.051	0.140	-0.145	-0.060	-0.177	0.114	0.155	-0.057	-0.111	-0.216	-0.068	-0.135	
p-value														
HAC	**0.017**	0.466	0.450	0.288	0.572	**0.045**	0.249	0.289	0.686	0.408	0.444	0.650	0.365	**0.038**
Bootstrap		0.496	0.473	0.312	0.593	0.069	0.295	0.319	0.705	0.424	0.467	0.667	0.394	0.197
Size (HAC)		0.064	0.066	0.065	0.072	0.074	0.071	0.068	0.058	0.059	0.061	0.065	0.062	0.228
Bootstrap		0.049	0.048	0.047	0.054	0.059	0.044	0.052	0.043	0.048	0.053	0.045	0.044	0.053
Power (HAC)		0.152	0.128	0.304	0.114	0.348	0.195	0.227	0.083	0.117	0.153	0.103	0.192	0.889
Bootstrap		0.112	0.107	0.260	0.087	0.299	0.153	0.203	0.066	0.094	0.125	0.088	0.152	0.679
B. Later sample, 2012:3-2023:2														
Coefficient	1.394	0.020	-0.059	-0.045	-0.031	-0.067	-0.289	-0.192	-0.212	0.266	-0.020	-0.116	-0.069	
p-value														
HAC	0.726	0.800	0.597	0.565	0.843	0.540	0.235	0.330	0.133	0.080	0.904	0.589	0.701	**0.000**
Bootstrap		0.810	0.646	0.612	0.855	0.593	0.291	0.393	0.196	0.151	0.920	0.622	0.742	0.211
C. Complete sample, 1985:7-2023:2														
Coefficient	2.491	0.070	-0.200	-0.086	0.083	-0.024	-0.033	0.018	-0.308	-0.370	-0.358	-0.202	0.199	
p-value														
HAC	**0.011**	0.145	0.078	0.417	0.259	0.749	0.791	0.879	**0.040**	**0.047**	0.092	0.196	0.101	0.084
Bootstrap		0.168	0.095	0.434	0.289	0.761	0.805	0.886	0.051	0.056	0.103	0.216	0.120	0.270

Note: Predictive regression for currency excess returns of the carry-trade strategy using only the developed countries currency basket ($y_{t+1} = rx^5_{net,t+1} - rx^1_{net,t+1}$). $y_{t+1} = \beta_0 + \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+1}$, where $x_{1t} = AFD_t$ and $x_{2t} = (H1_t, H2_t, ..., H9_t, G1_t, G2_t, G3_t)'$. Results in panel A are for the same period as studied by Filippou and Taylor (2017), Panel B for the sample 2012:3-2023:2, and Panel C for the complete sample period. HAC statistic and p-values are calculated using Newey-West standard errors with 4 lags. The column *Wald* reports p-values for the hypothesis that macro factors have no predictive power. p-values below 5% are in bold. Bootstrap indicates the Bootstrap procedure presented in Appendix B where we obtain bootstrap samples under $H_0 : \beta_2 = 0$. Under *Size*, we report estimates of the size of the tests, based on simulations from the simple bootstrap under the null hypothesis. Both tests have a nominal size of 5%.
Under *Power*, we report estimates of the power of the tests, based on simulations from the simple bootstrap under the alternative hypothesis.

level. Similar to the previous Section in which we discussed the determinants of bond risk premia, we try to identify the macro factors in order to get a better idea about the driving forces of currency risk premia. Figure 10 in Appendix F.2 graphs the marginal R^2 of regressing each of the series following the numbering of Appendix C.2 on the fifth domestic factor ($H5$).[10] The fifth factor loads especially heavily on the series of money and credit quantity aggregates and employment. The relation between unemployment and currency excess returns is studied by inter alia Berg and Mark (2018) and Colacito et al. (2020). Both studies document the importance of the unemployment gap between countries in determining the currency excess returns. In a similar fashion, we find that the employment category is an important predictor of carry trade excess returns. However, in contrast to these studies and to Filippou and Taylor (2017), we find that the global factors are not significant when controlling for the information already captured by the average forward discount.

Second, we move to panel B of Table 6 which shows the estimation results over a sample period that includes data since the publication by Filippou and Taylor (2017), starting from March 2012 and ending in February 2023. Using the more robust bootstrap Wald test we find that the macro factors are not jointly significant. Again, it is important to note that using the conventional Wald test would not have led to this same conclusion. During this sample period, none of the macro factors are individually statistically significant.

Lastly, we discuss panel C of Table 6 which includes the complete sample starting from July 1985 until February 2023. For this sample period, the macro factors are jointly significant at a 10% level by the common Wald test, whereas the bootstrap Wald test concludes that the macro factors are jointly insignificant. This again stresses the importance of using the Wald test since using the common Wald test would lead to an invalid conclusion. Over the complete sample, the eighth and ninth domestic factors are statistically significant at a 10% significance level. Figure 11 and 12 show the marginal R^2 of regressing the economic series on $H8$ and $H9$ respectively. From these figures, we find that $H8$ is related to the categories orders, stock price, and money and credit quantity aggregates. $H9$ loads most heavily on series from the category stock price. This suggests that the stock price category is the most important determinant of carry trade risk premia when including only currencies from developed countries over the sample period from July 1985 to February 2023. Similarly, Brunnermeier et al. (2008) documents that the VIX, the implied volatility of the S&P500 helps resolve the forward premium puzzle. This measures market stress and liquidity events.

Table 7 shows the goodness-of-fit of the model estimated over each sub-sample period. Although these goodness-of-fit measures are small, they relate to the magnitude other studies find, see for example Lustig et al. (2014) and Filippou and Taylor (2017). In the sample studied by Filippou and Taylor (2017), the macro factors improve the fit of the model such that the adjusted R^2 increases with 2.03 percentage points to 4.42%. The macro factors increase the fit over the complete sample period (Panel III) by 1.10 percentage points to a value of 3.46%, yet we observe that the macro factors deteriorate the fit sharply when estimating the model over data that appeared since the publication by Filippou and Taylor (2017) (Panel II). In addition to these goodness-of-fit measures as observed in the data, we also report the average adjusted R^2 and the 95% confidence interval of the bootstrap simulation where it is assumed that the macro factors are not part of constructing the currency excess returns. Although the adjusted R^2s change due to the addition of the macro factors to the model, we observe that this is not uncommon to appear when the macro factors are unimportant in explaining the currency risk premia. This can be concluded from the fact that the adjusted R^2s are included in the 95% confidence intervals.

Lastly, we evaluate the true out-of-sample performance of the macro factors by considering a rolling window regression. In the in-sample period from July 1985 until March 2012, the $RMSE$ decreases by 6.5%. This is in contrast to the out-of-sample period where we reestimate the model each month. The $RMSE$ increases with 13.6% when the macro factors are added to the model and this increase is

[10] Filippou and Taylor (2017) follow a similar procedure.

statistically significant by the DM test at a 5% level. The right panel of Figure 3 graphs the predicted payoff from the carry-trade strategy by both the restricted and unrestricted model, without and with the macro factors. It displays that the two models have a poor fit, as also indicated by the low adjusted R^2, such that it remains difficult to favor either one of the models.

Table 7: Currency Excess Return Predictability: Goodness-of-fit

A. In-sample performance

	R_1^2	R_2^2	$R_2^2 - R_1^2$
I. 1985:7-2012:3			
Data	2.39	4.32	1.93
Bootstrap	2.47	2.48	0.01
	(-0.14, 7.03)	(-1.59, 7.96)	(-2.39, 3.53)
II. 2012:3-2023:2			
Data	-0.60	-5.61	-5.01
Bootstrap	0.18	0.26	0.07
	(-0.77, 3.84)	(-6.62, 9.42)	(-6.32, 8.88)
III. 1985:7-2023:2			
Data	2.36	3.47	1.12
Bootstrap	2.45	2.45	0.00
	(0.09, 6.22)	(-0.74, 6.82)	(-1.70, 2.57)

B. Out-of-sample performance

	In-sample	Out-of-sample	DM-test p-value
RMSE-ratio	0.935	1.134	**0.022**

Note: Adjusted R^2 (in %) of predictive regression for currency carry-trade excess return (y_{t+1}) using the average forward discount ($x_{1t} = AFD_t$) and factors extracted from large US and global macroeconomic data sets ($x_{2t} = (H1_t, H2_t, .., H9_t, G1_t, G2_t, G3_t)'$). Results in Panel A.I are for the same period as studied by Filippou and Taylor (2017). Panel A.II uses newer data and Panel A.III contains all data available. The first row of each panel reports the statistics in the original data, followed by the bootstrap mean and 95% percentiles in parentheses. The bootstrap procedure assumes that the macro factors do not contain incremental predictive power. The first column corresponds to the predictive regression only including the average forward discount, the second column includes all regressors, and the third column is the difference in adjusted R^2 between these models. Panel B assesses the predictive power by the $RMSE$. The $RMSE$-ratio is the relative $RMSE$ of a model including all regressors and a model without the macro factors. The in-sample period is 1985:7-2012:3, and the out-of-sample period is 2012:4-2023:1. The out-of-sample predictions are estimated using an expanding window approach. The Diebold-Mariano (DM) test for equal forecasting accuracy considering the prediction errors out-of-sample for the unrestricted and restricted model.

To conclude, similar to the bond market illustration we find that conventional measures of fit and inference are not reliable for assessing the importance of macroeconomic information controlling for the yield information captured by the AFD. In the currency market illustration, we again find that the bootstrap tests have sizes closer to the nominal value of 5%. In fact, it is mainly the Wald test that is distorted since it accumulates the small-sample problems of the t-tests. Moreover, Table 2 shows that the endogeneity in the currency market is considerably smaller than in the bond market studies. Therefore we conclude that it is primarily the high persistence that distorts the Wald test. Also, when confronting the model with new data since the publication by Filippou and Taylor (2017), we find that the macro factors are jointly even less important. Therefore we conclude that both econometric problems and subsample stability raises concerns about the joint importance of the macro factors while controlling for the yield information.

5 Conclusion

When a regressor is correlated with the regression disturbance and is highly persistent, the OLS estimator can display finite-sample properties that differ substantially from those in the standard regression settings. Whereas formerly the stress was on the coefficient bias, we show that the standard error bias likewise changes the distribution of the OLS estimator. This negative bias shows a wider small-sample distribution which is unlike adopted by generally used t- and Wald tests. Using more robust inference to address these problematic features, the main aim of this paper is to investigate the importance of factors derived from large macroeconomic data sets over yield variables in predicting risk premia. In particular, we study this to improve our understanding of violations of the spanning hypothesis in the bond market and of the forward premium puzzle in the currency market.

First, by extending the data set used by Bauer and Hamilton (2018) we find that the distortion of the Wald test is not due to a particular sub-sample. In addition, similar to Bauer and Hamilton (2018) we document that the evidence to reject the spanning hypothesis by the macro factors is less convincing than shown in the original study of Ludvigson and Ng (2009). However, these factors appear as more significant in recent years, since the publication by Bauer and Hamilton (2018). When considering a rolling regression, we find that the model including the macro factors results in large forecasting errors. Therefore, similar to Bauer and Hamilton (2018), we conclude that again econometric problems and sub-sample stability raises concerns about the robustness of the results presented by Ludvigson and Ng (2009).

Second, in a similar fashion to Bauer and Hamilton (2018), we reinvestigate the study of Filippou and Taylor (2017). We find, similar to the bond market, that conventional measures of fit and hypothesis tests are unreliable for assessing the importance of macro factors in forecasting the excess return from a carry-trade strategy when controlling for the AFD. When controlling for the AFD, we find that the macro factors from the US, as well as the global economy, are not jointly significant. Although we repeatedly observe that the adjusted R^2 increases due to the addition of the macro factors to the model, we uncover by the bootstrap simulation that this increase is not abnormal when the macro factors do not contain predictive power. Moreover, we find that the prediction errors significantly increase due to the addition of the macro factors when using a rolling regression. However, we find some factors to be individually significant in predicting carry-trade risk premia on currencies of developed countries. For example, in the sample studied by Filippou and Taylor (2017), from July 1985 to March 2012, we find that the fifth US macro factor is significantly important. This latent factor is related to the categories of money and credit quantity aggregates and employment. When extending the sample to February 2023, we find that two US macro factors are statistically significant, which load most heavily on the series from the stock market category.

Finally, we discover that the Wald test has far from the true size of 5% with a value of 32.2% in the bond market and 22.3% in the currency market illustration. This results from the Wald test that accumulates size distortions of testing the significance of each macro factor. Moreover, the endogeneity that occurs in predicting currency risk premia is much less pronounced than in predicting bond risk premia. Therefore, we conclude that especially the persistence in the regressors drives the distortion. As exemplified by the motivating example, and real-life experiments in the bond and currency market, the parametric bootstrap procedure of Bauer and Hamilton (2018) leads to tests that have close-to-true size properties. That is why we recommend implementing the bootstrap inference in the presence of highly serially correlated regressor variables even when the endogeneity is extremely little.

For further research, we advise being attentive to not only the coefficient bias but also the standard error bias that moves the distributions of OLS estimates with consequences for inference. Although this bootstrap testing procedure leads to a close-to-true size property, it does not lead to the exact size. Therefore it may provide a useful future work direction to enhance the bootstrap procedure to generate a test with the exact true size and with better power properties. In fact, the motivating example shows that the bootstrap t-test has especially difficulty acquiring a close-to-true size when the regressors exhibit a trend. A possible solution is to include the trend in the VAR(1) specification when simulating the data in the bootstrap procedure. Additionally, we limited the investigation of the importance of macroeconomic information in predicting currency risk premia to only factors derived from macroeconomic data sets. However, investigating other macroeconomic variables similar to Bauer and Hamilton (2018) in the bond market, such as economic growth, inflation, or the trend in inflation, by the bootstrap procedure may provide a useful research direction.

References

Akram, Q. F., Rime, D. & Sarno, L. (2008). Arbitrage in the foreign exchange market: Turning on the microscope. *Journal of International Economics*, *76*(2), 237–253.

Bauer, M. D. & Hamilton, J. D. (2018). Robust bond risk premia. *The Review of Financial Studies*, *31*(2), 399–448.

Berg, K. A. & Mark, N. C. (2018). Global macro risks in currency excess returns. *Journal of Empirical Finance*, *45*, 300–315.

Bilson, J. F. (1980). *The 'speculative efficiency' hypothesis* (Tech. Rep.). National Bureau of Economic Research.

Brunnermeier, M. K., Nagel, S. & Pedersen, L. H. (2008). Carry trades and currency crashes. *NBER macroeconomics annual*, *23*(1), 313–348.

Campbell, J. Y. & Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, *81*(1), 27–60.

Chen, Y.-c. & Tsang, K. P. (2013). What does the yield curve tell us about exchange rate predictability? *Review of Economics and Statistics*, *95*(1), 185–205.

Cieslak, A. & Povala, P. (2015). Expected returns in treasury bonds. *The Review of Financial Studies*, *28*(10), 2859–2901.

Cochrane, J. H. (2017). Macro-finance. *Review of Finance*, *21*(3), 945–985.

Cochrane, J. H. & Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, *95*(1), 138–160.

Colacito, R., Riddiough, S. J. & Sarno, L. (2020). Business cycles and currency returns. *Journal of Financial Economics*, *137*(3), 659–678.

Corte, P. D., Riddiough, S. J. & Sarno, L. (2016). Currency premia and global imbalances. *The Review of Financial Studies*, *29*(8), 2161–2193.

Della Corte, P., Ramadorai, T. & Sarno, L. (2016). Volatility risk premia and exchange rate predictability. *Journal of Financial Economics*, *120*(1), 21–40.

Du, W., Tepper, A. & Verdelhan, A. (2018). Deviations from covered interest rate parity. *The Journal of Finance*, *73*(3), 915–957.

Fama, E. F. (1984). Forward and spot exchange rates. *Journal of Monetary Economics*, *14*(3), 319–338.

Filippou, I. & Taylor, M. P. (2017). Common macro factors and currency premia. *Journal of Financial and Quantitative Analysis*, *52*(4), 1731–1763.

Gabaix, X. & Maggiori, M. (2015). International liquidity and exchange rate dynamics. *The Quarterly Journal of Economics*, *130*(3), 1369–1420.

Hansen, L. P. & Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy*, *88*(5), 829–853.

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.

Joslin, S., Priebsch, M. & Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, *69*(3), 1197–1233.

Litterman, R. & Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, *1*(1), 54–61.

Liu, W. & Maynard, A. (2005). Testing forward rate unbiasedness allowing for persistent regressors. *Journal of Empirical Finance*, *12*(5), 613–628.

Ludvigson, S. C. & Ng, S. (2009). *A factor analysis of bond risk premia* (Tech. Rep.). National Bureau of Economic Research.

Lustig, H. & Richmond, R. J. (2017). *Gravity in fx r-squared: Understanding the factor structure in exchange rates* (Tech. Rep.). National Bureau of Economic Research.

Lustig, H., Roussanov, N. & Verdelhan, A. (2011). Common risk factors in currency markets. *The Review of Financial Studies*, *24*(11), 3731–3777.

Lustig, H., Roussanov, N. & Verdelhan, A. (2014). Countercyclical currency risk premia. *Journal of Financial Economics*, *111*(3), 527–553.

Lustig, H. & Verdelhan, A. (2007). The cross section of foreign currency risk premia and consumption growth risk. *American Economic Review*, *97*(1), 89–117.

Menkhoff, L., Sarno, L., Schmeling, M. & Schrimpf, A. (2012a). Carry trades and global foreign exchange volatility. *The Journal of Finance*, *67*(2), 681–718.

Menkhoff, L., Sarno, L., Schmeling, M. & Schrimpf, A. (2012b). Currency momentum strategies. *Journal of Financial Economics*, *106*(3), 660–684.

Ready, R., Roussanov, N. & Ward, C. (2017). Commodity trade and the carry trade: A tale of two countries. *The Journal of Finance*, *72*(6), 2629–2684.

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

Sarno, L. (2005). Towards a solution to the puzzles in exchange rate economics: Where do we stand? *Canadian Journal of Economics/Revue canadienne d'économique*, *38*(3), 673–708.

Stambaugh, R. F. (1988). The information in forward rates: Implications for models of the term structure. *Journal of Financial Economics*, *21*(1), 41–70.

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

Taylor, M. P. (1987). Covered interest parity: a high-frequency, high-quality data study. *Economica*, 429–438.

Villanueva, O. M. (2007). Forecasting currency excess returns: can the forward bias be exploited? *Journal of Financial and Quantitative Analysis*, *42*(4), 963–990.

Wei, M. & Wright, J. H. (2013). Reverse regressions and long-horizon forecasting. *Journal of Applied Econometrics*, *28*(3), 353–371.

A Derivation of Wald test

Next, we provide intuition on how the standard error bias arises as the focus in previous literature was mainly on the coefficient bias in b_1. We generalize Equation (4) to a regression model in matrix notation. Let $y = (y_{1+h}, y_{2+h}, ..., y_{T+h})'$ and stack x'_{1t} and x'_{2t} into $(T \times K_1)$ and $(T \times K_2)$ matrices denoted by X_1 and X_2. In addition, let $\varepsilon = (u_{1+h}, u_{2+h}, ..., u_{T+h})'$ represent the vector of forecast errors. To simplify, we assume that u_{t+h} is white noise. This results in the following model in matrix notation:

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon. \tag{7}$$

According to the Frisch-Waugh theorem, obtaining an estimate of β_2 (b_2) by a multiple regression as formulated in (7) is equivalent to estimating the partial effect of X_2 on y after removing the effect of X_1. To remove the effect of X_1, we first regress y on X_1 and X_2 on X_1. We obtain the cleaned y and X_2 as $\tilde{y} = M_1 y$ and $\tilde{x}_2 = M_1 X_2$ where $M_1 = I_T - X_1(X_1'X_1)^{-1}X_1'$. Second, we regress the cleaned $M_1 y$ on the cleaned $M_1 X_2$ to obtain b_2. Using the fact that M_1 is symmetric and idempotent,

$$X_2'M_1X_2 = (M_1X_2)'M_1X_2 = \sum_{t=1}^{T} \tilde{x}_{2t}\tilde{x}'_{2t}$$

and hence

$$b_2 = (X_2'M_1X_2)^{-1}X_2'M_1y = \sum_{t=1}^{T}(\tilde{x}_{2t}\tilde{x}'_{2t})^{-1}(\sum_{t=1}^{T}\tilde{x}_{2t}y_{t+h}).$$

Using equation (7), the orthogonality property of the residuals and again the idempotence of M_1,

$$b_2 = \beta_2 + \sum_{t=1}^{T}(\tilde{x}_{2t}\tilde{x}'_{2t})^{-1}(\sum_{t=1}^{T}\tilde{x}_{2t}u_{t+h}).$$

Then, the Wald test is for $h = 1$

$$W_T = (b_2 - \beta_2)'s^{-2}\sum_{t=1}^{T}\tilde{x}_{2t}\tilde{x}'_{2t}(b_2 - \beta_2) = (\sum_{t=1}^{T} u_{t+1}\tilde{x}'_{2t})(s^2\sum_{t=1}^{T}\tilde{x}_{2t}\tilde{x}'_{2t})^{-1}(\sum_{t=1}^{T}\tilde{x}_{2t}u_{t+1})$$

for $s^2 = (T - K_1 - K_2)^{-1}\sum_{t=1}^{T}(y_{t+1} - b_1'x_{1t} - b_2'x_{2t})$, where b_1 and b_2 are the OLS estimates of Equation (7). The validity of the Wald test depends on whether W_T converges to $\chi^2(K_2)$ distribution. However, if the regressors are highly persistent, the first step of regressing X_2 on X_1 acts like a spurious regression, which makes the coefficient behave like a random variable. In case the regressors are highly persistent and X_1 is not strictly exogenous, $\sum_{t=1}^{T}\tilde{x}_{2t}u_{t+1}$ has a nonstandard limiting distribution with a variance that is larger than that of $\sum_{t=1}^{T} x_{2t}u_{t+1}$. In fact, for near-unit-root processes, the small-sample distribution is quite different from $\chi^2(K_2)$.

B Bootstrap Design

B.1 Bond Market

Algorithm 1 shows the procedure that is followed to obtain the bootstrap p-value and size in the bond market application. Before the bootstrap starts, principal component analysis (PCA) is applied to the vector of yields of a bond of n-month maturity to obtain the weighting vector. The weighting vectors represent the eigenvectors that are normalized. The bootstrap design takes into account the high persistence of the regressor variables by first modeling the regressors as a VAR(1) process shown in lines 3 and 4. The persistence parameters are stored and reused in the bootstrap part to simulate the predictors with a similar serial correlation structure as in the data. In addition, we simulate under the null that variable in x_{2t} do not contain additional explanatory power, and thus the excess return y_{t+h} is completely determined by the first P principal components in x_{1t}. This can also be seen in lines 12 and 13, where the excess returns are determined by the artificial yield generated by $x_{1,t}$ in line 12. Moreover, this shows how the excess returns y_{t+h} and x_{1t} are simultaneously determined as both depend on the yield factor of a n-month maturity bond (i_{nt}). In addition, it should be noted that wherever a t-test is performed it uses Newey-West standard errors with 18 lags following other literature e.g. Cochrane and Piazzesi (2005); Ludvigson and Ng (2009).

B.2 Currency Market

Algorithm 2 shows how the algorithm of 1 is modified to make it more generally applicable. The main difference is the relation of x_{1t} to y_{t+h}. While for the bond market, the excess returns are a function of the yield curve, this is not generally the case for other financial markets such as the foreign exchange market. Therefore, we generalize the algorithm given in 1 by simulating the excess return in line 12 by reusing the coefficient estimates in the data (shown in line 2), while still modeling the regressors as a VAR(1) process (lines 10 and 11), by reusing the persistence parameters in the data (lines 5 and 6). In addition, it should be noted that wherever a t-test is performed it uses Newey-West standard errors with 4 lags following other literature e.g. Villanueva (2007).

Algorithm 1 Bootstrap Design for t-test: Bond Market

Calculate x_{1t} defined as the first P principal components of the observed yields (i_{nt}).
Estimate $i_{nt} = w_n^{'} x_{1t} + v_{nt}$, where i_{nt} is the yield on a n-month maturity bond and w_n represents the weighting vector computed from the PCA.
Estimate $y_{t+h} = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + u_{t+h}$
Obtain the t-statistic of the null hypothesis: $\beta_2 = 0$ and store it as t
Estimate $x_{1,t} = \phi_0 + \phi_1 x_{1,t-1} + e_{1,t}$ to obtain $\hat{\phi}_0$ and $\hat{\phi}_1$
Estimate $x_{2,t} = \alpha_0 + \alpha_1 x_{2,t-1} + e_{2,t}$ to obtain $\hat{\alpha}_0$ and $\hat{\alpha}_1$
for $b = 1$ to B **do**
    Obtain independent drawings of v_{nt}^b such that $v_{nt}^b \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$
    Obtain independent drawings of $(e_{1t}^{'b}, e_{2t}^{'b})$ from the joint empirical distribution of (e_{1t}, e_{2t})
    Construct an artificial time series: $x_{1t}^b = \hat{\phi}_0 + \hat{\phi}_1 x_{1,t-1}^b + e_{1t}^b$, starting at $x_{10}^b = x_{10}$
    Construct an artificial time series: $x_{2t}^b = \hat{\alpha}_0 + \hat{\alpha}_1 x_{2,t-1}^b + e_{2t}^b$, starting at $x_{20}^b = x_{20}$
    Construct an artificial time series: $i_{nt}^b = \hat{w}_n^{'} x_{1,t}^b + v_t^b$
    Construct an artificial time series: $y_{t+h}^b = n i_{nt}^b - h i_{ht}^b - (n-h) i_{n-h,t+h}^b$
    Estimate the model under $H_1 : \beta_2 \neq 0$: $y_{t+h}^b = \beta_0^b + \beta_1^b x_{1t}^b + \beta_2^b x_{2t}^b + \omega_{t+h}^b$
    Perform a t-test and store the t-statistics as t_b
    if Estimate Size = TRUE **then**
        Generate another bootstrap sample:
        Construct an artificial time series: $x_{1t}^* = \hat{\phi}_0 + \hat{\phi} 1 x_{1,t-1}^* + e_{1t}^*$, starting at $x_{10}^b = x_{10}$
        Construct an artificial time series: $x_{2t}^* = \hat{\alpha}_0 + \hat{\alpha}_1 x_{2,t-1}^* + e_{2t}^*$, starting at $x_{20}^b = x_{20}$
        Construct an artificial time series: $i_{nt}^* = \hat{w}_n^{'} x_{1,t}^* + v_t^*$
        Construct an artificial time series: $y_{t+h}^* = n i_{nt}^* - h i_{ht}^* - (n-h) i_{n-h,t+h}^*$
        Estimate the model under $H_1 : \beta_2 \neq 0$: $y_{t+h}^* = \beta_0^* + \beta_1^* x_{1t}^* + \beta_2^* x_{2t}^* + \omega_t^*$
        Perform a t-test and store the t-statistics as t_{boot}^b
    end if
end for
The p-value of the bootstrap test is given by $\frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{|t_b|>|t|}$
Use $t_{boot}^1, ..., t_{boot}^B$ to provide an estimate of the 95% asymptotic critical value and store as cv_{boot}
The size of the bootstrap test is given by $\frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{|t_b|>cv_{boot}}$

Algorithm 2 Bootstrap Design for t-test: General

Estimate $y_t = \gamma_0 + \gamma_1 x_{1,t-1} + v_t$
Obtain $\hat{v}_t = y_t - \hat{y}_t$, σ_v, $\hat{\gamma}_0$, and $\hat{\gamma}_1$
Estimate $y_t = \beta_0 + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + u_t$
Obtain the t-statistic of the null hypothesis: $\beta_2 = 0$ and store it as t
Estimate $x_{1,t} = \phi_0 + \phi_1 x_{1,t-1} + e_{1,t}$ to obtain $\hat{\phi}_0$ and $\hat{\phi}1$
Estimate $x_{2,t} = \alpha_0 + \alpha_1 x_{2,t-1} + e_{2,t}$ to obtain $\hat{\alpha}_0$ and $\hat{\alpha}_1$
for $b = 1$ to B **do**
    Obtain independent drawings of v_t^b such that $v_t^b \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$
    Obtain independent drawings of $(e_{1t}^{'b}, e_{2t}^{'b})$ from the joint empirical distribution of (e_{1t}, e_{2t})
    Construct an artificial time series: $x_{1t}^b = \hat{\phi}_0 + \hat{\phi}_1 x_{1,t-1}^b + e_{1t}^b$, starting at $x_{10}^b = x_{10}$
    Construct an artificial time series: $x_{2t}^b = \hat{\alpha}_0 + \hat{\alpha}_1 x_{2,t-1}^b + e_{2t}^b$, starting at $x_{20}^b = x_{20}$
    Construct an artificial time series: $y_t^b = \hat{\gamma}_0 + \hat{\gamma}_1 x_{1,t-1}^b + v_t^b$
    Estimate the model under $H_1 : \beta_2 \neq 0$: $y_t^b = \beta_0^b + \beta_1^b x_{1,t-1}^b + \beta_2^b x_{2,t-1}^b + \omega_t^b$
    Perform a t-test and store the t-statistics as t_b
    if Estimate Size = TRUE **then**
        Generate another bootstrap sample:
        Construct an artificial time series: $x_{1t}^* = \hat{\phi}_0 + \hat{\phi}1 x_{1,t-1}^* + e_{1t}^*$, starting at $x_{10}^b = x_{10}$
        Construct an artificial time series: $x_{2t}^* = \hat{\alpha}_0 + \hat{\alpha}_1 x_{2,t-1}^* + e_{2t}^*$, starting at $x_{20}^b = x_{20}$
        Construct an artificial time series: $y_t^* = \hat{\gamma}_0 + \hat{\gamma}_1 x_{1,t-1}^b + v_t^*$
        Estimate the model under $H_1 : \beta_2 \neq 0$: $y_t^* = \beta_0^* + \beta_1^* x_{1t}^* + \beta_2^* x_{2t}^* + \omega_t^*$
        Perform a t-test and store the t-statistics as t_{boot}^b
    end if
end for
The p-value of the bootstrap test is given by $\frac{1}{B} \sum_{b=1}^{B} \mathbb{1}_{|t_b|>|t|}$
Use $t_{boot}^1, ..., t_{boot}^B$ to provide an estimate of the 95% asymptotic critical value and store as cv_{boot}
The size of the bootstrap test is given by $\frac{1}{B} \sum_{b=1}^{B} \mathbb{1}_{|t_b|>cv_{boot}}$

C Data

Table 8 gives an overview of the variables that we use in this paper, where they come from (Source), a short description (Description), and the paper where you can find more information on these data sets (Reference paper). In addition, we report the sample period we use (Sample) and how we refer throughout this paper to these variables (Notation).

In the next sections, we give some more detailed information on the data sets we use in particular for the currency market application. Sections C.2 and C.3 provide an overview of variables that are contained in the large US data set and global data set respectively following Filippou and Taylor (2017).

Table 8: Data Sources

Panel A: Bond Market

Variable	Source	Description	Reference paper	Sample	Notation
Replication study: Economic Growth and Inflation			Joslin et al. (2014)	1985:1-2008:12	
Excess bond return	Anh Le's yields	Average from 2 to 10 years			y_{t+12}
Economic growth	3-month moving average of the Chicago Fed National Activity Index				GRO
Inflation	Blue Chip Financial Forecasts	1-year inflation expectations			INF
Replication study: Large Macro Data Sets			Ludvigson and Ng (2009)	1964:1-2022:12	
Excess bond return	Fama Bliss (WRDS)	Average from 2 to 5 years			y_{t+12}
Macro US data	Original data by Sydney C. Ludvigson Fred-MD database	131 macro variables			$F1, ..., F8$
Replication study: Trend Inflation			Cieslak and Povala (2015)	1971:1-2011:12	
Excess bond return	Anh Le's or Gurkaynak-Sack-Wright yields	Weighted average from 2 to 10 years			y_{t+12}
Trend Inflation	CPILFESL FRED	CPI All Items Less Food and Energy			τ
Replication study: Higher-Order PCs of Yields			Cochrane and Piazzesi (2005)	1964:1-2022:12	$PC1, ..., PC5$
Excess bond return	Fama Bliss (WRDS)	Average from 2 to 5 years			y_{t+12}

Panel B: Currency Market

Variable	Source	Description	Reference paper	Sample	Notation
Daily spot and 1 month forward exchange rates	Barclays and Reuters (BBI) WM/Refinitiv from Datastream	in US dollars	Lustig et al. (2011)	1985:7-2023:2	s and f
US Data	Datastream	125 monthly macroeconomic and financial series for US	Filippou and Taylor (2017)	1985:7-2023:2	$H1, ..., H9$
Global Data	Datastream	96 macroeconomic and financial variables from G10 countries	Filippou and Taylor (2017)	1985:7-2023:2	$G1, G2, G3$

Note: Panel A consists of four studies that were studied by Bauer and Hamilton (2018) and are replicated in this paper. The replication study on the large macro data set uses data from the FRED-MD database to introduce new data as a true out-of-sample period. Panel B contains the data sets that are used for the currency market application.

C.1 Panel

Our panel includes 43 countries collectively referred to as all countries. We include each of the following countries for the dates noted in parentheses: Australia (1984:12 - 2018:8), Austria (1996:12-1998:12), Belgium (1983:10-1998:12), Brazil (2000:7-2023:2), Bulgaria (2004:3-2023:2), Canada (1989:12-2023:2), Czech Republic (1996:12-2023:2), Denmark (1984:12-2023:2), Egypt (2004:3-2022:10), Euro area (1999:1-2023:2), Finland (1996:12-1998:12), France (1983:10-1998:12), Germany (1983:10-1998:12), Greece (1996:12-1998:12), Hong Kong (1983:10-2023:2), Hungary (1997:10-2023:2), India (1997:10-2023:2), Indonesia (1996:12-2023:2), Israel (2004:3-2023:2), Italy (1984:3-1998:12), Iceland (2004:3-2023:2), Japan (1983:10-2023:2), Kuwait (1990:5-2023:2), Malaysia (1984:12-2023:2), Mexico (1996:12-2023:2), the Netherlands (1983:10-1998:12), New Zealand (1984:12-2018:8), Norway (1984:12-2023:2), Philippines (1996:12-2023:2), Poland (1996:8-2023:5), Portugal (1996:12-1998:12), Russia (2004:3-2023:2), Saudi Arabia (1990:5-2023:2), Singapore (1984:12-2023:2), South Africa (1983:10-2023:2), South Korea (1999:8-2023:2), Spain (1986:10-1998:12), Sweden (1984:12-2023:2), Switzerland (1983:10-2023:2), Taiwan (1996:12-2023:2), Thailand (1995:3-2023:2), Ukraine (2004:3-2015:8), and the United Kingdom (1996:12-2023:2). The time period for each country is defined by data availability.

Our panel of developed countries includes 15 countries collectively referred to as developed countries. We include each of the following countries for the data noted in parentheses: Australia (1984:12 - 2018:8), Belgium (1983:10-1998:12), Canada (1989:12-2023:2), Denmark (1984:12-2023:2), Euro area (1999:1-2023:2), France (1983:10-1998:12), Germany (1983:10-1998:12), Italy (1984:3-1998:12), Japan (1983:10-2023:2), the Netherlands (1983:10-1998:12), New Zealand (1984:12-2018:8), Norway (1984:12-2023:2), Sweden (1984:12-2023:2), Switzerland (1983:10-2023:2), and the United Kingdom (1996:12-2023:2).

Following Lustig et al. (2011) and Filippou and Taylor (2017), we remove the following observations as they largely violate the CIP: South Africa (1985:07-1985:8), South Afric (2001:12-2004:5), Malaysia (1998:8-2005:6), Indonesia (1997:6-1998:3), Indonesia (2000:12-2007:5), Indonesia (2008:11-2009:2) and Kuwait (2001:3-2001:4).

C.2 US Data

Table 9: US Data

Series	Mnemonics	Transf	Description
			Real Output
1	870010061	3	US PRODUCTION - TOTAL INDUSTRY EXCL. CONSTRUCTION VOLA
2	870010074	3	US PROD IN TOTAL MFG VOLA
3	870010065*	3	US PROD OF TOTAL MFC CONSUMER GOODS VOLA
4	870010070*	3	US PROD OF TOTAL MFC INTERMEDIATE GOODS VOLA
5	870010058*	3	US PROD OF DWELLINGS(DISC.) CURN
6	USPERINCB	2	US PERSONAL INCOME (MONTHLY SERIES) (AR) CURA
7	USPILESTD	3	US PERSONAL INCOME LESS TRANSFER PAYMENTS (BCI 51) CONA
			Employment
8	870012315*	3	US EMPLOYEES: TOTAL (BUSINESS SURVEY)(DISC.) VOLA
9	870004508*	3	US CIVILIAN EMPLOYMENT: ALL PERSONS(DISC.) VOLA
10	870011929*	3	US CIVILIAN LABOUR FORCE: ALL PERSONS(DISC.) VOLA
11	870004623*	3	US UNEMPLOYMENT RATE: SURVEY-BASED (ALL PERSONS)(DISC.) SADJ
12	870004581	1	US WEEKLY HOURS WORKED: MFG VOLA
13	870004585	1	US WEEKLY OVERTIME HOURS: MFG VOLA
14	USUN%TOTQ	3	US UNEMPLOYMENT RATE SADJ
15	USUNPTOTO	3	US UNEMPLOYED (16 YRS & OVER) VOLA
16	USEMPTOTO	1	US TOTAL CIVILIAN EMPLOYMENT VOLA
			Consumption
17	USCP...CE	3	US CHAIN-TYPE PRICE INDEX FOR PERSONAL CONSMPTN.EXPENDITURE SADJ
18	USPERCONB	3	US PERSONAL CONSUMPTION EXPENDITURES (AR) CURA
19	USCNXFECE	3	US CHAIN-TYPE PRICE INDEX FOR PCE LESS FOOD & ENERGY SADJ
20	USCONDURD*	3	US PERSONAL CONSUMPTION EXPENDITURES - DURABLES (AR) CONA
21	USCONSRVB	3	US PERSONAL CONSUMPTION EXPENDITURES - SERVICES (AR) CURA
22	USCONNDRB	3	US PERSONAL CONSUMPTION EXPENDITURES - NONDURABLES (AR) CURA
			Housing Start
23	USHBRM..O	3	US HOUSING STARTED - MIDWEST (AR) VOLA
24	USHOUSE.O	2	US NEW PRIVATE HOUSING UNITS STARTED (AR) VOLA
25	USHBRN..O	2	US HOUSING STARTED - NORTHEAST (AR) VOLA
26	USHBRS..O	2	US HOUSING STARTED - SOUTH (AR) VOLA
27	USHBRW..O	2	US HOUSING STARTED - WEST (AR) VOLA
28	USHPERM.P	2	US HOUSING AUTHORIZED VOLN
29	USHB5ANDO	2	US HOUSING STARTED - 5 UNITS OR MORE (AR) VOLA
30	USHOUSATE	2	US NEW PRIVATE HOUSING UNITS AUTHORIZED BY BLDG.PERMIT (AR) VOLA
31	USNEWCONB	2	US CONSTRUCTION EXPENDITURES - TOTAL (AR) CURA
32	USEXHOMEO	2	US EXISTING HOME SALES: SINGLE-FAMILY & CONDO (AR) VOLA
33	USHC1...O	2	US HOUSING COMPLETED - 1 UNIT (AR) VOLA
34	USHC2TO4P	2	US HOUSING COMPLETED - 2 TO 4 UNITS VOLN
35	USHC1...P	2	US HOUSING COMPLETED - 1 UNIT VOLN
36	USHC5ANDO	2	US HOUSING COMPLETED - 5 UNITS OR MORE (AR) VOLA
37	USHC5ANDP	2	US HOUSING COMPLETED - 5 UNITS OR MORE VOLN
38	USHCRM..P	2	US HOUSING COMPLETED - MIDWEST VOLN
39	USHCRS..O	2	US HOUSING COMPLETED - SOUTH (AR) VOLA
40	USHCRW..O	2	US HOUSING COMPLETED - WEST (AR) VOLA
41	USHCRN..P	2	US HOUSING COMPLETED - NORTHEAST VOLN
42	USHCRS..P	2	US HOUSING COMPLETED - SOUTH VOLN
43	USHCRW..P	2	US HOUSING COMPLETED - WEST VOLN
44	USHCRM..O	2	US HOUSING COMPLETED - MIDWEST (AR) VOLA
45	USHCRN..O	2	US HOUSING COMPLETED - NORTHEAST (AR) VOLA
46	USHB1...P	2	US HOUSING STARTED - 1 UNIT VOLN
47	USHB2TO4P	2	US HOUSING STARTED - 2 TO 4 UNITS VOLN
48	USHB5ANDO	2	US HOUSING STARTED - 5 UNITS OR MORE (AR) VOLA
49	USHB5ANDP	2	US HOUSING STARTED - 5 UNITS OR MORE VOLN
50	USHBRM..P	2	US HOUSING STARTED - MIDWEST VOLN

Note: This Table provides a description of the US monthly data as well as the transformations applied to the series based on stationary tests. 0 = no transformation; 1 = first difference; 2 = logarithm; 3 = first difference of logarithm; 4 = second difference of logarithm. The data is available on Datastream and span the period 1985:7-2023:2. * indicates that the series has some missing observations over the sample.

Table 10: US Data (Continued)

Series	Mnemonics	Transf	Description
51	USHBRN..P	2	US HOUSING STARTED - NORTHEAST VOLN
52	USHBRS..P	2	US HOUSING STARTED - SOUTH VOLN
53	USHBRW..P	2	US HOUSING STARTED - WEST VOLN
54	USHU1...O	2	US HOUSING UNDER CONSTRUCTION - 1 UNIT (AR) VOLA
55	USHU1...P	2	US HOUSING UNDER CONSTRUCTION - 1 UNIT (EP) VOLN
56	USHU2TO4P	2	US HOUSING UNDER CONSTRUCTION - 2 TO 4 UNITS (EP) VOLN
57	USHU5ANDO	2	US HOUSING UNDER CONSTRUCTION - 5 UNITS OR MORE (AR) VOLA
58	USHU5ANDP	2	US HOUSING UNDER CONSTRUCTION - 5 UNITS OR MORE (EP) VOLN
59	USHURM..O	2	US HOUSING UNDER CONSTRUCTION - MIDWEST (AR) VOLA
60	USHURM..P	2	US HOUSING UNDER CONSTRUCTION - MIDWEST (EP) VOLN
61	USHURN..O	2	US HOUSING UNDER CONSTRUCTION - NORTHEAST (AR) VOLA
62	USHURN1.O	2	US HOUSING UNDER CONSTRUCTION - NORTHEAST (EP) VOLA
63	USHURS..O	2	US HOUSING UNDER CONSTRUCTION - SOUTH (AR) VOLA
64	USHURS1.O	2	US HOUSING UNDER CONSTRUCTION - SOUTH (EP) VOLA
65	USHURW..O	2	US HOUSING UNDER CONSTRUCTION - WEST (AR) VOLA
66	USHURW1.O	2	US HOUSING UNDER CONSTRUCTION - WEST (EP) VOLA
67	USHUNDERP	2	US HOUSING UNDER CONSTRUCTION AT END OF PERIOD (EP) VOLN
68	USPVH1UNE	2	US NEW PRIVATE HOUSING UNITS STARTED - 1 UNIT(AR) VOLA
69	USPVHOUCE	2	US NEW PRIVATELY OWNED HOUSING UNITS COMPLETED (AR) VOLA
70	USPVHCONE	2	US NEW PRIVATELY OWNED HOUSING UNITS UNDER CONSTRUCTION(AR) VOLA
			Orders
71	USNAPMNO	2	US ISM MANUFACTURERS SURVEY: NEW ORDERS INDEX SADJ
72	USCNORCGD	3	US NEW ORDERS OF CONSUMER GOODS & MATERIALS (BCI 8) CONA
			Stock Price
73	USSHRPRCF	3	US DOW JONES INDUSTRIALS SHARE PRICE INDEX (EP) NADJ
74	USNYSCOM	3	US NEW YORK STOCK EXCHANGE COMPOSITE SHARE PRICE INDEX NADJ
75	EMDJESUT	3	EM DOW JONES EUROSTOXX INDEX - UTILITIES NADJ
			Exchange Rates
76	SWKY3978F	3	SW SWISS FRANCS TO USD NADJ
77	741120006	3	UK EXCHANGE RATE: NATIONAL CURRENCY PER USD NADJ
78	741580006	3	JP EXCHANGE RATE: NATIONAL CURRENCY PER USD NADJ
			Interest Rates
79	870004511	0	US FEDERAL FUNDS RATE NADJ
80	870004512	0	US PRIME RATES NADJ
81	870009005	0	US YIELD 1+ YEAR FEDERAL GOVERNMENT BONDS NADJ
82	870009003	0	US IR OF THE 90-DAY DEPOSIT CERTIFICATE NADJ
83	870009004*	1	US RATE 3-MONTH EURO DEPOSITS NADJ
84	870009006	0	US YIELD 10-YEAR GOVERNMENT BONDS NADJ
85	741110441*	0	US INTEREST RATES: CENTRAL BANK POLICY RATE NADJ
86	741110450*	0	US INTEREST RATES: GOVERNMENT SECURITIES, TREASURY BILLS NADJ
87	741110465*	0	US 1-MONTH U.S. DEP. LIBOR(DISC.) NADJ
88	741110468*	1	US 3-MONTH U.S. DEP. LIBOR(DISC.) NADJ
89	741110471*	0	US 6-MONTH U.S. DEP. LIBOR(DISC.) NADJ
90	741110480*	0	US INTEREST RATES: GOVERNMENT SECURITIES, GOVERNMNET BONDS NADJ
91	741110483*	0	US INTEREST RATES: GOVT. SECURITIES, GOVT BONDS, SHORT-TERM NADJ
			Money and Credit Quantity Aggregates
92	870004548	3	US MONETARY AGGREGATE M1 CURA
93	870004544*	2	US MONETARY AGGREGATE M2(DISC.) CURA
94	870004546	3	US MONETARY AGGREGATE M3 CURA
95	741110057	3	US INTERNATIONAL RESERVES: OFFICIAL RESERVE ASSETS CURN
96	USBANKLPB	3	US COMMERCIAL BANK ASSETS - LOANS & LEASES IN BANK CREDIT CURA
97	USBCACI.B	3	US COMMERCIAL BANK ASSETS - COMMERCIAL & INDUSTRIAL LOANS CURA
98	USBCACI.Q	0	US COML BANK ASSETS-COMMERCIAL & INDL LOANS(BREAK ADJ,SAAR) SADJ
99	741110066	3	US INTERNATIONAL RESERVES: OFFICIAL RESERVE ASSETS, SDR CURN

Note: This Table provides a description of the US monthly data as well as the transformations applied to the series based on stationary tests. 0 = no transformation; 1 = first difference; 2 = logarithm; 3 = first difference of logarithm; 4 = second difference of logarithm. The data is available on Datastream and span the period 1985:7-2023:2. * indicates that the series has some missing observations over the sample.

Table 11: US Data (Continued)

Series	Mnemonics	Transf	Description
			Price Indices
100	870004479	3	US CPI ALL ITEMS SADJ
101	870004480	3	US CPI ALL ITEMS WAGE EARNERS NADJ
102	870006150	3	US CPI FOOD EXCL. RESTAURANTS NADJ
103	870006151	3	US CPI ENERGY NADJ
104	870006152	3	US CPI ALL ITEMS NON FOOD NON ENERGY NADJ
105	870004477	3	US CPI ALL ITEMS SYDNEY NADJ
106	USPFDOFGE	3	US PPI - FINISHED GOODS SADJ
107	USPFDGLEF	3	US PPI - FINISHED GOODS LESS FOODS AND ENERGY NADJ
108	USPFDNLFF	3	US PPI - NONDURABLE CONSUMER GOODS LESS FOODS AND ENERGY NADJ
109	USPPIOHDF	3	US PPI: OTHER HOUSEHOLD DURABLE GOODS NADJ
110	USPCIPSAF	3	US PPI - SPORTING & ATHLETIC GOODS NADJ
111	USPPITOYF	3	US PPI: TOYS, SPORTING GOODS, SMALL ARMS, ETC. NADJ
112	USPFDOCNF	3	US PPI - CONSUMER NONDURABLE GOODS LESS FOODS NADJ
113	USPPMLEYF	3	US PPI - PROCESSED MATERIALS LESS ENERGY NADJ
114	USPCX5SDE	3	US PPI: PROCESSED MATERIALS LESS ENERGY SADJ
115	USPSMNX.E*	3	US PPI - CRUDE NONFOOD MATERIALS EXCEPT FUEL(DISC.) SADJ
116	USPSFCPKF	3	US PPI-PORK PRODS,FRESH,FROZEN,OR PROCESSED, EXCEPT SAUSAGE NADJ
117	USPSPSYFE*	3	US PPI - MANUFACTURED ANIMAL FEEDS(DISC.) SADJ
118	870009105	3	US WEEKLY EARN: MFG SADJ
119	870004515	3	US WEEKLY EARN: MFG NADJ
120	870010200	3	US HOURLY EARN: PRIVATE SECTOR SADJ
121	870004629	3	US ITS IMPORTS C.I.F. TOTAL CURA
122	870004626	3	US ITS EXPORTS F.O.B. TOTAL CURA
123	870004632	1	US NET TRADE CURA
124	870006320*	0	US MFG - CONFIDENCE INDICATOR(DISC.) SADJ
125	USCAPUTLQ	0	US CAPACITY UTILIZATION RATE - ALL INDUSTRY SADJ

Note: This Table provides a description of the US monthly data as well as the transformations applied to the series based on stationary tests. 0 = no transformation; 1 = first difference; 2 = logarithm; 3 = first difference of logarithm; 4 = second difference of logarithm. The data is available on Datastream and span the period 1985:7-2023:2. * indicates that the series has some missing observations over the sample.

C.3 Global Data

Table 12: Global Data

Series	Mnemonics	Transf	Description
			Real Output
1	CN2PTOTCD*	3	CN GDP - INDUSTRIAL PRODUCTION (AR)(DISC.) CONA
2	AUSTEELPP*	3	AU AUSTRALIA - STEEL PRODUCTION VOLN
3	UKIPTOT.G	3	UK INDEX OF PRODUCTION - ALL PRODUCTION INDUSTRIES VOLA
4	SDIPTOT5G*	3	SD INDUSTRIAL PRODN-MINING & MANUFACTURING (CAL ADJ)(DISC.) VOLA
5	BDIP0093G*	3	BD INDUSTRIAL PRODUCTION: MANUFACTURING (CAL ADJ)(DISC.) VOLA
			Employment
6	DKUNPTOTP	3	DK UNEMPLOYMENT NET (METHDOLOGY BREAK JANUARY 2007) VOLN
7	CNUN%TOTQ	3	CN UNEMPLOYMENT RATE (15 YRS & OVER) SADJ
8	JPUN%TOTQ	1	JP UNEMPLOYMENT RATE (METHO BREAK OCT 2010) SADJ
9	AUUN%TOTQ	1	AU UNEMPLOYMENT RATE (LABOUR FORCE SURVEY ESTIMATE) SADJ
10	NZMLM005P*	3	NZ REGISTERED UNEMPLOYMENT: LEVEL (ALL PERSONS)(DISC.) VOLN
11	UKUN%TOTQ	1	UK CLAIMANT COUNT RATE SADJ
12	SWUN%TOTR	1	SW UNEMPLOYMENT RATE (METHOD BREAK JAN 2014) NADJ
13	OEUN%TOTR	2	OE UNEMPLOYMENT RATE % NADJ
14	NWUN%TOTQ*	1	NW UNEMPLOYMENT RATE (% OF LFS)(DISC.) SADJ
			Consumption
15	NWPERCGDG	3	NW PRIVATE CONSUMPTION - GOODS VOLA
16	BCPIEXC	3	BOC. Weekly Excluding Energy - PRICE INDEX
17	AUIMPCSGB	2	AU IMPORTS FOB - CONSUMPTION GOODS N.E.S. CURA
18	JPCCEPCSE*	1	JP ELECTRIC POWER CONSUMPTION - LARGE CORPORATIONS(DISC.) SADJ
19	CNPPOCOMP*	2	CN PETROLEUM PRODUCTS: ALL PRODUCTS - OWN CONSUMPTION VOLN
20	AUIMPCGDA	2	AU IMPORTS FOB - CONSUMPTION GOODS CURN
21	UKHYELECG	3	UK CONSUMPTION OF HYDRO ELECTRICITY VOLA
22	SDECTOTLP	3	SD CONSUMPTION OF ELECTRICITY VOLN
23	NWPERCGDG	3	NW PRIVATE CONSUMPTION - GOODS VOLA
24	EUCNMCOIP*	3	EU CONSUMPTION - CRUDE OIL(DISC.) VOLN
25	DKESEIWBP	1	DK ENERGY: TOTAL CONSUMPTION OF NATURAL GAS VOLN
			Stock Price
26	JPSHRPRCF	3	JP TOKYO STOCK EXCHANGE - TOPIX (EP) NADJ
27	CNSHRPRCF	3	CN S&P/TSX COMPOSITE SHARE PRICE INDEX (EP) NADJ
28	TOTXTER	1	EUROPE-DS DS-MARKET EX TMT - PRICE INDEX
29	HLTHCDK	3	DENMARK-DS Health Care - PRICE INDEX
30	TOTMKAU	1	AUSTRALIA-DS Market - PRICE INDEX
31	FINANUK	3	UK-DS Financials - PRICE INDEX
32	MSSWDNL	1	MSCI SWEDEN - PRICE INDEX
33	MSSWITL	1	MSCI SWITZERLAND - PRICE INDEX
			Price Indices
34	CNCONPRCF	3	CN CPI NADJ
35	JPCONPRCF	1	JP CPI: NATIONAL MEASURE NADJ
36	AUCPANNL	1	AU INFLATION RATE (DS CALCULATED QUARTERLY) NADJ
37	NZCPANNL	1	NZ INFLATION RATE NADJ
38	UKOCP009R	2	UK CPI ALL ITEMS NADJ
39	SWCONPRCF	1	SW CPI (2020M12=100) NADJ
40	SDCONPRCF	3	SD CPI NADJ
41	NWCONPRCF	1	NW CPI NADJ
42	EUOCP009F	3	EU CPI ALL ITEMS NADJ
43	DKCONPRCF	2	DK CPI NADJ
44	CNMPIFG1F*	3	CN TOTAL PPI FINISHED GOODS(DISC.) NADJ
45	JPOPIFG2F*	3	JP DOMESTIC PPI FINISHED GOODS NADJ
46	UKPROPRCF	3	UK PPI - OUTPUT OF MANUFACTURED PRODUCTS EXCLUDING DUTY NADJ
47	SWPROPRCE	1	SW PPI (2020M12=100) SADJ
48	NWPROPRCF	3	NW PPI (LINKED & REBASED) NADJ
49	EUOPIMP2F	3	EU DOMESTIC PPI MFG - PROXY NADJ
50	DKESPPINF*	3	DK PPI: MIG - NON-DURABLE CONSUMER GOODS, 2010=100(DISC.) NADJ

Note: This Table provides a description of the global monthly data as well as the transformations applied to the series based on stationary tests. 0 = no transformation; 1 = first difference; 2 = logarithm; 3 = first difference of logarithm; 4 = second difference of logarithm. The data is available on Datastream and span the period 1985:7-2023:2. * indicates that the series has some missing observations over the sample.

Table 13: Global Data (Continued)

Series	Mnemonics	Transf	Description
			Interest Rates
51	ECCAD1M	4	CANADIAN DOLLAR 1M DEPOSIT (FT/RFV) - MIDDLE RATE
52	ECJAP1M	4	JAPANESE YEN 1M DEPOSIT (FT/RFV) - MIDDLE RATE
53	ECUKP1M	1	UK STERLING 1M DEPOSIT (FT/RFV) - MIDDLE RATE
54	ECWGM1M	1	BD EU-MARK 1M DEPOSIT (FT/RFV) - MIDDLE RATE
55	ECSWF1M	1	SWISS FRANC 1M DEPOSIT (FT/RFV) - MIDDLE RATE
56	ECDKN1M	1	DANISH KRONE 1M DEPOSIT (FT/RFV) - MIDDLE RATE
57	ECUSD1M	1	US DOLLAR 1M DEPOSIT (FT/RFV) - MIDDLE RATE
58	ECCAD3M	4	CANADIAN DOLLAR 3M DEPOSIT (FT/RFV) - MIDDLE RATE
59	ECJAP3M	4	JAPANESE YEN 3M DEPOSIT (FT/RFV) - MIDDLE RATE
60	ECWGM3M	1	BD EU-MARK 3M DEPOSIT (FT/RFV) - MIDDLE RATE
61	ECSWF3M	1	SWISS FRANC 3M DEPOSIT (FT/RFV) - MIDDLE RATE
62	ECDKN3M	1	DANISH KRONE 3M DEPOSIT (FT/RFV) - MIDDLE RATE
63	ECUSD3M	4	US DOLLAR 3M DEPOSIT (FT/RFV) - MIDDLE RATE
			International Trade
64	CNVISBOPB	0	CN VISIBLE TRADE BALANCE (BALANCE OF PAYMENTS BASIS) CURA
65	JPVISGDSA	0	JP VISIBLE TRADE BALANCE CURN
66	AUBALGOSA	0	AU BALANCE OF TRADE IN GOODS & SERVICES (BOP BASIS) CURN
67	NZVISGDSA	0	NZ VISIBLE TRADE BALANCE CURN
68	UKVISBOPB	1	UK VISIBLE TRADE BALANCE - BALANCE OF PAYMENTS BASIS CURA
69	SWTA2891E	0	SW TRADE BALANCE TOTAL 1 (SA) CURA
70	SDVISGDSA	0	SD VISIBLE TRADE BALANCE CURN
71	NWVISGDSA	0	NW VISIBLE TRADE BALANCE CURN
72	BDVISGDSB	1	BD VISIBLE TRADE BALANCE CURA
73	DKVISGDSA	0	DK VISIBLE TRADE BALANCE CURN
74	USVISGDSB	1	US VISIBLE TRADE BALANCE F.A.S.-F.A.S. CURA
			Reserves
75	870008751*	4	DK SDR RESERVE ASSETS(DISC.) CURN
76	498012588	4	JP FOREIGN CURRENCY RESERVES CURN
77	360790010*	4	SW OFFICIAL RESERVES MINUS GOLD (US$) CURN
78	USRESCURA	3	US FOREIGN CURRENCY RESERVES CURN
79	CNB3802.	4	CN OFFICIAL INTERNATIONAL RESERVES:CONVERTIBLE NON-U.S.$ CURRENC
80	100700010	4	AU OFFICIAL RESERVE ASSETS (METHOD BREAK JAN2015) CURN
81	109998872	3	AU AUSTRALIAN $ EFFECTIVE EXCHANGE RATE INDEX NADJ
82	USNLTSECA	0	US FOREIGN NET LONG TERM FLOWS IN SECURITIES CURN
83	116600110	3	NZ CREDIT AGGREGATES: DOM. CREDIT- CREDIT TO PRIVATE SECTOR CURN
84	116600740	3	NZ TOTAL OFFICIAL RESERVES CURN
85	870008981*	3	NW RESERVE ASSETS(DISC.) CURN
86	SDRESERVA	3	SD BANK OF SWEDEN: ASSETS - GOLD & FOREIGN EXCHANGE RESERVE CURN
			G7 Countries
87	G7MPI009R*	3	G7 DOMESTIC PPI MFG - PROXY NADJ
88	G7MPI009R*	1	G7 DOMESTIC PPI MFG - PROXY NADJ
89	G7MXT008Q	1	G7 ITS EXPORTS F.O.B. TOTAL SADJ
90	505676793*	1	G7 NET TRADE(DISC.) CURA
91	502621288	0	G7 PRODUCTION - TOTAL INDUSTRY EXCL. CONSTRUCTION SADJ
92	503351909	3	G7 CPI ALL ITEMS NON FOOD NON ENERGY NADJ
93	503547075	1	G7 CPI FOOD NADJ
94	504352258*	0	G7 WEEKLY EARN: MFG SADJ
95	502120123	0	G7 TOTAL RETAIL TRADE (VOLUME) SADJ
96	MSCIG7$	3	MSCI G7 U$ - PRICE INDEX

Note: This Table provides a description of the global monthly data as well as the transformations applied to the series based on stationary tests. 0 = no transformation; 1 = first difference; 2 = logarithm; 3 = first difference of logarithm; 4 = second difference of logarithm. The data is available on Datastream and span the period 1985:7-2023:2. * indicates that the series has some missing observations over the sample.

C.4 Autocorrelation of Factors from Large Macro Data Sets

Table 14: Currency Market: Factors of Large Macro Data Sets

	G1	G2	G3	H1	H2	H3	H4	H5	H6	H7	H8	H9
ACF1	0.94	0.24	0.54	0.98	0.63	0.70	0.46	0.84	0.55	0.50	0.35	0.12
ACF2	0.92	0.11	0.32	0.97	0.43	0.55	0.22	0.76	0.36	0.46	0.29	-0.11
% Var	0.09	0.09	0.08	0.31	0.12	0.11	0.06	0.04	0.03	0.03	0.02	0.02
Cum. % Var	0.09	0.17	0.25	0.31	0.43	0.53	0.59	0.64	0.66	0.69	0.71	0.73

Note: This table reports the first and second-order autocorrelation coefficients and the variation explained by each factor. The factors are computed over the sample period July 1985 - February 2023 for currencies of developed countries. The factors starting with "G" represent global factors, while the factors starting with "H" represent domestic factors. The last column shows the first and second-order autocorrelation for the average forward discount on the Developed countries' currency basket.

C.5 Cumulative Returns from Carry-Trade Strategy

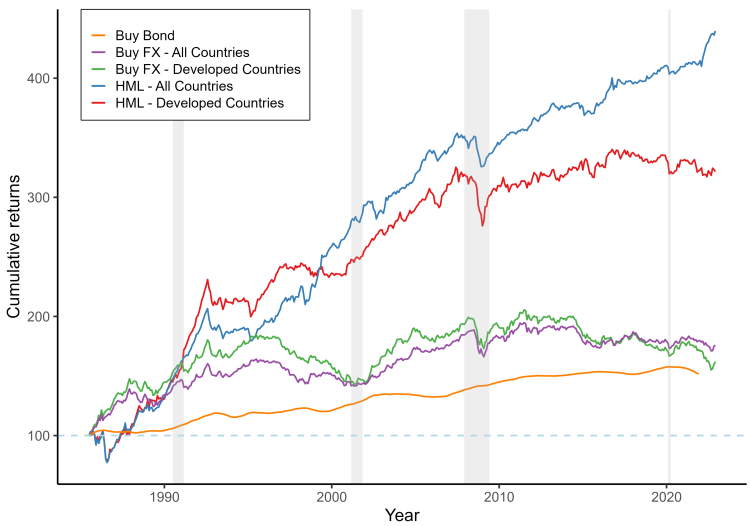

Figure 4: Cumulative excess returns (monthly) of the HML strategy versus equal-weight buy FX strategies and equal-weight buy bond strategy as a benchmark during the period 1985:1-2022:12. The initial investment (January 1985) is USD 100.

D Additional Results - Bond Market

D.1 Factors of Large Macro Data Set

Table 15: Principal Components Analysis Results

Maturity in Months	PC1	PC2	PC3	PC4	PC5
12	0.45	-0.70	0.52	-0.18	0.07
24	0.46	-0.26	-0.51	0.51	-0.45
36	0.45	0.07	-0.44	-0.13	0.76
48	0.45	0.36	-0.07	-0.68	-0.46
60	0.43	0.55	0.52	0.48	0.07
% Var	0.99	0.01	0.00	0.00	0.00
ACF1	0.99	0.96	0.80	0.66	0.55
ACF2	0.97	0.91	0.66	0.55	0.36

Note: This table reports the coefficients of the principal components and the variation explained by each factor. The factors and autocorrelations are computed over the sample period January 1985 - December 2016.

Table 16: Bond Market: Factors of Large Macro Data Set

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
ACF1	0.20	0.09	0.78	0.56	0.12	0.76	0.43	-0.02
ACF2	-0.10	-0.08	0.74	0.34	-0.02	0.69	0.23	0.05

Note: This table reports the first and second-order autocorrelation coefficients and the variation explained by each factor. The factors and autocorrelations are computed over the sample period January 1985 - December 2016.

D.2 Interpretation of Macro Factors

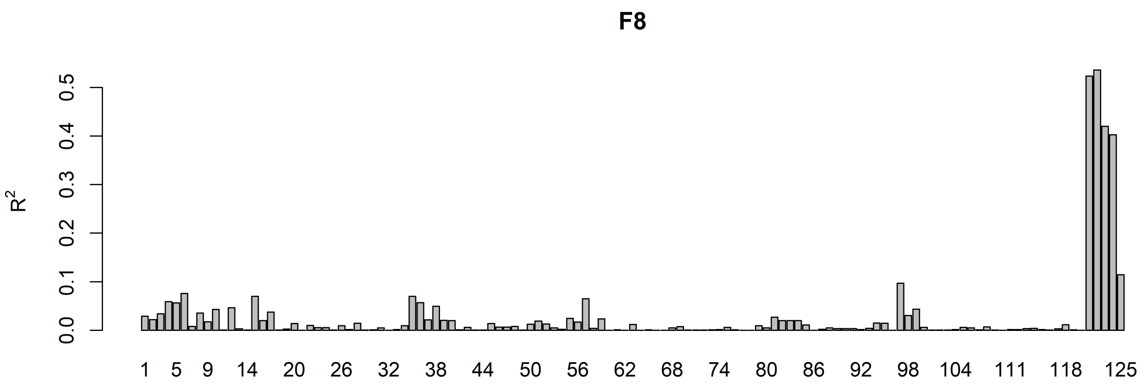

Figure 5: Figure shows the R^2 from regressing the series number given on the x-axis onto the estimated factor named in the heading. The factor is estimated using data from 1964:1-2007:12.

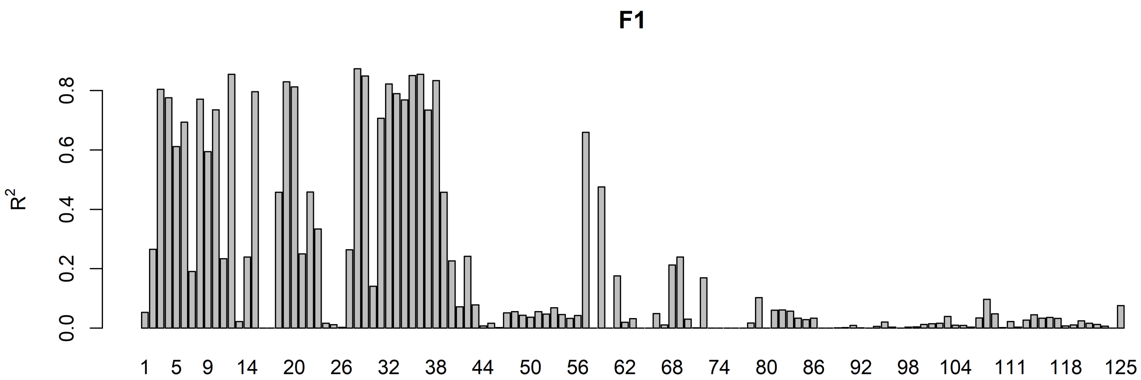

Figure 6: Figure shows the R^2 from regressing the series number given on the x-axis onto the estimated factor named in the heading. The factor is estimated using data from 1985:1-2022:12.

E Additional Predictors - Bond Market

E.1 Economic Growth and Inflation

Joslin et al. (2014) study the significance of inflation and economic growth as macro variables in forecasting bond excess returns. They quantify that variation in economic activity and inflation in the US influences the market prices of level, slope, and curvature risks in Treasury markets. Moreover, they conclude that macro risks are not completely spanned by the information in the yield curve and that they have predictive content for excess returns. They include measures of real economic activity (GRO) and inflation (INF) as macro variables which suggest a link to the Taylor rule. In particular, GRO is the three-month moving average of the Chicago FED National Activity Index (CFNAI), as a measure of current real economic conditions. INF is the expected rate of inflation over the coming year computed from the surveys of professional forecasters by Blue Chip Financial Forecasts. Hence, x_{2t} in equation (4) is the parsimonious choice of $x_{2t} = (GRO_t, INF_t)'$. The reason to investigate this study is due to the high persistence in the regressors and the endogeneity due to the PCs shown in Table 2. In addition, Table 17 shows the substantial first and second-order autocorrelation of the yield factors, and the first-order autocorrelation for the variables GRO and INF is $0.946, 0.986$ respectively. Since the persistence of GRO and INF is high, it may be important to adjust for small-sample bias in the VAR estimates, therefore next to the simple bootstrap Bauer and Hamilton (2018) include the bias-corrected bootstrap.

Table 18 shows the output of estimating equation (4) which is similar to Bauer and Hamilton (2018).[11] Table 18 is divided into two panels, where panel A concerns the sample period as studied by Joslin et al. (2014) and panel B the sample period as studied by Bauer and Hamilton (2018). The first row of both panels shows the coefficient estimates. Overall, the signs of the coefficient estimates remain the same when including more recent data. In addition, Table 18 shows the p-values of the t-test using HAC standard errors, simple bootstrap test, and the bootstrap test with bias correction. Also, it displays the true size and power of the t-test using HAC standard errors and the bootstrap procedure.

First, we focus on the original sample as studied by Joslin et al. (2014). While the variable GRO appears as significant at a 5% significance level by the t-test, it is not according to the bootstrapping procedures. In contrast, INF emerges as significantly explaining the bond excess returns by all tests considered at a 5% significance level. The bootstrap p-value for the Wald test is slightly below 5% for the simple version, and slightly above for the bias-corrected version. In conclusion, the evidence against

[11]These results relate to Table 3 by Bauer and Hamilton (2018).

Table 17: Joslin-Priebsch-Singleton: Principal components

Maturity in Months	PC1	PC2	PC3	PC4	PC5
6	0.31	-0.52	0.48	-0.51	0.27
12	0.32	-0.46	0.20	0.33	-0.43
24	0.33	-0.26	-0.26	0.42	-0.06
36	0.32	-0.11	-0.36	0.20	0.29
48	0.32	0.01	-0.37	-0.13	0.14
60	0.31	0.11	-0.26	-0.22	0.12
72	0.30	0.19	-0.15	-0.31	0.12
84	0.29	0.25	0.01	-0.19	-0.59
96	0.28	0.29	0.13	-0.05	-0.12
108	0.28	0.33	0.24	-0.01	-0.19
120	0.26	0.36	0.48	0.46	0.45
% Var	0.98	0.02	0.00	0.00	0.00
ACF1	0.99	0.97	0.87	0.85	0.67
ACF2	0.97	0.93	0.76	0.74	0.46

Note: This Table reports the principal components' coefficients, the variation explained by each factor, and the autocorrelation. The factors and autocorrelations are computed over the sample period 1985:1-2016:12.

the spanning hypothesis is much weaker than would appear from the conventional t-test.

Table 18: Joslin-Priebsch-Singleton: GRO and INF

A. Original sample, 1985:1-2008:12	PC1	PC2	PC3	GRO	INF	Wald
Coefficient	1.090	1.793	2.874	-2.200	-6.052	
p-value						
HAC	**0.000**	**0.000**	0.425	**0.014**	**0.000**	**0.000**
Simple bootstrap				0.112	**0.034**	**0.037**
BC bootstrap				0.128	**0.043**	0.053
Size						
HAC				0.196	0.275	0.363
Bootstrap				0.054	0.060	0.061
Power						
HAC				0.418	0.990	0.994
Bootstrap				0.200	0.925	0.902
B. Later sample, 1985:1-2016:12						
Coefficient	0.371	1.741	1.542	-0.429	-2.420	
p-value						
HAC	**0.022**	**0.001**	0.542	0.592	0.073	0.187
Simple bootstrap				0.708	0.278	0.507
BC bootstrap				0.713	0.320	0.549

Note: Predictive regression for annual bond excess returns (y_{t+12}). $y_{t+12} = \beta_0 + \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+12}$, where $x_{1t} = (PC1_t, PC2_t, PC3_t)'$ and $x_{2t} = (GRO_t, INF_t)'$. Results in panel A are for the same period as studied by Joslin et al. (2014), panel B includes the same period as studied by Bauer and Hamilton (2018). HAC statistic and p-values are calculated using Newey-West standard errors with 18 lags. The column *Wald* reports p-values for the hypothesis that GRO and INF have no predictive power. p-values below 5% are in bold. Under *Size* we report estimates of the size of the tests, based on simulations from the simple bootstrap under the null hypothesis. Both tests have a nominal size of 5%. Under *Power* we report estimates of the power of the tests, based on simulations from the simple bootstrap under the alternative hypothesis.

Table 19 shows the adjusted R^2 by estimating regression (4). The first row corresponds to an estimation by OLS, followed by the average adjusted R^2 over the bootstrap samples of a simple and bias-corrected procedure. The addition of x_{2t} in the predictive regression increases the adjusted R^2 by 19 percentage point from 19% to 38%. JPS also found an increase for the adjusted R^2 of the two-year bond that increased from 14% to 48% when x_{2t} is included. While this increase is substantial, it does not provide evidence against the spanning hypothesis. The bootstrapping procedures assume that the spanning hypothesis holds true, however, the Table shows that an adjusted R^2 of 38% is not uncommon when including x_{2t}. In other words, adding the regressors x_{2t} leads to considerable increases in the R^2, although x_{2t} has no predictive power in population by construction.

Table 19: Joslin-Priebsch-Singleton: GRO and INF

A. In-sample performance

	R_1^2	R_2^2	$R_2^2 - R_1^2$
I. Original sample, 1985:1-2008:12			
Data	0.19	0.38	0.19
Simple bootstrap	0.32	0.38	0.06
	(0.10, 0.55)	(0.15, 0.60)	(0.00, 0.20)
BC bootstrap	0.36	0.42	0.06
	(0.09, 0.63)	(0.15, 0.68)	(0.00, 0.22)
II. Later sample, 1985:1-2016:12			
Data	0.17	0.21	0.04
Simple bootstrap	0.28	0.32	0.05
	(0.08, 0.49)	(0.13, 0.53)	(-0.00, 0.17)
BC bootstrap	0.29	0.34	0.05
	(0.07, 0.53)	(0.11, 0.57)	(-0.00, 0.19)

B. Out-of-sample performance

	In-sample	Out-of-sample	DM-test p-value
RMSE-ratio	0.759	2.156	**0.005**

Note: Adjusted R^2 of predictive regression for annual bond excess returns (y_{t+12}) using three PCs ($x_{1t} = (PC1_t, PC2_t, PC3_t)'$) and economic growth ($GRO$) and expected inflation (INF) which are contained in x_{2t} ($x_{2t} = (GRO_t, INF_t)'$). Results in panel A.I are for the same period as studied by Joslin et al. (2014), and panel A.II includes the same period as studied by Bauer and Hamilton (2018). The first row of panels A.I and A.II reports the statistics in the original data. The following reports bootstrap mean and 95% percentiles in parentheses. The bootstrap procedure assumes that x_{2t} contains no incremental predictive power. The first column corresponds to the predictive regression that includes x_{1t} and the second column to the inclusion of x_{1t} and x_{2t}. Panel B assesses the predictive power for excess bond returns averaged across maturities by the $RMSE$. The $RMSE$-ratio is the relative $RMSE$ of a model including x_{1t} and x_{2t} and a model without x_{2t}. The in-sample period is 1985:1-2007:12, and the out-of-sample period is 2008:1-2015:12. The out-of-sample predictions are estimated using an expanding window approach. The Diebold-Mariano (DM) test for equal forecasting accuracy considering the prediction errors out-of-sample for the unrestricted and restricted model.

Next, we consider the later sample (1985-2016). Interestingly, while INF is significantly important at a 10% significance level using the HAC inference, it is certainly not for the bootstrap procedures with a p-value of 27.8% and 32% respectively. Similarly, the variable GRO obtains a considerably higher p-value for all tests in the later sample and therefore we conclude that this variable does not have significant explanatory power when controlling for the information in the yield curve. As a result of the lack of significance, the inclusion of these variables does not lead to a substantial increase in R^2 (from 17% to 21%). Apart from this, the increased R^2 observed in the data is included in the 95% confidence interval of the bootstrapping procedures. Therefore, this increase is not uncommon, even when GRO and INF are not contributing to the excess bond returns by construction.

Lastly, panel B of Table 19 shows the $RMSE$ of the unrestricted model which includes x_{1t} and x_{2t} relative to the $RMSE$ of the restricted model which includes only x_{1t}. The forecasts out-of-sample are computed by using an expanding window, where the model is reestimated each month to include the newest available data. While the $RMSE$-ratio is smaller than one for the in-sample period, it is substantially above one for the out-of-sample period. This indicates that the addition of x_{2t} to the model considerably increases the $RMSE$. Moreover, the Diebold-Mariano (DM) test shows the rejection of equal forecasting accuracy in the out-of-sample period. This rejection and the magnitude of the $RMSE$-ratio again show that the addition of x_{2t} does not lead to a significantly better predictive performance in this true out-of-sample period.

To conclude, this real-time experiment shows that inference based on a conventional t-test and Wald test is unreliable. For example, the variable concerning economic growth appeared as significant by the t-test in the original sample, while its significance disappeared when using the bootstrapping procedures. In addition, economic growth and inflation appeared as significant in the original sample. However, the evidence found by the conventional Wald test is not as convincing as resulting from the bootstrap procedures. Using these procedures to profit from the better size and power, we conclude that inflation is a significant contributor to bond excess returns in the original sample. However, economic growth does not contain significant additional predictive power in both samples considered. By including more recent data, the evidence against the spanning hypothesis disappears.

E.2 Trend Inflation

Additionally, Cieslak and Povala (2015) decompose Treasury yields into inflation expectations and maturity-specific interest-rate cycles. They state that highly persistent expected inflation dynamics, often called trend inflation, determine the level of interest rates in the long run and across maturities. They capture the trend inflation by a discounted moving average (DMA) of past core inflation which reflects that people update their inflation expectations only slowly over time. In fact, they construct a DMA of past CPI inflation:

$$\tau_t^{CPI} = (1-\nu)\sum_{i=0}^{t-1} \nu^i \pi_{t-i},$$

with the year-over-year inflation, $\pi_t = ln\frac{CPI_t}{CPI_{t-1}}$, monthly sampling and $\nu = 0.987$. They account for small-sample biases by relying on conservative standard errors from reverse regressions (Hodrick, 1992; Wei & Wright, 2013). In addition, they identify that because of the bias from the persistent interest rates, the predictive R^2's overstate the true degree of predictability. Therefore, they compute the small sample distribution of R^2's under the null of no predictive power by any of the variables considered which is commonly known as the expectations hypothesis. Cieslak and Povala (2015) find that the inflation trend does not predict the excess return on its own, but when added to a regression that includes yields, the trend becomes highly significant as do the yields. We replicate the results of Bauer and Hamilton (2018) in which the dependent variable is the weighted average of annual excess returns on two- to fifteen-year bonds and $x_{2t} = \tau_t$, the inflation trend.

Table 20 shows that the first three PCs indeed represent level, slope, and curvature components of the yields as commonly assumed. In addition, the first three factors, explain almost all variations in the observed yields. Interestingly, the factors are highly persistent, as is the trend inflation with autocorrelation of 0.99 and 0.98 for the first and second order respectively. While the variables in x_{1t} and x_{2t} are highly persistent, they are stationary and both exhibit a trend as shown in Figure 7 and 8 with a positive trend until 1983 and a negative trend afterward. The motivating example in Section 2 shows that the standard error bias becomes worse in the presence of trends, and therefore the predictive power of the inflation trend and of the yield factors could be spurious. This is the reason to investigate this study and compare it to conventional t-tests. As the inflation trend is highly persistent, the bootstrap procedure is bias-corrected to account for the small-sample bias that arises in the VAR estimates.

Table 21 shows the output of estimating the predictive regression, where panel A displays the sample

Table 20: Cieslak and Povala: Trend Inflation

Maturity in Months	PC1	PC2	PC3	PC4	PC5
1	0.24	-0.53	0.73	-0.35	0.06
12	0.28	-0.44	-0.06	0.62	-0.42
24	0.28	-0.31	-0.25	0.25	0.23
36	0.28	-0.19	-0.29	-0.05	0.36
48	0.27	-0.10	-0.27	-0.21	0.25
60	0.27	-0.02	-0.23	-0.28	0.06
72	0.26	0.04	-0.17	-0.27	-0.11
84	0.25	0.09	-0.11	-0.23	-0.23
96	0.25	0.13	-0.05	-0.16	-0.29
108	0.24	0.16	0.01	-0.08	-0.29
120	0.24	0.19	0.06	-0.01	-0.24
132	0.23	0.21	0.10	0.06	-0.15
144	0.23	0.23	0.14	0.12	-0.03
156	0.23	0.24	0.17	0.17	0.11
168	0.22	0.26	0.20	0.21	0.26
180	0.22	0.27	0.22	0.24	0.42
% Var	0.97	0.03	0.00	0.00	0.00
ACF1	0.99	0.97	0.78	0.80	0.83
ACF2	0.97	0.93	0.62	0.68	0.70

Note: This Table reports the principal components' coefficients, the variation explained by each factor and the autocorrelation. The factors and autocorrelations are computed over the sample period 1971:11-2016:12.

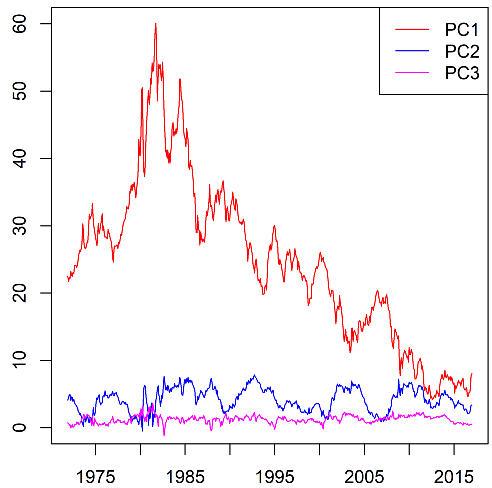

Figure 7: First three *PCs* computed over the period 1971:11-2016:12

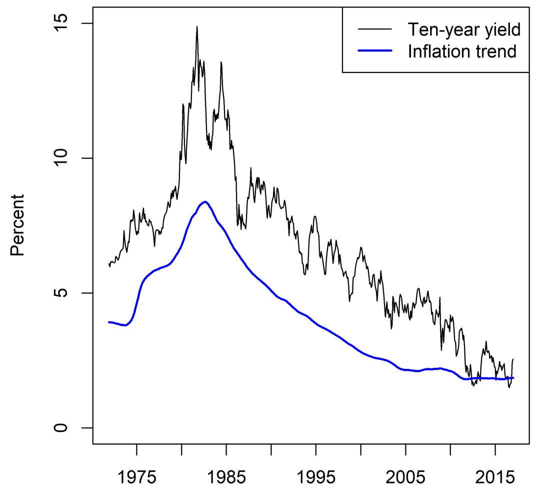

Figure 8: Inflation trend and yield on a ten-year Treasury bond over the period 1971:11-2016:12

studied by Cieslak and Povala (2015) and panel B the extended sample studied by Bauer and Hamilton (2018).[12] The first row displays the estimation results when only including the first three yield factors in the model and the second row when only including the inflation trend in the model. Notably, when the inflation trend (τ_t) is added to the yield factors, it is highly significant and it increases the significance of the PCs. However, the inflation trend and the yield factors are not close to the significance level of 5% when included individually. A similar analysis applies to the estimation output in panel B of Table 21. The high R^2, the trend in the variables and the latter fact that the variables are insignificant on their own, but are significant when included together leads to the conclusion that this increased significance

[12]The results are equivalent to the results by Bauer and Hamilton (2018) displayed in Table 6 and Figure 2.

may be spurious. In addition, the RR approach only partially alleviates the size distortion as its size is smaller than the size of the HAC t-test, but still not close to the nominal value of 5%. This shows that these tests reject the null too frequently while it is actually true. However, the power of each test is close to 1. Initializing the bootstrap samples at the population means improves the RR size substantially to a value of 15.8%, but it decreases the power even more to 31.9%.

Table 21: Cieslak and Povala: Trend Inflation

A. Original sample, 1974:11-2011:12	PC1	PC2	PC3	τ
Coefficient				
RR - only yields	0.003	0.240	-0.127	
RR - only trend				-0.051
RR - all	0.160	0.429	-0.059	-0.962
p-value				
RR - only yields	0.654	**0.013**	0.529	
RR - only trend				0.859
RR - all	**0.000**	**0.000**	0.748	**0.000**
HAC	**0.000**	**0.000**	0.654	**0.000**
Bootstrap RR				**0.000**
Bootstrap HAC				**0.001**
Size				
RR				0.436
HAC				0.551
Bootstrap				0.086
Power				
RR				0.998
HAC				1.000
Bootstrap				0.978
B. Later sample, 1985:1-2016:12				
Coefficient				
RR - only yields	0.019	0.180	-0.056	
RR - only trend				0.057
RR - all	0.106	0.297	0.061	-0.607
p-value				
RR - only yields	0.069	0.102	0.978	
RR - only trend				0.611
RR - all	**0.000**	**0.000**	0.541	**0.000**
Bootstrap RR				**0.035**

Note: Predictive regression for annual bond excess returns (y_{t+12}), averaged over two- to fifteen-year bond maturities. $y_{t+12} = \beta_0 + \beta_1^{'} x_{1t} + \beta_2^{'} x_{2t} + u_{t+12}$, where $x_{1t} = (PC1_t, PC2_t, PC3_t)^{'}$ and $x_{2t} = \tau_t$. Results in panel A are for the same period as studied by Cieslak and Povala (2015), panel B includes the same period as studied by Bauer and Hamilton (2018). p-values based on RR and HAC approach using Newey-West standard errors with 18 lags. p-values below 5% are in bold. Under *Size* we report estimates of the size of the tests, based on simulations from the BC bootstrap under the null hypothesis. All tests have a nominal size of 5%. Under *Power* we report estimates of the power of the tests.

Table 22 shows how the in-sample fit measured by the adjusted R^2 changes when the inflation trend is added to the model containing the yield factors. The R^2 increases by 33 percentage point to 50% due to the addition of the inflation trend. This implies that the inflation trend actually improves the in-sample fit. The second row of Panel A.I shows the average adjusted R^2 when adding the inflation trend to the model, while it does not contain explanatory power by construction. The increased R^2 of 50% is not included in the 95% confidence interval of the bootstrap adjusted R^2 and therefore the R^2 is too large to attribute to the yield factors only. However, when considering the later sample, the increase in R^2 due to the addition of the inflation trend is less pronounced. Moreover, this R^2 of 34% is not uncommon when the inflation trend actually does not contain explanatory power as it is contained in the 95% confidence interval resulting from the bootstrap procedure. Lastly, we use the period starting from January 2011 as a true out-of-sample period, by reestimating the model each month and evaluating the model performance by the $RMSE$. The $RMSE$-ratio is the $RMSE$ of the model including all variables relative to the $RMSE$ of the model only including the yield factors. While the in-sample ratio is smaller than one, indicating that the model including all variables has better predictive performance, the out-of-sample ratio is larger than one. Further, the DM-test tests the null of equal forecasting ability of the restricted and unrestricted model in the out-of-sample period. The low p-value and large $RMSE$-ratio point to the inflation trend exhibiting no predictive power for the new data when correcting for the information contained in the yield curve.

Table 22: Cieslak and Povala: Trend Inflation

A. In-sample performance

	R_1^2	R_2^2	$R_2^2 - R_1^2$
I. Original sample, 1971:11-20011:12			
Data	0.16	0.50	0.33
Bootstrap	0.18	0.25	0.07
	(0.02, 0.40)	(0.08, 0.45)	(-0.00, 0.21)
II. Later sample, 1985:1-2016:12			
Data	0.17	0.34	0.17
Bootstrap	0.28	0.34	0.06
	(0.06, 0.53)	(0.12, 0.56)	(-0.00, 0.22)

B. Out-of-sample performance

	In-sample	Out-of-sample	DM-test p-value
RMSE-ratio	0.603	3.213	**0.006**

Note: Adjusted R^2 of predictive regression for annual bond excess returns (y_{t+12}) using three PCs ($x_{1t} = (PC1_t, PC2_t, PC3_t)'$) and the inflation trend (τ_t) which are contained in x_{2t} ($x_{2t} = \tau_t$). Results in panel A.I are for the same period as studied by Cieslak and Povala (2015), and panel A.II includes the same period as studied by Bauer and Hamilton (2018). The first row of panels A.I and A.II reports the statistics in the original data. The following reports bootstrap mean and 95% percentiles in parentheses. The bootstrap procedure assumes that x_{2t} contains no incremental predictive power. The first column corresponds to the predictive regression that includes x_{1t} and the second column to the inclusion of x_{1t} and x_{2t}. Panel B assesses the predictive power for excess bond returns averaged across maturities by the $RMSE$. The $RMSE$-ratio is the relative $RMSE$ of a model including x_{1t} and x_{2t} and a model without x_{2t}. The in-sample period is 1971:11-20010:12, and the out-of-sample period is 20011:1-2015:12. The out-of-sample predictions are estimated using an expanding window approach. The Diebold-Mariano (DM) test for equal forecasting accuracy considering the prediction errors out-of-sample for the unrestricted and restricted model.

In conclusion, commonly used tests to take care of overlapping data and serially correlated errors lead to poor size, whereas the bias-corrected bootstrap test results in a better size. However, due to the endogeneity, persistence, and trend in the right-side variables, the significance of the yield factors and trend inflation may have arisen spuriously. Moreover, the increase in the adjusted R^2 resulting from the addition of the inflation trend to the model in the sample studied by Cieslak and Povala (2015) is too large to be attributed to the yield factors alone. However, considering a later subsample period leads to an increased R^2 that can appear when the inflation trend actually does not consist of explanatory power. Interestingly, when considering this later subsample as a true out-of-sample period, the predictive power of a model included the inflation trend deteriorates significantly.

E.3 Higher-Order PCs of Yields

In contrast to various studies proposing new macroeconomic predictors of excess bond returns, Cochrane and Piazzesi (2005) study the time variation in expected excess bond returns by using a single factor, a tent-shaped linear combination of forward rates. They find that this factor predicts excess returns on one- to five-year maturity bonds with R^2 up to 44% when including additional lags and up to 37% without lags. An important component of this factor is unrelated to the level, slope, and curvature movements described by most term structure models. They conclude that their forecasts are statistically significant even taking into account the small-sample properties of test statistics. They identify the inference problem that overlapping data and highly cross-correlated and autocorrelated right-hand variables may give. Therefore, they use a Hansen-Hodrick correction to handle overlapping observations, a Newey-West correction with 18 lags and compute the parameter covariance matrix using regressions with nonoverlapping data. Further, they compute three small-sample distributions for their test statistics but focus on the null of no predictability under the expectations hypothesis.

Table 23: Cochrane-Piazzesi: Higher-order PCs of yields

A. Original sample, 1964:1-2003:12

	PC1	PC2	PC3	PC4	PC5	Wald
Coefficient	0.127	2.740	-6.307	-16.128	-2.038	
p-value						
HAC	0.085	**0.000**	**0.003**	**0.000**	0.455	**0.000**
Bootstrap				**0.000**	0.511	**0.000**
Size						
HAC				0.086	0.081	0.103
Bootstrap				0.049	0.047	0.048
Power						
HAC				0.996	0.147	0.993
Bootstrap				0.989	0.104	0.983

B. Later sample, 1985:1-2022:12

	PC1	PC2	PC3	PC4	PC5	Wald
Coefficient	0.159	1.337	2.985	-8.264	-7.301	
p-value						
HAC	**0.008**	0.053	0.383	0.213	0.367	0.272
Simple bootstrap				0.299	0.447	0.418

Note: Predictive regression for annual bond excess returns (y_{t+12}), averaged over one- to five-year bond maturities. $y_{t+12} = \beta_0 + \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+12}$, where $x_{1t} = (PC1_t, PC2_t, PC3_t)'$ and $x_{2t} = (PC4_t, PC5_t)'$. Results in panel A are for the same period as studied by Cochrane and Piazzesi (2005), panel B includes the same period as studied by Bauer and Hamilton (2018). HAC statistic and p-values are calculated using Newey-West standard errors with 18 lags. The column *Wald* reports p-values for the hypothesis that $PC4$ and $PC5$ have no predictive power. p-values below 5% are in bold. Under *Size* we report estimates of the size of the tests, based on simulations from the simple bootstrap under the null hypothesis. Both tests have a nominal size of 5%. Under *Power* we report estimates of the power of the tests, based on simulations from the simple bootstrap under the alternative hypothesis.

Cochrane and Piazzesi (2005) use the same CP factor as Ludvigson and Ng (2009), however similar to Bauer and Hamilton (2018) we extract principal components from the Fama-Bliss yield data on one- to five-year maturity discount bonds. Therefore, Table 15 shows the factors resulting from the PCA, the variation explained by each factor, and the first and second-order autocorrelation. While the fourth and the fifth account for a minuscule variation Cochrane and Piazzesi (2005) identify that they capture dynamics that cannot be captured by the first three factors. In addition, this Table displays the high persistence of especially the first two factors. Similarly to the real-time experiment proposing macro factors to contain additional predictive power, we compare the original sample to an extended data set.[13]

Table 23 contains the regression output similar to previously discussed studies. It shows that the fourth PC is significant both by the commonly used t-test, as well as by the bootstrap procedure. However, the fifth factor does not appear as significant by both tests. In addition, the size of the bootstrap test is closer to the nominal size of 5%, than the conventional t-test showing its size distortion. The size distortions are only small, which is likely due to the low persistence of higher-order PCs. While the Wald tests identify the fourth and fifth factors as jointly significant in the original sample, they are not significant in the later sample shown in Panel B. In addition, the fourth PC is not significant anymore.

Table 24 reports the adjusted R^2 for a model including only the first three yield factors, and a model including all five. Panel A.I shows that the addition of the fourth and the fifth factor makes the R^2 increase by 9 percentage point to 35%. However, this increase is not uncommon when the fourth and the fifth factor do not actually contain any explanatory power as the R^2 of 35% is contained in the 95% confidence interval. A similar pattern can be observed for the later sample in Panel A.II, where the increase in the adjusted R^2 in the data is slightly smaller (2 percentage point). Panel B gives the ratio of $RMSE$ when including all five factors relative to only the first three factors. In-sample, the ratio is smaller than one indicating that the model including all five factors has more explanatory power. This is in contrast to the out-of-sample period, where the ratio larger than one implies that the model including five PCs actually performs worse. However, this deterioration in $RMSE$ by including the fourth and the fifth factor is only statistically significant at a 10% level.

Table 24: Cochrane-Piazzesi: Higher-order PCs of yields

A. In-sample performance

	R_1^2	R_2^2	$R_2^2 - R_1^2$
I. Original sample, 1964:1-2003:12			
Data	0.26	0.35	0.09
Bootstrap	0.21	0.22	0.01
	(0.06, 0.40)	(0.06, 0.41)	(0.00, 0.02)
II. Later sample, 1985:1-2022:12			
Data	0.16	0.18	0.02
Bootstrap	0.32	0.34	0.01
	(0.13, 0.52)	(0.14, 0.53)	(0.00, 0.04)

B. Out-of-sample performance

	In-sample	Out-of-sample	DM-test p-value
RMSE-ratio	0.890	1.141	0.091

Note: Adjusted R^2 of predictive regression for annual bond excess returns (y_{t+12}) using three PCs ($x_{1t} = (PC1_t, PC2_t, PC3_t)'$) the fourth and fifth PCs which are contained in x_{2t} ($x_{2t} = (PC4_t, PC5_t)'$). Results in panel A.I are for the same period as studied by Cochrane and Piazzesi (2005), and panel A.II includes the same period 1985:1-2022:12. The first row of panels A.I and A.II reports the statistics in the original data. The following reports bootstrap mean and 95% percentiles in parentheses. The bootstrap procedure assumes that x_{2t} contains no incremental predictive power. The first column corresponds to the predictive regression that includes x_{1t} and the second column to the inclusion of x_{1t} and x_{2t}. Panel B assesses the predictive power for excess bond returns averaged across maturities by the $RMSE$. The $RMSE$-ratio is the relative $RMSE$ of a model including x_{1t} and x_{2t} and a model without x_{2t}. The in-sample period is 1964:1-2002:12, and the out-of-sample period is 2003:1-2021:12. The out-of-sample predictions are estimated using an expanding window approach. The Diebold-Mariano (DM) test for equal forecasting accuracy considering the prediction errors out-of-sample for the unrestricted and restricted model.

[13] These results relate to Table 7 by Bauer and Hamilton (2018). We also include estimation results for an extended sample that includes observations that appeared since the publication by Bauer and Hamilton (2018).

In conclusion, the fourth PC is identified as significant at a 5% level by both the conventional t-test as well as by the bootstrap procedure. Similarly, the fourth and fifth PC are significant by the Wald tests. The persistence in higher-order PCs is low and that is why the size distortions are only small. When the model including the higher-order PCs is confronted with a newer data set, the PCs are not significant anymore, both individually as well as jointly. Therefore, the claim Cochrane and Piazzesi (2005) make that their factor captures a component that is unrelated to the first three PCs does not seem to hold. This factor is mainly a robust factor caused by the first three PCs. This is particularly shown by confronting the model with new data and reassessing the increased R^2.

F Additional Results - Currency Market

F.1 Currency Return Predictability - All Countries

We discuss the estimation output when estimating the regression for the currency basket including all currencies displayed in Table 25. First, we discuss the estimation results over the sample as originally studied by Filippou and Taylor (2017) displayed in Panel A. Again, we find that the t-tests are not as sharply distorted as the Wald test. However, for both the t-tests as well as the Wald test we observe that the distortion is not sufficient to overturn the conclusions from the tests. For example, the first domestic factor ($H1$) is indicated as statistically significant at a 5% level by both tests. Figure 13 shows the marginal R^2 when regressing the numbered series onto the factor. This factor represents mostly the housing category. Similarly, the macro factors are jointly significant at a 5% level. However, by the common Wald test even at a 0.01% level.

Next, we discuss panel B of Table 25 which contains the results of estimating the model over data that has appeared since the publication by Filippou and Taylor (2017). First, the first domestic and global factors are statistically significant at a 5% level. Second, it is important to use the bootstrap Wald test as the common test would lead to the conclusion that the macro factors are jointly significant even at a 1% level. The bootstrap Wald test indicates that this is not the case, even not at a 10% significance level.

Lastly, over the complete sample period as displayed in panel C of Table 25 the first, second, and third domestic factor and first global factor are statistically significant at a 5% significance level. Moreover, the macro factors are not jointly significant when using the bootstrap Wald test. Figure 14, 15 and 16 displays the marginal R^2 for $H2$, $H3$ and $G1$ resp. The second domestic factor loads mostly on the real output, employment, and housing category. The third domestic factor is very similar. The global factor loads most heavily on series from consumption, price indices, and international trade categories. One source of differences across countries is the composition of their trade which is found to be a determinant of the carry trade return (Ready, Roussanov & Ward, 2017). The importance of consumption as a driving force between currency returns is also numerously shown in previous studies, for example by Lustig and Verdelhan (2007) and Lustig et al. (2014).

Table 25: Carry-Trade Excess Return Predictability: All Countries

A. Early sample, 1985:7-2012:3

	AFD	H1	H2	H3	H4	H5	H6	H7	H8	H9	G1	G2	G3	Wald
Coefficient	1.845	0.104	-0.061	-0.164	-0.022	-0.180	0.127	-0.012	-0.131	-0.151	-0.219	-0.047	0.042	
p-value														
HAC	0.071	**0.026**	0.644	0.083	0.773	0.057	0.196	0.918	0.163	0.268	0.253	0.628	0.698	**0.000**
Bootstrap		**0.036**	0.657	0.102	0.787	0.082	0.235	0.924	0.178	0.289	0.276	0.642	0.712	**0.021**
Size (HAC)		0.064	0.066	0.065	0.072	0.074	0.071	0.068	0.058	0.059	0.061	0.065	0.062	0.228
Bootstrap		0.049	0.048	0.047	0.054	0.059	0.044	0.052	0.043	0.048	0.053	0.045	0.044	0.053
Power (HAC)		0.437	0.081	0.417	0.074	0.409	0.250	0.065	0.206	0.191	0.172	0.086	0.084	0.978
Bootstrap		0.374	0.065	0.364	0.055	0.359	0.207	0.057	0.174	0.159	0.146	0.072	0.061	0.908

B. Later sample, 2012:3-2023:2

	AFD	H1	H2	H3	H4	H5	H6	H7	H8	H9	G1	G2	G3	Wald
Coefficient	-9.670	0.180	-0.035	0.070	-0.188	0.093	-0.063	-0.029	-0.141	0.106	-0.318	-0.205	-0.084	
p-value														
HAC	**0.000**	**0.001**	0.669	0.334	0.069	0.298	0.809	0.793	0.188	0.322	**0.006**	0.212	0.515	**0.023**
Bootstrap		**0.005**	0.712	0.391	0.115	0.355	0.822	0.819	0.260	0.407	**0.018**	0.268	0.573	0.493

C. Complete sample, 1985:7-2023:2

	AFD	H1	H2	H3	H4	H5	H6	H7	H8	H9	G1	G2	G3	Wald
Coefficient	0.975	0.086	-0.195	-0.184	0.119	0.034	-0.160	-0.100	-0.122	-0.199	-0.439	-0.035	0.029	
p-value														
HAC	0.276	**0.007**	**0.025**	**0.027**	**0.039**	0.630	0.118	0.290	0.261	0.177	**0.018**	0.768	0.740	0.055
Bootstrap		**0.012**	**0.035**	**0.035**	0.050	0.645	0.135	0.310	0.271	0.196	**0.021**	0.771	0.744	0.215

Note: Predictive regression for currency excess returns of the carry-trade strategy using only the all countries currency basket ($y_{t+1} = rx^6_{net,t+1} - rx^1_{net,t+1}$). $y_{t+1} = \beta_0 + \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+1}$, where $x_{1t} = AFD_t$ and $x_{2t} = (H1_t, H2_t, ..., H9_t, G1_t, G2_t, G3_t)'$. Results in panel A are for the same period as studied by Filippou and Taylor (2017), Panel B for the sample 2012:3-2023:2, and Panel C for the complete sample period. HAC statistic and p-values are calculated using Newey-West standard errors with 4 lags. The column *Wald* reports p-values for the hypothesis that macro factors have no predictive power. p-values below 5% are in bold. Bootstrap indicates the Bootstrap procedure presented in Appendix B where we obtain bootstrap samples under $H_0 : \beta_2 = 0$.
Under *Size*, we report estimates of the size of the tests, based on simulations from the simple bootstrap under the null hypothesis. Both tests have a nominal size of 5%.
Under *Power*, we report estimates of the power of the tests, based on simulations from the simple bootstrap under the alternative hypothesis.

Table 26: Currency Excess Return Predictability

A. In-sample performance

	R_1^2	R_2^2	$R_2^2 - R_1^2$
I. 1985:7-2012:3			
Data	-0.30	8.56	8.86
Bootstrap	0.01	0.02	0.01
	(-0.31, 1.35)	(-2.58, 3.78)	(-2.44, 3.61)
II. 2012:3-2023:2			
Data	5.12	7.87	2.76
Bootstrap	5.54	5.60	0.06
	(-0.46, 15.41)	(-4.03, 17.87)	(-5.99, 8.47)
III. 1985:7-2023:2			
Data	0.05	1.88	1.82
Bootstrap	0.27	0.27	0.00
	(-0.22, 1.89)	(-1.71, 3.20)	(-1.73, 2.66)

B. Out-of-sample performance

	In-sample	Out-of-sample	DM-test p-value
RMSE-ratio	0.957	1.012	0.736

Note: Adjusted R^2 (in %) of predictive regression for currency excess returns using the average forward discount ($x_{1t} = AFD_t$) and factors extracted from a large macro data set ($x_{2t} = (H1_t, H2_t, .., H9_t, G1_t, G2_t, G3_t)^{'}$). Results in Panel A.I are for the same period as studied by Filippou and Taylor (2017). Panel A.II uses newer data and Panel A.III contains all data available. The first row of each panel reports the statistics in the original data, followed by the bootstrap mean and 95% percentiles in parentheses. The bootstrap procedure assumes that the macro factors do not contain incremental predictive power. The first column corresponds to the predictive regression only including the average forward discount, the second column includes all regressors, and the third column is the difference in adjusted R^2 between these models. Panel B assesses the predictive power by the $RMSE$. The $RMSE$-ratio is the relative $RMSE$ of a model including all regressors and a model without the macro factors. The in-sample period is 1983:11-2012:3, and the out-of-sample period is 2012:4-2023:1. The out-of-sample predictions are estimated using an expanding window approach. The Diebold-Mariano (DM) test for equal forecasting accuracy considering the prediction errors out-of-sample for the unrestricted and restricted model.

The right panel of Table 26 contains the regression R^2 of the models estimated over the different subsamples. In contrast to currency basket including only the developed countries, the addition of the macro factors to the models estimated for all countries currency basket leads to a substantial increase in model fit. This is especially apparent in the sample considered by Filippou and Taylor (2017), shown in panel I with an increase of 8.81 percentage points. Panel II documents the model fit for the model estimated over the data since the publication by Filippou and Taylor (2017). In contrast to the developed countries currency basket, we now observe that the adjusted R^2 increases by 2.75 percentage points to a value of 7.87%. Panel B of Table 26 shows the $RMSE$-ratio of a model including all variables relative to a restricted model including only the AFD. For the in-sample period, starting from July 1985 and ending in March 2012, the macro factors reduce the $RMSE$ by 4.3%. For the out-of-sample period, starting from April 2012, the $RMSE$ increases with 1.2% using a rolling-window regression. However, this increase in $RMSE$ is not statistically significant as shown by the DM-test. Similar to the currencies of developed countries, Figure 9 graphs the forecasts of the model with and without the macro factors compared to observed carry-trade risk premia. We generally observe that both models have difficulty capturing the high variability of the risk premia.

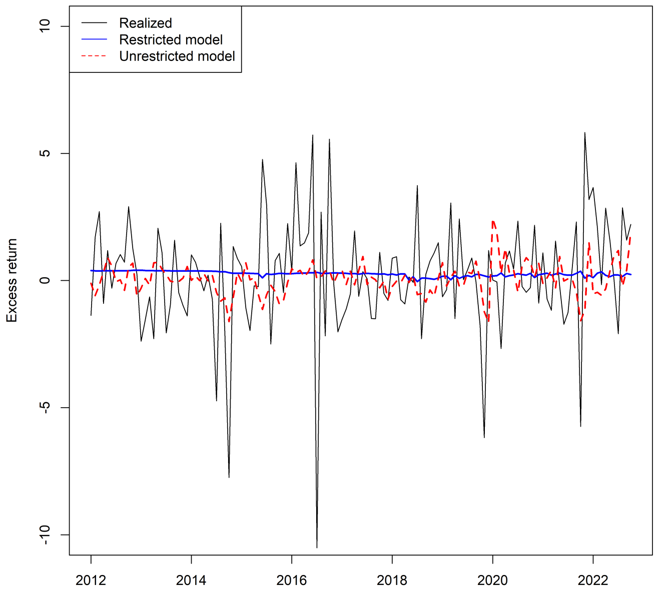

Figure 9: Out-of-sample prediction of the payoff from a carry trade strategy on all countries. The out-of-sample period starts in April 2012 and ends in January 2023. The model is reestimated each month using an expanding window. The restricted model uses only the average forward discount of the developed countries' basket, whereas the unrestricted model extends this by including nine macro factors on the US economy and three macro factors on the global economy.

F.2 Interpretation of Macro Factors

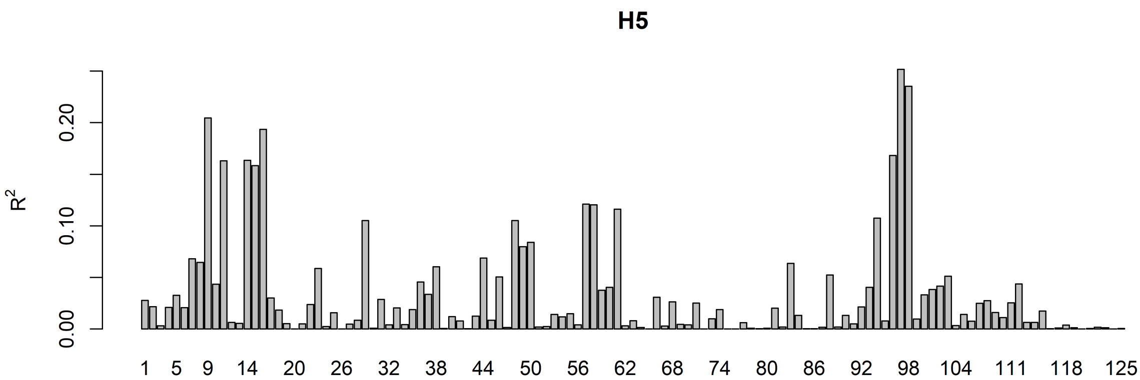

Figure 10: Figure shows the R^2 from regressing the series number given on the x-axis onto the estimated factor named in the heading. The factor is estimated using data from 1985:7-2012:3.

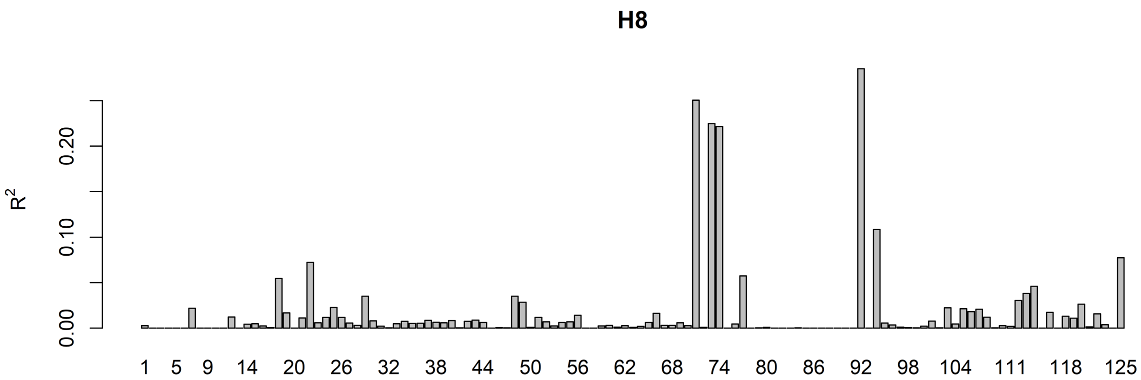

Figure 11: Figure shows the R^2 from regressing the series number given on the x-axis onto the estimated factor named in the heading. The factor is estimated using data from 1985:7-2023:2.

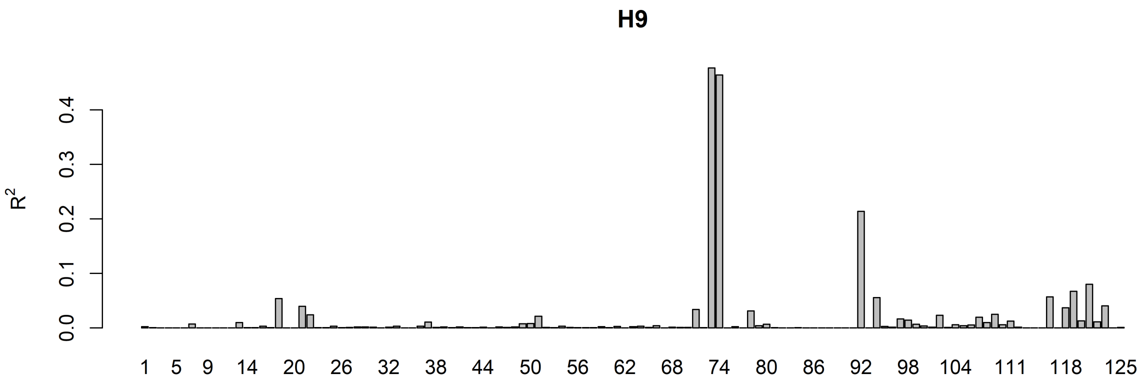

Figure 12: Figure shows the R^2 from regressing the series number given on the x-axis onto the estimated factor named in the heading. The factor is estimated using data from 1985:7-2023:2.

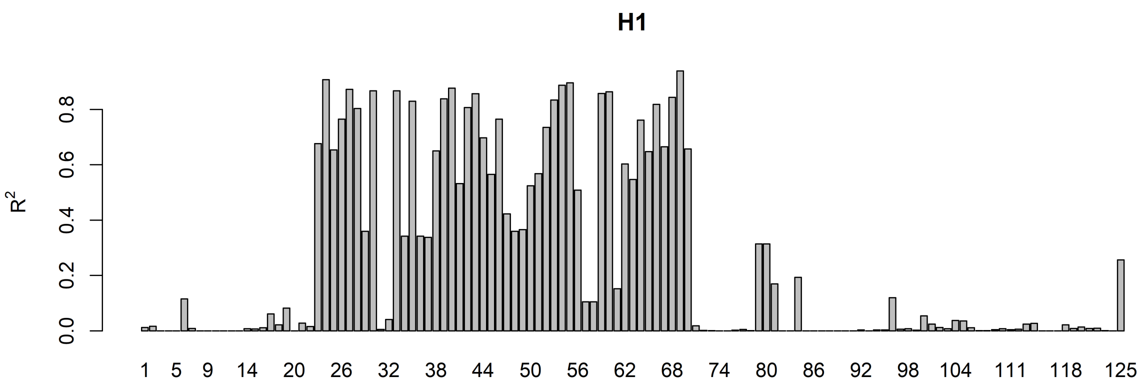

Figure 13: Figure shows the R^2 from regressing the series number given on the x-axis onto the estimated factor named in the heading. The factor is estimated using data from 1985:7-2023:2.

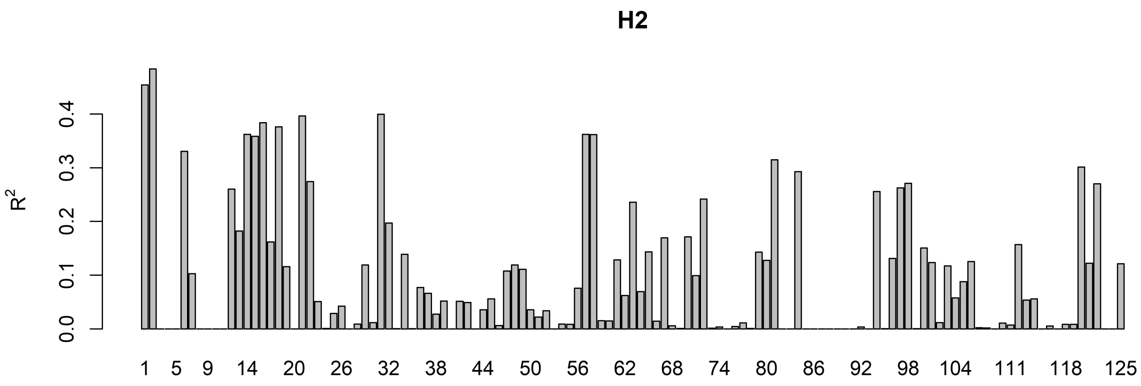

Figure 14: Figure shows the R^2 from regressing the series number given on the x-axis onto the estimated factor named in the heading. The factor is estimated using data from 1985:7-2023:2.

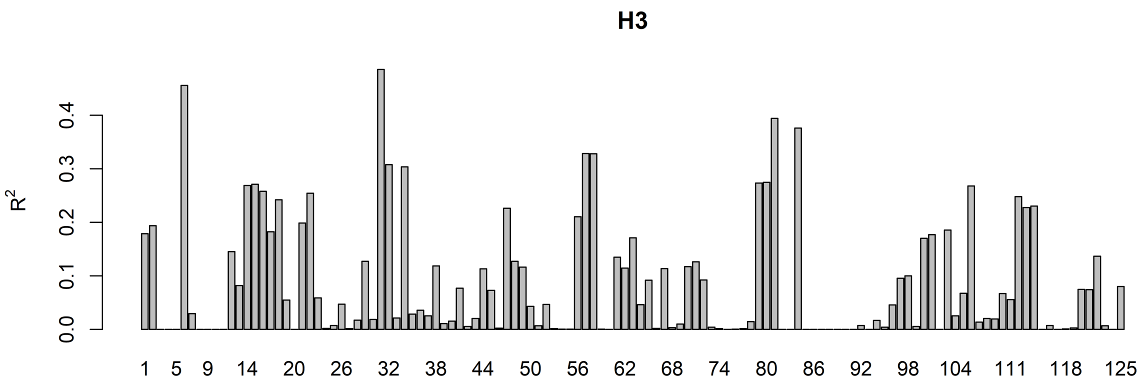

Figure 15: Figure shows the R^2 from regressing the series number given on the x-axis onto the estimated factor named in the heading. The factor is estimated using data from 1985:7-2023:2.

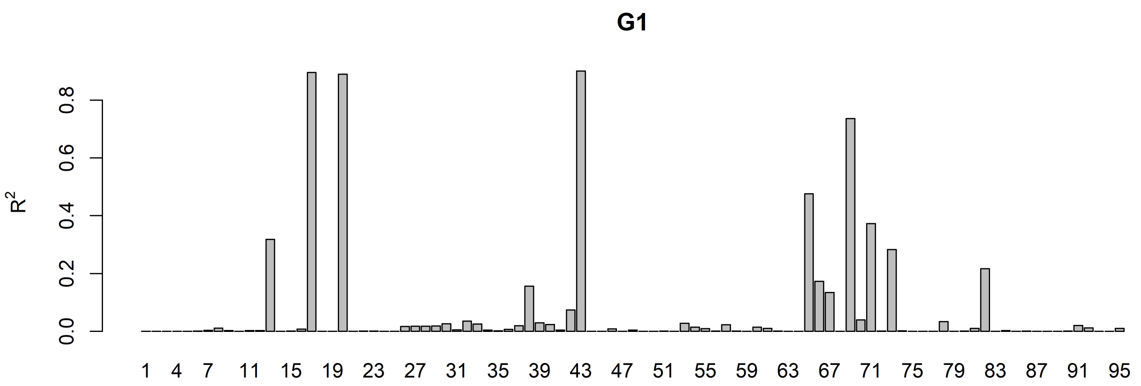

Figure 16: Figure shows the R^2 from regressing the series number given on the x-axis onto the estimated factor named in the heading. The factor is estimated using data from 1985:7-2023:2.

G Code

The code, written in R, follows the same set-up as Bauer and Hamilton (2018).[14] Therefore it should be sufficiently self-explanatory and self-contained that you should be able to replicate the results simply by following these steps: 1. Unpack the ZIP file into a folder 2. Start an R session 3. Set the working directory to that folder using setwd() 4. Run the code in each of the scripts in the folder, e.g. source('jps.r')

Note: You may need to install additional R packages, e.g. install.packages('sandwich')

The folder is divided into two parts. The folder 'replication' contains the scripts and data to replicate each of the six studies in the bond market. The folder 'Extension' contains the scripts and data to replicate and extend the study in the currency market.

The folder 'replication' created by Bauer and Hamilton (2018) contains scripts to replicate each of the six studies in the bond market. This folder contains a folder 'data' which contains the data sets used to replicate these studies. We extend the data set to reinvestigate the study by Ludvigson and Ng (2009) (LN) by adding new Fama-Bliss bond data and data from the FRED-MD database. In addition, figures are added to most replicated studies which are stored in the folder 'Figures'.

The script 'functions.R' contains additional functions that are loaded when the other scripts in the folder 'Extension' are run.

The folder 'Data' in the folder 'Extension' contains the data we use in the empirical case study of explaining carry trade returns.

Furthermore, each script contains a header indicating to which table and/or figure it belongs. In addition, the folder 'Extension' contains a file 'readme.txt' with an outline to run the code to obtain results in this paper.

[14] The code and data can be downloaded here Code and Data