How puzzling is the forward premium puzzle? A meta-analysis

The study uses meta-analysis to investigate why different papers report conflicting evidence on the forward premium puzzle and finds that the puzzle is much less prevalent than commonly thought.

Diana Zigraiova
European Stability Mechanism
Tomas Havranek
Charles University, Prague
Jiri Novak
Charles University, Prague

European Stability Mechanism

How puzzling is the forward premium puzzle? A meta-analysis

Diana Zigraiova [1] European Stability Mechanism
Tomas Havranek [2] Charles University, Prague
Jiri Novak [3] Charles University, Prague

Abstract

A key theoretical prediction in financial economics is that under risk neutrality and rational expectations, a currency's forward rates should form unbiased predictors of future spot rates. Yet scores of empirical studies report negative slope coefficients from regressions of spot rates on forward rates, which is inconsistent with the forward rate unbiasedness hypothesis. We collect 3,643 estimates from 91 research articles and using recently developed techniques investigate the effect of publication and misspecification biases on the reported results. Correcting for these biases yields slope coefficients of 0.31 and 0.98 for developed and emerging currencies respectively, which implies that empirical evidence is in line with the theoretical prediction for emerging economies and less puzzling than commonly thought for developed economies. Our results also suggest that the coefficients are systematically influenced by the choice of data, numeraire currencies, and estimation methods. The findings can be applied to calibrating carry trade strategies for individual currencies.

Keywords: Forward rate bias, uncovered interest parity, meta-analysis, publication bias, model uncertainty

JEL codes: C83, F31, G14

[1] d.zigraiova@esm.europa.eu
[2] tomas.havranek@ies-prague.org
[3] jiri.novak@fsv.cuni.cz

Disclaimer

ISSN 2443-5503
ISBN 978-92-95085-96-1

doi:10.2852/ 438843
EU catalog number DW-AB-20-005-EN-N

How puzzling is the forward premium puzzle? A meta-analysis*

Diana Zigraiova[a,b], Tomas Havranek[b], and Jiri Novak[b]

[a]European Stability Mechanism, Luxembourg
[b]Charles University, Prague

July 29, 2020

Abstract

A key theoretical prediction in financial economics is that under risk neutrality and rational expectations a currency's forward rates should form unbiased predictors of future spot rates. Yet scores of empirical studies report negative slope coefficients from regressions of spot rates on forward rates. We collect 3,643 estimates from 91 research articles and using recently developed techniques investigate the effect of publication and misspecification biases on the reported results. Correcting for these biases yields slope coefficients of 0.31 and 0.98 for developed and emerging currencies respectively, which implies that empirical evidence is in line with the theoretical prediction for emerging economies and less puzzling than commonly thought for developed economies. Our results also suggest that the coefficients are systematically influenced by the choice of data, numeraire currency, and estimation method. The findings can be applied to calibrating carry trade strategies for individual currencies.

Keywords: Forward rate bias, uncovered interest parity, meta-analysis, publication bias, model uncertainty

JEL Codes: C83, F31, G14

1 Introduction

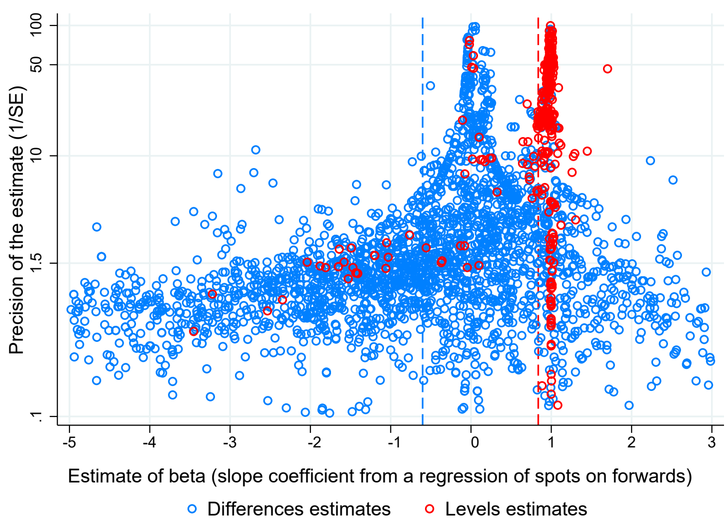
If forward exchange rates systematically differ from future spot rates, money can be made on the difference: a risk-neutral agent with rational expectations can exploit the inefficiency and hence, supposedly, the anomaly should disappear. It is therefore puzzling that the forward anomaly has been found again and again for dozens of different currencies, time periods, and identification designs. Yet the exact results in the literature vary, and the null hypothesis is not rejected universally. The anomaly is commonly labeled "forward premium puzzle," because

*An online appendix with data and code is available at `meta-analysis.cz/forward`. Corresponding author: Jiri Novak, `jiri.novak@fsv.cuni.cz`. We thank the participants of the MAER-Net Colloquium in Greenwich, 2019, and seminar participants at the European Stability Mechanism and the Crawford School of Public Policy, Australian National University, for their helpful comments. This project received support from the European Union's 2020 Research and Innovation Staff Exchange programme under the Marie Sklodowska-Curie grant agreement #681228 and from the Czech Science Foundation (grant #18-02513S). The views expressed here are ours and not necessarily those of our employers.

most studies do not estimate the relationship in levels (future spot rates on forward rates), but subtract current spot rates from both sides of the regression. Thus one obtains currency depreciation on the left-hand side and the forward discount on the right-hand side. The puzzle is that, according to many studies, depreciation is positively associated with forward *premium*, not a discount. So not only do researchers typically reject the hypothesis of the coefficient (we will call it β) being equal to one, but often they find a statistically significant negative coefficient. If the covered interest parity holds, this result is equivalent to the finding that currencies with higher interest rates tend to appreciate. Thus tests of forward rate unbiasedness are closely related to tests of the uncovered interest parity.

The forward premium puzzle is a traditional problem in international economics and finance; as such, it has attracted the attention of dozens of researchers during the last four decades. Yet still no clear-cut consensus emerges on whether the puzzle really exists or whether it represents a statistical artifact, how large the departure from the null hypothesis is, and how material the implications are in practice. Important prospective solutions to the forward premium puzzle put forward in the last decade include infrequent portfolio decisions (Bacchetta & van Wincoop, 2010), investor overconfidence (Burnside *et al.*, 2011), omitted variables (Pippenger, 2011), sentiment (Yu, 2013), sovereign default risk (Coudert & Mignon, 2013), order flow (Breedon *et al.*, 2016), and inflation targeting (Coulibaly & Kempf, 2019), a string of efforts that highlights persistent research activity in the field. What the literature lacks is a quantitative synthesis, or meta-analysis, that would take stock of the enormous body of work and shed light on potential biases and patterns that are impossible to detect in individual studies considered separately. That is what we attempt to achieve in this paper.

Figure 1: *Are positive estimates underreported?*



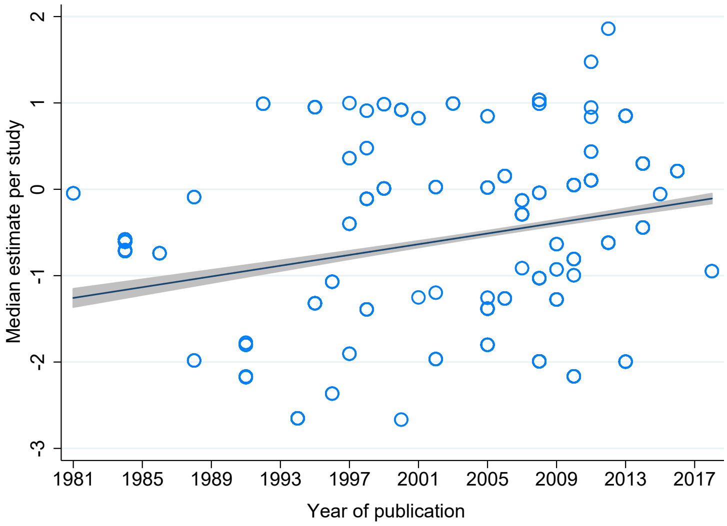
Notes: In the absence of publication bias, the scatter plot should resemble an inverted funnel that is symmetrical around the most precise estimates. Red circles (darker in grayscale) represent estimates extracted from the levels equation (Equation 2); blue circles (lighter in grayscale) show estimates from the differences equation (Equation 3). Outlying observations are cut from the figure for ease of exposition but included in all statistical tests. The red dashed vertical line shows the mean estimate from the levels specification (0.84); the blue dashed vertical line shows the mean from the differences specification (-0.60).

Figure 1 illustrates one type of insight a meta-analysis can bring on top of the results of individual studies. The insight concerns publication bias, i.e. the potential tendency of authors, editors, and referees to prefer results that are statistically significant and consistent with previous findings or underlying theory. The bias has been discussed, among others, by Havranek (2015), Brodeur *et al.* (2016), Bruns & Ioannidis (2016), Christensen & Miguel (2018), Brodeur *et al.* (2020), and Blanco-Perez & Brodeur (2020). Ioannidis *et al.* (2017) show that publication bias looms large in economics and finance, exaggerating the mean reported coefficient twofold. That is not to say the bias arises intentionally: for better or worse, many researchers use statistical significance as an indicator of importance, and select the results for publication accordingly. A useful analogy appears in McCloskey & Ziliak (2019), who compare publication bias to the Lombard effect in biology, when speakers increase their effort with increasing noise. With large imprecision given by noisy data or inefficient estimation techniques, researchers may strive for larger estimates that would still produce sufficiently large t-statistics. In consequence, a correlation between estimates and the corresponding standard errors arises.

Figure 1 shows the traditional visual test, originating from medical research, of the correlation between estimates and standard errors (here precision, the inverse of standard error), and thus of publication bias. The figure is called a funnel plot, because in the absence of publication bias, the observations should form a symmetrical inverted funnel. Intuitively, the most precise estimates should be close to the underlying mean value of the parameter in question, while less precise estimates should be more dispersed, giving rise to the funnel shape. The funnel should be symmetrical because there is no reason for imprecise negative and positive estimates to have a different probability of publication. Figure 1 shows two groups of estimates: red ones (darker in grayscale) are derived from levels regressions, spots on forwards. Blue ones (lighter in grayscale) are derived from differences regressions, depreciation on the forward discount. Two observations stand out. First, levels estimates do not form a funnel, but are almost always very close to 1, with little dispersion irrespective to precision. The observation is consistent with levels estimates being attracted to 1 via spurious regression. Second, the funnel for differences estimates is asymmetrical, which is consistent with publication bias. The most precise estimates are around zero or mildly positive, but many imprecise positive estimates seem to be missing from the literature, which leads to the overall observed mean of -0.6.

We test for publication bias formally using the recently developed nonlinear techniques due Ioannidis *et al.* (2017, weighted average of adequately powered estimates), Andrews & Kasy (2019, selection model), Bom & Rachinger (2019, endogenous kink), and Furukawa (2019, stem-based technique), which are all, to some extent, based on the Lombard effect, but allow for a nonlinear relationship between the magnitude of publication bias and the size of the standard error. Our results based on these techniques suggest substantial publication bias. The corrected mean β estimates vary between 0.2 and 0.8 depending on the method, far from the simple arithmetic average of -0.6 computed from all the reported estimates. The pattern we observe in the literature is thus consistent with a type of publication bias called confirmation bias, i.e. the tendency to publish results that corroborate the famous finding on the negative coefficient by Fama (1984) rather than estimates that point in the opposite direction. Thus, correcting the literature for publication bias makes the forward premium puzzle look much less puzzling than previously suggested.

Figure 2: *Do estimates of β increase with time?*



Notes: The horizontal axis denotes the year when the first version of the paper appeared in Google Scholar. The solid line represents a linear trend and the surrounding shaded area shows the corresponding 95% confidence band. Only differences estimates (Equation 3) are included.

We further explore how the published β estimates vary with the choice of data samples and estimation methodology. Different studies use different data, different techniques, and are published in journals of different reputation. For example, Froot & Thaler (1990) collected 75 such estimates published until the end of the 1980s and reported their mean to equal -0.88. Figure 2 shows that even without any correction for a potential publication bias, the β estimates from the differences regression exhibit a tendency to increase over time, starting with values around -1 in the 1980s and approaching values close to 0 at the end of the 2010s. To capture the context in which individual estimates are obtained, we collect 43 corresponding variables and then regress the reported estimates on these variables. Because of model uncertainty inherent in such an exercise, we cannot place all the 43 variables into one regression, but have to use model averaging techniques that run millions of different regressions with various combinations of the 43 variables and then weight these models according to fit and complexity. We employ both Bayesian (Eicher *et al.*, 2011; Steel, 2020) and frequentist (Hansen, 2007; Amini & Parmeter, 2012) model averaging.

Our results suggest that several data, method, and publication characteristics systematically affect the reported estimates. The most robust findings concern differences among individual currencies. The estimates for the currencies of emerging economies tend to be much larger than estimates for developed economies, even if we control for other features in which studies vary. This finding corroborates that of Frankel & Poonawala (2010), who also report that much less evidence exists for the forward premium puzzle in emerging economies compared to developed countries. Moreover, we also find substantially above-average estimates for the former French franc and Italian lira, while substantially below-average estimates for the euro, British pound, Japanese yen, and Swiss franc. Thus even among the currencies of developed countries, the less risky ones are associated with more evidence of the puzzle. Taken together, our results support the conclusion of Frankel & Poonawala (2010) that a time-varying exchange risk premium does not represent a perspective explanation of the forward premium puzzle, because larger risk premia are typically more volatile.

As the bottom line of our analysis, we compute a synthetic β that uses all the available results reported in the literature but, aside from correcting for publication bias, gives more weight to estimates that are based on arguably more reliable and larger datasets, employ modern estimation techniques, and are published in the best journals. The synthetic β is constructed using the parameters from model averaging and choosing values for each variable (for example, sample maximum for data size and the impact factor of the journal in which the study was published). We obtain a coefficient of 0.31 for the currencies of developed economies and 0.98 for emerging economies. Thus, exploiting the heterogeneity of published studies and correcting for the publication selection bias produces β estimates which suggest that for many currencies the forward premium puzzle is less puzzling than previously thought. For emerging economy currencies the estimated value of 0.98 is very close to the theoretical prediction of 1. For developed economy currencies the estimated value of 0.31 is well below 1; nevertheless, in contrast to the common interpretation of prior findings, it is positive. Even after correcting for publication and misspecification biases we document negative β estimates for the Swiss franc (-0.03), the Japanese yen (-0.39), and especially for the euro (-0.71). Meta-analysis is thus no panacea, and there remains scope for other explanations to the puzzle.

2 Testing Forward Rate Unbiasedness

In this section we briefly describe how the coefficient β is typically estimated in the literature; further details are provided in Section 5 and in the studies quoted in this section. We start with the straightforward theoretical relationship between forward and future spot rates. The forward rate should differ from the expected spot exchange rate by a premium rp_{t+k}, which is a compensation for the perceived risk of holding different currencies based on information available at time t. This can be written in logarithms as

$$f_{t,t+k} = E_t s_{t+k} + rp_{t+k}, \tag{1}$$

where $f_{t,t+k}$ is the forward value of the spot exchange rate s_t for a contract signed in period t that expires k periods in the future.

Since the expectation term in Equation 1 is not directly observable, researchers typically invoke rational expectations. Coupled with the assumption of risk neutrality, we arrive at the following regression:

$$s_{t+k} = \alpha + \beta * f_{t,t+k} + \upsilon_t. \tag{2}$$

In practice, however, researchers often subtract s_t from both sides of Equation 2:

$$s_{t+k} - s_t = \alpha + \beta * (f_{t,t+k} - s_t) + \nu_t, \tag{3}$$

which has two benefits: i) both sides of the equation can now be typically considered stationary, ii) both sides also have an intuitive interpretation in percentage points, the left-hand side denoting depreciation, the right-hand side representing the forward discount. For the forward rate unbiasedness hypothesis to hold, and thus for the absence of the forward premium puzzle, α should equal 0, β should equal 1, and ν_t should be serially uncorrelated. In practice, most researchers focus on the slope coefficient β, and β is also our focus in this meta-analysis. A large body of literature has had troubles confirming that β equals 1. What is more, researchers frequently find that β is zero or even negative (e.g., Backus *et al.*, 1993; Hai *et al.*, 1997; Bekaert, 1995; Byers & Peel, 1991; MacDonald & Taylor, 1990; McFarland *et al.*, 1994). Froot & Thaler (1990), on the basis of 75 published regressions, compute that the average β is equal to -0.88. Moreover, under covered interest parity we have

$$f_{t,t+k} - s_{t,k} = i_{t,t+k} - i^*_{t,t+k}, \tag{4}$$

where $i_{t,t+k}$ denotes the logarithm of one plus the interest rate paid on domestic assets for k periods while $i^*_{t,t+k}$ applies to the rate paid on foreign assets. So Equation 3 can be also thought of as a test of uncovered interest parity, similarly rejected by a large body of research.

The literature has attempted to explain the frequent rejection of the null hypothesis that β equals 1 in various ways. First, some authors attribute it solely to statistical considerations. It is only correct to regress the change in the spot exchange rate on the forward premium in Equation 3 if both variables are stationary. The forward premium thus needs to be integrated of order zero as well. Goodhart *et al.* (1997) attest that Equation 3 would be misspecified and the measured value β would be biased towards zero if forward premium was not *I(0)*. Crowder (1994) fails to reject the presence of a unit root in the forward premium series while Baillie & Bollerslev (1994) find forward premiums to be fractionally integrated processes. Second, Fama (1984) explains the frequent rejection of the null hypothesis by highly variable rational expectations risk premia in Equation 3. Numerous other studies support this result: for example, Domowitz & Hakkio (1985), Wolf (1987) and Baillie & Bollerslev (1989). On the other hand, Frankel (1982), Frankel (1986) and Frankel & Froot (1987) do not confirm the presence of significant risk premia and report instead that the empirical rejection of the unbiasedness hypothesis implies that expectations are generally not rational due to excessive speculations.

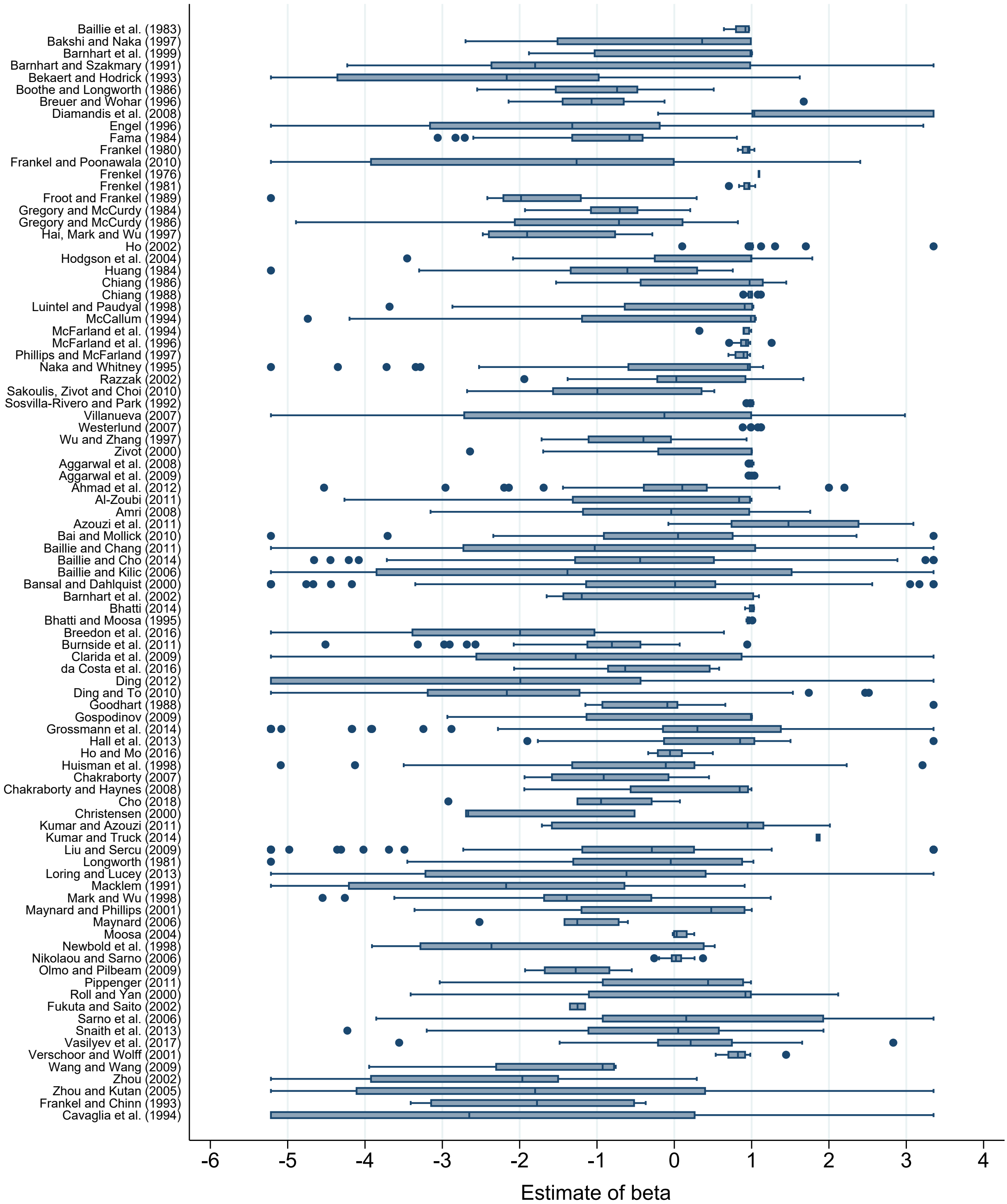
While the aforementioned explanations for the finding of the forward premium puzzle are the most prominent ones, others have been put forward in the literature. McCallum (1994) attributes the rejection of the hypothesis to the fact that monetary authorities aim to avoid sudden exchange rate changes and thus smooth interest rates. As a result, tests of forward unbiasedness suffer from the absence of an equation that would take into account the behavior of the monetary authority. Furthermore, the theory behind the forward rate unbiasedness does not indicate whether long-term or short-term interest rates should be used or whether using T-bill rates as opposed to rates on commercial papers should matter. To address this issue, Razzak (2002) performs the test using one-year forward exchange rates instead of one-month rates and finds support for the null hypothesis when exchange rates are measured in US dollars. Nevertheless, no such support for the hypothesis emerges when other currencies are used for measuring exchange rates. In a similar vein, studies by Mussa (1979), Chinn & Meredith (2004), and Nadal De Simone & Razzak (1999) corroborate that long-term rates are more suitable for explaining the movements of spot exchange rates in tests of forward rate unbiasedness.

The exchange rate regime, time period, stage of a country's development, and data contamination have also appeared to factor in the testing of the hypothesis. Flood & Rose (1996) offer evidence that negative estimates of β hold only for floating exchange rate regimes. Using data for the European Monetary System they show that a large part of the forward discount puzzle vanishes for fixed exchange rate regimes. Frankel & Poonawala (2010) show that for emerging market currencies, there is a smaller bias in the estimated β. The coefficient is on average positive and never significantly less than zero. The study by Chiang (1988) indicates that empirical evidence against the forward rate unbiasedness hypothesis might disappear if the regression parameters are allowed to vary in time. As for data contamination, Cornell (1989) argues that estimates of the slope coefficient are biased towards $\beta < 1$ due to data mismatch. He suspects that most studies do not find exactly the future spot exchange rate that corresponds to the forward rate in their data and proposes the use of a lagged forward discount as the right-hand-side variable in Equation 3 to deal with this problem. Using this technique Cornell (1989) cannot reject the forward rate unbiasedness for the Canadian dollar/US dollar rate. He, however, rejects the hypothesis for other currencies relative to the US dollar. Moreover, following the recommendation put forward by Cornell (1989), Bekaert & Hodrick (1993) investigate the unbiasedness hypothesis for the mark, pound and yen relative to the US dollar and find β significantly negative. Given the conflicting findings in the literature, it is surprising that no quantitative synthesis of the empirical evidence has been conducted. In the next section we describe the first step in such a meta-analysis, data collection.

3 Data

We use Google Scholar to search for studies estimating the forward rate unbiasedness hypothesis. Google's algorithm goes through the full text of studies, thus increasing the coverage of suitable published estimates, irrespective of the precise formulation of the study's title, abstract, and keywords. This is the key advantage in contrast to other databases commonly used in research synthesis, such as the Web of Science. Our search query contains expressions "forward rate unbiasedness," "forward premium puzzle," "forward discount puzzle," "forward premium anomaly," "foreign exchange efficiency," and "forward rate spot rate." We inspect the papers returned by the search for the lists of references to check whether we can find usable studies not returned by our baseline search, a method called "snowballing" in the literature on research

Figure 3: *Estimates vary both within and across studies*

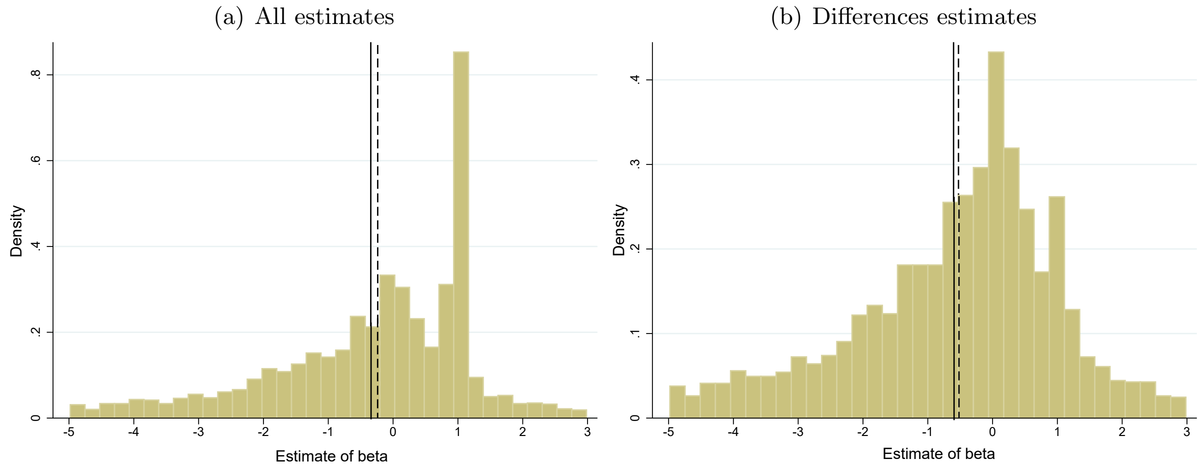


Notes: The figure shows a box plot of estimates of β reported in individual studies. The box plot shows interquartile range (IQR) (25th -75th percentile) and the median. Whiskers cover (25th percentile - 1.5*IQR) to (75th percentile + 1.5*IQR). The dots denote outlying estimates.

synthesis. We terminate the search on July 31, 2018, and do not add any new studies beyond that date.

To be included in our dataset, a study must meet three criteria. First, at least one estimate in the study must originate from an equation regressing either s_{t+k} on $f_{t,t+k}$ (Equation 2) or regressing $s_{t+k} - s_t$ on $f_{t,t+k} - s_t$ (Equation 3), as described in Section 2. That is, we do not collect estimates from studies that focus on uncovered interest parity and replace the forward discount with the interest rate differential. While such estimates are comparable to our dataset under the assumption of covered interest parity, the covered interest parity does not have to hold for all markets, especially after the financial crisis. Second, the study must be published. This criterion is mostly due to feasibility since even after restricting our efforts to published studies, the dataset involves a manual collection of hundreds of thousands of data points. In any case, studies published in journals can be expected to contain fewer typos and be, on average, of higher quality due to peer review. Third, the study must report standard errors of the estimated β or other statistics from which the standard error can be computed. This requirement is necessary for tests of publication bias.

Figure 4: *Distribution of the estimates*



Notes: The solid vertical line indicates the mean reported estimate of β; the dashed vertical line shows the mean of the median estimates reported per study.

Using the search queries and the study inclusion criteria specified above, we obtain 3,643 estimates of the slope coefficient β from 91 published studies, which makes our paper one of the largest meta-analyses ever conducted in economics and finance. For the list of primary studies included in our meta-analysis, please see Appendix B. All data and codes are available in an online appendix at `meta-analysis.cz/forward`. To ensure that outliers do not drive our results, we winsorize the collected estimates and their standard errors at the 5% level. The main results, however, are not sensitive to the chosen level of winsorization. Figure 3 shows a box plot of the estimates. We can observe the estimates vary greatly both within and across studies, with most studies reporting both negative and positive estimates of β. The mean of all estimates is -0.34, which confirms that the finding of the forward premium puzzle is common in the literature. The histogram of all the estimated coefficients (the left panel of Figure 4) has two peaks: 0 and 1, while its left tail is much longer than the right tail, suggesting a relatively greater representation of negative than positive estimates in the literature. This seems to be in line with the prevalence of negative estimates of β reported in primary studies and could represent a type of confirmation bias, in which the findings consistent with the prevalent view are more likely to get selected for reporting and publication. In their survey, Jongen *et al.* (2008) observe that a negative association between currency depreciation and forward discount constitutes common wisdom in the literature.

In the preceding paragraph we discussed the distribution of all the estimates of β from the studies included in our dataset. Consequently, we did not differentiate between estimates originating from the two main tests of forward rate unbiasedness conducted in the literature, estimates computed in levels and differences. There are only 654 estimates obtained from Equation 2 in our dataset (levels estimates), of which only 5.4% are negative and representing the puzzling result that the forward rate is negatively related to the future spot rate. On the other hand, differences estimates extracted from Equation 3 are much more numerous: there are 2,989 of these estimates in our dataset (their distribution is depicted in the right-hand panel of Figure 4). Out of these 58.4% are negative, and the mean is -0.6, which is in line with the ongoing quest in the literature to explain the predominance of puzzling results. As we already discussed in the Introduction, the levels estimates are problematic because of the likely unit root, and few modern studies use Equation 2. For this reason, in the remainder of the analysis we focus exclusively on the differences estimates.

Table 1: *Results for different currencies vary widely*

Country	**Mean**	**Weighted mean**	**Estimates**
All	-0.602***	-0.840***	2,989
Advanced countries	-0.507**	-0.697***	1,159
Emerging countries	0.364*	0.759*	407
Japanese yen	-1.587***	-1.629***	309
German mark	-0.980***	-0.937***	181
British pound	-0.981***	-1.003***	269
French franc	-0.266	-0.300	133
Italian lira	-0.272	-0.169	117
Swiss franc	-0.825***	-1.048***	168
European currencies	-0.822***	-0.834***	1,434
Asian currencies	-0.922***	-1.118***	520
Euro	-2.371***	-2.261***	102
non-European/non-Asian currencies	-0.864***	-0.572**	416

Notes: Weighted means are calculated using the inverse of the number of estimates reported per study as the weight. Only differences estimates (Equation 3) are included. Significance levels: *** p-value $<$ 0.01, ** p-value $<$ 0.05, * p-value $<$ 0.1.

Apart from the estimates of β and their standard errors, we collect 43 variables that capture different aspects of how the studies are designed. In consequence, we have to collect more than

156,000 data points from the primary studies. The data collection was performed by one of the co-authors while another double-checked random portions of the data to minimise potential mistakes made during the data coding process.

Table 1 presents mean estimates of reported β for various currencies or groups of currencies. The reported mean overall coefficient is negative, in line with the forward rate unbiasedness puzzle. The most striking observation obtained from the table, however, is the apparent difference between the mean coefficients reported for the currencies of advanced and emerging economies, respectively. The mean estimate for the former is negative, while the mean estimate for the latter is positive, both being statistically significantly different from zero at least at the 10% level. Thus our data corroborate the results of Frankel & Poonawala (2010): the bias in the forward discount as a predictor of future changes in the spot exchange rate is less severe in emerging market currencies than in advanced country currencies. Frankel & Poonawala (2010) observe that the coefficient for emerging market currencies is on average slightly above zero, and even if negative, it is rarely significantly less than zero. Nevertheless, all of these results can be influenced by publication bias, an issue to which we turn next.

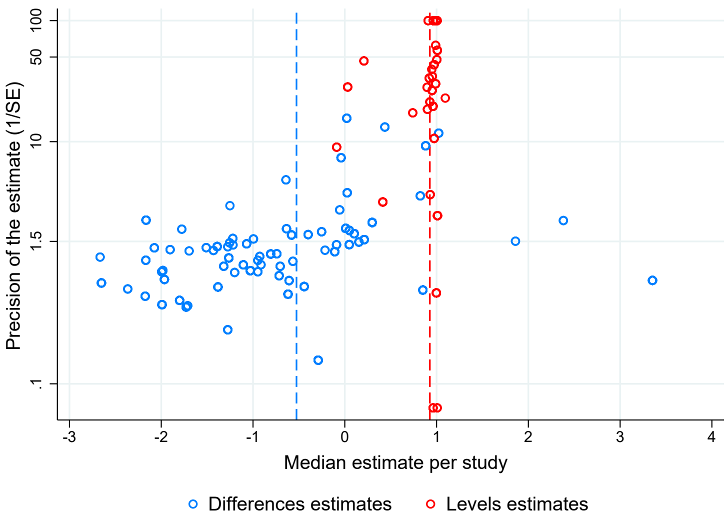
4 Publication Bias

Publication bias is the empirical observation in many sciences, including economics (Ioannidis *et al.*, 2017), that the reported results constitute a biased reflection of the universe of results obtained by researchers before they write up their papers. Why should reported results be biased? One reason is that statistical significance is sometimes perceived as evidence of scientific importance. This perception might represent a valid principal in some cases, but in general it means that the published results will exaggerate the true underlying effect unless the true effect is zero. Estimates that are, simply by chance, much larger than the true effect (in absolute value) will be statistically significant. Estimates that are, also by chance, much smaller than the true value will probably be insignificant. If statistical significance is taken for scientific importance, the former estimates will become overrepresented in the literature. Another reason for publication bias is a simple preference for a particular sign of the regression parameter, perhaps given by underlying theory or previous influential findings. A more precise label for the problem is "selective reporting", because there is no reason why the problem should be confined to journals: researchers write their working papers with the intention to publish. We use the term "publication bias" to keep consistency with previous studies on the topic.

Our main identification assumption in testing for publication bias (we will relax the assumption later) is that in the absence of the bias, the estimate and its standard error should be statistically independent quantities. This assumption follows from the properties of the econometric methods used to estimate β in the literature. In almost all the cases, the econometric techniques suggest that the ratio of the estimated β to its standard error has a symmetrical distribution (typically a t-distribution). In consequence, we should observe zero correlation between estimates and their standard errors. But if, for example, researchers strive to report statistically significant results, they will, given some standard error, search for estimates that are large enough to bring the t-statistic above 2 in absolute value. This search can be conducted by choosing a subset of the entire dataset available, different estimation technique, or different control variables. A similar correlation between estimates and standard errors arises if estimates of a particular sign are discriminated against; the correlation follows from the observation that a regression of estimates on standard errors is heteroskedastic.

The first step in identifying the potential presence of publication bias is investigation of a funnel plot, which we have already discussed in the Introduction. Figure 5 presented in this section focuses on median estimates from each study, which has the benefit of giving each study the same weight and removing outliers that may simply represent unimportant robustness checks not central to the main results of the study. Nevertheless, the funnel tells the same story as the one in Figure 1 from the Introduction. The estimates obtained from regressions estimated in levels form are almost always close to 1 and do not form a funnel. This observation corroborates the potential of spurious regression in the levels and bias towards one; that is why we exclude these estimates from further meta-analysis tests. In contrast, the differences estimates clearly form a funnel, albeit an asymmetrical one: imprecise positive estimates are mostly missing from the funnel. The shape of the funnel plot is thus consistent with strong publication bias against positive estimates of β, and can be interpreted as confirmation bias due to the prevalence of negative estimates since the earliest studies in the field.

Figure 5: *Funnel plot for median estimates reported per study*



Notes: In the absence of publication bias, the scatter plot should resemble an inverted funnel that is symmetrical around the most precise estimates. Red circles (darker in grayscale) represent estimates extracted from the levels equation (Equation 2); blue circles (lighter in grayscale) show estimates from the differences equation (Equation 3). Outlying observations are cut from the figure for ease of exposition but included in all statistical tests. The red dashed vertical line shows the mean estimate from the levels specification; the blue dashed vertical line shows the mean from the differences specification.

In the next step we quantify the extent of publication bias numerically. As we have noted, if there is no bias, there should be no correlation between estimates and their standard errors. In the presence of bias, we will observe a correlation consistent with the Lombard effect, as researchers will increase their effort to find larger estimates of β in response to noise (Stanley, 2005):

$$\beta_{i,j} = \gamma_0 + \gamma_1 * SE(\beta_{i,j}) + \epsilon_{i,j}, \tag{5}$$

where $\beta_{i,j}$ is the i-th slope coefficient estimate in study j collected from the differences specification, as detailed in Equation 3. $SE(\beta_{i,j})$ is the standard error of this estimate. γ_1 captures the severity of publication bias in the literature while γ_0 measures the mean efect beyond bias. Nevertheless, the specification in Equation 5 is heteroskedastic since the right-hand-side variable (SE) captures the variance of the left-hand-side variable (coefficient estimate β). To correct for this heteroskedasticity, we divide Equation 5 by the standard error of the estimate and obtain

$$t_{i,j} = \gamma_1 + \gamma_0 * \frac{1}{SE(\beta_{i,j})} + \omega_i, \tag{6}$$

where $t_{i,j}$ is t-statistic of the i-th estimate of β from study j, γ_1 captures publication selectivity, and γ_0 measures the corrected effect beyond bias (the mean β conditional on maximum precision, and therefore no publication bias). The weighted specification has the additional allure of giving more precise estimates greater weight.

Table 2: *Linear tests of publication bias*

	FE	**OLS**	**IV**
Each estimate has the same weight			
1/SE (mean beyond bias)	0.258	0.306	0.00859
	(0.177)	(0.188)	(0.127)
Constant (publication bias)	-0.878	-1.094**	0.251
	(0.800)	(0.472)	(0.972)
Observations	2,989	2,989	2,989
R-squared	0.215	0.274	0.015
Number of studies	74	74	74
Each study has the same weight			
1/SE (mean beyond bias)	0.611***	0.605***	0.215
	(0.170)	(0.153)	(0.222)
Constant (publication bias)	-2.057***	-2.034***	-0.613
	(0.618)	(0.386)	(0.707)
Observations	2,989	2,989	2,989
R-squared	0.494	0.531	0.311
Number of studies	74	74	74

Notes: The table presents the results of regression in Equation 6. Standard errors of the parameters are clustered at the study level and shown in parentheses. OLS = ordinary least squares, FE = study-level fixed effects, IV = instrumental variables regression with the inverse of the square root of the number of observations used as an instrument of the standard error. Only differences estimates (Equation 3) are included. Significance levels: *** p-value $<$ 0.01, ** p-value $<$ 0.05, * p-value $<$ 0.1.

Table 2 shows the results of running Equation 6. We consider both a version that gives each estimate the same weight (the top half of the table) and a version that gives each study the same weight by additionally weighting the regression by the inverse of the number of estimates reported per study (the bottom half of the table). It makes little sense to give more weight to studies that report more estimates, so we prefer the second version of the equation, which also yields more precise estimates of both publication bias and the mean effect corrected for publication bias. We further prefer including dummy variables for each study, which controls for unobserved study-level characteristics that can be related to quality; the resulting specification is labeled "FE" for fixed effects. We observe that in this preferred specification, publication bias is robustly negative, and the mean β corrected for the bias is around 0.6. Nevertheless, some method choices in the primary studies could potentially influence the estimates and their standard errors in the same direction, which would make our estimates of publication bias spurious. Therefore, as a robustness check, we use the inverse of the square root of the number of observations as an instrument for the standard error. This quantity is correlated with the standard error by definition, but should not be much correlated with method choices. Our results still show negative publication bias and positive mean beyond the bias, but with much less precision.

So far we have assumed that publication bias is directly proportional to the size of the standard error. But in principle, this does not have to be the case. Next, we estimate a quadratic model of publication bias, the so-called PEESE (precision-effect estimate with standard error) developed by Stanley & Doucouliagos (2012) and Stanley & Doucouliagos (2014):

$$t_{i,j} = \gamma_1 * SE(\beta_{i,j}) + \gamma_0 * \frac{1}{SE(\beta_{i,j})} + \xi_i, \tag{7}$$

where γ_1 again captures publication bias in the literature, while γ_0 denotes the mean β corrected for publication bias, i.e. the slope coefficient of regression the changes in the spot exchange rate on the forward discount. Table 3 presents the results of this testing, which are similar to those of the linear specification. Because fixed effects yield results essentially identical to those of OLS, we do not report them. Once again, if each study is given the same weight (the bottom panel of the table), we obtain evidence of publication bias towards negative estimates of β. The corrected mean estimate is 0.55, only slightly smaller than the one implied by the linear specification. The IV estimation again brings coefficients which display the same signs, but are smaller and much less precise. In our case, unfortunately, the inverse of the square root of the number of observations is a relatively weak instrument for the standard error. But even the imprecise IV estimation suggests that the underlying β is likely positive, contrary to the naive mean obtained from the reported estimates.

Nonlinear techniques more sophisticated than PEESE have recently been developed. We use a battery of these advanced tests to evaluate the robustness of our results regarding the mean β corrected for publication bias, especially in light of the insignificant estimates obtained from the IV specifications. We employ five methods: the weighted average of adequately powered estimates by Ioannidis *et al.* (2017), the stem-based method by Furukawa (2019), the selection model by Andrews & Kasy (2019) and the endogenous kink technique by Bom & Rachinger (2019). First, Ioannidis *et al.* (2017), using a survey of more than 60,000 estimates published in economics, find that the median statistical power among the published results in economics is 18%. They investigate how power is associated with publication bias and propose a correction technique that employs the estimates with power above 80%. Furthermore, using Monte Carlo simulations, Ioannidis *et al.* (2017) show that their technique outperforms the commonly used meta-analysis estimators. The results of their model, shown in Table 4, are close to that of our IV estimation.

Table 3: *Quadratic correction for publication bias (PEESE)*

	OLS	**IV**
	Each estimate has the same weight	
SE (publication bias)	-0.0439	0.0837
	(0.0287)	(0.376)
1/SE (mean beyond bias)	0.279	0.0224
	(0.177)	(0.105)
Observations	2,989	2,989
R-squared	0.257	0.035
	Each study has the same weight	
SE (publication bias)	-0.197***	-0.460
	(0.0674)	(0.443)
1/SE (mean beyond bias)	0.552***	0.228
	(0.148)	(0.203)
Observations	2,989	2,989
R-squared	0.491	0.314

Notes: The table presents the results of regression in Equation 7. Standard errors of the parameters are clustered at the study level and shown in parentheses. OLS = ordinary least squares, IV = instrumental variables regression with the inverse of the square root of the number of observations used as an instrument of the standard error. Only differences estimates (Equation 3) are included. Significance levels: *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Second, the approach by Furukawa (2019) relies on the the assumption that the most precise estimates suffer from little bias: authors do not find it difficult to produce statistically significant estimates when the standard errors are very small. Previous researchers in meta-analysis focused on a fraction of the most precise estimates in meta-analysis—such as top-ten method by Stanley *et al.* (2010). Furukawa (2019) finds a new way of estimating this fraction of estimates by exploiting the trade-off between bias and variance. His technique delivers a large estimate of the mean β corrected for publication bias, 0.8. Third, Andrews & Kasy (2019) apply the observation reported by many researchers (e.g. Brodeur *et al.*, 2016) that standard cut-offs for the p-value are associated with jumps in the distribution of reported estimates. Consequently, they build on Hedges (1992) in order to construct a selection model of publication probability for each estimate in the literature given its p-value. Using their technique, we obtain an estimate of 0.19, which is statistically significant at the 1% level. Fourth, Bom & Rachinger (2019) account for the case that estimates get reported only when they cross a particular precision threshold. In their method they estimate this threshold and introduce an "endogenous kink" to extend the linear test of publication bias. The technique gives us an estimate of 0.31.

Table 4: *Advanced nonlinear corrections for publication bias*

	Mean beyond bias	**Std. Error**
Ioannidis *et al.* (2017)	0.204	0.434
Furukawa (2019)	0.795	0.314
Andrews & Kasy (2019)	0.194	0.028
Bom & Rachinger (2019)	0.306	0.009

Notes: The table presents results of four recently introduced nonlinear corrections for publication bias. Ioannidis *et al.* (2017) focuses on estimates with adequate power, Furukawa (2019) exploits the trade-off between precision and bias, Andrews & Kasy (2019) use a selection model, and Bom & Rachinger (2019) search for a precision threshold beyond which publication bias is unlikely.

On balance, we find strong evidence of publication bias in the literature on the forward rate unbiasedness hypothesis. All 4 recently proposed nonlinear techniques suggest that the mean β corrected for publication bias is positive, which contrasts with the naive mean of -0.6 obtained by averaging the reported estimates of β. In the next part of the manuscript we consider the effects of differences in study design on the reported estimates.

5 Heterogeneity

We aim to identify the aspects of estimation context that systematically influence the reported estimates of the slope coefficient β from the differences specification of the forward rate unbiasedness hypothesis tests presented in Equation 3. To this end, we collect 43 variables that reflect country scope, data characteristics, estimation characteristics, regimes capturing different market conditions, databases used, and publication characteristics. In the following paragraphs we describe these variables.

5.1 Variables

Country scope Previous studies have hinted on potential differences between the reported β's for countries in different stages of development. For instance, Bansal & Dahlquist (2000) observe that the puzzling finding of a negative β is systematically related to the use of data from advanced economies. Summary statistics of our data corroborate this finding: the mean β for advanced economy currencies equals -0.51, while the mean for developing economy currencies is 0.36, both significantly different from zero. For this reason we include dummy variables for advanced economy currencies (38.8% of the estimates), emerging economy currencies (13.6%), estimates arising from both advanced and developing currency samples (4.8%), estimates specifically obtained for the former German mark (6.1%), French franc (4.4%), British pound (9%), Italian lira (4%), Japanese yen (10.3%), Swiss franc (5.6%), euro (3.4%), and for currencies from different geographical regions: Europe (48%), Asia (17.4%), non-European and non-Asian countries (14%), and for estimates from geographically mixed datasets (20.7%). In addition, we control for the numeraire currency against which other currencies are tested. Flood & Rose (1996) find a larger β for European currencies tested versus the German mark than for those tested versus the USD. Bansal & Dahlquist (2000) hypothesize that this finding could be due to the fact that the economies within the European Monetary System synchronised their monetary policy with Germany. The proportions of estimates in our dataset with different numeraire currencies are the following: USD (70.7%), British pound (10.5%), Japanese yen (2%), Swiss franc (1.4%), euro (8.4%), German mark (7%).

Data characteristics This category comprises additional information regarding the data that were used to produce the estimates of β. In particular, we collect information on the forward rate horizon; i.e., the number of periods k after which settlement will occur, the frequency of the data used in the estimation, the size of the time series component of the data (the number of time periods used for estimation), the size of the cross-sectional component (the number of currencies included in a panel), total sample size, and sources of data in primary studies. As for the horizon of the rates used to test the unbiasedness hypothesis, Razzak (2002) finds support for the unbiasedness hypothesis when one-year forward exchange rates are used instead of one-month contracts. His finding is consistent with the literature on uncovered interest rate parity, for which some authors also find support when using long-term interest rates (e.g., Chinn & Meredith, 2004; Nadal De Simone & Razzak, 1999). Next, we collect dummies for the frequencies of data as follows: daily (12.3%), weekly (13%), monthly (72%), quarterly (2.4%), and other frequency (0.2%). We further account for the so-called overlapping samples problem, where data frequency is higher than the maturity horizon of instruments, which introduces serial correlation in the error term of the regression. In our sample, 38% of the estimates originate from data that suffer from this problem, and many authors introduce corrections in the methods they apply in order to mitigate the issue. For instance, Goodhart (1988) applies an adjustment to the OLS covariance matrix proposed by Hansen (1982).

Estimation Researchers apply different estimation techniques to test the validity of the forward rate unbiasedness hypothesis. The most frequently used method is OLS regression (66.5% of estimates) followed by the regime switching model (13.8%), seemingly unrelated regressions (7.1%), and fixed effects regression (3.1%). Other methods, such as instrumental variables regression, error correction model and maximum likelihood, are also used, and generate 9.4% estimates of β. Some researchers advocate the use of the seemingly unrelated regressions technique which allows them to account for cross-correlation across currencies in their samples, arising from, among other things, the use of a common numeraire currency. For instance, Fama (1984) applies both OLS and seemingly unrelated regressions and finds that joint estimation of β improves the precision of the estimate. Moreover, he observes that the estimated slope coefficients from seemingly unrelated regressions are closer to zero compared to the estimates from OLS.

Other studies associate the existence of the forward premium anomaly with different market conditions or regimes. For this reason, they apply a variety of regime switching techniques to model this transition. For instance, Baillie & Chang (2011) apply logistic smooth transition dynamic regression models with interest differentials and foreign exchange market volatility as transition variables between two regimes. In one regime, they observe exchange rate movements that are characterized by persistent deviations from the uncovered interest rate parity, while in the other regime reversions to the parity occur. They show that the forward premium anomaly ceases to manifest when foreign exchange market volatility is high. Moreover, we account for any additional variables that researchers may include in their regressions. For this reason, we control for the inclusion of lagged values of the forward rate (0.13%), interest rate differentials (0.13%), forward discount terms to the power of two and three (2.7%), and other controls (1.7%). Since these categories of controls do not comprise enough cases separately, we aggregate them into one variable "Controls". Overall, 4.5% of the estimates in our dataset include additional control variables.

We also account for the units in which Equation 3 is specified. Typically, the forward discount and the change in the spot exchange rate are expressed as the difference between the natural logarithm of the forward rate and the natural logarithm of the spot rate at time t, and the difference between the natural logarithm of the spot rate at $t+k$ and the logarithm of the spot rate at time t, respectively. This specification is the most frequent in the literature: over 95% of the estimates of *beta* are obtained from this specification. Alternatively, about 5% of the estimates arise from the specification where the change in spot rates and the forward discount are expressed as a percentage of the spot rate at time t. More precisely, the percentage change in the spot exchange rate is expressed as $\frac{(S_{t+k}-St)}{S_t}$ and the change in the forward discount is written as $\frac{(F_t-St)}{S_t}$, where S_t and F_t is the spot and the forward rate at t, respectively.

Regimes Researchers have investigated whether the apparent presence of the forward rate bias is subject to different market conditions; that is, if the unbiasedness hypothesis holds in some so-called regimes while it is violated in others. For instance, Baillie & Kiliç (2006) find using the logistic smooth transition dynamic regression model that the forward premium anomaly is more likely to occur during the periods of high volatility in US money growth while the periods of relative stability in terms of US money growth volatility are associated with forward rate unbiasedness. They also find that the growth of foreign money relative to US money leads to a higher likelihood of unbiasedness condition not being rejected. Therefore, money supply differentials serve as important transition variables between regimes in their study. Furthermore, Grossmann *et al.* (2014), using a sample of advanced economy currencies vis-a-vis the euro and the British pound, find that a significant forward premium anomaly exists for advanced country currencies during crisis periods when the numeraire currency sells at a premium or is overvalued. On the other hand, Zhou & Kutan (2005) do not find evidence for any forward premium asymmetry between the US dollar and the six currencies in their sample between 1977 and 1998. We include controls for large and positive forward premium (equals one for 5.7% of the estimates in our sample), negative forward premium (5.9%), overvalued currency (3.5%), undervalued currency (3.5%), large differential, which comprises controls for positive interest differential and other positive differentials (e.g., money growth differentials), for 3.3% of estimates, small differential, which includes controls for negative interest rate differential and other small/negative differentials, altogether for 3.5% of estimates, and a dummy for other regimes, which comprises both high and low foreign exchange volatility regimes among others, for 5.7% of collected estimates.

Data sources We control for the different sources of the data that researchers employ in primary studies. For example, some researchers advocate the use of survey data to address the issue whether the forward premium is due to systematic expectational errors or the risk premium (e.g., Froot & Frankel, 1989). Researchers use Datastream (51% of estimates), various bank data sources (14%), Data Resources, Inc. (3.2%), survey data (1.5%), and other minor sources (30%) to calculate the estimates of β reported in primary studies.

Publication characteristics To capture the potential aspects of study quality which are not reflected by the differences in data and methods across studies outlined above, we include three study-level variables: the year when the first draft of the paper appeared in Google Scholar (we opt for this measure instead of publication year due to increasing publication lags in economics and finance), the recursive discounted impact factor of the journal from RePEc, and the number of citations per year since the paper first appeared in Google Scholar. Table 5 presents the summary statistics of the aforementioned variables.

Table 5: *Description and summary statistics of regression variables*

Variable	**Description**	**Mean**	**SD**	**WM**
β	The reported estimates of the slope coefficient β	-0.60	2.13	-0.84
SE	The reported standard error of the coefficient β	2.29	3.85	1.39
Country scope				
Advanced_currencies	=1 if advanced economy currencies are used	0.39	0.49	0.32
Emerging_currencies	=1 if emerging economy currencies are used	0.14	0.34	0.08
German_mark	=1 if German mark is used	0.06	0.24	0.09
French_franc	=1 if French franc is used	0.04	0.21	0.05
GBP	=1 if British pound is used	0.09	0.29	0.15
Italian_lira	=1 if Italian lira is used	0.04	0.19	0.03
JPY	=1 if Japanese yen is used	0.10	0.30	0.15
Swiss_franc	=1 if Swiss franc is used	0.06	0.23	0.07
Euro	=1 if euro is used	0.03	0.18	0.03
geo_Europe	=1 if European currencies are used	0.48	0.50	0.53
geo_Other	=1 if non-European and non-Asian currencies are used	0.14	0.35	0.21
GBP_base	=1 if British pound is used as the numeraire currency	0.10	0.31	0.03
Euro_base	=1 if euro is used as the numeraire currency	0.08	0.28	0.01
German_mark_base	=1 if German mark is used as the numeraire currency	0.07	0.26	0.05
Data characteristics				
Less_1month	=1 if forward contract maturity is less than 1 month	0.06	0.23	0.03

Continued on next page

Table 5: *Description and summary statistics of regression variables (continued)*

Variable	Description	Mean	SD	WM
Onemonth	=1 if forward contract maturity is 1 month	0.65	0.48	0.70
Onemonth_to_1year	=1 if forward contract maturity is between 1 month and 1 year	0.19	0.39	0.21
Oneyear	=1 if forward contract maturity is 1 year	0.08	0.27	0.05
Daily	=1 if data frequency is daily	0.12	0.33	0.10
Weekly	=1 if data frequency is weekly	0.13	0.34	0.15
Monthly	=1 if data frequency is monthly	0.72	0.45	0.68
Time_diff	The logarithm of the number of observations in the forward contract maturity horizon	1.01	1.30	0.76
N	The logarithm of the number of currencies used in the estimation	0.55	1.06	0.15
Sample_size	The logarithm of the total number of observations used in the estimation	5.61	1.31	5.21
Overlapping_problem	=1 if the overlapping samples problem is present	0.38	0.49	0.31
Estimation				
OLS	=1 if OLS is used	0.66	0.47	0.75
FE	=1 if fixed effects regression is used	0.03	0.17	0.02
Regime_switching	=1 if a regime switching/transition model is used	0.14	0.34	0.05
SUR	=1 if seemingly unrelated regressions model is used	0.07	0.26	0.07
Controls	=1 if there are additional control variables included	0.04	0.21	0.05
Diff_percent	=1 if spot rate change and forward premium are expressed in percentage of the spot rate	0.05	0.22	0.07
Regimes				
Large_differential	=1 if estimation period is characterized by large differentials (in interest rates, money growth, etc.)	0.03	0.18	0.01
Small_differential	=1 if estimation period is characterized by small differentials (in interest rates, money growth, etc.)	0.03	0.18	0.01
Large_positive_premium	=1 if estimation period is characterized by a positive forward premium	0.06	0.23	0.02
Low_negative_premium	=1 if estimation period is characterized by a negative forward premium	0.06	0.24	0.02
Overvalued_currency	=1 if estimation period is characterized by overvaluation of the currency	0.03	0.18	0.00
Undervalued_currency	=1 if estimation period is characterized by undervaluation of the currency	0.03	0.18	0.00
Data sources				
Datastream	=1 if data from Datastream is used in the estimation	0.51	0.50	0.29
Bank_data_sources	=1 if data from various bank sources is used in the estimation	0.14	0.35	0.19
Data_Resources_Inc	=1 if data from Data Resources, Inc. is used in the estimation	0.03	0.18	0.13
Publication characteristics				
IF_recursive	The recursive discounted impact factor from RePEc	0.46	0.61	0.51

Continued on next page

Table 5: *Description and summary statistics of regression variables (continued)*

Variable	Description	Mean	SD	WM
Citations	The logarithm of the number of Google Scholar citations normalized by the number of years since the first draft of the paper appeared in Gooogle Scholar	1.70	1.80	1.72
firstpub	The year when the first draft of the study appeared in Google Scholar minus the year when the first study was published	30.11	8.03	26.24

Notes: SD = standard deviation, WM = mean weighted by the inverse of the number of estimates reported per study. The impact factor was extracted from RePEc and the number of citations from Google Scholar. The remaining variables were collected from studies estimating β.

5.2 Methodology

To investigate which variables systematically explain the differences among the reported estimates of β extracted from Equation 3, the natural method would be to regress the reported βs on the variables capturing the context in which βs are calculated in primary studies. In other words, one wishes to estimate the following equation:

$$\beta_{ij} = \gamma_0 + \gamma_1 * SE(\beta_{ij}) + \sum_{l=1}^{43} \gamma_l * X_{l,ij} + \zeta_{ij}, \tag{8}$$

where β_{ij} denotes estimates of the slope coefficient i in study j obtained from regressing changes in spot foreign exchange rates on the forward premium, as detailed in Equation 3. $X_{l,ij}$ represents the set of control variables that we introduced in Subsection 5.1, γ_1 measures the severity of publication bias conditional on the inclusion of controls, and γ_0 is the mean β estimate corrected for publication bias but also conditional on the variables included in X.

Nevertheless, including all the variables in X into one regression is problematic because of model uncertainty. While there is a strong rationale to include some of the variables, there are others which we would like to include as controls because they can also affect the slope coefficient β but for which little theory exists that would firmly justify their inclusion ex ante. Estimating Equation 8 as a single regression would negatively affect the precision of the coefficient estimates due to a large number of variables. There are several ways one can approach this problem, the most commonly traveled one being stepwise regression. Nevertheless, sequential t-testing does not properly account for the conditionality on the results of the previous t-test, and could thus accidentally exclude a useful variable at some stage. A more appropriate solution to model uncertainty is Bayesian Model Averaging (BMA). For more details on the technique, see Eicher *et al.* (2011) and Steel (2020).

BMA estimates many regressions using different subsets of the variables from the model space. In our case, since we consider 43 variables, this yields 2^{43} possible models to estimate. To run all the models would be infeasible even with a modern computer. For this reason, we use Markov chain Monte Carlo (MCMC; Madigan *et al.*, 1995) algorithm that approximates the model space and walks through the part of the model space that contains models with the highest posterior model probabilities (PMP). In frequentist terms, PMP is an analogue to information criteria, thus measuring how well the model performs compared to other models of similar complexity. BMA reports the posterior mean coefficient and posterior standard deviation of the coefficient, which are based on the weighted average of the coefficients from all the estimated models with weights being the PMP. Furthermore, for each variable BMA reports its posterior inclusion probability (PIP), which is equal to the sum of the PMPs of all the models in which this variable is included. In the baseline specification we apply the uniform model prior (assigning each model the same prior probability) and unit information g-prior (the prior that all regression coefficients equal zero obtains the same weight as one data point). We apply alternative priors as well and report the results in Appendix A, and the main results are compared in Figure 7.

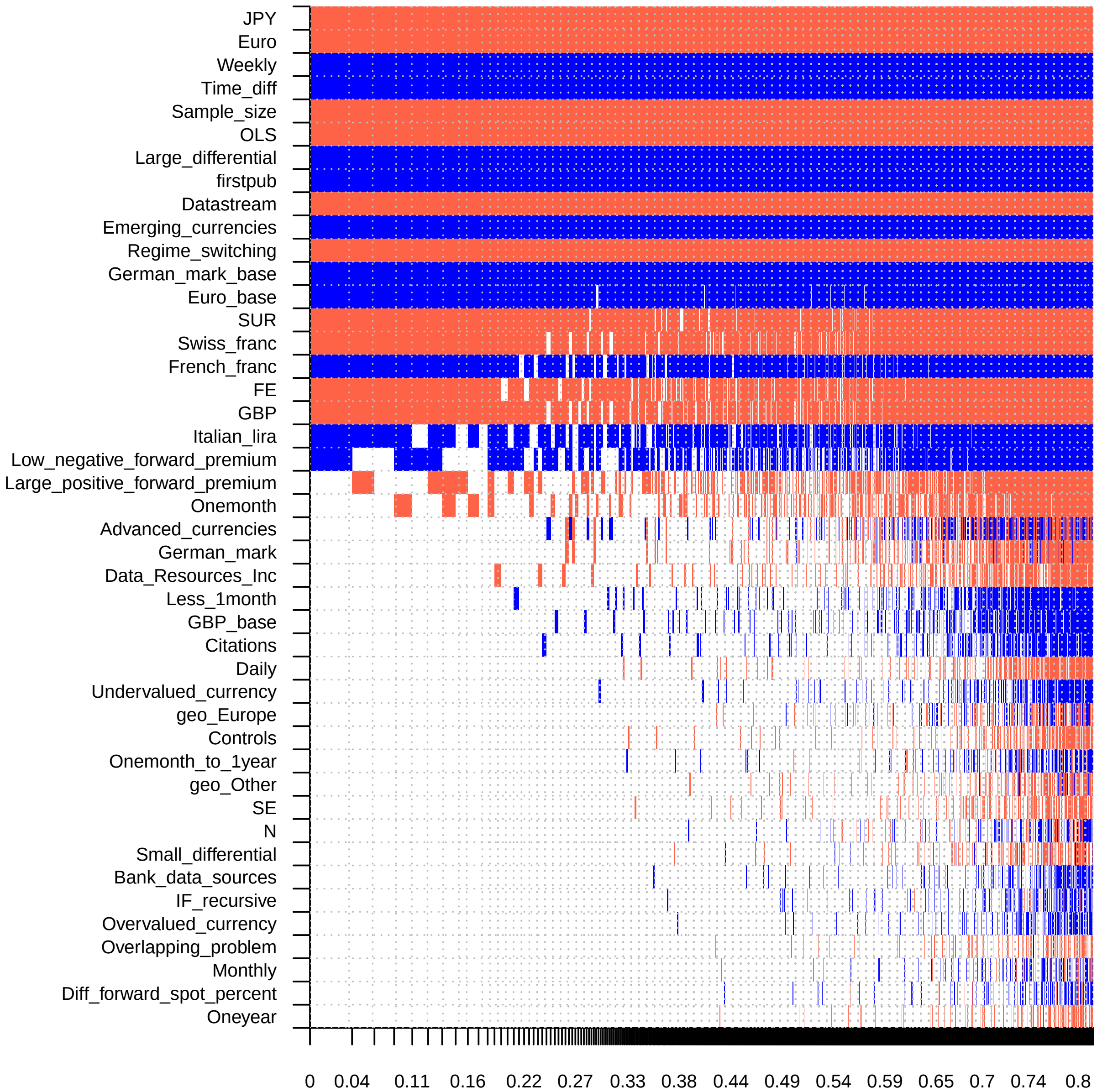
Because the results for different priors vary, we perform further robustness checks by estimating the OLS model and study fixed effects model based on the results of BMA, i.e. by including only variables with PIP > 0.5: the variables that are not irrelevant for the explanation of the differences in the estimates of β. We choose the threshold of 0.5 for PIP based on Jeffreys (1961), who categorizes the values of PIP below 0.5 as irrelevant, between 0.5 and 0.75 as weak, values between 0.75 and 0.95 as positive, values between 0.95 and 0.99 as strong, and values above 0.99 as showing decisive evidence of an effect. In both OLS and fixed effects we cluster standard errors at the level of studies. Last but not least, we run frequentist model averaging (FMA). In our implementation of FMA we use Mallows' criteria as weights since they were shown to be asymptotically optimal (Hansen, 2007). Nevertheless, by using a frequentist approach there is no immediate alternative to the MCMC, and we find it infeasible to estimate all the 2^{43} models. For this reason we follow Amini & Parmeter (2012) and orthogonalize the covariate space.

5.3 Results

The results of BMA are first summarized visually in Figure 6. The vertical axis shows the explanatory variables ordered by their PIP from the highest to the lowest value, and the horizontal axis shows individual models ordered by their posterior model probabilities; the best models are on the left. The blue color indicates that a variable is included in the model and its estimated coefficient sign is positive, while the red color stands for a negative coefficient sign. Blank cells indicate a variable is not included in the corresponding model.

Overall, there are 20 variables that can explain the variation in the reported estimates; 18 of them have PIP above 0.8, which means there is at least positive evidence for their effect on the estimated coefficient. Table 6 presents the numerical results of the BMA exercise as well as the results of the complementary OLS, fixed effects, and frequentist model averaging. In OLS and fixed effects we include only the 20 variables with PIP > 0.5. The standard error variable appears unimportant, but that is due to the concurrent inclusion of more nuanced variables related to sample size which are correlated with the standard error by definition. In the next paragraphs we describe the results for individual variables.

Figure 6: *Model inclusion in BMA*



Notes: The response variable is an estimate of slope coefficient *beta* from Equation 3. Columns show individual models, and variables are listed in descending order by their posterior inclusion probabilities. The horizontal axis shows cumulative posterior model probabilities from the 5,000 best models. To ensure convergence of the Markov Chain Monte Carlo sampler, we use 5,000,000 iterations with 1,000,000 burn-ins to allow the sampler to converge to the part of the model space with high posterior probability models. Blue color (darker in grayscale) = the variable is included in the model with a positive sign. Red color (lighter in grayscale) = the variable is included in the model with a negative sign. No color = the variable is missing from the model. The data is weighted by the inverse of the number of estimates reported per study. For description of all variables see Table 5.

Country scope The choice of the currency has strong implications for the estimated β coefficient. Estimates obtained from testing the unbiasedness hypothesis on emerging economy currencies are substantially larger than the estimates for advanced economies, and in BMA the effect is classified as *decisive* for explaining β. Larger estimates are also reported for the former French franc and Italian lira, but to a lesser degree. For some of the advanced economy currencies, especially the British pound, Japanese yen, and euro, the reported β estimates are even smaller than the mean for advanced economies as a whole. The choice of the numeraire currency also has an impact on the reported β. Our results indicate that employing euro as the numeraire currency increases the reported coefficient by about 1.5 on average. The finding of a smaller bias in the estimates of β for emerging country currencies corroborates the results of Frankel & Poonawala (2010), who show that the coefficient for these currencies is on average positive and never significantly less than zero.

Data characteristics Our results suggest that if the spot rates and the forward rates are sampled with weekly frequency the estimated β tends to be larger. Similarly, if longer horizon rates are used to test the forward unbiasedness hypothesis the reported coefficients are larger by around 1.2 (the coefficient equals 0.2, but note that the variable is used in logs). In a similar vein, Nadal De Simone & Razzak (1999) find that long-term interest rates can explain a larger portion of the spot exchange rate movements. Furthermore, larger sample sizes also yield larger estimates of β by about 0.8 (again, the coefficient itself is -0.2, but the variable is used in logs).

Table 6: *Why do estimates of beta vary?*

Response variable: estimated β	**BMA**			**OLS**	**FE**	**FMA**
	PIP	*Post Mean*	*Post SD*			
SE	0.03	0.00	0.00			-0.012
						(0.014)
Country scope						
Advanced_currencies	0.17	-0.01	0.13			-0.162
						(0.269)
Emerging_currencies	1.00	1.05	0.20	1.061***	0.665	0.959***
				(0.360)	(0.465)	(0.270)
German_mark	0.11	-0.04	0.14			-0.370
						(0.3)
French_franc	0.84	0.46	0.25	0.497**	0.621***	0.261
				(0.213)	(0.213)	(0.313)
GBP	0.82	-0.31	0.20	-0.329**	-0.171	-0.630**
				(0.139)	(0.132)	(0.294)
Italian_lira	0.71	0.43	0.32	0.556**	0.625**	0.337
				(0.224)	(0.265)	(0.329)
JPY	1.00	-0.90	0.16	-0.885***	-0.697***	-1.149***
				(0.162)	(0.153)	(0.275)
Swiss_franc	0.87	-0.43	0.24	-0.471*	-0.439**	-0.784**
				(0.250)	(0.184)	(0.307)
Euro	1.00	-1.91	0.23	-1.889***	-1.444***	-2.245***
				(0.466)	(0.397)	(0.334)
geo_Europe	0.05	0.00	0.04			0.050
						(0.192)
geo_Other	0.04	0.00	0.03			-0.144
						(0.168)
GBP_base	0.09	0.03	0.12			0.450**
						(0.215)
Euro_base	0.96	1.50	0.52	1.555***	0.250	1.701***

Continued on next page

Table 6: *Why do estimates of beta vary? (continued)*

Response variable: estimated β	BMA *PIP*	BMA *Post Mean*	BMA *Post SD*	OLS	FE	FMA
				(0.416)		(0.464)
German_mark_base	0.98	0.56	0.16	0.577*	0.555	0.563***
				(0.302)	(0.468)	(0.15)
Data characteristics						
Less_1_month	0.10	0.04	0.14			0.809*
						(0.469)
Onemonth	0.26	-0.05	0.10			0.024
						(0.306)
Onemonth_to_1year	0.04	0.00	0.03			0.037
						(0.252)
Oneyear	0.02	0.00	0.03			-0.144
						(0.264)
Daily	0.06	-0.02	0.08			-0.537
						(0.377)
Weekly	1.00	0.87	0.11	0.864***	0.470**	0.654**
				(0.322)	(0.233)	(0.286)
Monthly	0.02	0.00	0.02			-0.077
						(0.173)
Time_diff	1.00	0.20	0.04	0.210**	0.112	0.286***
				(0.086)	(0.112)	(0.069)
N	0.04	0.00	0.03			-0.090
						(0.112)
Sample_size	1.00	-0.20	0.04	-0.200***	-0.007	-0.163***
				(0.066)	(0.109)	(0.045)
Overlapping_problem	0.02	0.00	0.02			-0.053
						(0.136)
Estimation						
OLS	1.00	-0.72	0.12	-0.762***	-0.621*	-0.836***
				(0.285)	(0.343)	(0.113)
FE	0.83	-0.64	0.36	-0.759	-0.189	-0.924***
				(0.472)	(0.404)	(0.303)
Regime_switching	0.99	-1.03	0.25	-1.187***	-1.269***	-1.006***
				(0.404)	(0.440)	(0.246)
SUR	0.92	-0.50	0.21	-0.554	-0.109	-0.701***
				(0.35)	(0.338)	(0.168)
Controls	0.04	-0.01	0.04			-0.074
						(0.152)
Diff_percent	0.02	0.00	0.02			0.283*
						(0.144)
Regimes						
Large_differential	1.00	2.58	0.35	2.718***	2.701***	2.33***
				(0.623)	(0.650)	(0.384)
Small_differential	0.03	-0.01	0.10			-0.412
						(0.364)
Large_positive_premium	0.40	-0.24	0.32			-0.534**
						(0.232)
Low_negative_premium	0.54	0.35	0.36	0.668	0.757	0.543**
				(0.492)	(0.854)	(0.236)
Overvalued_currency	0.03	0.01	0.12			0.485
						(0.626)

Continued on next page

Table 6: *Why do estimates of beta vary? (continued)*

Response variable: estimated β	BMA *PIP*	BMA *Post Mean*	BMA *Post SD*	OLS	FE	FMA
Undervalued_currency	0.05	0.04	0.23			0.796 (0.626)
Data sources						
Datastream	1.00	-0.38	0.09	-0.400* (0.219)	0.293*** (0.004)	-0.483*** (0.096)
Bank_data_sources	0.03	0.00	0.02			0.020 (0.093)
Data_Resources_Inc	0.10	-0.02	0.06			-0.062 (0.116)
Publication characteristics						
IF_recursive	0.03	0.00	0.01			0.079 (0.060)
Citations	0.06	0.00	0.01			0.044** (0.018)
firstpub	1.00	0.03	0.00	0.0341*** (0.011)		0.040*** (0.006)
Constant	1.00	-0.02	NA	-0.0204 (0.025)	-0.0641 (0.087)	-0.033*** (0.012)
Number of obs.		2,989		2,989	2,989	2,989
Number of studies		74		74	74	74

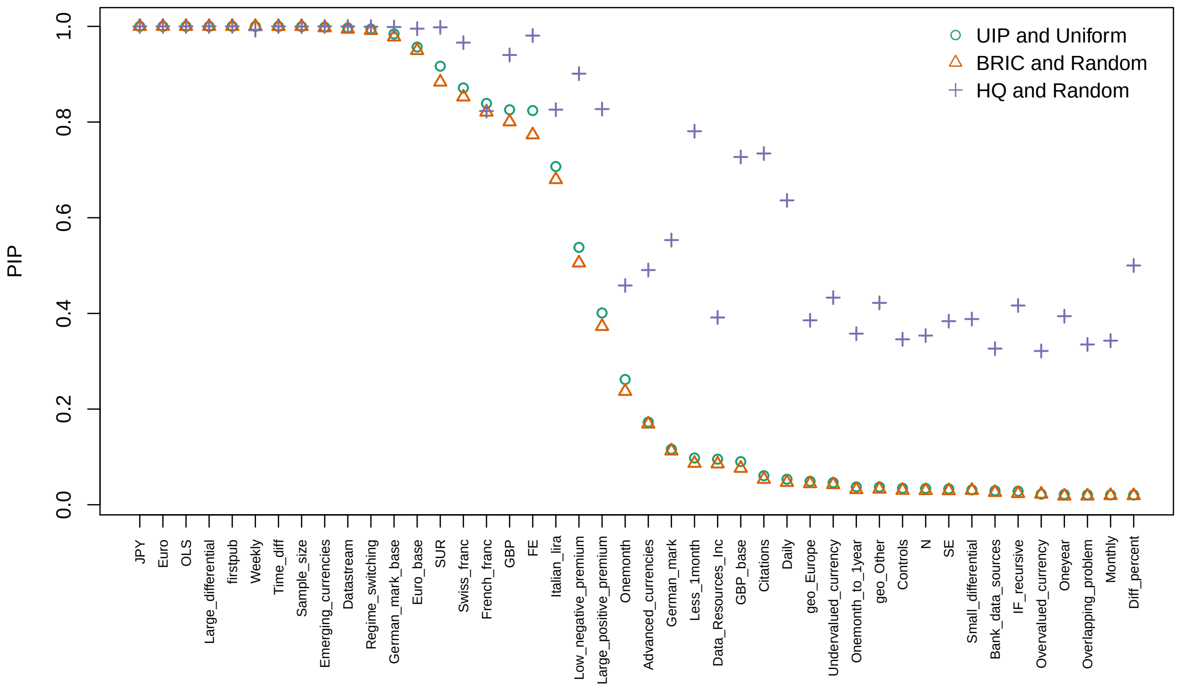
Notes: Response variable = estimate of β from Equation 3. SD = standard deviation, OLS = ordinary least squares regression, FE = fixed effects regression, BMA = Bayesian Model Averaging, FMA = frequentist model averaging, PIP = posterior inclusion probability. In OLS and FE we include only variables with PIP $>$ 0.5. The standard errors in OLS and FE are clustered at the study level. The data is weighted by the inverse of the number of estimates reported per study. For detailed description of all the variables see Table 5.

Estimation method Our results suggest that estimates arising from regime switching models tend to be on average twice smaller than those from seemingly unrelated regressions. In addition, simple OLS and panel fixed effects models yield estimates that are on average larger than those from the regime switching approaches but smaller than those from seemingly unrelated regressions. This is in line with Fama (1984), who applies both OLS and seemingly unrelated regressions to test for the unbiasedness hypothesis and finds that the estimated slope coefficients from the seemingly unrelated regressions are closer to zero (less negative) compared to the estimates from OLS. The omitted estimation category comprises other techniques, most prominently instrumental variables. Thus the results are consistent with the observation that methods that try to account for potential endogeneity tend to yield larger estimates.

Regimes We find that periods characterized by large differentials between (for example) interest rates or money growth coincide with less forward premium bias. The variable is decisive for explaining the reported β, and the implied effect estimate for this method choice is greater by about 2.5. Studies by Baillie & Kiliç (2006) and Baillie & Chang (2011) are consistent with our results, because they observe that in periods characterized by large money supply differentials, large interest rate differentials, and when foreign money growth is in excess of

US money growth, the uncovered interest rate parity is likely to hold. This observation has direct implications for the smaller bias in coefficient β in these periods. In addition, we find weak evidence for β being larger when the forward premium is negative. The existence of an asymmetric effect is confirmed also by Grossmann *et al.* (2014), who find that a significant forward premium anomaly exists for advanced country currencies when the numeraire currency sells at a premium.

Figure 7: *Sensitivity of BMA to different priors*



Notes: UIP (unit information prior) and Uniform = priors recommended by Eicher *et al.* (2011). BRIC and Random = the benchmark g-prior for parameters with the beta-binomial model prior for model space; this ensures that each model size has equal prior probability. HQ prior behaves asymptotically like the Hannan-Quinn criterion. PIP stands for posterior inclusion probability.

Data sources & Publication characteristics The source of the data that is used to test the null hypothesis matters. We find decisive evidence that data extracted from Datastream are associated with 0.4 smaller estimates of β. Furthermore, estimates of β have an increasing time trend, which corroborates the simple graph presented in the Introduction. We find decisive evidence that the reported β coefficients increase by about 0.03 per year. On the other hand, publication quality, as captured by the number of citations and impact factor of the outlet, do not affect the size of reported estimates of β. Consequently, there is not much evidence that estimates extracted from studies of higher quality are different from those originating from studies of lower quality after accounting for other aspects of the context in which the estimates were obtained.

5.4 Implied Estimates

In this subsection we calculate the mean β implied by the literature after we take into account the effect that currency, method, and data decisions have on the estimates of β. For this purpose, we create a synthetic study in which we include all the 2,989 estimates of β extracted from the differences specification. Nevertheless, we assign different weights to different aspects of data, methodology, and publication quality. Effectively, we calculate a "best practice" estimate of β by using the results of FMA in Subsection 5.3 and calculating a linear combination of FMA coefficients and the chosen values for each variable. We use the results from FMA instead of those from BMA because FMA allows us to construct confidence intervals around the implied β. In the linear combination of the coefficients we plug in sample maxima for variables that are in line with best practice in the literature, sample minima for variables that depart from best practice, and sample means for variables where best practice is not obvious. Of course, any definition of best practice is subjective, but we do our best to follow the consensus in the literature.

Table 7: *Best-practice estimates by currency*

	Implied β	**95% confidence interval**		**Correction vs. mean**
All	0.383	-1.026	1.793	1.223
Advanced currencies	0.312	-1.047	1.670	1.009
Emerging currencies	0.979	-0.370	2.328	0.220
Japanese yen	-0.388	-1.496	0.720	1.241
German mark	0.207	-1.123	1.537	1.144
British pound	0.059	-1.289	1.406	1.062
French franc	0.543	-0.846	1.931	0.843
Italian lira	0.564	-0.805	1.933	0.733
Swiss franc	-0.028	-1.395	1.340	1.020
European currencies	0.454	-1.003	1.911	1.288
Euro	-0.705	-2.163	0.753	1.556
non-European/non-Asian currencies	0.323	-1.146	1.792	0.895

Notes: The table presents mean estimates of β conditional on selected data, estimation, and publication characteristics. The exercise is akin to constructing a synthetic study that uses all estimates in the literature but puts more weight on selected aspects of study design. The results in the column 'Implied β' are conditional on our definition of best practice (see the text for more details). Individual rows show best practice estimates for different currencies in our sample. The last column shows difference between a best practice estimate and the mean estimate for the corresponding currency.

We prefer precise estimates, estimates using foreign exchange spot and forward rates over longer horizons, controlling for overlapping samples problem in the data, and using more complex estimation techniques than OLS; so we plug sample minima for the standard error, less-than-one-month dummy, overlapping-problem dummy, and OLS dummy. We also prefer the use of foreign exchange rates of longer maturities, larger sample size, estimation methods allowing for different regimes or cross-sectional dependence across currencies, estimates extracted from more recent studies and studies of higher publication quality. In line with these best practice preferences in the literature we plug sample maxima one-year horizon dummy, time difference, sample size, regime-switching and SUR dummies, impact factor, citation count, and publication year. For all the remaining variables included in the FMA we cannot reliably discern best practice and thus plug in their sample means.

The results of the best practice exercise for different currencies are shown in Table 7. Aggregating across all countries we observe a best practice β estimate of 0.38, which is in line with our baseline result obtained after correcting for publication bias. This estimate is below the theoretically predicted value of 1; however, in contrast to the conclusion in numerous prior studies, it is positive rather than negative. Furthermore, distinguishing between more and less developed economies we obtain a best practice β estimate of 0.31 for advanced economy currencies and of 0.98 for emerging economy currencies. While the wide confidence intervals of the implied estimates do not allow us to conclude that the best practice estimates are statistically different from zero at the 5% level, all these estimates are clearly larger than the simple mean estimates of β, which are reported in Table 1.

6 Conclusion

We present the first quantitative synthesis of the extensive empirical literature on the forward premium puzzle. We collect 3,643 estimates of β from 91 studies, which makes this synthesis one of the largest meta-analyses ever conducted in economics and finance. Our results suggest that, after correction for various biases, the average slope coefficient β in the literature is positive but smaller than 1.

Furthermore, we exploit the heterogeneity in data samples and estimation methodologies and examine the impact of 43 study and estimation characteristics on the reported β estimates. To address the problem of model uncertainty arising from the presence of many potential explanatory variables we use the Bayesian and frequentist model averaging techniques. We observe systematic differences between currencies of more and less advanced economies with higher β estimates for emerging country currencies, and to a lesser extent the former French franc and Italian lira. We also find that the β estimates tend to be larger when using weekly observations, longer-term exchange rates, larger samples, and sophisticated estimation methods that account for potential endogeneity. Furthermore, we find evidence that β estimates are regime-dependent as they differ across different time periods and they tend to increase over time. On the contrary, we document no systematic effect of publication quality proxies on the results.

As the bottom line of our analysis we use 2,989 estimates of β from the differences specification to construct a synthetic study based on weighted study characteristics. We obtain a best practice β estimate of 0.38, which is close to our estimates that correct for the publication bias without any judgment on the relative desirability of data samples and methodology. Nevertheless, our β estimates based on the synthetic control study exercise exhibit wide confidence intervals, which imply that the estimates are not statistically different from 0 at the conventional 5% significance level.

Three qualifications of our results are in order. First, we only collect data from studies published in peer-reviewed journals and focusing on forward rates. In principle, one could also include unpublished papers and studies using the interest rate differential on the right-hand side of the regression. We prefer published studies from unpublished ones because the former are likely to be of a higher quality due to the peer-review process. Studies using the interest rate differential produce estimates comparable to our β only if the covered interest parity holds, which does not have to be the case for all markets, especially after the financial crisis. It is almost inevitable that a meta-analysis will miss some of the studies that could have been included, but that causes no bias unless the studies are omitted selectively based on their results. Second, estimates reported within one study are unlikely to be independent. We attempt to tackle this issue by clustering the standard errors in our regressions at the study level, but we admit that clustering is not an ultimate solution to the problem of sample overlap. Third, our best-practice analysis is inevitably subjective, because judgment is required on various aspects of study design. Different authors may choose a somewhat different set of characteristics. Nevertheless, plausible changes to the definition of best practice keep our results qualitatively unchanged.

References

Amini, S. M. & C. Parmeter (2012): "Comparison of model averaging techniques: Assessing growth determinants." *Journal of Applied Econometrics* **27(5)**: pp. 870–876.

Andrews, I. & M. Kasy (2019): "Identification of and correction for publication bias." *American Economic Review* **109(8)**: pp. 2766–94.

Bacchetta, P. & E. van Wincoop (2010): "Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle." *American Economic Review* **100(3)**: pp. 870–904.

Backus, D. K., A. W. Gregory, & C. I. Telmer (1993): "Accounting for Forward Rates in Markets for Foreign Currency." *Journal of Finance* **48(5)**: pp. 1887–1908.

Baillie, R. T. & T. Bollerslev (1989): "Common Stochastic Trends in a System of Exchange Rates." *Journal of Finance* **44(1)**: pp. 167–81.

Baillie, R. T. & T. Bollerslev (1994): "Cointegration, Fractional Cointegration, and Exchange Rate Dynamics." *Journal of Finance* **49(2)**: pp. 737–45.

Baillie, R. T. & S. S. Chang (2011): "Carry trades, momentum trading and the forward premium anomaly." *Journal of Financial Markets* **14(3)**: pp. 441 – 464.

Baillie, R. T. & R. Kiliç (2006): "Do asymmetric and nonlinear adjustments explain the forward premium anomaly?" *Journal of International Money and Finance* **25(1)**: pp. 22 – 47.

Bansal, R. & M. Dahlquist (2000): "The forward premium puzzle: different tales from developed and emerging economies." *Journal of International Economics* **51(1)**: pp. 115–144.

Bekaert, G. (1995): "The Time Variation of Expected Returns and Volatility in Foreign-Exchange Markets." *Journal of Business & Economic Statistics* **13(4)**: pp. 397–408.

Bekaert, G. & R. J. Hodrick (1993): "On biases in the measurement of foreign exchange risk premiums." *Journal of International Money and Finance* **12(2)**: pp. 115–138.

Blanco-Perez, C. & A. Brodeur (2020): "Publication Bias and Editorial Statement on Negative Findings." *Economic Journal* **(forthcoming)**.

Bom, P. R. D. & H. Rachinger (2019): "A kinked meta-regression model for publication bias correction." *Research Synthesis Methods* **10(4)**: pp. 497–514.

BREEDON, F., D. RIME, & P. VITALE (2016): "Carry Trades, Order Flow, and the Forward Bias Puzzle." *Journal of Money, Credit and Banking* **48(6)**: pp. 1113–1134.

BRODEUR, A., N. COOK, & A. HEYES (2020): "Methods Matter: P-Hacking and Causal Inference in Economics." *American Economic Review* **(forthcoming)**.

BRODEUR, A., M. LÉ, M. SANGNIER, & Y. ZYLBERBERG (2016): "Star wars: The empirics strike back." *American Economic Journal: Applied Economics* **8(1)**: pp. 1–32.

BRUNS, S. B. & J. P. A. IOANNIDIS (2016): "p-Curve and p-Hacking in Observational Research." *PloS ONE* **11(2)**: p. e0149144.

BURNSIDE, C., B. HAN, D. HIRSHLEIFER, & T. Y. WANG (2011): "Investor Overconfidence and the Forward Premium Puzzle." *Review of Economic Studies* **78(2)**: pp. 523–558.

BYERS, J. D. & D. A. PEEL (1991): "Some evidence on the efficiency of the sterling-dollar and sterling-franc forward exchange rates in the interwar period." *Economics Letters* **35(3)**: pp. 317–322.

CHIANG, T. C. (1988): "The Forward Rate as a Predictor of the Future Spot Rate–A Stochastic Coefficient Approach." *Journal of Money, Credit and Banking* **20(2)**: pp. 212–32.

CHINN, M. & G. MEREDITH (2004): "Monetary policy and long-horizon uncovered interest parity." *IMF Staff Papers* **51(3)**: pp. 409–430.

CHRISTENSEN, G. & E. MIGUEL (2018): "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature* **56(3)**: pp. 920–980.

CORNELL, B. (1989): "The impact of data errors on measurement of the foreign exchange risk premium." *Journal of International Money and Finance* **8(1)**: pp. 147–157.

COUDERT, V. & V. MIGNON (2013): "The "forward premium puzzle" and the sovereign default risk." *Journal of International Money and Finance* **32(C)**: pp. 491–511.

COULIBALY, D. & H. KEMPF (2019): "Inflation targeting and the forward bias puzzle in emerging countries." *Journal of International Money and Finance* **90(C)**: pp. 19–33.

CROWDER, W. J. (1994): "Foreign exchange market efficiency and common stochastic trends." *Journal of International Money and Finance* **13(5)**: pp. 551–564.

DOMOWITZ, I. & C. S. HAKKIO (1985): "Conditional variance and the risk premium in the foreign exchange market." *Journal of International Economics* **19(1-2)**: pp. 47–66.

EICHER, T. S., C. PAPAGEORGIOU, & A. E. RAFTERY (2011): "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

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

FLOOD, R. P. & A. K. ROSE (1996): "Fixes: Of the Forward Discount Puzzle." *The Review of Economics and Statistics* **78(4)**: pp. 748–52.

FRANKEL, J. & K. FROOT (1987): "Using survey data to test standard propositions regarding exchange rate expectations." *American Economic Review* **77(1)**: pp. 133–53.

FRANKEL, J. & J. POONAWALA (2010): "The forward market in emerging currencies: Less biased than in major currencies." *Journal of International Money and Finance* **29(3)**: pp. 585–598.

FRANKEL, J. A. (1982): "In search of the exchange risk premium: A six-currency test assuming mean-variance optimization." *Journal of International Money and Finance* **1(1)**: pp. 255–274.

FRANKEL, J. A. (1986): "The implications of mean-variance optimization for four questions in international macroeconomics." *Journal of International Money and Finance* **5(1)**: pp. S53–S75.

FROOT, K. A. & J. A. FRANKEL (1989): "Forward Discount Bias: Is it an Exchange Risk Premium?" *The Quarterly Journal of Economics* **104(1)**: pp. 139–161.

Froot, K. A. & R. H. Thaler (1990): "Foreign Exchange." *Journal of Economic Perspectives* **4(3)**: pp. 179–92.

Furukawa, C. (2019): "Publication bias under aggregation frictions: Theory, evidence, and a new correction method." *Working paper*, MIT.

Goodhart, C. (1988): "The foreign exchange market: A random walk with a dragging anchor." *Economica* **55(220)**: pp. 437–460.

Goodhart, C. A. E., P. C. McMahon, & Y. L. Ngama (1997): "Why Does the Spot-Forward Discount Fail to Predict Changes in Future Spot Rates." *International Journal of Finance & Economics* **2(2)**: pp. 121–29.

Grossmann, A., A. A. Lee, & M. W. Simpson (2014): "Forward premium anomaly of the british pound and the euro." *International Review of Financial Analysis* **34(C)**: pp. 140 – 156.

Hai, W., N. C. Mark, & Y. Wu (1997): "Understanding Spot and Forward Exchange Rate Regressions." *Journal of Applied Econometrics* **12(6)**: pp. 715–34.

Hansen, B. E. (2007): "Least Squares Model Averaging." *Econometrica* **75(4)**: pp. 1175–1189.

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators." *Econometrica* **50(4)**: pp. 1029–1054.

Havranek, T. (2015): "Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting." *Journal of the European Economic Association* **13(6)**: pp. 1180–1204.

Hedges, L. V. (1992): "Modeling publication selection effects in meta-analysis." *Statistical Science* **7(2)**: pp. 246–255.

Ioannidis, J. P. A., T. D. Stanley, & H. Doucouliagos (2017): "The Power of Bias in Economics Research." *The Economic Journal* **127(605)**: pp. 236–265.

Jeffreys, H. (1961): *Theory of Probability (3rd edn)*. Oxford University Press.

Jongen, R., W. F. Verschoor, & C. C. Wolff (2008): "Foreign Exchange Rate Expectations: Survey And Synthesis." *Journal of Economic Surveys* **22(1)**: pp. 140–165.

MacDonald, R. & M. P. Taylor (1990): "The Term Structure of Forward Foreign Exchange Premia: The Inter-war Experience." *The Manchester School of Economic & Social Studies* **58(1)**: pp. 54–65.

Madigan, D., J. York, & D. Allard (1995): "Bayesian graphical models for discrete data." *International Statistical Review* **63(2)**: pp. 215–232.

McCallum, B. T. (1994): "A reconsideration of the uncovered interest parity relationship." *Journal of Monetary Economics* **33(1)**: pp. 105–132.

McCloskey, D. N. & S. T. Ziliak (2019): "What Quantitative Methods Should We Teach to Graduate Students? A Comment on Swann's Is Precise Econometrics an Illusion?" *Journal of Economic Eduation* **50(4)**: pp. 356–361.

McFarland, J. W., P. C. McMahon, & Y. Ngama (1994): "Forward exchange rates and expectations during the 1920s: A re-examination of the evidence." *Journal of International Money and Finance* **13(6)**: pp. 627–636.

Mussa, M. (1979): "Empirical regularities in the behavior of exchange rates and theories of the foreign exchange market." *Carnegie-Rochester Conference Series on Public Policy* **11(1)**: pp. 9–57.

Nadal De Simone, F. & W. Razzak (1999): "Nominal Exchange Rates and Nominal Interest Rate Differentials." *IMF Working Papers 99/141*, International Monetary Fund.

Pippenger, J. (2011): "The solution to the forward-bias puzzle." *Journal of International Financial Markets, Institutions and Money* **21(2)**: pp. 296–304.

Razzak, W. A. (2002): "The Forward Rate Unbiasedness Hypothesis Revisited." *International Journal of Finance & Economics* **7(4)**: pp. 293–308.

Stanley, T. & H. Doucouliagos (2012): *Meta-Regression Analysis in Economics and Business.* London: Routledge.

Stanley, T. & H. Doucouliagos (2014): "Meta-regression approximations to reduce publication selection bias." *Research Synthesis Methods* **5(1)**: pp. 60–78.

Stanley, T. D. (2005): "Beyond Publication Bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Stanley, T. D., S. B. Jarrell, & H. Doucouliagos (2010): "Could It Be Better to Discard 90% of the Data? A Statistical Paradox." *The American Statistician* **64(1)**: pp. 70–77.

Steel, M. F. J. (2020): "Model averaging and its use in economics." *Journal of Economic Literature* **(forthcoming)**.

Wolf, C. C. P. (1987): "Forward Foreign Exchange Rates, Expected Spot Rates, and Premia: A Signal-Extraction Approach." *Journal of Finance* **42(2)**: pp. 395–406.

Yu, J. (2013): "A sentiment-based explanation of the forward premium puzzle." *Journal of Monetary Economics* **60(4)**: pp. 474–491.

Zhou, S. & A. M. Kutan (2005): "Does the forward premium anomaly depend on the sample period used or on the sign of the premium?" *International Review of Economics & Finance* **14(1)**: pp. 17 – 25.

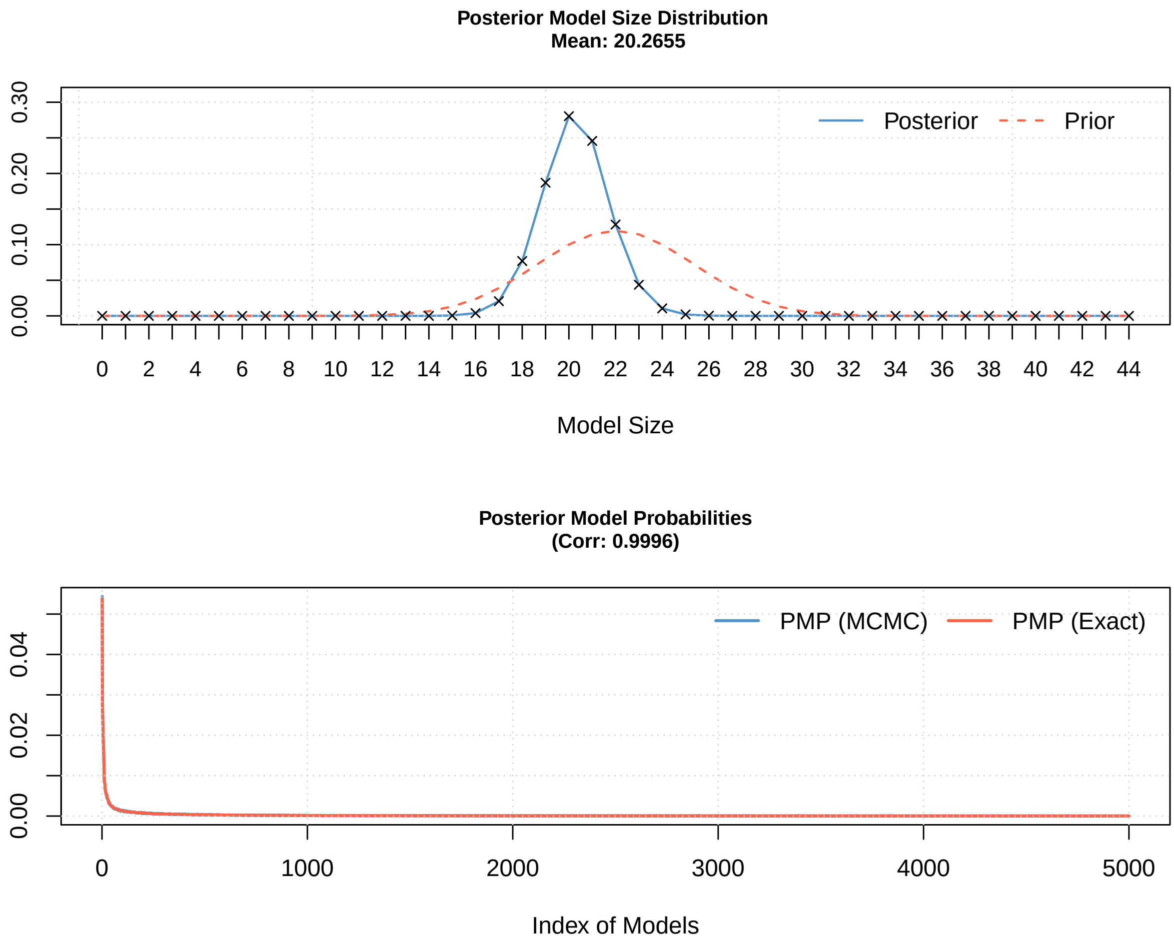
A BMA Diagnostics and Robustness Checks

Table A1: Diagnostics of the baseline BMA estimation

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
20.2664	5,000,000	1,000,000	4.195397 mins	720,348
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
1.80E+13	0.0000041%	81%	0.9996	2,989
Model prior	*g-prior*	*Shrinkage-stats*		
Uniform	UIP	Av = 0.9997		

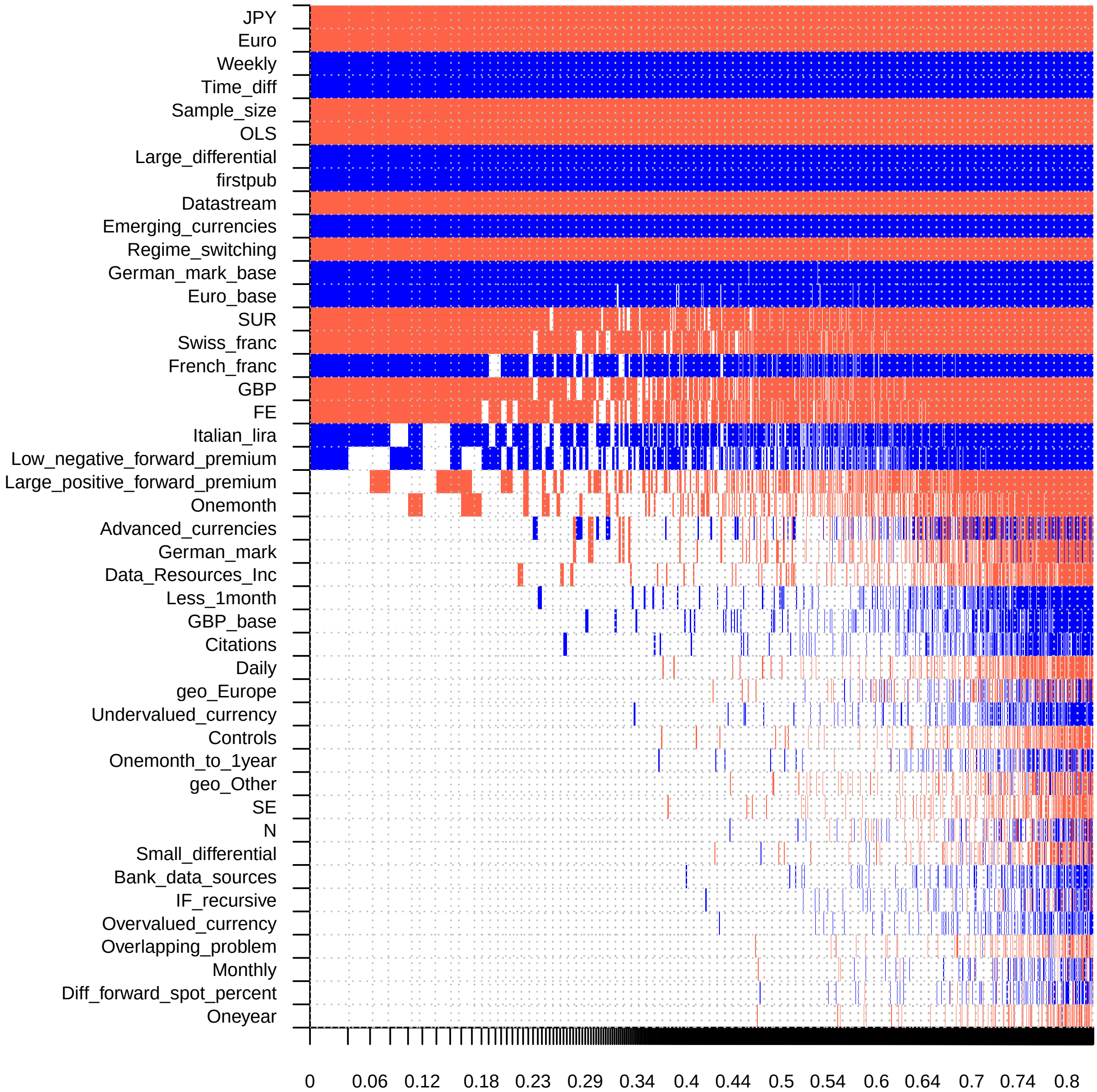
Notes: We use the uniform model prior, which gives each model the same prior probability, and the unit information prior, which contains the same amount of information as one observation in the data. These priors are proposed by Eicher *et al.* (2011). The results of this BMA specification are reported in Table 6.

Figure A1: *Baseline BMA - Model size and convergence*



Notes: The figure shows the posterior model size distribution and the posterior model probabilities of our baseline BMA analysis reported in Table 6.

Figure A2: *Model inclusion in BMA with alternative priors*



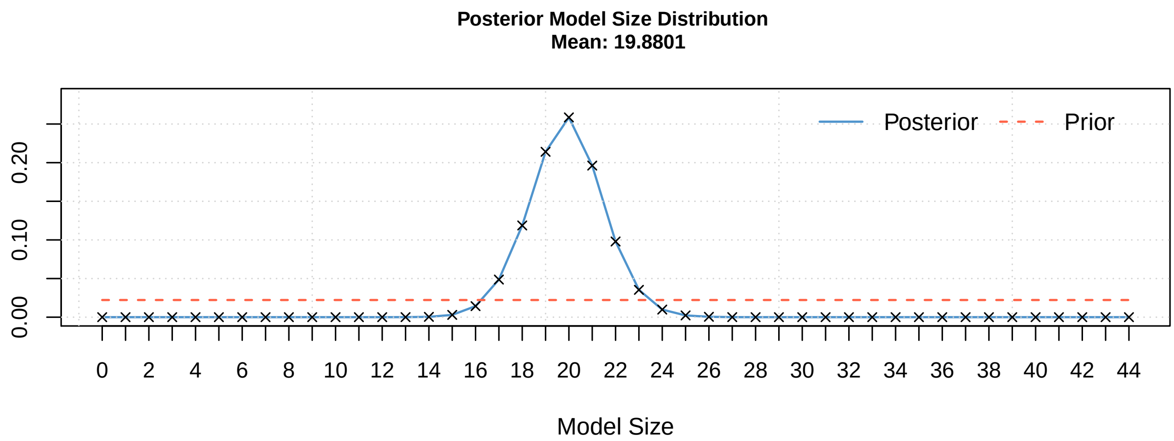
Notes: The response variable is the estimate of slope coefficient *beta* from Equation 3. Columns show individual models and variables are listed in descending order by their posterior inclusion probabilities. The horizontal axis shows cumulative posterior model probabilities from the 5,000 best models. The priors used are BRIC, the benchmark g-prior for parameters, and the beta-binomial model prior for model space, which ensures that each model size has equal prior probability. Blue color (darker in grayscale) = the variable is included in the model with a positive sign. Red color (lighter in grayscale) = the variable is included in the model with a negative sign. No color = the variable is missing from the model. The data is weighted by the inverse of the number of estimates reported per study. For description of all variables see Table 5.

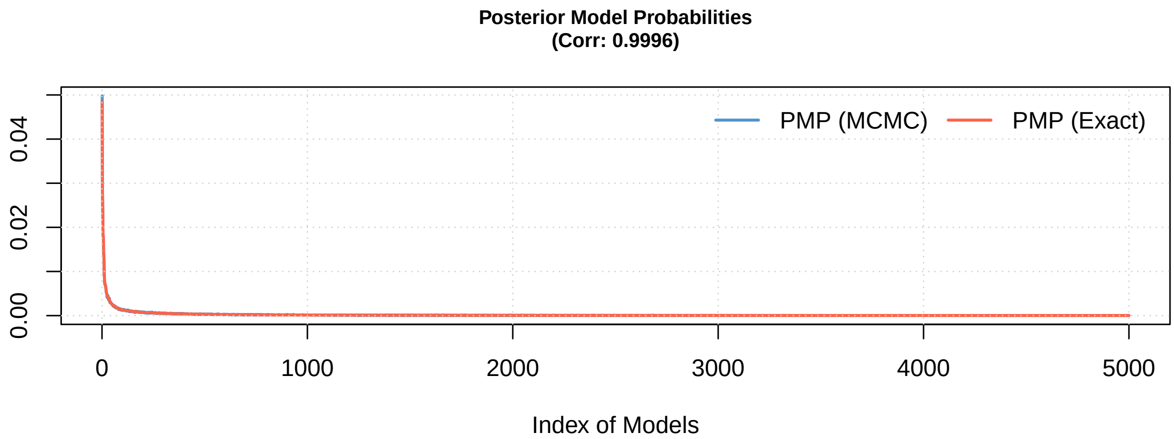
Table A2: BMA diagnostics with alternative priors

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
19.8722	5,000,000	1,000,000	4.495317 mins	724,011
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
1.80E+13	0.0000041%	82%	0.9997	2,989
Model prior	*g-prior*	*Shrinkage-stats*		
Random	BRIC	Av = 0.9997		

Notes: We use BRIC, the benchmark g-prior for parameters, and the beta-binomial model prior for model space, which ensures that each model size has equal prior probability.

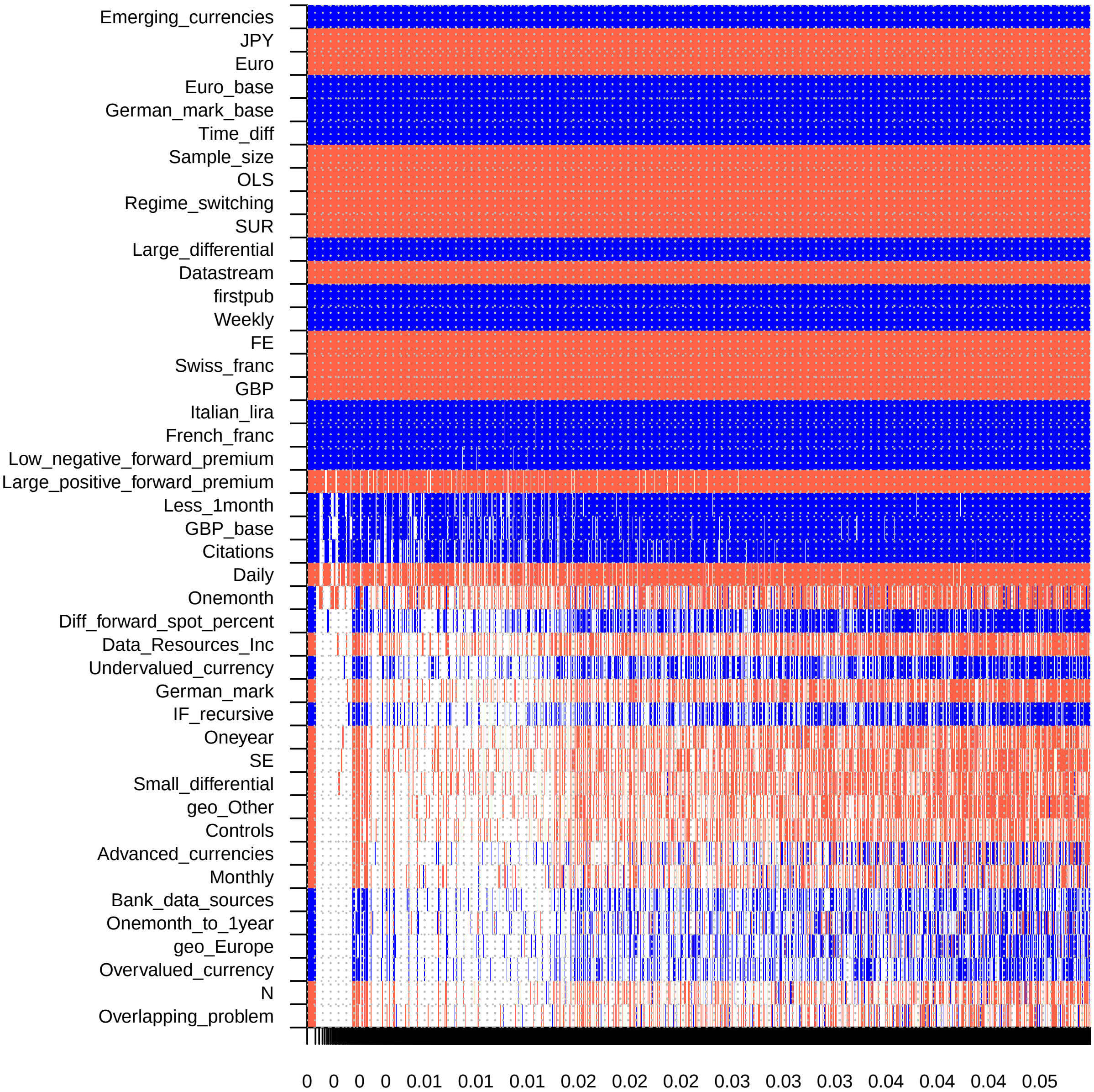
Figure A3: *Alternative BMA priors - Model size and convergence*





Notes: The figure shows the posterior model size distribution and the posterior model probabilities of the BMA exercise with BRIC prior for parameters and the beta-binomial model prior for model space.

Figure A4: *Model inclusion in BMA with alternative priors*



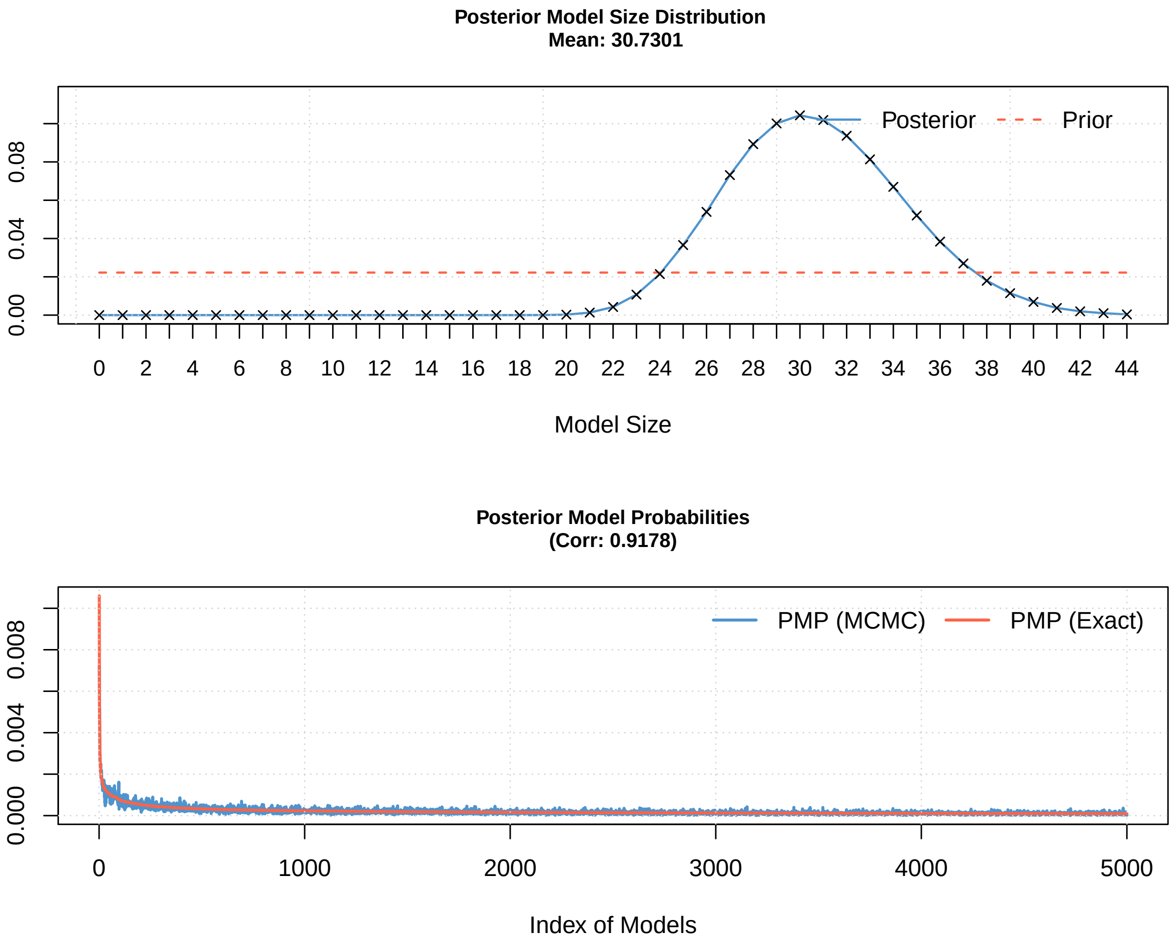
Notes: The response variable is the estimate of slope coefficient *beta* from Equation 3. Columns show individual models and variables are listed in descending order by their posterior inclusion probabilities. The horizontal axis shows cumulative posterior model probabilities from 5,000 best models. The priors used are HQ for parameters, which behaves asymptotically akin to Hannan-Quinn criterion, and the beta-binomial model prior, which ensures that each model size has equal prior probability. Blue color (darker in grayscale) = the variable is included in the model with a positive sign. Red color (lighter in grayscale) = the variable is included in the model with a negative sign. No color = the variable is missing from the model. The data is weighted by the inverse of the number of estimates reported per study. For description of all variables see Table 5.

Table A3: BMA diagnostics with alternative priors

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*Models visited*
30.7568	5,000,000	1,000,000	11.09895 mins	2,118,410
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
1.80E+13	0.00000012%	5%	0.953	2,989
Model prior	*g-prior*	*Shrinkage-stats*		
Random	hyper (a=2.001)	Av = 0.9752		
		St. dev.= 0.007		

Notes: We use HQ prior for parameters, which behaves asymptotically akin to Hannan-Quinn criterion, and the beta-binomial model prior for model space, which ensures that each model size has equal prior probability.

Figure A5: *Alternative BMA priors - Model size and convergence*



Notes: The figure shows the posterior model size distribution and the posterior model probabilities of the BMA exercise with HQ prior for parameters and the beta-binomial model prior for model space.

B Studies Included in the Meta-analysis

Aggarwal, R., W. T. Lin, & S. K. Mohanty (2008): "Are Forward Exchange Rates Rational Forecasts of Future Spot Rates? An Improved Econometric Analysis for the Major Currencies." *Multinational Finance Journal* **12(1-2)**: pp. 1–20.

Aggarwal, R., B. M. Lucey, & S. K. Mohanty (2009): "The Forward Exchange Rate Bias Puzzle Is Persistent: Evidence from Stochastic and Nonparametric Cointegration Tests." *The Financial Review* **44(4)**: pp. 625–645.

Ahmad, R., S. G. Rhee, & Y. M. Wong (2012): "Foreign exchange market efficiency under recent crises: Asia-Pacific focus." *Journal of International Money and Finance* **31(6)**: pp. 1574–1592.

Al-Zoubi, H. A. (2011): "A new look at the forward premium "puzzle"." *Journal of Futures Markets* **31(7)**: pp. 599–628.

Amri, S. (2008): "Analysing the forward premium anomaly using a Logistic Smooth Transition Regression model." *Economics Bulletin* **6(26)**: pp. 1–18.

Azouzi, D., R. V. Kumar, & C. Aloui (2011): "Forward Rate Unbiasedness Hypothesis in the Tunisian Exchange Rate Market." *International Journal of Academic Research in Business and Social Sciences* **1(2)**: pp. 17–44.

Bai, S. & A. V. Mollick (2010): "Currency crisis and the forward discount bias: Evidence from emerging economies under breaks." *Journal of International Financial Markets, Institutions and Money* **20(5)**: pp. 556–574.

Baillie, R. T. & S. S. Chang (2011): "Carry trades, momentum trading and the forward premium anomaly." *Journal of Financial Markets* **14(3)**: pp. 441–464.

Baillie, R. T. & D. Cho (2014): "Time variation in the standard forward premium regression: Some new models and tests." *Journal of Empirical Finance* **29(C)**: pp. 52–63.

Baillie, R. T. & R. Kilic (2006): "Do asymmetric and nonlinear adjustments explain the forward premium anomaly?" *Journal of International Money and Finance* **25(1)**: pp. 22–47.

Baillie, R. T., R. E. Lippens, & P. C. McMahon (1983): "Testing Rational Expectations and Efficiency in the Foreign Exchange Market." *Econometrica* **51(3)**: pp. 553–563.

Bakshi, G. S. & A. Naka (1997): "Unbiasedness of the Forward Exchange Rates." *The Financial Review* **32(1)**: pp. 145–162.

Bansal, R. & M. Dahlquist (2000): "The forward premium puzzle: different tales from developed and emerging economies." *Journal of International Economics* **51(1)**: pp. 115–144.

Barnhart, S., R. McNown, & M. Wallace (2002): "Some answers to puzzles in testing unbiasedness in the foreign exchange market." *Applied Financial Economics* **12(10)**: pp. 687–696.

Barnhart, S. W., R. McNown, & M. S. Wallace (1999): "Non-Informative Tests of the Unbiased Forward Exchange Rate." *Journal of Financial and Quantitative Analysis* **34(2)**: pp. 265–291.

Barnhart, S. W. & A. C. Szakmary (1991): "Testing the Unbiased Forward Rate Hypothesis: Evidence on Unit Roots, Co-Integration, and Stochastic Coefficients." *Journal of Financial and Quantitative Analysis* **26(2)**: pp. 245–267.

Bekaert, G. & R. J. Hodrick (1993): "On biases in the measurement of foreign exchange risk premiums." *Journal of International Money and Finance* **12(2)**: pp. 115–138.

Bhatti, R. & I. Moosa (1995): "An alternative approach to testing uncovered interest parity." *Applied Economics Letters* **2(12)**: pp. 478–481.

Bhatti, R. H. (2014): "The existence of uncovered interest parity in the CIS countries." *Economic Modelling* **40(C)**: pp. 227–241.

Boothe, P. & D. Longworth (1986): "Foreign exchange market efficiency tests: Implications of recent empirical findings." *Journal of International Money and Finance* **5(2)**: pp. 135–152.

Breedon, F., D. Rime, & P. Vitale (2016): "Carry Trades, Order Flow, and the Forward Bias Puzzle." *Journal of Money, Credit and Banking* **48(6)**: pp. 1113–1134.

Breuer, J. B. & M. E. Wohar (1996): "The Road Less Travelled: Institutional Aspects of Data and Their Influence on Empirical Estimates with an Application to Tests of Forward Rate Unbiasedness." *Economic Journal* **106(434)**: pp. 26–38.

Burnside, C., B. Han, D. Hirshleifer, & T. Y. Wang (2011): "Investor Overconfidence and the Forward Premium Puzzle." *Review of Economic Studies* **78(2)**: pp. 523–558.

Cavaglia, S. M. F. G., W. F. C. Verschoor, & C. C. P. Wolff (1994): "On the Biasedness of Forward Foreign Exchange Rates: Irrationality or Risk Premia?" *The Journal of Business* **67(3)**: pp. 321–343.

Chakraborty, A. (2007): "Learning, Forward Premium Puzzle and Exchange Rate Fundamentals under Sticky Prices." *Economics Bulletin* **6(34)**: pp. 1–13.

Chakraborty, A. & S. E. Haynes (2008): "Econometrics of the Forward Premium Puzzle." *Economics Bulletin* **6(42)**: pp. 1–17.

Chiang, T. C. (1986): "Empirical Analysis On The Predictors Of Future Spot Rates." *Journal of Financial Research* **9(2)**: pp. 153–162.

Chiang, T. C. (1988): "The Forward Rate as a Predictor of the Future Spot Rate–A Stochastic Coefficient Approach." *Journal of Money, Credit and Banking* **20(2)**: pp. 212–232.

Cho, D. (2018): "On the persistence of the forward premium in the joint presence of nonlinearity, asymmetry, and structural changes." *Economic Modelling* **70(C)**: pp. 310–319.

Christensen, M. (2000): "Uncovered interest parity and policy behavior: new evidence." *Economics Letters* **69(1)**: pp. 81–87.

Clarida, R., J. Davis, & N. Pedersen (2009): "Currency carry trade regimes: Beyond the Fama regression." *Journal of International Money and Finance* **28(8)**: pp. 1375–1389.

da Costa, C. E., J. de Jesus Filho, & P. Matos (2016): "Forward-premium puzzle: is it time to abandon the usual regression?" *Applied Economics* **48(30)**: pp. 2852–2867.

Diamandis, P. F., D. A. Georgoutsos, & G. P. Kouretas (2008): "Testing the forward rate unbiasedness hypothesis during the 1920s." *Journal of International Financial Markets, Institutions and Money* **18(4)**: pp. 358–373.

Ding, L. (2012): "The Thursday effect of the forward premium puzzle." *International Review of Economics & Finance* **21(1)**: pp. 302–318.

Ding, L. & L. To (2010): "The Forward Premium Puzzle Across Maturities." *Economics Bulletin* **30(2)**: pp. 1113–1119.

Engel, C. (1996): "The forward discount anomaly and the risk premium: A survey of recent evidence." *Journal of Empirical Finance* **3(2)**: pp. 123–192.

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

Frankel, J. & J. Poonawala (2010): "The forward market in emerging currencies: Less biased than in major currencies." *Journal of International Money and Finance* **29(3)**: pp. 585–598.

Frankel, J. A. (1980): "Tests of rational expectations in the forward exchange market." *Southern Economic Journal* **46(4)**: pp. 1083–1101.

Frankel, J. A. & M. D. Chinn (1993): "Exchange Rate Expectations and the Risk Premium: Tests for a Cross Section of 17 Currencies." *Review of International Economics* **1(2)**: pp. 136–144.

Frenkel, J. A. (1976): "A monetary approach to the exchange rate: Doctrinal aspects and empirical evidence." *The Scandinavian Journal of Economics* **78(2)**: pp. 200–224.

Frenkel, J. A. (1981): "Flexible Exchange Rates, Prices, and the Role of "News": Lessons from the 1970s." *Journal of Political Economy* **89(4)**: pp. 665–705.

Froot, K. A. & J. A. Frankel (1989): "Forward Discount Bias: Is it an Exchange Risk Premium?" *The Quarterly Journal of Economics* **104(1)**: pp. 139–161.

Fukuta, Y. & M. Saito (2002): "Forward Discount Puzzle and Liquidity Effects: Some Evidence from Exchange Rates among the United States, Canada, and Japan." *Journal of Money, Credit and Banking* **34(4)**: pp. 1014–1033.

G., H. S., K. Amangeldi, S. P. A. V. B., & T. G. S. (2013): "The forward rate premium puzzle: a case of misspecification?" *Studies in Nonlinear Dynamics & Econometrics* **17(3)**: pp. 265–279.

Goodhart, C. (1988): "The foreign exchange market: A random walk with a dragging anchor." *Economica* **55(220)**: pp. 437–460.

Gospodinov, N. (2009): "A New Look at the Forward Premium Puzzle." *Journal of Financial Econometrics* **7(3)**: pp. 312–338.

Gregory, A. W. & T. H. McCurdy (1984): "Testing the unbiasedness hypothesis in the forward foreign exchange market: A specification analysis." *Journal of International Money and Finance* **3(3)**: pp. 357–368.

Gregory, A. W. & T. H. McCurdy (1986): "The unbiasedness hypothesis in the forward foreign exchange market: A specification analysis with application to France, Italy, Japan, the United Kingdom and West Germany." *European Economic Review* **30(2)**: pp. 365–381.

Grossmann, A., A. A. Lee, & M. W. Simpson (2014): "Forward premium anomaly of the British pound and the euro." *International Review of Financial Analysis* **34(C)**: pp. 140–156.

Hai, W., N. C. Mark, & Y. Wu (1997): "Understanding Spot and Forward Exchange Rate Regressions." *Journal of Applied Econometrics* **12(6)**: pp. 715–734.

Ho, T.-W. (2002): "The Forward Rate Unbiasedness Hypothesis revisited." *Applied Financial Economics* **12(11)**: pp. 799–804.

Ho, T.-W. & W.-S. Mo (2016): "Testing the Persistence of the Forward Premium: Structural Changes or Misspecification?" *Open Economies Review* **27(1)**: pp. 119–138.

Hodgson, D. J., O. Linton, & K. Vorkink (2004): "Testing forward exchange rate unbiasedness efficiently: a semiparametric approach." *Journal of Applied Economics* **7(2)**: pp. 325–353.

Huang, R. D. (1984): "Some alternative tests of forward exchange rates as predictors of future spot rates." *Journal of International Money and Finance* **3(2)**: pp. 153–167.

Huisman, R., K. Koedijk, C. Kool, & F. Nissen (1998): "Extreme support for uncovered interest parity." *Journal of International Money and Finance* **17(1)**: pp. 211–228.

Kumar, R. V. & D. Azouzi (2011): "Tunisian and Indian Forex Markets: A Comparision on Forward Rate Unbiased Hypothesis." *Romanian Economic Journal* **14(40)**: pp. 81–98.

Kumar, S. & S. TrĂĽck (2014): "Unbiasedness and risk premiums in the Indian currency futures market." *Journal of International Financial Markets, Institutions and Money* **29(C)**: pp. 13–32.

Liu, F. & P. Sercu (2009): "The Forex Forward Puzzle: The Career Risk Hypothesis." *The Financial Review* **44(3)**: pp. 371–404.

Longworth, D. (1981): "Testing the Efficiency of the Canadian-U.S. Exchange Market under the Assumption of no Risk Premium." *Journal of Finance* **36(1)**: pp. 43–49.

Loring, G. & B. Lucey (2013): "An analysis of forward exchange rate biasedness across developed and developing country currencies: Do observed patterns persist out of sample?" *Emerging Markets Review* **17(C)**: pp. 14–28.

Luintel, K. B. & K. Paudyal (1998): "Common stochastic trends between forward and spot exchange rates." *Journal of International Money and Finance* **17(2)**: pp. 279–297.

Mark, N. C. & Y. Wu (1998): "Rethinking Deviations from Uncovered Interest Parity: The Role of Covariance Risk and Noise." *Economic Journal* **108(451)**: pp. 1686–1706.

Maynard, A. (2003): "Testing for Forward-Rate Unbiasedness: On Regression in Levels and in Returns." *The Review of Economics and Statistics* **85(2)**: pp. 313–327.

Maynard, A. (2006): "The forward premium anomaly: statistical artefact or economic puzzle? New evidence from robust tests." *Canadian Journal of Economics* **39(4)**: pp. 1244–1281.

Maynard, A. & P. C. B. Phillips (2001): "Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly." *Journal of Applied Econometrics* **16(6)**: pp. 671–708.

McCallum, B. T. (1994): "A reconsideration of the uncovered interest parity relationship." *Journal of Monetary Economics* **33(1)**: pp. 105–132.

McFarland, J. W., P. C. McMahon, & Y. Ngama (1994): "Forward exchange rates and expectations during the 1920s: A re-examination of the evidence." *Journal of International Money and Finance* **13(6)**: pp. 627–636.

Moosa, I. A. (2004): "An empirical examination of the Post Keynesian view of forward exchange rates." *Journal of Post Keynesian Economics* **26(3)**: pp. 395–418.

Naka, A. & G. Whitney (1995): "The unbiased forward rate hypothesis re-examined." *Journal of International Money and Finance* **14(6)**: pp. 857–867.

Newbold, P., M. E. Wohar, T. Rayner, N. Kellard, & C. Ennew (1998): "Two puzzles in the analysis of foreign exchange market efficiency." *International Review of Financial Analysis* **7(2)**: pp. 95–111.

Nikolaou, K. & L. Sarno (2006): "New evidence on the forward unbiasedness hypothesis in the foreign-exchange market." *Journal of Futures Markets* **26(7)**: pp. 627–656.

Olmo, J. & K. Pilbeam (2009): "Uncovered Interest Parity: Are Empirical Rejections of It Valid?" *Journal of Economic Integration* **24(2)**: pp. 369–384.

Phillips, P. C. B. & J. W. McFarland (1997): "Forward exchange market unbiasedness: the case of the Australian dollar since 1984." *Journal of International Money and Finance* **16(6)**: pp. 885–907.

Phillips, P. C. B., J. W. McFarland, & P. C. McMahon (1996): "Robust Tests of Forward Exchange Market Efficiency with Empirical Evidence from the 1920s." *Journal of Applied Econometrics* **11(1)**: pp. 1–22.

Pippenger, J. (2011): "The solution to the forward-bias puzzle." *Journal of International Financial Markets, Institutions and Money* **21(2)**: pp. 296–304.

Razzak, W. A. (2002): "The Forward Rate Unbiasedness Hypothesis Revisited." *International Journal of Finance & Economics* **7(4)**: pp. 293–308.

Roll, R. & S. Yan (2000): "An explanation of the forward premium 'puzzle'." *European Financial Management* **6(2)**: pp. 121–148.

Sakoulis, G., E. Zivot, & K. Choi (2010): "Structural change in the forward discount: Implications for the forward rate unbiasedness hypothesis." *Journal of Empirical Finance* **17(5)**: pp. 957–966.

Sarno, L., G. Valente, & H. Leon (2006): "Nonlinearity in Deviations from Uncovered Interest Parity: An Explanation of the Forward Bias Puzzle." *Review of Finance* **10(3)**: pp. 443–482.

Snaith, S., J. Coakley, & N. Kellard (2013): "Does the forward premium puzzle disappear over the horizon?" *Journal of Banking & Finance* **37(9)**: pp. 3681–3693.

Sosvilla-Rivero, S. & Y. B. Park (1992): "Further tests on the forward exchange rate unbiasedness hypothesis." *Economics Letters* **40(3)**: pp. 325–331.

Tiff Macklem, R. (1991): "Forward exchange rates and risk premiums in artificial economies." *Journal of International Money and Finance* **10(3)**: pp. 365–391.

Vasilyev, D., V. Busygin, & S. Busygin (2017): "Testing and interpreting uncovered interest parity in Russia." *Russian Journal of Economics* **3(2)**: pp. 158 – 173.

Verschoor, W. F. C. & C. C. P. Wolff (2001): "Scandinavian forward discount bias risk premia." *Economics Letters* **73(1)**: pp. 65–72.

Villanueva, O. M. (2007): "Spot-forward cointegration, structural breaks and FX market unbiasedness." *Journal of International Financial Markets, Institutions and Money* **17(1)**: pp. 58–78.

Wang, P. & P. Wang (2009): "Does a "correct" parameter estimate tell a better story about foreign exchange market efficiency?" *Journal of International Money and Finance* **28(2)**: pp. 183 – 197.

Westerlund, J. (2007): "Estimating Cointegrated Panels with Common Factors and the Forward Rate Unbiasedness Hypothesis." *Journal of Financial Econometrics* **5(3)**: pp. 491–522.

Wu, Y. & H. Zhang (1997): "Forward premiums as unbiased predictors of future currency depreciation: a non-parametric analysis." *Journal of International Money and Finance* **16(4)**: pp. 609–623.

Zhou, S. (2002): "The forward premium anomaly and the trend behavior of the exchange rates." *Economics Letters* **76(2)**: pp. 273–279.

Zhou, S. & A. M. Kutan (2005): "Does the forward premium anomaly depend on the sample period used or on the sign of the premium?" *International Review of Economics & Finance* **14(1)**: pp. 17–25.

Zivot, E. (2000): "Cointegration and forward and spot exchange rate regressions." *Journal of International Money and Finance* **19(6)**: pp. 785–812.

European Stability Mechanism

6a Circuit de la Foire Internationale
L-1347 Luxembourg
Tel: +352 260 292 0
www.esm.europa.eu
info@esm.europa.eu