EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Anastasia Allayioti, Anthony Garratt

Herding in the foreign exchange market

No 3243

Abstract

Using a recent and comprehensive data set covering nine of the most actively traded currencies on a monthly basis from 1995 to 2024, this paper explores the presence and potential drivers of herding behaviour in foreign exchange rate forecasts. The dataset features an average of 40–50 forecasters per currency, representing a broader range of currencies, a longer time frame, and a larger cross section of forecasters than is commonly found in the FX herding literature. Our results provide mixed evidence on herding, where the balance tends towards anti-herding conclusions. While some revision-based tests suggest herding when current consensus forecasts are used, this evidence weakens considerably when lagged information is employed. In contrast, forecast-error based tests, Bernhardt et al. statistics, and over-reaction regressions more often point to anti-herding, particularly at longer horizons. Overall, we interpret the findings as suggesting that differences among forecasters are largely attributable to heterogeneous views, noise, or idiosyncratic error rather than systematic convergence toward the consensus. When alternative explanations for expectation formation or revisions are considered, the main findings remain unchanged across a wide range of measures, including different types of uncertainty and FX predictors such as the forward premium, the real exchange rate, and the depreciation rate.

JEL codes: C10; C22; F31; F47; G17.

Keywords: Herding, anti-herding, consensus forecasts, individual forecasts, exchange rates, panel estimation.

Non-technical summary

In recent years, much innovative work on expectations formation has been conducted, and various theories of forecaster behaviour have been proposed to explain the observed heterogeneity in expectations; a feature inconsistent with the full-information rational expectations hypothesis, under which there is no scope for persistent disagreement among forecasters. Unlike macroeconomic expectations, such as those for inflation or output, exchange rate expectations are often highly volatile, and the way agents form beliefs about them can affect capital flows, inflation expectations, and, as a result, the transmission of monetary policy.

This paper investigates one potential mechanism of the expectation formation process in the foreign exchange (FX) market, namely herding. Herding can be defined as the tendency of individual forecasters to align their expectations with those of other forecasters rather than independently assess past and incoming information. The opposite possibility is anti-herding, where forecasters move their expectations away from the consensus view. This behaviour may arise because forecasters have incentives to differentiate themselves from the crowd in order to gain reputational benefits or signal superior information to clients and market participants.

Testing for herding behaviour is challenging, since individual forecasters may cluster together for a variety of reasons unrelated to how they perceive or respond to the consensus. For this reason, the paper considers a range of complementary reduced-form tests and extends the analysis to a panel setting that explicitly allows for cross-sectional dependence across individual forecasters. We also examine whether alternative explanations for expectation formation can help rationalise the observed heterogeneity in exchange rate forecasts.

We employ a dataset that spans a maximum period from March 1995 to December 2024, with an average of 40–50 forecasters per currency. This rich dataset provides a wider range of currencies, a longer time period, and a more diverse cross section of forecasters than is typically employed in the existing literature on FX herding, offering a robust foundation for analysis. Our findings do not point to a consistent or robust presence of herding behaviour. Although some revision-based tests suggest herding when the current consensus forecast is used, this result weakens substantially when lagged information is employed. Where the analysis uncovers clearer patterns, the evidence is more suggestive of anti-herding, whereby forecasters move their reported forecasts away from, rather than toward, the consensus. This interpretation is reinforced by the over-reaction results, which indicate that forecast revisions tend to overshoot rather than adjust gradually. More broadly, the results suggest that differences across forecasters mainly reflect heterogeneity in beliefs, noise, and idiosyncratic error. Finally, despite the common view that uncertainty may encourage market participants to look more closely at the behaviour of others, we find little evidence that a broad range of uncertainty measures plays a decisive role in explaining herding or anti-herding in the FX market.

1 Introduction

Herding, characterised by the tendency of individuals to align their actions or expectations with others rather than independently assessing information, and anti-herding, where individuals deliberately act or form expectations contrary to others, can significantly influence market dynamics, pricing mechanisms, and overall market efficiency in foreign exchange markets. Interest in herding in FX markets lies in its potential to exacerbate market volatility, distort price discovery mechanisms, and amplify systemic risks. Therefore, testing whether herding or anti-herding is prevalent, and understanding the factors that may lie behind these behaviours, is useful for policymakers, market participants, and academics alike.

This paper focuses on testing whether herding behaviour occurs in foreign exchange markets. Our empirical framework is based on individual-level forecasts and how they relate to the consensus or average forecast, using empirical formulations described and motivated in macro-forecasting papers by Gallo et al., (2002) and Clements (2018). In addition, we calculate the test developed by Bernhardt et al. (2006), which complements one of the regression-based tests. The approach uses both individual forecaster-level regressions and panel mean group estimation, allowing for cross-sectional dependence. We take advantage of an extensive dataset collected from the Eikon Refinitiv platform, consisting of individual market participants' nominal exchange rate forecasts for 1-month, 3-month, 6-month, and 12-month horizons across nine major currencies relative to the US dollar, covering the G10. The dataset spans a maximum period from March 1995 to December 2024, with an average of 40–50 forecasters per currency. This rich dataset provides a broader range of currencies, a longer time period, and a more diverse cross section of forecasters than is typically employed in the existing literature on FX herding, offering a robust foundation for analysis.

In addition, we investigate alternative explanations for the formation and revision of expectations, in the context of the regression-based approach used to test herding. If the observed evidence on herding changes, or is robust to the inclusion of a range of factors measuring alternative explanations of expectations formation, we gain some understanding of the various influences on expectations formation and, in turn, on herding-type behaviour. We examine a range of factors largely based on uncertainty, volatility, and well-known FX predictor variables. The motivation for doing so is twofold. First, theories of herding emphasise that when the environment is uncertain, agents may place greater weight on the behaviour or beliefs of others, either because they believe others possess superior information or because the cost of processing information independently is high. Second, anti-herding may arise when forecasters seek to differentiate themselves from the consensus or when they overreact to new information. For this reason, the interpretation of the regression-based tests is central to the analysis, and in particular to understanding whether evidence suggestive of herding in some specifications should instead be read as reflecting noise, informational frictions, or over-reaction.

When uncertainty in the FX market is high, due for example to geopolitical events, economic crises, or policy changes, investors may find it more difficult to predict future exchange rate movements. In such periods, market participants may look for signals from other traders, analysts, or broader market sentiment when forming their own expectations. At the same time, uncertainty may also amplify differences in judgement, models, and information sets across forecasters. We therefore investigate the effects of including the following variables: the cross-sectional standard deviation of forecasts, global FX volatility, equity market volatility, forecast uncertainty, US economic policy uncertainty, global economic policy uncertainty, trade policy uncertainty, a macro uncertainty measure following Jurado, Ludvigson, and Ng (2015), geopolitical risk, inflation expectations, and FX predictors following Dahlquist and Söderlind (2023), namely the forward premium, the real exchange rate, and the depreciation rate.

Applying these methods to the FX market distinguishes the paper from much of the existing literature on herding in exchange rate forecasts. The applications in Gallo, Granger and Jeon (2002) and Clements (2018) are to macroeconomic prediction datasets, whereas our application is to a large panel of individual exchange rate forecasts. As in Clements (2018), and in a number of papers analysing herding in FX markets, such as Fritsche et al. (2015) and Frenkel et al. (2020), we also compute the Bernhardt et al. (2006) statistic. Our analysis extends the literature in three ways. First, it examines the regression-based herding tests in a panel context using mean group estimation that allows for cross-sectional dependence across forecasts. Second, it measures the effects of volatility, uncertainty, and policy uncertainty using a wide range of indices and measures, as well as the conventional cross-sectional standard deviation of individual forecasts. Third, it tests for herding conditional on exchange rate predictor variables linked to carry, value, and momentum.

The range of tests examined overall does not provide robust evidence of herding. Revision-based tests analysing the relationship between individual forecast revisions and deviations from the consensus suggest herding behaviour in 86% of cases when current consensus forecasts are used. However, this evidence weakens considerably when lagged consensus forecasts are employed, with only 44% of cases rejecting the null. By contrast, the regression-based test that uses the forecast error as the dependent variable suggests that in 51% of cases there is neither herding nor anti-herding, while, where the null is rejected, the evidence more often points to anti-herding. This finding is reinforced by the Bernhardt et al. statistic and by the over-reaction regressions in Section 5.3. In the pooled over-reaction regressions, only one of the nine currencies is significant at the 1-month horizon, but this rises to five of the nine currencies at the 3-month horizon and all nine currencies at the 6-month horizon, with coefficients mostly negative, indicating over-reaction. Taken together, these results suggest that the stronger indications of herding from some revision-based specifications should be treated with caution, and that the balance of evidence is more consistent with heterogeneity, noise, and over-reaction than with genuine herding.

The results of the investigation of alternative explanations for expectation formation or revisions in expectations, other than deviations between individual forecasters and consensus forecasts, suggest that the main conclusions remain largely unchanged. The wide range of measures covering different types of uncertainty, potentially relevant in determining herding-type behaviour, appears to have little effect. The three FX predictors – the forward premium, the real exchange rate, and the depreciation rate – exert very little influence, where their effects are modest and do not overturn the broader interpretation of the results.

The remainder of this paper is organised as follows. Section 2 describes the dataset. Section 3 takes a first look at forecasts by examining unconditional and conditional bias and forecast accuracy at the aggregate and individual levels. Section 4 describes the herding tests and Section 5 presents the results of the tests using both individual forecaster regressions and mean group-based estimation, together with an interpretation of the findings. Section 6 investigates the effects of alternative explanations, other than herding, on expectation formation. Section 7 concludes.

2 The Data

Our main data set, collected from the Eikon Refinitiv platform, consists of individual market participants 1-month, 3-month, 6-month and 12-month ahead (nominal) exchange rate forecasts. The maximum sample period is January 1994 to December 2024 (372 observations), but varies depending on the currency.[1] The nine currencies (covering the G10) are: the Australian Dollar (AUD), New Zealand Dollar (NZD), Canadian Dollar (CAD), Swiss Franc (CHF), the Euro (EUR) (DEM before 1999), British Pound (GBP), Japanese Yen (JPY), Norwegian Krona (NOK) and the Swedish Krona (SEK). The home currency is the US dollar, where exchange rates are defined as the number of US dollars per one foreign currency unit, i.e. an increase in the spot exchange rate corresponds to a US dollar depreciation relative to the foreign currency. [2]

We retrieve monthly spot and forward exchange rates for the January 1994- December 2024 period from *Datastream*, and consumer price indices (CPIs) from the Organisation for Economic Co-operation and Development (OECD).

The number of individual forecasters and time series observations vary by currency, but for each currency, remain relatively consistent across the four forecast horizons. Therefore, in Table 1, we present the distribution statistics (mean, standard deviation, minimum, and maximum) for both the number of individual forecasters (panel A) and their corresponding observations (panel B) *averaged across the four forecast horizons* $h = 1, 3, 6$, and 12.

The mean number of forecasters for most currencies lies within the range 38 to 49, with the exceptions of NOK and SEK, which average 25. The variability overtime is reasonably high, with standard deviations of the number of participants mostly in the range 6-11, but is up to twice as large for the major currencies, GBP, EUR and JPY. For the six currency pairs, AUD, CAD, CHF, EUR, GBP, and JPY, the maximum number of forecasters contributing lies within the range 58-70 . For the three remaining currency pairs, the number of forecasters available ranges from 38 to 54.

In Panel B we provide summary statistics regarding the number of forecast observations. For each currency and forecast horizon we take the number of individual forecasts available in the cross section at time t. We then compute the average over time and forecast horizons, of the four summary statistics given in the columns. Individual forecasters for the main traded currencies, the GBP, EUR, and JPY, have an average number of observations of 133, 129, and 135, respectively. Hence, the proportion of missing observations, relative to the full sample period (372 observations), is a little over two-thirds. For the AUD, CAD, NZD and CHF we observe an average of 108, 110, 84 and 107 observations, respectively. For the NOK and SEK currencies, the average number of observations is much lower at 44, since the sample begins in 2013. The variability among individual forecasters is

[1] The sample period for each currency, noting that the number of forecasters differs across periods, is as follows: March 2005 to December 2024 (238 observations) for AUD; July 2009 to December 2024 (186 observations) for NZD; December 2002 to December 2024 (265 observations) for CAD; January 2004 to December 2024 (252 observations) for CHF; January 1999 to December 2024 (312 observations) for EUR; March 1995 to December 2024 (358 observations) for GBP; January 1994 to December 2024 (372 observations) for JPY; December 2013 to December 2024 (133 observations) for NOK; and December 2013 to December 2024 (133 observations) for SEK.

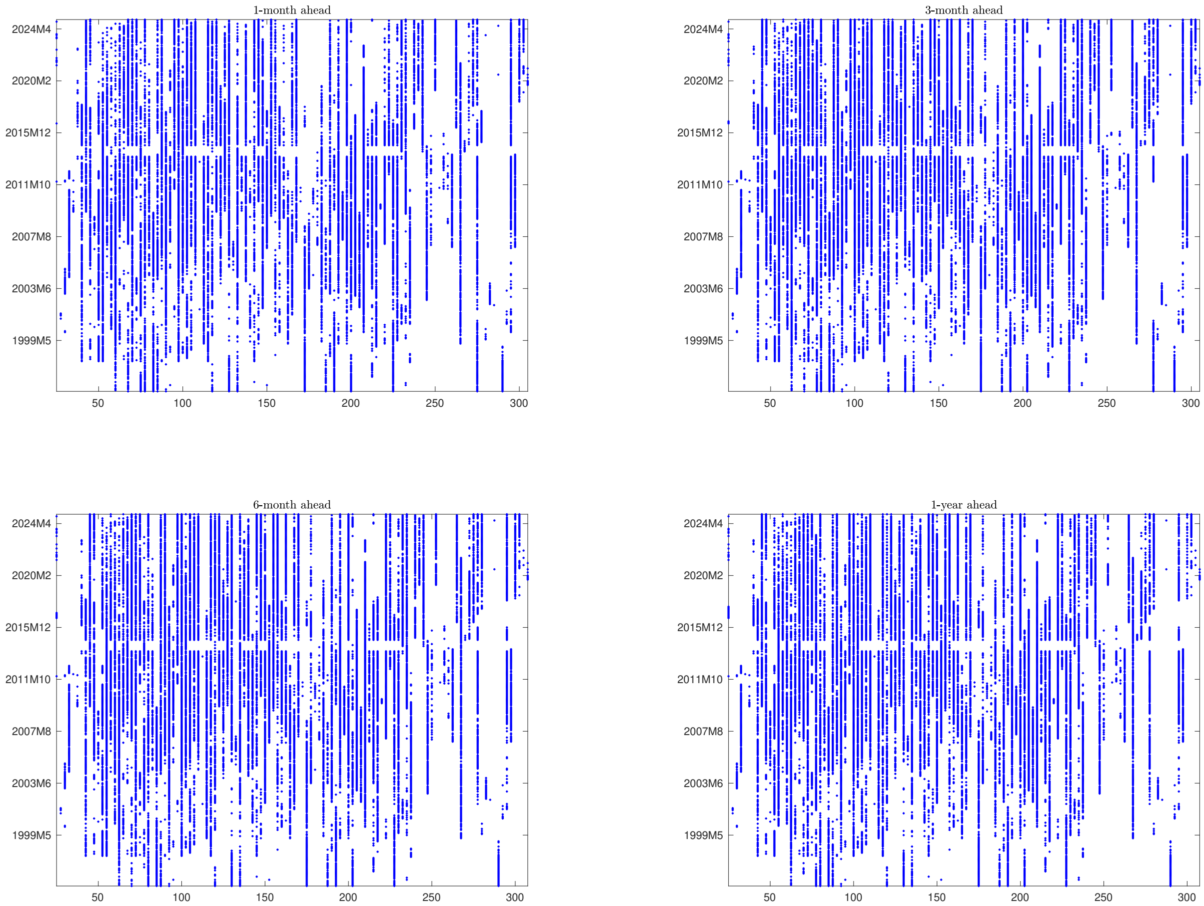
[2] The forecasters are mostly affiliated with investment banks (e.g., BNP Paribas, Commerzbank, Citigroup, Goldman Sachs, Deutsche Bank, Royal Bank of Canada, Royal Bank of Scotland, Santander, Société Générale, etc.), but also consultancies (e.g., Oxford Economics, EIU) and research institutes (such as WIIW, NIESR). Contributors to FX Poll from Refinitiv/Reuters include economists, buy/sell-side research analysts, strategists, and research think-tanks.

Table 1: Distributional Statistics of the Number of Forecaster's and Forecast Observations

Panel A: Number of forecasters				
	Mean	Std. Dev	Min	Max
UK	42	16	9	70
Australia	43	8	26	58
Canada	40	9	21	58
Switerland	42	11	21	65
Euro	49	14	15	70
Japan	43	17	9	70
New Zealand	38	8	22	54
Norway	25	6	13	39
Sweden	25	6	14	38
Panel B: Number of Forecast observations				
	Mean	Std. Dev	Min	Max
UK	133	83	1	295
Australia	108	71	1	229
Canada	110	76	1	246
Switerland	107	66	1	238
Euro	129	85	1	285
Japan	135	86	1	312
New Zealand	84	59	1	183
Norway	44	42	1	132
Sweden	44	42	1	132

Notes: Panel A reports the average, taken over the four forecast horizons $h = 1, 3, 6$ and 12, of the mean, standard deviation, minimum and maximum number of forecasters, for the (maximum) sample period 1994m1- 2024m12 (372 observations) - the sample period differs for each currency. Panel B reports the average, taken over the four forecast horizons $h = 1, 3, 6$ and 12, of the mean, standard deviation, minimum and maximum number of forecast observations of all forecasters.

Figure 1: Data patterns for individual forecasters of USD/GBP



Notes: The Y-axis measures the full sample period for the USD/GBP exchnage rate and the X-axis assigns a number to each of the individual forecasters observed during the full sample period. A blue dot denotes a recorded forecast observation made at the time period on the Y-axis, by each forecaster on the X-axis, with respect to 1, 3, 6 and 12 months horizons, for each of the four sub-plots respectively. No dot denotes no forecast from the forecaster on the X-axis for the time period on the Y-axis. For example, a continuous blue vertical line would denote that a forecaster had observations for all periods in the sample. A blue block would denote a complete or balanced panel, with all forecasters providing forecasts at all time periods in the sample.

large, with standard deviations ranging from 59 to 86 for the seven major currencies, dropping to 42 for the NOK and SEK. The number of forecast observations for some individual forecasters can be high, with a maximum value as large as 312 for the JPY and around 290 for other major currencies, but also can be very low with some forecasters only appearing once in the sample (the subsequent regressions we report require individual forecasters to have at least 10 observations, with a robustness check requiring a minimum of 30 observations).

As an example, in Figure 1, we plot the number of active forecasters over time for the USD/GBP currency pair at forecast horizons $h = 1, 3, 6,$ and 12 . This visually illustrates the broad characteristics of the type of data we have (identical plots using other currencies illustrate a similar pattern). The data set's missing value characteristics, forecasters' absences across the whole sample period, forecasters' shifting identities, and the fact that there are usually more forecasters and observations starting in 2005 are all highlighted in the figure.

3 A first look at Exchange Rate Forecasts

There is a large body of literature that evaluates the properties of professional forecasts, both at an aggregate level and individual level (see for example, Farmer *et al.* (2024), Bachman *et al.* (2022), Bordalo *et al.* (2020), Born *et al.* (2022), Coibion and Gorodnichencko (2015), Lahari and Sheng (2008, 2010), and in the context of foreign exchange markets Ince *et al.* (2017) and Jongen *et al.* (2012), to name but a few. Based around the idea of testing if forecasts are rational, most often results in the literature suggest professional forecasts are biased, the forecast errors are autocorrelated and are predictable by revisions. These features, whilst not directly a test of, are consistent with the occurrence of herding or anti-herding behaviour. Therefore, in order to evaluate our data in terms of these properties and relevance to the testing of herding, we document some basic properties of the forecasts, at both an aggregate and individual level. We examine bias, forecast error auto-correlation, rationality via a Minzer-Zarnowitz regression, accuracy relative to a random walk and in addition the properties of the difference between aggregate and individual forecasts, over currency and forecast horizons.[3]

3.1 Consensus and Individual Forecast Properties

Table 2 reports the results, for the mean consensus forecasts (in columns 2-5) and the individual forecasts (columns 6-13), of the tests for (unconditional) bias, autocorrelated forecast errors, the coefficient on the forecast term from a standard Mincer-Zarnowitz regression, and the root mean square error (RMSE) ratios, relative to a random walk, examining forecast accuracy.[4]

We define consensus and individual forecast errors, respectively, as follows:

$$e_{t+h} = s_{t+h} - \bar{s}_{t+h|t}, \tag{1a}$$

$$e_{j,t+h} = s_{t+h} - s^{j}_{t+h|t}, \tag{1b}$$

where $s_{t+h} = log(S_{t+h})$ and S_{t+h} is the nominal spot exchange rate for each currency in period $t+h$ (where for ease of notation we suppress i, $i = 1, 2, ..., N$, where $N = 9$, the number of currencies) , and the consensus (mean) forecast, $\bar{s}_{t+h|t}$, is given by $(1/J_t)\sum_{j=1}^{J_t} s^{j}_{t+h|t}$. The term $s^{j}_{t+h|t}$ denotes the forecast of any currency, made by forecaster j at time t, for the period $t+h$. Individual forecasters are denoted by $j = 1, 2, ..., J_t$, where J_t is the number of forecasts available for any currency i at time t. A *positive* forecast error implies that consensus/individual forecasts *under-predict* the actual observed outcome, whereas a *negative* forecast error is associated with consensus/individual forecasts that *over-predict*.

We examine bias by regressing the forecast errors, defined in equation (1), on a constant ($\hat{\alpha}$ in Table 2) and testing, using Newey-West heteroskedastic corrected standard errors, the null that the constant equals zero. Table 2, column 2, reports the estimated values for the consensus forecasts, and column

[3]In addition, in Section 6 and in the appendix, we report Minzer-Zarnowitz regressions where we include a range of uncertainty/additional information variables. Also, in Section 5, we estimate over-confidence regressions, as described in Bordalo *et al.* (2020).

[4]Here, we evaluate the forecast errors implicitly assuming a quadratic loss function and that they are independent of each other.

6 the *mean* of the estimated values of the individual forecasters. The tests of bias for the consensus forecasts suggest that most currencies, across all forecast horizons, are unbiased, with the exception of the Swiss currency, which shows a significant positive bias or tendency to under predict and the Norwegian and Swedish currencies, which show a significant negative bias or tendency to over predict. For the shorter horizons the errors are mostly positive, suggesting a tendency to under-predict, whilst at longer horizons there is a tendency toward negative forecast errors and under-prediction.

The bias tests for individual forecasters are conducted for each currency and forecast horizon. Given the nature of our data set, we encounter issues around missing observations and, in some cases, relatively few observations, as documented in the data section. Hence we require that a forecaster has a minimum of 10 observations to be included in the calculations (but where a large majority have considerably more). The mean values of $\hat{\alpha}$ are similar in terms of size and sign to the estimated bias values for the consensus forecasts, and as with the consensus results, the bias tests suggest that most individual forecasters are unbiased across currencies and forecast horizons. The proportions of p-values less than 0.1 (reported in column 7), rejecting the null of unbiased forecasts, with some exceptions for the Swiss Franc, Swedish and Norwegian Krona's, are (typically) considerably lower than 50%.

An additional prediction for rational forecasts is that the forecast errors should be serially uncorrelated. To evaluate this prediction and to gain some understanding of persistence of forecast errors in this context, we regress the h-period ahead forecast error on a constant and their own past value h periods earlier (i.e., non-overlapping h-period forecasts). In a complete information setting, forecast rationality requires that the constant and the coefficient of the lagged forecast error $\hat{\delta_1}$ (where $e_t = \delta_0 + \delta_1 e_{t-h} + \epsilon_t$) in Table 2 are equal to zero, that is, the forecast error should not be predictable by known information at time t when the forecast is made. Table 2, column 3, reports the estimated values for the consensus forecasts, and column 8 the mean of the estimated values of the individual forecasters. Consensus forecasts show a sizeable and significantly different from zero positive autocorrelation for horizons $h = 1$ and $h = 3$ (denoted in bold). For horizons $h = 6$ and $h = 12$, they remain mostly positive but are closer to zero (occasionally they take small negative values) but are not significantly different from zero. This pattern is largely mimicked at the individual forecast level, which for $h = 1$ and $h = 3$ have slightly lower positive autocorrelation and where the proportions of rejections of the zero null are high (column 9). The pattern then changes to one of smaller degrees of autocorrelation, often negative, but where the proportion of forecasters rejecting the null of a zero coefficient is much lower.

The forecast bias test evaluates the mean forecasts. To further assess forecast properties, we implement standard joint (conditional) rationality tests—covering both bias and efficiency—using the conventional Mincer–Zarnowitz regression:

$$(s_{t+h} - s_t) = \beta_0 + \beta_1(\bar{s}_{t+h|t} - s_t) + u_{t+h}, \tag{2}$$

For individual forecasts, we replace $\bar{s}_{t+h|t}$ with $s^j_{t+h|t}$. Estimates of β_0 differ from α, measuring conditional bias, and β_1 measures the degree of response of the forecast to actual outcomes, so $\beta_1 = 1$ suggests no scaling or efficiency problem. Table 2 reports the estimates of β_1, for the consensus forecasts (column 4) and the mean estimates for the individual forecasts (column 10), and documents when the coefficient is part of rejecting the *joint* test of the null hypothesis of rationality: $\beta_0 = 0 \cap \beta_1 = 1$, using Newey-West heteroscedastic corrected standard errors. In the case of individual forecasts we present this as a proportion of individual forecasts that reject the null at the 10% level of significance.[5]

For both the consensus and individual forecasts, the results provide strong evidence that the forecasts are not rational. The estimated values of β_1 are mostly close to zero, sometimes negative, or take values below 0.5. This is in nine currencies and four forecast horizons. The joint null, $\beta_0 = 0 \cap \beta_h = 1$, is rejected for the forecast horizons $h = 1, 3, 6$ and 12 for the consensus forecasts and the proportions of individual forecasters rejecting the null is consistently very high, with the lowest proportion at 87.4%.[6] Hence, we have mostly unbiased (unconditionally) forecasts when considering both consensus and individual forecasts. But the Mincer-Zarnowitz regression rejects the null of rationality. Typically, this occurs when $\alpha = 0$ (which is not the same as $\beta_0 = 0$) but when $\beta_1 = 0$. The forecasts are unbiased on average (no constant over or under prediction), but they under react to actual variation.

Related to the Mincer-Zarnowitz rationality regressions are the statistics reported in Table 3, which examine the means and standard deviations of the realised and expected depreciation, the dependent and independent variables in Equation (2) respectively, for all currencies and forecast horizons. In general, three main patterns are apparent from the table. First, the *absolute* value of the mean expected depreciation increases with the forecast horizon. For example, the average absolute value of the expected depreciation across all currencies is 0.36, 0.58, 0.95 and 1.66 for $h = 1, 3, 6$ and 12, respectively. Hence, forecasters do not believe that exchange rates follow a mean-reverting process within the 12-month horizon, a view that requires the observed pattern to be a *decline* in the absolute value of expected depreciation (see Ince and Molodtsova, 2017). The mean of the realised depreciation also increases with the forecast horizon, an observation consistent with the expectations formed but notable by smaller orders of magnitude (consistent with the value of $\hat{\beta}_1$ mainly being < 1.0). The average absolute value of realised depreciation across all currencies is 0.06, 0.18, 0.35 and 0.75 for $h = 1, 3, 6$ and 12, respectively. Second, the standard deviation of the expected and actual depreciation increases with the forecast horizon.[7] Third, we can see that, in general, the standard deviation of the realised depreciation is larger than that of the expected depreciation, while the pattern for the absolute mean realised depreciation is mixed.

Finally, we examine the accuracy of the forecasts, as measured by the conventional root mean square error (RMSE) measure, using the forecast errors defined in Equation (1). For consensus forecasts we report these as ratios relative to the RMSE calculated using a benchmark forecast generated by a random walk without drift (column 5), where ratios greater than one suggest a random walk

[5]The form of the Mincer-Zarnowitz regression, as the actual change of the exchange rate on the expected change, reflects that the (log) levels of observed and forecast exchange rates are non-stationary, whilst the changes are not. Note also that the outcomes and forecasts form a cointegrating vector and as such move together in the long-run. These results are available on request.

[6]For robustness and comparison purposes, we also examine bias and accuracy of the mean exchange rate forecasts from a survey conducted by Consensus Economics and published in *Foreign Exchange Consensus Forecasts*, for the same set of currencies over the sample period 1995m3-2021m9, for forecast horizons $h = 1, 3$ and 12. The results are very similar, suggesting robustness and consistency with our results despite the many missing values. See the Appendix for the full set of results.

[7]We observe the absolute mean and standard deviation of the forecast errors increasing with the forecast horizon, results consistent with Ince and Molodtsova (2017).

forecast is preferred. For individual forecasts, we report the proportion of RMSE ratios greater than one (column 12). The consensus forecasts show poor accuracy, particularly in short-term forecast horizons of $h = 1$ and $h = 3$, with RMSE ratios significantly greater than one, implying that a random walk-out performs market forecasters. At the longer forecast horizons, $h = 6$ and $h = 12$, there is more of an equivalence in performance with the random walk forecasts, but in only one case, for the Swiss Franc at forecast horizon $h = 12$, market forecasts outperform the random walk. The same result holds for individual forecasters. For forecast horizons $h = 1, 3, 6$, for all currencies, 95% and higher report an RMSE ratio larger than one, indicating continued low forecast accuracy compared to the random walk benchmark without drift. At the forecast horizon $h = 12$ this result is somewhat less pronounced, but where the lowest proportions are still high at 72.4%.

Table 2: Consensus and Individual Forecast Bias and Accuracy

	Consensus Forecasts				Individual Forecasts							
	$\hat{\alpha}$	$\hat{\delta}_1$	$\hat{\beta}_1$	RMSE	Mean $\hat{\alpha}$	Prop. p<0.1	Mean $\hat{\delta}_1$	Prop. p<0.1	Mean $\hat{\beta}_1$	Prop. p<0.1	Prop. RMSE > 1	p<0.1
Panel A: 1-month horizon												
UK	0.200	**0.572**	**-0.018**	**1.416**	0.093	0.157	0.428	0.951	-0.026	1.000	1.000	0.922
Australia	0.252	**0.582**	**0.019**	**1.405**	0.402	0.120	0.406	0.904	0.025	0.988	1.000	0.807
Canada	0.010	**0.548**	**0.068**	**1.343**	-0.020	0.190	0.359	0.893	0.074	0.988	0.988	0.857
Switzerland	**0.751**	**0.446**	**0.166**	**1.304**	0.857	0.584	0.321	0.888	0.113	1.000	1.000	0.820
Euro	0.195	**0.553**	**-0.072**	**1.483**	0.299	0.184	0.396	0.922	-0.018	1.000	1.000	0.903
Japan	-0.345	**0.594**	**-0.032**	**1.449**	-0.256	0.176	0.460	0.961	-0.012	0.990	1.000	0.922
New Zealand	0.430	**0.521**	**0.167**	**1.301**	0.648	0.219	0.365	0.922	0.138	1.000	1.000	0.906
Norway	**-1.217**	**0.584**	**0.076**	**1.474**	-1.411	0.571	0.483	0.939	0.047	1.000	1.000	0.878
Sweden	-0.787	**0.660**	**-0.010**	**1.532**	-1.084	0.375	0.530	0.938	0.007	1.000	1.000	0.896
Panel B: 3-month horizon												
UK	0.266	**0.342**	**-0.131**	**1.215**	0.347	0.167	0.195	0.765	-0.061	1.000	0.990	0.922
Australia	0.512	**0.368**	**-0.053**	**1.202**	0.833	0.167	0.202	0.702	0.012	1.000	0.988	0.845
Canada	0.128	**0.321**	**0.198**	**1.103**	0.163	0.155	0.166	0.690	0.152	0.976	0.976	0.750
Switzerland	**1.541**	**0.273**	**0.212**	**1.193**	1.806	0.674	0.137	0.539	0.122	1.000	1.000	0.775
Euro	0.178	**0.310**	**-0.105**	**1.244**	0.578	0.204	0.170	0.738	-0.011	0.990	0.990	0.893
Japan	-0.386	**0.278**	**-0.110**	**1.232**	-0.263	0.186	0.187	0.608	-0.017	0.990	1.000	0.912
New Zealand	0.701	**0.244**	**0.207**	**1.160**	1.113	0.203	0.151	0.516	0.144	1.000	1.000	0.844
Norway	**-2.555**	**0.193**	**0.312**	**1.160**	-2.509	0.538	0.139	0.404	0.168	1.000	0.962	0.596
Sweden	**-1.794**	**0.394**	**-0.011**	**1.246**	-1.739	0.442	0.192	0.673	0.004	0.981	0.981	0.865
Panel C: 6-month horizon												
UK	0.029	0.052	**0.024**	**1.091**	0.058	0.157	-0.051	0.167	0.040	0.922	0.941	0.882
Australia	0.518	0.050	**-0.025**	**1.108**	0.968	0.169	-0.042	0.301	-0.014	0.988	1.000	0.867
Canada	0.016	0.092	**0.334**	**1.031**	0.149	0.224	-0.020	0.306	0.193	0.929	0.918	0.753
Switzerland	**2.369**	0.001	**0.359**	**1.115**	2.666	0.742	-0.075	0.337	0.195	0.978	0.966	0.798
Euro	-0.087	0.192	**0.002**	**1.149**	0.573	0.183	-0.027	0.144	0.054	0.971	0.952	0.894
Japan	-0.637	0.285	**0.060**	**1.132**	-0.335	0.168	0.201	0.634	0.047	0.980	0.950	0.911
New Zealand	0.429	0.309	**0.121**	**1.167**	1.135	0.175	0.153	0.492	0.101	1.000	0.984	0.952
Norway	**-5.365**	0.283	**0.324**	**1.232**	-5.162	0.745	0.062	0.510	0.131	1.000	0.961	0.863
Sweden	**-4.171**	0.210	**-0.094**	**1.260**	-3.727	0.569	-0.015	0.353	-0.032	1.000	0.980	0.882
Panel D: 12-month horizon												
UK	**-8.920**	-0.193	**0.469**	**1.231**	-7.737	0.796	-0.260	0.426	0.340	0.981	0.852	0.889
Australia	-0.740	-0.051	**0.476**	**1.004**	-1.064	0.245	-0.193	0.392	0.298	0.951	0.814	0.912
Canada	0.083	-0.033	**0.434**	**1.009**	0.420	0.207	-0.096	0.195	0.324	0.897	0.782	0.793
Switzerland	-0.399	0.100	**0.752**	**0.965**	-0.287	0.299	-0.019	0.241	0.364	0.874	0.724	0.828
Euro	**3.696**	-0.023	**0.622**	1.046	4.224	0.722	-0.061	0.211	0.364	1.000	0.867	0.833
Japan	-0.656	0.032	**0.294**	**1.055**	0.160	0.314	-0.133	0.343	0.286	0.962	0.857	0.933
New Zealand	-1.329	**0.311**	**-0.064**	**1.153**	-0.706	0.184	0.214	0.456	0.007	0.981	0.932	0.951
Norway	-0.391	0.085	**0.158**	**1.102**	0.542	0.269	-0.031	0.284	0.217	0.985	0.910	0.910
Sweden	**-10.745**	-0.039	**0.521**	**1.282**	-9.898	0.815	-0.109	0.241	0.230	0.963	0.889	0.889

Notes: **Columns 2–5 report diagnostics for the consensus forecasts**: the estimated values of the bias coefficients, AR(1) parameter for autocorrelated forecast errors and the estimated coefficient from the Mincer-Zarnowitz regression, $(s_{i,t+h} - s_{i,t}) = \beta_0 + \beta_1(\bar{s}_{i,t+h|t} - s_{i,t}) + u_{i,t+h}$. Bold numbers indicate significance at the 10% level, using Newey-West t-statistic testing significant differences from zero, and the Newey-West adjusted p-value of the F-test of the the joint hypothesis $\beta_0 = 0$ and $\beta_1 = 1$. Root Mean Square Error (RMSE) are ratios, relative to a random walk, and bold numbers indicate significance using the Giacomini and White (GW) (2006) test of significance. **Columns 6–13 report diagnostics for individual forecasts**. Columns 6–11, report the mean of the individual forecasters estimated values and the proportion that have p-values on the nulls less than 0.1 (labelled p<0.1), for the bias, AR(1) and Minzer-Zarnowitz estimates respectively. Column 12 reports the proportion of individual forecast Root Mean Square Error (RMSE) ratios relative to a random walk that are greater than 1. Column 13 reports the proportion where the Giacomini and White (GW) (2006) test suggests that the RMSE ratio is significantly different from one.

Table 3: Summary Statistics for Realised and Expected Depreciation

Panel A: 1-month horizon

	Realised Depreciation		Expected Depreciation	
	Mean	Std. deviation	Mean	Std. deviation
UK	-0.05	2.40	-0.26	2.38
Australia	-0.04	3.34	-0.32	3.63
Canada	0.02	2.29	0.04	2.46
Switzerland	0.13	2.88	-0.60	2.75
Euro	-0.02	2.64	-0.21	2.74
Japan	-0.10	3.04	0.19	2.99
New Zealand	0.00	3.52	-0.49	3.82
Norway	-0.11	3.23	0.77	3.82
Sweden	-0.09	3.04	0.39	3.15

Panel B: 3-month horizon

	Realised Depreciation		Expected Depreciation	
	Mean	Std. deviation	Mean	Std. deviation
UK	-0.13	4.23	-0.43	2.63
Australia	-0.08	5.84	-0.69	3.92
Canada	-0.04	3.85	0.00	2.57
Switzerland	0.39	4.69	-1.03	3.09
Euro	-0.07	4.65	-0.18	3.17
Japan	-0.31	5.53	-0.06	3.55
New Zealand	0.02	5.91	-0.92	3.89
Norway	-0.33	5.55	1.23	4.07
Sweden	-0.26	5.36	0.64	3.36

Panel C: 6-month horizon

	Realised Depreciation		Expected Depreciation	
	Mean	Std. deviation	Mean	Std. deviation
UK	-0.25	6.28	-0.33	2.88
Australia	-0.14	8.41	-0.80	4.16
Canada	-0.02	5.39	0.18	2.65
Switzerland	0.79	6.38	-1.38	3.62
Euro	-.13	6.73	0.18	3.84
Japan	-0.61	7.78	-0.19	4.42
New Zealand	0.06	8.66	-0.87	4.28
Norway	-0.67	7.72	2.67	4.38
Sweden	-0.50	7.90	1.94	3.78

Panel D: 12-month horizon

	Realised Depreciation		Expected Depreciation	
	Mean	Std. deviation	Mean	Std. deviation
UK	-0.62	8.52	0.13	3.54
Australia	-.37	11.87	-0.65	4.86
Canada	0.01	7.89	0.63	3.05
Switzerland	1.44	8.86	-1.86	4.63
Euro	-0.38	9.63	0.89	5.00
Japan	-1.35	10.65	-0.11	5.82
New Zealand	0.04	12.59	-0.50	5.04
Norway	-1.44	10.91	5.48	5.26
Sweden	-1.08	11.31	4.65	4.83

Notes: The table reports the mean and standard deviation of realised and expected depreciation calculated for the maximum sample period 1994m1-2024m12 (372 observations). The sample period differs for each currency. The realized depreciation is defined as $(s_{i,t+h} - s_{i,t})$, and the expected depreciation as $(\bar{s}_{i,t+h|t} - s_{i,t})$, defined in the main text.

3.2 Dispersion in foreign exchange rate expectations

In addition to the performance based evaluation of the forecasts, a relevant empirical feature when examining herding is the extent of the differences between consensus and individual forecasts. If forecasts across individuals are broadly similar, this could reflect herding behaviour which revise forecasts when using information on consensus forecasts, bringing individual forecasts closer together. If they differ, then forecasts might be considered heterogeneous not subject to herding effects, or they have been subject to anti-herding behaviour.

To examine the existence of heterogeneity in the forecasts, we follow the methodology developed by Ito (1991), MacDonald and March (1996) as applied in Jongen *et al.* (2012). An individual forecaster is assumed to form forecasts based on a common structural part based on public information $f(\Omega_t)$, an individual effect g_j and an idiosyncratic component ϵ_j. We define the consensus forecast as the cross-sectional average of the J_t forecasts, which implies that these forecasts are determined by the same common structural part for individual forecasts and cross-sectional averages of individual effects and idiosyncratic components, $\bar{g}$ and $\bar{\epsilon}$, respectively. Therefore, when calculating the dispersion between the individual and consensus forecasts, the common structural term drops out, and we therefore define the dispersion as follows:

$$d_{j,t+h} = s^j_{t+h|t} - \bar{s}_{t+h|t} = (g_j - \bar{g}) + (\epsilon_j - \bar{\epsilon}). \tag{3}$$

If individual forecasts are homogeneous, there is no room for individual effects and all information is captured in the common structural part. Hence $d_{j,t+h}$ should show no particular pattern, and when regressed on a constant, should not reject the null of zero dispersion.

In Table 4 we provide summary statistics of the distributional properties of the dispersion variable $d_{j,t+h}$, for each currency and forecast horizon.[8] In addition, the table reports the proportion of individual regressions rejecting the null hypothesis of a zero constant when regressing $d_{j,t+h}$ on a constant, at the 10% significance level. Consistent with the results of Jongen et al. (2012), we find evidence of significant elements of heterogeneity across FX forecasters.

The standard deviation, which may serve as a rough proxy for the dispersion in the forecasts, increases as the forecast increases. The deviations are positively autocorrelated and are larger for longer forecast horizons.[9] Although the individual differences do not appear skewed in any direction, their distributions are consistently leptokurtic.[10] The Hill statistic indicates that the distribution of $d_{j,t+h}$ is heavy tailed. That is, the forecasts differ from each other. Additional evidence in favour of heterogeneity is given by the percentages of significant deviations from zero. For 1-month ranging from 17.9% to 30.6%, for 3-months 19.6% to 34.6%, 6-months 26.5% to 47.1%, and 12-months 33.3% to 43.3% across currencies.

Figure 2 plots the coefficient of variation of exchange rate forecasts across currencies and forecast horizons.[11] A clear and consistent pattern emerges: for all currencies, forecast dispersion increases

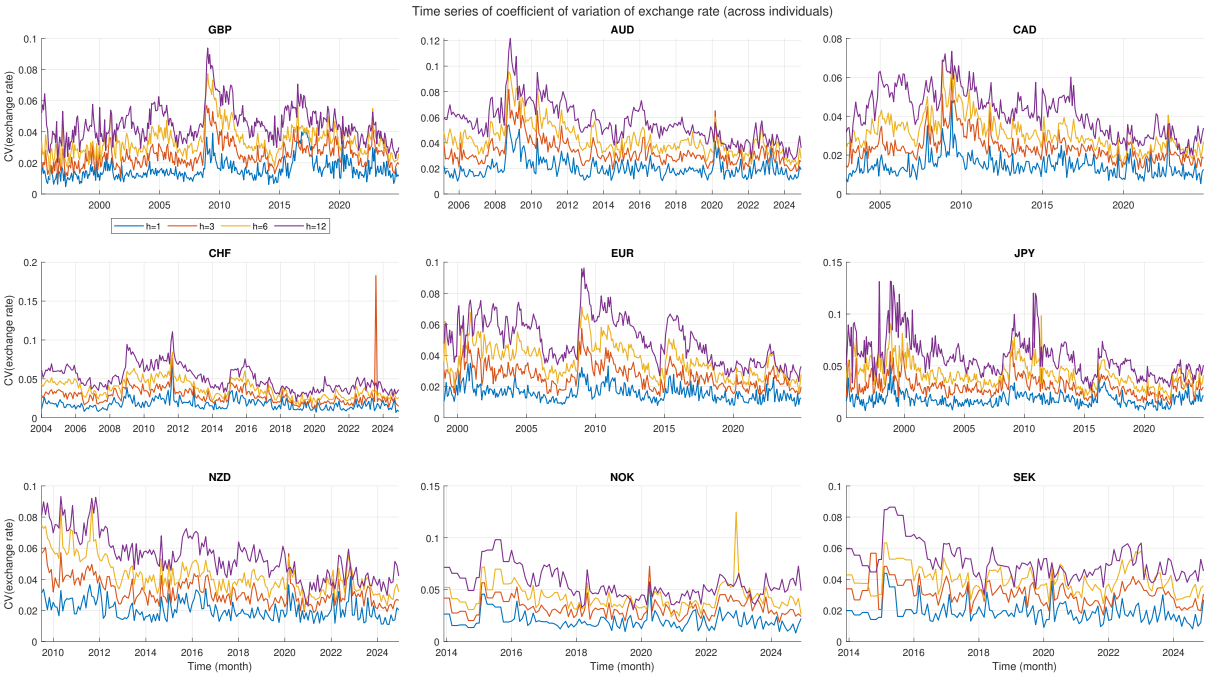
[8] The statistics are calculated over both the cross-section and time, for each currency and forecast horizon.

[9] Here we allow for just one lag and this may reflect overlapping forecasts.

[10] As well as the skew statistics in Table 4 suggesting small skew, if you were to plot the distributions of $d_{j,t+h}$ for each currency and horizon, they suggest normal looking distributions.

[11] This is defined, at time t, as the cross sectional standard deviation divided by the cross-sectional mean of the

Figure 2: Coefficient of Variation (across individual forecasts) for Exchange Rate Forecasts



Notes: For each currency and forecast horizon, we calculate the coefficient of variation for each time period t by dividing the cross-sectional standard deviation by the cross-sectional mean, where the cross-section is over individual forecasts.

monotonically with the forecast horizon, with each sub-figure displaying an ordering from the largest dispersion at ($h = 12$) to the smallest at ($h = 1$). Interpreting forecast dispersion as a proxy for uncertainty, this pattern suggests that market participants agree more about near-term exchange rate movements than about outcomes further into the future.

In addition, there are distinct periods of elevated dispersion. For all currencies spanning the Global Financial Crisis of 2008, forecast dispersion increases markedly. For the Japanese yen, an additional increase in dispersion is observed around 2001, coinciding with the Asian Financial Crisis. In the post-GFC period, dispersion generally declines on average, with a notable exception for sterling, which exhibits increased dispersion around the 2016 Brexit referendum. These patterns raise the question of the underlying sources of forecast dispersion; in what follows, we investigate whether herding behaviour can help account for these dynamics.[12]

deviation of forecasts. Specially: $cv_t^{ih} = \sigma_t^{ih}/\bar{d}_t^{ih}$, where $\sigma_t^{ih} = 1/J_{it}\sum_{j=1}^{J_{it}}(d_j^{ih} - \bar{d}_t^{ih})$ and $\bar{d}_t^{ih} = 1/J_{it}\sum_{j=1}^{J_{it}} d_j^{ih}$.

[12] Jongen et al. (2012) introduce additional variables capturing value, carry, and momentum effects, which we later incorporate into our herding regressions.

Table 4: Summary Statistics of Forecast Dispersion d_{jt}

	Min	Max	Std.	Skew	Kurt	JBp	Hill	AC	Prop. Reject
Panel A: 1-month horizon									
UK	31.349	-16.395	2.617	0.971	17.750	0.000	3.328	0.407	0.216
Australia	15.309	-14.251	1.772	-0.256	8.864	0.000	3.042	0.361	0.229
Canada	12.087	-13.927	1.380	-0.036	8.880	0.000	3.309	0.331	0.179
Switzerland	21.561	-32.800	1.971	-1.261	26.081	0.000	2.853	0.303	0.270
Euro	19.103	-14.753	2.009	-0.135	7.041	0.000	3.807	0.355	0.233
Japan	0.135	-0.148	0.018	0.115	8.656	0.000	3.048	0.387	0.088
New Zealand	8.676	-11.482	1.535	-0.403	6.527	0.000	3.278	0.354	0.203
Norway	1.261	-2.767	0.260	-0.820	15.456	0.000	3.337	0.526	0.306
Sweden	1.347	-2.652	0.235	-1.374	19.267	0.000	3.233	0.509	0.229
Panel B: 3-month horizon									
UK	24.129	-24.464	4.152	-0.008	4.643	0.000	4.180	0.546	0.284
Australia	14.135	-17.079	2.794	-0.260	4.784	0.000	3.892	0.573	0.262
Canada	13.344	-13.344	2.223	-0.018	5.132	0.000	3.871	0.546	0.238
Switzerland	25.680	-106.426	3.136	-3.651	127.380	0.000	3.800	0.484	0.348
Euro	19.227	-16.852	3.340	0.016	4.365	0.000	4.686	0.574	0.291
Japan	0.331	-0.161	0.029	0.566	7.305	0.000	3.663	0.580	0.196
New Zealand	12.446	-13.539	2.390	-0.223	4.588	0.000	4.187	0.566	0.328
Norway	1.928	-2.959	0.402	-0.077	6.401	0.000	3.279	0.631	0.346
Sweden	1.525	-3.975	0.377	-1.245	16.024	0.000	3.953	0.638	0.346
Panel C: 6-month horizon									
UK	29.950	-28.666	5.623	0.071	4.384	0.000	4.363	0.678	0.265
Australia	16.537	-23.707	3.663	-0.289	4.621	0.000	4.135	0.687	0.301
Canada	19.669	-16.708	2.903	-0.050	4.794	0.000	4.343	0.679	0.365
Switzerland	26.549	-34.254	3.891	0.019	5.565	0.000	4.304	0.672	0.416
Euro	25.109	-21.833	4.625	0.063	4.188	0.000	4.595	0.721	0.394
Japan	0.820	-0.174	0.040	1.224	19.018	0.000	3.914	0.696	0.327
New Zealand	26.766	-19.677	3.193	-0.235	5.264	0.000	4.303	0.699	0.365
Norway	5.240	-3.003	0.540	0.372	7.920	0.000	3.283	0.732	0.412
Sweden	2.449	-2.941	0.469	-0.045	5.547	0.000	3.443	0.741	0.471
Panel D: 12-month horizon									
UK	47.977	-33.795	7.418	0.094	4.564	0.000	4.138	0.770	0.333
Australia	27.647	-26.209	4.796	-0.224	4.555	0.000	4.205	0.777	0.379
Canada	26.389	-22.292	3.883	0.052	4.304	0.000	4.410	0.767	0.345
Switzerland	32.486	-34.549	5.216	0.172	5.004	0.000	3.878	0.781	0.433
Euro	31.657	-28.834	6.323	-0.025	3.839	0.000	5.291	0.831	0.390
Japan	0.884	-0.256	0.056	1.343	18.934	0.000	4.093	0.789	0.408
New Zealand	23.248	-18.277	4.170	-0.214	4.001	0.000	4.580	0.798	0.418
Norway	3.824	-3.320	0.755	0.170	5.065	0.000	3.689	0.824	0.407
Sweden	3.425	-3.406	0.644	0.130	5.119	0.000	4.844	0.808	0.389

Notes: The table reports summary statistics of forecast dispersions for nine currencies at horizons $h = 1, 3, 6,$ and 12 months. Columns report the minimum, maximum, standard deviation, skewness, kurtosis, Jarque–Bera p-value (JBp), Hill tail index (Hill), first-order autocorrelation (AC), and the proportion of individual regressions rejecting the null hypothesis of a zero constant, when regressing the individual dispersion term on a constant, at the 10% significance level. All statistics are computed using monthly data over the maximum available sample period.

3.3 Implications of the preliminary results

The results in Sections 3.1 and 3.2 provide an informative starting point for the formal herding analysis that follows. Taken together, they show that FX expectations are neither fully rational nor homogeneous, but they do not by themselves establish whether the observed patterns arise because forecasters herd toward the consensus, move deliberately away from it, or simply differ because of idiosyncratic noise. These are the issues Sections 4, 5 and 6 investigate in detail.

A first implication is that, at both the consensus and individual levels, exchange rate forecasts are mostly *unconditionally* unbiased, although there are some notable exceptions for particular currencies and horizons. However, this should not be taken as evidence in favour of rational expectations. The more informative result is that forecast errors are positively serially correlated at the short horizons, especially at $h = 1$ and $h = 3$, while the magnitude and significance of this persistence decline markedly at $h = 6$ and $h = 12$. The presence of predictable forecast errors at short horizons is inconsistent with full-information rationality and suggests that forecasters revise beliefs only gradually, or that they repeatedly make similar mistakes in response to incoming information. In the context of herding, such persistence is suggestive because it is consistent with forecasters not fully processing information independently, and instead responding to common signals, market narratives, or the published views of others. At the same time, serial correlation on its own is not a sufficient statistic for herding, since it may also arise from sluggish information updating, noisy information, or other forms of expectation rigidity.

This interpretation is reinforced by the Mincer–Zarnowitz results. For both consensus and individual forecasts, the joint null of rationality is strongly rejected across currencies and horizons. Hence, although forecasts are often unbiased on average in the unconditional sense, they are conditionally inefficient: forecasted exchange rate changes do not move one-for-one with realised outcomes, and the estimated slope coefficients are typically far below unity. In economic terms, forecasters appear to under-respond to the variation in realised exchange rate movements. This wedge between unconditional unbiasedness and conditional inefficiency is important for the later analysis, because it implies that departures from rationality are more likely to show up in the way forecasts are revised and scaled, rather than in a simple tendency to systematically over- or under-predict.

The forecast accuracy evidence points in the same direction. The RMSE comparisons show that both consensus and individual forecasts perform poorly relative to a random walk benchmark, particularly at short horizons. This is a well-known feature of exchange rate forecasting, but in the present context it has an additional implication: if professional forecasters have only limited success in forecasting the exchange rate itself, then similarity in reported forecasts cannot automatically be interpreted as superior information aggregation. Weak forecasting performance leaves open the possibility that any apparent clustering of forecasts may reflect imitation, common noise, or strategic reporting rather than genuinely informative agreement.

Section 3.2 adds a second key element to the interpretation: there is substantial and persistent heterogeneity across individual forecasts. Dispersion is sizeable, varies over time, rises with the forecast horizon, and becomes particularly pronounced during episodes of market stress. The distribution of deviations from the consensus is also non-Gaussian and heavy-tailed, indicating that disagreement is not driven by a few trivial departures from a common view, but instead reflects economically meaningful differences across forecasters. This finding is important because it shows that there is considerable room for heterogeneity in expectations, consistent with the broader literature that emphasises dispersed beliefs in financial markets, such as Ottaviani and Sorensen (2006). In other words, the cross section of FX forecasts is far from collapsing onto a single representative expectation.

This heterogeneity has two implications for the remainder of the paper. First, it motivates the herding tests in Section 4. If forecasters were essentially homogeneous, there would be little scope for detecting movements toward or away from the consensus. By contrast, the sizeable dispersion documented here means that we can study whether revisions systematically narrow the distance between individual and consensus forecasts, as herding would predict, or widen it, as anti-herding would imply. Second, the time variation in dispersion suggests that disagreement itself may contain information about the environment in which forecasts are formed. Periods of heightened uncertainty or market stress may alter the incentives to imitate others, but they may also simply amplify differences in models, information sets, and judgement. Section 6 provides a detailed examination of the effect of measuring uncertainty.

Overall, the evidence from this preliminary analysis suggests a nuanced interpretation. FX forecasts are not fully rational, especially at short horizons, yet they are also far from uniform across forecasters. Thus, the data contain both the kind of dependence in forecast errors that could be associated with herding and the type of cross-sectional disagreement that points to substantial heterogeneity in beliefs. The next sections therefore try to test these competing interpretations using formal tests based on the relationship between individual forecasts, revisions, and the consensus forecast.

4 Tests of Herding in Exchange Rate Forecasting

In this section, we describe the methods we use to test if herding occurs among foreign exchange rate forecasters. The main set of tests below are based on a series of regressions, using individual-level forecasts and how they relate to the consensus or mean forecast, building on formulations described and motivated in Gallo, Granger and Jeon (2002) (GGJ) and Clements (2018).

4.1 Regression Based Tests

Regression based tests for herding use successive individual and consensus (mean/median) forecasts of the same event. Our main regression based approach is one we adopt from Clements (2018), which in turn is founded on the **S** test developed in Bernhardt, Campello and Kutsoati (2006) (BCK). Their method was implemented in the same context that we use here, the behavior of professional financial analysts (and is explored in other papers examining herding in FX markets; see, for example, Fritsche *et al.* (2015) and Frenkel *et al.* (2020)). As such a comparison of a regression method based on the BCK **S** test, and the **S** test itself, is useful as it provides a robustness check on the regression based approach we adopt in a financial market context. A regression based test allows for the advantages when testing of a parametric approach which is less demanding in terms of (large) sample size requirements.[13]

The underlying idea in BCK assumes that analysts form a private forecast using their own research, public information, and previously published forecasts by other analysts. This private forecast corresponds to the centre of their posterior belief distribution, usually the mean or median. If the analyst reports this value truthfully, the forecast is unbiased.[14] However, analysts may publish forecasts that

[13]The non-parametric test of BCK is asymptotically normally distributed, under the null of no herding. We compute and report the non-parametric BCK test. Let τ be an index for an analyst's forecast for a given month. If we select conditioning events z_τ^+ and z_τ^-, where z_τ^+ implies that the analysts' forecast exceeds the extant consensus forecast, and z_τ^- implies that the forecast fell short of the extant consensus forecast, at the moment the analyst reported. Then, defining the conditioning indicator functions, γ_τ^+ and γ_τ^-, where

$$\gamma_\tau^+ = 1, \text{ if } z_\tau^+ \text{ occurred}, \gamma_\tau^+ = 0 \text{ otherwise}; \gamma_\tau^- = 1, \text{ if } z_\tau^- \text{ occurred}, \gamma_\tau^- = 0 \text{ otherwise}.$$

Likewise, define the overshooting functions, δ_τ^+ and δ_τ^-, where

$$\begin{aligned}\delta_\tau^+ &= 1, \text{ if } z_\tau^+ \text{ occurred and } s_{t+h|t}^j > s_{i,t+h}, \delta_\tau^+ = 0 \text{ otherwise};\\ \delta_\tau^- &= 1, \text{ if } z_\tau^- \text{ occurred and } s_{t+h|t}^j < s_{i,t+h}, \delta_\tau^- = 0 \text{ otherwise}.\end{aligned}$$

The test statistic used is given by:

$$S(z^-, z^+) = \frac{1}{2}\left[\frac{\sum_\tau \delta^+}{\sum_\tau \gamma_\tau^+} + \frac{\sum_\tau \delta^-}{\sum_\tau \gamma_\tau^-}\right],$$

to estimate the probability that exchange rate forecasts overshoot the actual exchange rate in the same direction as they overshoot the consensus. $\sum_\tau \delta^+ / \sum_\tau \gamma_\tau^+$ is our estimate of the conditional probability of overshooting the actual exchange rate given that the forecast exceeds the consensus forecast, while $\sum_\tau \delta^- \sum_\tau \gamma_\tau^-$ is our estimate of the conditional probability of falling short of the actual exchange rates given that the forecast falls short of the consensus. **S** is the average of the two conditional overshooting probability estimates.

[14]A reported forecast is unbiased if it truthfully reflects the analyst's private posterior belief, typically summarized by the posterior mean or median. But this is only a conditional notion of unbiasedness: it holds relative to the analyst's model and information set. If that model is misspecified, then even an honestly reported forecast can be biased relative to the actual outcome. The key distinction, therefore, is between bias arising from model error and bias arising from strategic behavior such as herding. A forecaster can be honest but wrong because their model is bad, or strategically biased because they herd or anti-herd. The empirical challenge is separating those two. This is why these tests focus on

differ from their private best estimate. Herding occurs when an analyst shifts the published forecast toward the earlier consensus forecast. Anti-herding occurs when the analyst deliberately shifts the published forecast away from that consensus.

If the reported forecast is unbiased and equals the posterior median, then the actual outcome should be equally likely to lie above or below it, conditional on the information available at the time. Therefore, this probability should be 0.5 regardless of whether the forecast is above or below the previous consensus. In herding, the reported forecast is pulled towards the consensus relative to the private estimate of the analyst. As a result, forecasts above consensus are biased downward and forecasts below consensus are biased upward. This implies that the probability patterns will differ from 0.5 in a predictable way. Under anti-herding, the opposite occurs: forecasts are pushed away from consensus, and the probabilities move in the opposite direction.[15]

The regression approach takes advantage of this implication by examining whether forecast errors are systematically related to the extent to which a reported forecast lies above or below the consensus. The regression is based on the BCK idea that if a forecaster's current reported forecast exceeds the consensus, then a forecaster who herds (or anti-herds) will have moved their forecast towards (away from) the consensus, relative to a "private"or unreported forecast, making it more likely that the forecast error will be positive (negative). Hence, following Clements (2018), the dependent variable is the forecast error, and this translates into a test using the parametric regression of the following form:

$$(s_{t+h} - s^j_{t+h|t-(h+a)}) = \rho_0 + \rho_1(s^j_{t+h|t-(h+a)} - \bar{s}_{t+h|t-(h+b)}) + u_{t+h}. \tag{4}$$

The term s_{t+h} denotes the observed exchange rate at time $t+h$ (to simplify the notation, we do not distinguish between individual currencies and omit the i subscripts from the parameters). The term $s^j_{t+h|t-(h+a)}$, is the j^{th} forecasters forecast, made at time $t-(h+a)$, of the exchange rate at $t+h$ ($j = 1, 2, 3, ..., J_t$). The final term, $\bar{s}_{t+h|t-(h+b)}$, is the consensus forecast (in our case the mean of the J_t forecasters), made at time $t-(h+b)$, of the exchange rate at $t+h$. We assume it is this forecast against which individual forecasters (potentially) benchmark their forecasts, and then adjust, according to whether they exhibit (anti-) herding or no herding behaviour. Rejecting the no herding null of $\rho_1 = 0$, in favour of the alternative hypothesis $\rho_1 > 0$ indicates herding, or if the alternative hypothesis is $\rho_1 < 0$ anti-herding. Therefore, high proportions of rejecting the null of $\rho_1 = 0$ in favour

whether errors are related specifically to the forecast's position relative to the consensus. A bad model alone does not necessarily create that exact pattern, whilst herding does.

[15]Under herding, suppose the analyst's true private estimate is higher than the consensus, and the analyst herds downward toward that consensus before publishing. Then the forecast reported is now too low relative to the analyst's genuine belief. Due to that downward bias, even if the reported forecast is still above consensus, it is now less likely than 50% to exceed the actual outcome. Why? Because the analyst pulled the forecast down from where they really thought it should be. Likewise, if the analyst's private estimate is below the consensus, and the analyst herds upward toward consensus, the reported forecast becomes too high relative to the analyst's true belief. Then, if the reported forecast is below the consensus, it is less likely than 50% to be below the actual outcome. So under herding, forecasts on either side of consensus are "pulled inward," and this creates a predictable pattern in forecast errors.

Anti-herding is the reverse. The analyst pushes the published forecast further away from consensus than their true private estimate. That makes forecasts more "extreme" than the analyst really believes. So: a forecast above consensus is now more likely than 50% to end up above the actual outcome, and a forecast below consensus is more likely than 50% to end up below the actual outcome. That is why probabilities become greater than half under anti-herding. Herding pulls forecasts toward consensus, making them too moderate, anti-herding pushes forecasts away from consensus, making them too extreme.

of $\rho_1 > 0$ ($\rho_1 < 0$) provide strong evidence of herding (anti-herding).

We introduce a and b into equation (4) to allow us to denote the timings of when a forecast is made in accordance with the constraints of our fixed horizon (as opposed to fixed event required in the regressions) data set, and the assumption we wish to make on the information individual forecasters have on the consensus forecasts at the time they make their forecast. The timings and horizons we examine are in accordance with our data being fixed-horizon forecasts, as opposed to forecasts for a fixed period in the future and not having forecasts at all forecast horizons. For example, for the forecast horizon $h = 1$, as we only have forecasts three steps ahead but not two steps, we use $s^j_{t+1|t}, s^j_{t+1|t-2}$ and $\bar{s}_{t+1|t-2}$. As such, the forecast revision uses information accrued in the two months between $t-2$ and t.

For the BCK S test, the null hypothesis that analysts issue unbiased forecasts implies that **S** should be one-half. The alternative that analysts herd implies that **S** should be *less* than one-half, or that analysts anti-herd implies that **S** should *greater* than one-half. For consistency between the S test and the regression approach, we expect to see higher proportions of S exceeding 0.5 (anti-herding) and large proportions of the null $\rho_1 = 0$, rejected in favour of $\rho_1 < 0$ (anti-herding). In contrast, higher proportions of **S** less than 0.5 (herding) were associated with the rejection of null $\rho_1 = 0$ in favour of $\rho_1 > 0$ (herding).

In a second regression based test, we adopt an arguably more direct intuitive approach and test whether $(s^j_{t+h|t} - s^j_{t+h|t-(h+a)})$, the forecast revision of forecaster j, is negatively correlated with $(s^j_{t+h|t-(h+a)} - \bar{s}_{t+h|t-(h+b)})$, the difference between the individual forecast j^{th} and the consensus forecast. For example, if an individual forecast for $t+1$ exceeds the consensus forecast for $t+1$, when both are made in, say, period $t-1$, the individual forecast made in period t of $t+1$ is lowered relative to the forecast made in $t-1$ for $t+1$. This corresponds to $\phi_1 < 0$ in the regression:

$$(s^j_{t+h|t} - s^j_{t+h|t-(h+a)}) = \phi_0 + \phi_1(s^j_{t+h|t-(h+a)} - \bar{s}_{t+h|t-(h+b)}) + u_{t+h}. \tag{5}$$

The term $s^j_{t+h|t}$ denotes the forecast of j^{th} forecasters, made at time t, of the exchange rate at $t+h$. Rejecting the null of no herding $\phi_1 = 0$, in favour of the alternative hypothesis $\phi_1 < 0$ indicates herding, or if the alternative hypothesis is $\phi_1 > 0$ anti-herding. The high proportions of individual regressions that reject the null of $\phi_1 = 0$ in favour of $\phi_1 < 0$ provide strong evidence of herding. This regression is similar to efficiency testing, as future forecast revisions should not be predictable from information available at that time if information is being used efficiently. It is a test for herding because we test in the direction of the revision being related to the divergence of the earlier forecast from the consensus.

Note that in this example, we use forecasts 3-months ahead dated at $t-2$ for $t+1$, for both individual and consensus forecasts, which assumes that each forecaster knows the forecasts of all others forecasters when making their forecast. Setting $a = b = 1$, we allow for current information with respect to the consensus forecast. However, Clements (2018) and others argue that this is a potentially too strong informational assumption and opt to be more conservative by using the most recent *lagged* consensus forecast (of $t+1$) relative to when individual forecasts of $t+1$ are made. We also examine the effects of a weaker informational assumption, where for this $h = 1$ example, the closest lagged consensus forecast is the 6-month ahead forecasts made at time $t-5$ for time $t+1$. Hence, in this case $a=1$ and $b=4$.

All regressions are run for each individual forecaster, for the forecast horizon h, using the variation they exhibit over time. This allows individual forecasters to have heterogeneous behaviour (coefficients) and for the behaviour to vary over forecast horizons such that a given forecaster may consider other forecasters at some horizons, but not at others. Note that in section 5.2 we contrast these results with the estimation using mean group panel methods. As such, we analyse evidence on the overall tendency (for each currency and horizon) to herd, taking advantage of using a larger number of observations.

5 Empirical Implementation

Given the extent of missing data and, in a limited number of cases, sample periods that do not overlap, we place the primary emphasis on the approach of Clements (2018), where separate regressions are estimated for each forecaster, currency pair and forecast horizon. This framework allows for heterogeneity in the coefficients between forecasters. As such, the formal tests described in the previous section are implemented at the individual forecaster level, where we report the proportion of forecasters for whom the null is rejected in favour of the specified one-sided alternative. Although panel estimation remains feasible and is pursued later in the section, these data limitations make it comparatively more difficult.[16]

We run regressions (4) and (5), for all individual forecasts for each forecast horizon and currency pair. In order to address the issue of small numbers of observations and the noise that may be associated with this estimation, we require 10 or more forecast observations for any given horizon.[17] This requirement is met by around 80 to 100 individual forecasts, for the UK pound, Australian dollar, Canadian dollar, Swiss franc, the Euro and the Japanese yen for all horizons. For the remaining three currencies, the New Zealand dollar, Norwegian krona, and Swedish krona, the number is around 50 to 65 at all horizons.[18]

5.1 Individual Regression Results

In Table 5 we report the results of tests based on Equation (4). For the $h=1$ forecast horizon we set $a=-1$ and $b=1$, for the $h=3$ we set $a=-3$ and $b=0$ and for the forecast horizon $h=6$, we set $a=-6$ and $b=0$. These values reflect the latest observed consensus forecast available to the individual forecasters when forming their forecasts of the future period, and are used to construct the forecast error defining the dependent variable.

[16] The issue of what constitutes a "large" or "small" proportion of tests rejecting the null, indicating the strength of evidence regarding herding, can be partially mitigated by cross-checks with the formal statistical tests on panel as opposed to individual estimates.

[17] We examine the robustness of the results to this requirement, where a minimum of 30 forecasts for any given horizon is considered. We find that all the patterns and percentage of regressions rejecting the nulls, across all tests, horizons, and currencies are very similar.

[18] In Table 1 the average number of forecasters for the major currencies at any time ranges from 40-49. Here, the 80-100 forecasters cover the full range of forecasters which potentially move in and out of the sample.

The results suggest that the forecast-error-based test provides weak evidence supporting herding, and where no herding is rejected the strongest evidence favours anti-herding. Across all currencies and horizons, the null of no herding is not rejected in 51% of cases, so the dominant result is one of weak or absent evidence for systematic herding. This is especially true at the 1-month horizon, where "neither" herding nor anti-herding accounts for roughly 63% of cases. That share falls to about 48% at 3 months and 42% at 6 months, implying that departures from the null become somewhat more common as the horizon lengthens, but even then the evidence does not shift strongly toward herding. Instead, when the null is rejected, it is more often in the direction of anti-herding than herding. Averaging across currencies and horizons, around 31% of outcomes point to anti-herding, compared with only about 18% pointing to herding (the sum of the 49% of cases rejecting the null). In other words, the forecast-error specification suggests that forecasters are more likely to end up on the opposite side of the consensus than to move toward it.

The cross-currency pattern in Table 5 is also informative. At the 1-month horizon, anti-herding is pronounced for Canada, Switzerland, New Zealand and Norway, where the share of rejections in favour of anti-herding is around one-half or more, while the UK, Australia, the Euro area and Sweden are dominated by "neither" outcomes. Japan is especially notable because the evidence is weakest there: only 7.8% of regressions support anti-herding and 6.9% support herding, leaving 85.3% in the "neither" category. At the 3-month horizon, anti-herding strengthens further, especially for Switzerland (72.7%), New Zealand (85.7%) and Norway (67.3%), while herding remains rare. By 6 months, the anti-herding signal becomes very strong for several currencies, including the UK (89.2%), Switzerland (82.9%), Canada (69.0%) and Sweden (62.0%), although Japan, New Zealand and Norway still show a significant proportion of "neither" outcomes. So the broad pattern in Table 5 is not just that herding is weak on average, but that evidence increasingly moves toward anti-herding as the horizon lengthens.

The Bernhardt et al. S-statistics reinforce this interpretation. Overall, about 77% of forecasters have S-statistics above 0.5, which is the region associated with anti-herding rather than herding. This is true at each horizon: roughly 73% at 1 month, 81% at 3 months, and 78% at 6 months. The currency-level values in Table 5 tell the same story. For example, the S-statistic points strongly to anti-herding for Norway and the UK at 6 months, with 98.0% of UK forecasters and 62.5% of Norwegian forecasters above 0.5, and for Switzerland the proportion is 98.9%. Even where the regression-based forecast-error results are more mixed, the S-statistics usually lean against herding. This makes the anti-herding interpretation more convincing, because it is supported by two distinct methods rather than only one.

Table 6 reports the evidence for anti-herding/herding based on equation (5), where we report results for two instances, one assuming that forecasters can observe contemporaneous or current consensus forecasts, and a second more conservative assumption that they only observe lagged consensus forecasts. For the $h = 1$ forecast horizon we set $a = 1$, and $b = 1$ and 4 for current and lagged information, respectively. For the $h = 3$ forecast horizon we set $a = 0$, and $b = 0$ and 6 for current and lagged information, respectively. For the forecast horizon $h = 6$, where we have only the current information, we set $a = 0$ and $b = 0$. The Table reports the rejection rate of a one-sided test of $\phi_1 = 0$ versus $\phi_1 < 0$ indicating herding, the rejection frequency of $\phi_1 = 0$ versus $\phi_1 > 0$ indicating anti-herding, and the proportion for which neither is found.

The results differ from those reported in Table 5, where the revision-based test using current consensus information produces strong evidence of herding. The null is rejected in favour of herding in about 89% of cases overall, with particularly high rejection rates at the 1-month horizon and still very high values at 3 and 6 months. In many currency-horizon combinations, the herding share is above 0.9 and anti-herding is essentially zero. Therefore, this appears to be different from Table 5, where the modal outcome is no herding and the balance of rejections favours anti-herding.

Although Table 6 appears to provide strong support for herding under the current-consensus assumption, this contrast with Table 5 weakens substantially under the lagged-information assumption. When forecasters are assumed to observe only lagged, rather than contemporaneous, consensus forecasts, the proportion rejecting in favour of herding in Table 6 declines sharply from around 89% to 46% overall, with corresponding averages of 44% at the 1-month horizon and 47% at the 3-month horizon. At the same time, the share of cases classified as "neither" rises markedly, while anti-herding remains limited. Under this more realistic informational assumption, the revision-based evidence therefore becomes considerably less decisive and aligns more closely with the broad message of Table 5. Rather than providing overwhelming support for herding, the results point to a more mixed pattern in which many forecasters cannot be clearly classified as herding. The apparent conflict between the two tables is driven primarily by the current-consensus specification; once lagged consensus information is assumed, the revision-based test becomes more consistent with the forecast-error test and the S-statistics, both of which suggest that disagreement across forecasters is to some extent due to deliberate differentiation, or anti-herding, from the consensus.

Table 5: Proportion of individual forecaster regressions for which we reject null hypothesis of the forecast error tests of herding and the S test of Bernhardt et al (2006)

Panel A: 1-month horizon

	Forecast Error			S Tests	
	$\rho_1 < 0$	$\rho_1 > 0$	*Neither*	$S < 0.5$	$S > 0.5$
UK	0.176	0.039	0.784	0.304	0.696
Australia	0.193	0.012	0.831	0.193	0.807
Canada	0.488	0.000	0.512	0.133	0.867
Switzerland	0.602	0.000	0.398	0.135	0.865
Euro	0.194	0.010	0.796	0.427	0.573
Japan	0.078	0.069	0.853	0.564	0.436
New Zealand	0.609	0.000	0.438	0.141	0.859
Norway	0.490	0.000	0.510	0.020	0.980
Sweden	0.229	0.000	0.771	0.500	0.500

Panel B: 3-month horizon

	Forecast Error			S Tests	
	$\rho_1 < 0$	$\rho_1 > 0$	*Neither*	$S < 0.5$	$S > 0.5$
UK	0.490	0.020	0.490	0.147	0.853
Australia	0.393	0.012	0.595	0.369	0.631
Canada	0.571	0.000	0.429	0.155	0.845
Switzerland	0.727	0.000	0.273	0.090	0.910
Euro	0.485	0.029	0.485	0.359	0.641
Japan	0.392	0.010	0.598	0.118	0.882
New Zealand	0.857	0.016	0.127	0.094	0.906
Norway	0.673	0.000	0.327	0.176	0.824
Sweden	0.308	0.019	0.673	0.240	0.760

Panel C: 6-month horizon

	Forecast Error			S Tests	
	$\rho_1 < 0$	$\rho_1 > 0$	*Neither*	$S < 0.5$	$S > 0.5$
UK	0.892	0.000	0.108	0.020	0.980
Australia	0.615	0.024	0.361	0.349	0.651
Canada	0.690	0.000	0.310	0.153	0.847
Switzerland	0.829	0.000	0.170	0.011	0.989
Euro	0.587	0.010	0.404	0.183	0.817
Japan	0.267	0.020	0.713	0.390	0.610
New Zealand	0.500	0.061	0.484	0.226	0.774
Norway	0.420	0.040	0.540	0.375	0.625
Sweden	0.620	0.040	0.340	0.255	0.745

Notes: In the first three columns, the table presents the proportion of the individual regressions, for each currency and forecast horizon, which reject the no herding null hypothesis suggested by the forecast error test. The restrictions stated in the column headings are those under the alternative hypothesis, implying a range of herding or anti behaviour. The last two columns report the "S" statistic of Bernhardt et al (2006), where we report the proportion of tests less than and greater that (or equal to) 0.5. Less than 0.5 is consistent with herding, whilst greater than 0.5 suggests anti-herding behaviour.

Table 6: Proportion of individual forecaster regressions rejecting the revision based no-herding null

	Current consensus			Lagged consensus		
	$\phi_1 < 0$	$\phi_1 > 0$	Neither	$\phi_1 < 0$	$\phi_1 > 0$	Neither
Panel A: 1-month horizon						
UK	0.948	0.000	0.052	0.396	0.000	0.604
Australia	0.822	0.000	0.178	0.329	0.000	0.671
Canada	0.961	0.000	0.039	0.566	0.000	0.434
Switzerland	1.000	0.000	0.000	0.659	0.000	0.341
Euro	0.937	0.011	0.053	0.453	0.011	0.537
Japan	0.917	0.000	0.083	0.229	0.010	0.760
New Zealand	0.983	0.000	0.017	0.800	0.000	0.233
Norway	0.881	0.000	0.119	0.429	0.048	0.524
Sweden	0.842	0.000	0.158	0.158	0.132	0.711
Panel B: 3-month horizon						
UK	0.918	0.000	0.082	0.526	0.010	0.464
Australia	0.897	0.000	0.103	0.410	0.038	0.551
Canada	0.918	0.000	0.082	0.466	0.000	0.534
Switzerland	0.952	0.000	0.048	0.831	0.000	0.169
Euro	0.899	0.000	0.101	0.444	0.010	0.545
Japan	0.899	0.000	0.101	0.315	0.034	0.652
New Zealand	0.967	0.000	0.033	0.576	0.000	0.407
Norway	0.826	0.000	0.174	0.349	0.000	0.651
Sweden	0.689	0.022	0.289	0.286	0.000	0.714
Panel C: 6-month horizon						
UK	0.854	0.000	0.146	–	–	–
Australia	0.724	0.013	0.263	–	–	–
Canada	0.831	0.000	0.169	–	–	–
Switzerland	0.939	0.000	0.061	–	–	–
Euro	0.853	0.000	0.147	–	–	–
Japan	0.849	0.000	0.151	–	–	–
New Zealand	0.948	0.000	0.052	–	–	–
Norway	0.714	0.048	0.238	–	–	–
Sweden	0.692	0.000	0.308	–	–	–

Notes: Entries are the proportions of individual forecaster regressions, by currency and forecast horizon, that reject the no-herding null under test in Equation (5). Columns 1–3 use contemporaneous (current) consensus forecasts. Columns 4–6 use lagged consensus forecasts. For the 6-month horizon, lagged-consensus results are not available in the original data.

5.2 Panel Estimation

To strengthen our testing of the herding hypotheses (as defined in the two individual regression based tests), we consider evidence which uses the full cross-section of individual forecasters, at each point in time, as well as the time-series. However, given the appropriate form of panel estimation in this context is complicated by the sparse nature of the panel, we first test if the estimated parameters across forecasters can be assumed to be homogeneous, as is standard when using fixed effect estimation methods.[19] For this we construct the test of slope homogeneity, in unbalanced panels, proposed by Pesaran and Yamagata (2008), which compares estimates using a pooled approach with those from individual regressions for each cross sectional unit.[20] For all currencies, forecast horizons and across the range of specifications for the revision and forecast error based tests, we strongly reject the null of homogeneous parameters.[21]

Given that we have a reasonable number of time series observations (see Table 1), and reject homogeneous parameters, we choose to use the Mean Group (MG) panel estimation, as proposed by Pesaran and Smith (1995).[22] We apply the MG panel approach to each currency and forecast horizon, where the cross section is over individual forecasters. We allow for cross sectional dependence between forecasters, which seems likely in FX markets, and implement the regressions as suggested in Pesaran (2006).[23] We test for cross-sectional dependence using the Pesaran (2004, 2015) cross-sectional dependence test based on correlations of the residuals.[24] For all currencies, forecast horizons and regression specifications we strongly reject the null of no cross- sectional dependence. We also address the issue of outliers (possibly caused, in some limited cases, by a smaller number of time-series observations) by trimming the estimates and removing the largest and smallest 5% of the estimated parameters from the calculations. Hence, we can make inferences with respect to the range of herding nulls, using a larger set of observations where the tests are based on panel estimates. This complements the evidence based on the proportion of rejections used in the individual regression-based approach, where instead we use p-values which provide formal statistical tests of the various herding null hypothesis.[25]

Table 7 reports the estimated coefficients $\hat{\phi}_1$ and $\hat{\rho}_1$, along with the p-values of a two-sided alternative testing against the null that the coefficients are significantly different from zero. We use the signs of the estimated coefficients to label the result as providing evidence for either herding or anti-herding, where negative coefficients suggest evidence in favour of herding and anti-herding, for $\hat{\phi}_1$ and $\hat{\rho}_1$ respectively. Where we previously observed high proportions of individual regressions rejecting

[19]Fixed effects panel estimation is commonly used across a wide range of finance applications, for example, when analysing individual foreign exchange forecasters in Dahlquist and Soderland (2023).

[20]Also see page 742 in Pesaran (2015).

[21]The size of the test statistics range from around 5 through to 40, with the p-values close to zero.

[22]MG estimation is suitable when the assumption is that slope coefficients are heterogeneous across cross-sectional units (i.e., individual forecasters). It allows for individual slope coefficients for each unit and averages them to obtain the overall estimate, allowing for heterogeneity when estimating the variance of the estimates and in turn effecting the outcomes of hypothesis testing. When the true model has heterogeneous slopes, the estimation of fixed effects can produce biased and inconsistent estimates.

[23]If there is cross-sectional dependence in the residuals, then the panel estimates can be both biased and inefficient.

[24]Also see page 794 in Pesaran (2015).

[25]We cross check the parameter estimates from a fixed effects panel estimation, with those from the MG with cross sectional dependence and in most cases - across currency, forecast horizon, and for the various regression specifications - the sign and size of the coefficients are similar. However, the difference in the size of the standard errors can be large.

the null in favor of the alternative, if the panel estimation is consistent with the individual regression results, we would observe low p values, rejecting or providing weak evidence in favour of the null. And vice versa.

The general observation is that the proportions of rejections at the individual level and the size and signs of the estimated coefficients and p-values from the panel estimation are consistent, and as such corroborate each other. For the null of $\phi_1 = 0$ in equation (5) the p-values clearly reject the null, and the estimated coefficients are all strongly negative, providing confirmation of evidence in favour of herding behaviour. This is the case across currencies, forecast horizons and whether current or lagged consensus forecasts are used. This result provides much stronger evidence for herding compared to the evidence when using lagged consensus forecast from individual regressions, which typically have fewer than 50% rejecting the null. For the null of $\rho_1 = 0$ in equation (4), the results are similar, where p-values clearly reject the null, and the estimated coefficients are all negative (the coefficients are small for Japan, Norway and Sweden, suggesting a small anti herding effect), but where in contrast this provides confirmation of evidence in favour of anti-herding behaviour. Hence, as with the individual regression based evidence, the results present a mixed picture.

Table 7: Panel herding tests: coefficients and p-values (Mean Group, cross-sectional dependence)

	Revision (Current)		Revision (Lagged)		Error (Current)	
Currency	$\hat{\phi}_1$	$p(\phi_1 = 0)$	$\hat{\phi}_1$	$p(\phi_1 = 0)$	$\hat{\rho}_1$	$p(\rho_1 = 0)$
Panel A: 1-month horizon						
UK	-0.764	0.000	-0.879	0.000	-0.158	0.000
Australia	-0.926	0.000	-0.881	0.000	-0.158	0.000
Canada	-0.807	0.000	-0.862	0.000	-0.178	0.000
Switzerland	-0.800	0.000	-0.857	0.000	-0.199	0.000
Euro	-0.757	0.000	-0.909	0.000	-0.144	0.000
Japan	-0.860	0.000	-0.864	0.000	-0.150	0.000
New Zealand	-0.934	0.000	-0.875	0.000	-0.226	0.000
Norway	-0.790	0.000	-0.870	0.000	-0.150	0.000
Sweden	-0.827	0.000	-0.891	0.000	-0.165	0.000
Panel B: 3-month horizon						
UK	-0.788	0.000	-0.876	0.000	-0.239	0.000
Australia	-0.788	0.000	-0.882	0.000	-0.245	0.000
Canada	-0.788	0.000	-0.863	0.000	-0.284	0.000
Switzerland	-0.829	0.000	-0.826	0.000	-0.304	0.000
Euro	-0.778	0.000	-0.897	0.000	-0.217	0.000
Japan	-0.861	0.000	-0.856	0.000	-0.210	0.000
New Zealand	-0.865	0.000	-0.867	0.000	-0.321	0.000
Norway	-0.908	0.000	-0.878	0.000	-0.193	0.000
Sweden	-0.885	0.000	-0.910	0.000	-0.196	0.000
Panel C: 6-month horizon						
UK	-0.752	0.000	–	–	-0.215	0.000
Australia	-0.956	0.000	–	–	-0.199	0.000
Canada	-0.823	0.000	–	–	-0.253	0.000
Switzerland	-0.826	0.000	–	–	-0.319	0.000
Euro	-0.774	0.000	–	–	-0.204	0.000
Japan	-0.890	0.000	–	–	-0.172	0.000
New Zealand	-0.969	0.000	–	–	-0.296	0.000
Norway	-0.857	0.000	–	–	-0.168	0.000
Sweden	-0.887	0.000	–	–	-0.146	0.000

Notes: The table presents the estimated coefficient and p-values for the mean group with cross-sectional dependence panel estimation. The null for the tests is no herding i.e. the estimated coefficients are zero, versus a two-sided alternative, where the sign of the estimated coefficient guides rejecting the null in favour of either herding or anti-herding. The panel estimates trim or remove the top and bottom 5% of parameters estimates when ranked in order of size, in order to alleviate sensitivity to outliers.

5.3 Intepretation

Contrasting the forecast error based test in equation (4) with one based on revisions in equation (5), naturally leads to regressions designed to test over- or under reaction, at either the individual or aggregate level (see, for example, Bordalo *et al.* (2020)). In the context of herding, forecast errors or revisions, are related to deviations of individual forecasts from the consensus forecast. Whereas tests of over or under reactions regress the forecast error, the dependent variable of equation (4), on forecast revisions, the dependent variable of (5).

Therefore, as in Bordalo *et al.* (2020), we run two sets of regressions to analyse individual forecasts. First, we run the forecaster-by-forecaster regressions:

$$FE_{t+h}^{j} = \gamma_{j0} + \gamma_{j1} REV_{t}^{j} + \varepsilon_{t+h} \tag{6}$$

where $FE_{t+h}^{j} = (s_{t+h} - s_{t+h|t-(h+a)}^{j})$ and $REV_{t}^{j} = (s_{t+h|t-(h+a)}^{j} - s_{t+h|t-(h+b)}^{j})$. We then report, in the first two columns of Table 8, summary estimation results for individual forecasters. At the individual level, the proportion of individual forecasters showing significant positive or negative estimates of γ_{j1} is low (see the first column), and hence the evidence suggests neither over nor under reaction. However, the proportion of rejections increases with the forecast horizon, with a strong tendency for the proportions of rejections in favour of the alternative hypothesis being positive (see column 2) to be low. This indicates that when significant, the coefficient γ_{j1} has a strong tendency to be negative, therefore individual forecasters seem to overreact. Because $\rho_1 < 0$ in equation (4), which implies that forecast errors and deviations from consensus move in opposite directions, the finding that γ_{j1} is negative is consistent with the anti-herding evidence from equation (4). Thus, the negative overreaction coefficients reinforce this interpretation provided by ρ_1.

A second alternative regression approach is to pool forecasters and estimate a common coefficient γ_{p1} from the regression:

$$FE_{t+h}^{j} = \gamma_{p0} + \gamma_{p1} REV_{t}^{j} + \varepsilon_{t+h} \tag{7}$$

A negative $\hat{\gamma}_{p1}$ indicates over reaction (overshoot of revisions).[26] The estimated coefficients are reported in the last two columns in Table 8. For the forecast horizon $h = 1$, only for Sweden we observe $\hat{\gamma}_{p1}$ being significantly different from zero. As the forecast horizon increases, we observe stronger degrees of significance (5 of 9 currencies at $h = 3$ at the 10% level of significance and all currencies at $h = 6$) and mostly negative coefficients, again providing evidence in favour of over reaction.

As a link to the herding hypothesis examined, and as a starting point to the analysis in Section 6, following Amedo (2025) we add into the individual regression the deviation between the individual and consensus forecast (the herding term) :

$$FE_{t+h}^{j} = \delta_{j0} + \delta_{j1} REV_{t}^{j} + \delta_{j2} DEV_{t}^{j} + \varepsilon_{t+h} \tag{8}$$

[26]The literature also estimates a panel version that allows for a forecaster fixed effect.

where $DEV_t^j = (s^j_{t+h|t-(h+a)} - \bar{s}_{t+h|t-(h+b)})$. The results (not reported) strongly suggest that including the difference between individual and consensus forecasts, a herding motivated deviation term, makes very little difference to the conclusions regarding over and under reaction. The proportion of rejections for $\delta_{j1} = 0$ and $\delta_{j1} > 0$ null hypothesis is very similar in size and pattern across currencies and forecast horizons to those reported in columns 1 and 2 in Table 8. The estimated coefficients on the deviation term, $\hat{\delta}_{j2}$, are mostly negative, small in size, and significant in 20-30% of cases across forecast horizons and currency. Results are consistent with those reported in Table 6.

In general, the evidence does not support the presence of meaningful herding in the FX forecasts. Although revision regression estimates of ϕ_1 frequently produce negative coefficients that would mechanically be interpreted as herding, this pattern is most likely driven by transitory noise, timing mismatches, and the mechanical pull of revisions toward the mean—issues that Clements (2018) identifies as producing "pseudo-herding" even when agents behave independently.

In contrast, the more robust BCK-based tests using forecast errors ρ_1 and the overreaction regressions both predominantly point to anti-herding, not herding. In the FX market specifically, this is economically plausible: FX strategists face strong incentives to differentiate their forecasts, to reflect institution-specific views, trade recommendations, and house positions, and to signal private information in a highly competitive environment with little reputational gain from clustering around consensus. Moreover, when consensus information is lagged to reflect a more realistic information set for FX forecasters—who do not observe consensus in real time—the apparent herding largely disappears. This strongly suggests that the herding detected ϕ_1 arises from econometric artifacts rather than genuine behavioural convergence.

Taken together, the weight of the evidence points toward either (i) no herding, or (ii) a tendency toward anti-herding, where forecasters deliberately differentiate themselves from consensus. In this case, forecasters may have stronger incentives to differentiate themselves from the consensus in order to gain visibility, attract clients, or establish a reputation for unique insights. Being distinctly correct can generate substantially larger reputational gains than issuing a forecast close to the consensus. As a result, forecasters may intentionally place greater weight on private information or exaggerate deviations from market expectations. Evidence supporting anti-herding, instead of herding, may also reflect genuine informational heterogeneity, particularly during periods of heightened uncertainty, when agents interpret macroeconomic news, monetary policy signals, or geopolitical developments differently.

This conclusion is consistent with the institutional features of FX forecasting and need not mirror findings from macroeconomic survey data, where incentives to herd are stronger and the Clements critique is more directly applicable. In FX, heterogeneity of views, institutional positioning, and noise dominate, leaving little role for genuine herding.[27]

[27] Why does this differ from macro-prediction? Macro forecasters face reputational incentives to cluster (avoiding big errors, professional norms). FX strategists face incentives to differentiate, publish bold directional calls, or align with trading strategies. FX forecasts are updated more frequently, are noisier, and reflect market positioning rather than pure information processing. The FX consensus is less stable, so moving "to consensus" is not a dominant behaviour.

Table 8: Overreaction regression results by horizon

Currency	Prop. reject $\gamma_{j1} = 0$	Prop. reject $\gamma_{j1} > 0$	$\hat{\gamma}_{p1}$	p-value
Panel A: 1-month horizon				
UK	0.101	0.034	0.126	0.142
Australia	0.164	0.045	0.949	0.420
Canada	0.188	0.000	0.092	0.782
Switzerland	0.304	0.000	-0.159	0.358
Euro	0.140	0.023	-0.078	0.325
Japan	0.202	0.169	0.001	0.972
New Zealand	0.317	0.000	-0.056	0.901
Norway	0.500	0.026	2.317	0.124
Sweden	0.314	0.029	2.367	0.054
Panel B: 3-month horizon				
UK	0.207	0.011	0.055	0.479
Australia	0.279	0.015	-1.004	0.098
Canada	0.250	0.000	0.334	0.533
Switzerland	0.519	0.000	-0.311	0.000
Euro	0.226	0.011	0.216	0.479
Japan	0.209	0.023	-0.082	0.424
New Zealand	0.424	0.000	-1.247	0.064
Norway	0.400	0.000	-1.647	0.062
Sweden	0.216	0.000	-1.632	0.061
Panel C: 6-month horizon				
UK	0.678	0.000	-0.160	0.087
Australia	0.360	0.000	-0.956	0.000
Canada	0.278	0.000	-0.859	0.000
Switzerland	0.608	0.000	-1.004	0.000
Euro	0.427	0.011	-0.983	0.000
Japan	0.181	0.036	-1.024	0.000
New Zealand	0.246	0.000	-1.100	0.000
Norway	0.270	0.054	-0.978	0.000
Sweden	0.364	0.030	-0.980	0.000

Notes: The table reports overreaction regression results by currency and forecast horizon. The first column is the proportion of individual regressions rejecting the null of $\gamma_{j1} = 0$; the second column is the proportion rejecting in the positive direction; column three reports the estimated value of γ_{p1}, the pooled consensus slope and column four its Newey–West p-value;

6 Alternative Explanations of Changing Expectations Formation and Revisions

An individually-published forecast reflects the model and/or methods specific to the forecaster and will change in relation to previous recent forecasts following the arrival of new information. For example, new data, developments in the FX market, the macro economic environment and uncertainty, as well as information of what others have or are currently forecasting for the same target variable at the same point in the future. As such, it is difficult for the econometrician to disentangle in a new forecast what is new information and what is an imitation or herding effect.

Therefore, in this section, we investigate alternative explanations of the formation and revisions of expectations in the context of the regression-based approach examined. If the observed evidence on herding changes or is robust to the inclusion of a range of factors measuring alternative explanations of expectations formation, we gain some understanding of the various influences on expectations formation.

6.1 Regressions with Additional Variables

Here we re-estimate equations (4) and (5), for all individual forecasts for each forecast horizon and currency pair, but where we include an additional term(s). Hence they take the form:

$$(s_{t+h} - s^j_{t+h|t-(h+a)}) = \rho_0 + \rho_1(s^j_{t+h|t-(h+a)} - \bar{s}_{t+h|t-(h+b)}) + \rho_2 X_t + u_{t+h}. \tag{9}$$

$$(s^j_{t+h|t} - s^j_{t+h|t-(h+a)}) = \phi_0 + \phi_1(s^j_{t+h|t-(h+a)} - \bar{s}_{t+h|t-(h+b)}) + \phi_2 X_t + u_{t+h}. \tag{10}$$

where X_t is either a scalar or a vector of variables that could potentially influence expectation formation.[28] Here we investigate if the inference on ρ_1 and ϕ_1 changes with the inclusion of X_t. We use X at time t, as it is this point at which individual forecasts are formed, either in terms of forecast errors or in terms of the revision of the previous period. We examine six different types of X_t variables.

First, we introduce the cross-sectional standard deviation of the individual forecasts from the previous period, described in GGJ as a general form of uncertainty, which measures an effect due to a desire to move closer together as the time horizon decreases. Dahlquist and Soderland (2024) include such a term in their regressions, and this is one measure of disagreement typically used in the literature.[29]

Second, we consider the global FX volatility measure of Menkoff et al. (2012). This requires we calculate the absolute daily log return, $|r^i_\tau| = |\Delta s_\tau|$, for each currency i on each day τ in our sample.[30] We then average over all currencies available on any given day (nine in our case) and then average

[28] For example, the deviation from consensus term in Section 5.3

[29] Other measures of disagreement include the inter-quartile and highest-lowest ranges.

[30] Daily exchange rate data are sourced from LSEG Datastream. We collect "dollar into US rates", code: XUSL**D, where ** denotes the country code.

the average daily values over the monthly frequency. The global volatility proxy in month t is defined as:

$$\sigma_t^{FX} = \frac{1}{T_t} \sum_{\tau \in T_t} \left[\sum_{i \in N_\tau} \left(\frac{|r_\tau^i|}{N_\tau} \right) \right], \tag{11}$$

where N_τ denotes the number of currencies available on day τ and T_t denotes the number of trading days in month t.

Third, we analyze a measure of uncertainty for each currency and forecast horizon, constructed following Rossi and Sekhposyan (2015), using the information embedded within a predictive density. Specifically, we construct an uncertainty index based on the cumulative density of the forecast errors evaluated at the actual realized forecast error. In our application, as in Rossi and Sekhposyan (2015), we compare the forecast error to the unconditional distribution from historical data.[31] If the realized forecast error lies in the tail of the distribution, we conclude that the realization was difficult to predict, and therefore the environment with respect to that variable is very uncertain.

Fourth, broadening our perspective on what constitutes uncertainty, we examine a variety of well-known uncertainty measures from the literature. Specifically, we include, one at a time, monthly data for the following indices: (i) Economic Policy Uncertainty (EPU) Index for the United States, (ii) Global Economic Policy Uncertainty Index, PPP adjusted by GDP, (iii) Economic Policy Uncertainty Index for Trade Policy, (iv) Jurado, Ludvigson and Ng's (2015) (JLN) 1-month, 3-month, and 1-year ahead Macroeconomic Uncertainty Indexes, (v) Jurado, Ludvigson, and Ng's (2015) (JLN) 1-month, 3-month and 1-year ahead Financial Uncertainty Indexes, (vi) Equity Market Volatility Tracker: Overall Tracker and (separately) for Macroeconomic News and Outlook on Business Investment and Sentiment, (vii) the Caldara and Iacoviello (2022) newspaper-based Geopolitical Index, [32] and (viii) as in Yang et al. (2025), the innovation from a stochastic volatility AR(1) process of the JLN 1-month conditional volatilities, for macro and financial uncertainty.[33]

Fifth, following Dahlquist and Soderland (2024), we examine the effect of including three common foreign exchange predictors used in the literature, which respectively form the basis of carry, value,

[31] We do not distinguish upside from downside uncertainty, although the constructed index would also allow for such type of non-linearity to be incorporated in our analysis.

[32] See https://fredhelp.stlouisfed.org for: EPU in index for US (USEPUINDXD); EPU Index trade policy (EPU-TRADE); equity market voaltility tracker overall index (EMVOVERALLEMV); Equity market volatility tracker macro economic news (EMVMACROBUS).

For global economic policy uncertainty see PolicyUncertainty.com, "Global Economic Policy Uncertainty Index," https://www.policyuncertainty.com/global$_{m}$*onthly.html*.

Data for the Geopolitical Risk (GPR) Index are available at Matteo Iacoviello's website, https://www.matteoiacoviello.com/gpr.htm. The EPU-based uncertainty indices put forward by Baker, Bloom and Davis (2016) rely on newspaper coverage frequency.

For the JLN indices see Sydney C. Ludvigson, "Macro and Financial Uncertainty Indexes," https://www.sydneyludvigson.com/macro-and-financial-uncertainty-indexes.

[33] We follow Yang et al. (2025), who focus on shocks to a state variable that contains predictive information for the cross-section of returns. They in turn follow literature such as Bali et al. (2017). Ang et al. (2006) who rely on first differences of the VIX to limit the high persistence typically observed in volatility indices. Yang et al. (2025) estimate a stochastic volatility specification to capture the dynamics of the conditional volatilities used in JLN. Specifically, they assume that

$$\log(\sigma_{t+1}^j)^2 = \gamma_0^j + \gamma_1^j \log(\sigma_t^j)^2 + \eta_{t+1}^j, \qquad \eta_{t+1}^j \sim \text{i.i.d. } N(0, \tau^j),$$

where σ_{t+1}^j denotes the conditional volatility of the forecast error generated by a FAVAR model as in JLN. The innovations η_{t+1}^j extracted from this volatility process are the measures of uncertainty used.

and momentum/reversal strategies: the forward premium, the log of the real exchange rate, and the recent USD depreciation rate.[34] Sixth, we use a range of inflation expectations and forecasts of macro variables, where we consider four variables; the overall Michigan inflation expectations forecast, their 1-year inflation expectations measure, and the 4-quarter Survey of Professional Forecasters (SPF) forecasts of real output growth and (price deflator) inflation.

Finally, we undertake a simple principle components analysis (PCA), constructing three sets of principal component representations capturing the co-movements of the conditioning variables examined. Having constructed the principal components, we include the first three (for each set) directly in individual regressions as an alternative to including the conditioning variables individually. The first set of variables considered in the PCA are: the economic uncertainty policy indices for the US and trade, equity market volatility overall and macro indices, the JLN 1-month, 3-month and 1-year ahead macro and financial uncertainty indices, the global volatility measure of FX, the Michigan inflation expectations indices overall and 1-year, and current and 1- to 4-quarter ahead SPF forecasts for output growth and PGDP inflation. The second set of variables considers the three FX predictor variables, across currencies and forecast horizon. The total number of variables considered is 108. The final set combines the first two sets of variables and then computes the principal components. In each case, we use the first three principal components.[35]

6.2 Results

An issue in this context is whether any of the alternative variables we consider have predictive power for exchange rate changes. If not, then the inclusion of them in the regressions would be unlikely to make any difference. If they have predictive power, then in principle they might influence exchange rate fluctuations, or it may be the case that the differences between individual and consensus forecasts alone drive herding behaviour as they contain all or most of the information contained in such variables. Therefore, before proceeding to the next step, we examine the predictability of exchange rate changes by adding to the Minzer-Zarnowitz regressions reported in Section 3 the X_t variables.[36]

Table A2 in the appendix reports the proportion of individual forecasters who reject the null of the estimated coefficients of additional variables in the MZ regression being significantly different from zero, i.e. do they have predictive power. The general observation is that, as the forecast horizon increases, the forecasts become less efficient, where the proportion of individual forecasts that have p-values less than 0.1 (if we were to adopt a 10% significance criterion) increases. For example, at $h = 1$, low proportions of individual forecasters reject the null of forecast efficiency when adding additional variables (the highest average value across currencies is 0.29 for the Michigan 1 year expected inflation

[34] If we let $f_{k,t}$ denote the log k-period forward exchange rate at date t, we can define $f_{k,t} - s_t$ as the forward premium. We define the real exchange rate as the spot exchange rate multiplied by the foreign CPI level divided by the US CPI level. The depreciation of the USD relative to the foreign currency is defined as $s_{t+k} - s_t$.

[35] The first principal component explains 78%, 98.7% and 78% of the movement for the three sets, respectively, the second 19%, 0.5% and 19%, and the third 2.9%, 0.4% and 2.9%.

[36] We estimate the standard joint (conditional) tests of unbiasedness or rationality, based on the following augmented conventional Mincer-Zarnowitz regression:

$$(s_{i,t+h} - s_{i,t}) = \alpha_h + \beta_h(\bar{s}_{i,t+h|t} - s_{i,t}) + \gamma_h X_t + u_{i,t+h},$$

and 0.26 for the FX predictor and volatility variables). As h increases, we observe inefficient forecasts or strong predictability from a wide range of additional variables. For $h = 3$, the lowest proportion is 0.22 for the cross-sectional standard deviation, the highest 0.68 for the FX predictor and volatility variables (where most are at least 0.50). As we move to $h = 6$ and $h = 12$, we observe very high proportions (the lowest being 0.38 or 0.40 for Geopolitical risk), which at $h = 12$ are as high as 0.95 for the FX predictors and volatility and 0.91 for JLN Financial uncertainty variables. The additional variables that perform the best overall are the FX predictor and volatility variables, JLN financial uncertainty measures, macro inflation surveys, and SPF output growth and inflation forecasts. These results suggest that this information has predictive power, and therefore the forecasts are not rational or inefficient (bias was examined in section 3) and as such, potentially, they may be relevant when examining herding.[37]

We focus on the evidence from the individual regression based approach. Table 9 reports the *average* proportions in the nine currencies of individual forecasters who reject the null hypothesis of no herding for the herding and anti-herding alternatives described in the forecast error and revision tests. In the first row of each panel, we report the averages of Tables 5 and 6, where no conditioning variables are included, which act as a set of benchmarks. The remaining rows are the proportion of rejections for the same set of null and alternative hypotheses, but where the regressions include, in turn, the conditioning variables.

When we introduce the (lagged) cross-sectional standard deviation (disagreement) into the range of regressions, we observe that it makes very little difference to the proportions of individual regressions which reject the null hypothesis on the two tests; when using current and lagged consensus forecasts, and across different forecast horizons and individual currencies (not reported). Therefore, the conclusions we have drawn in the previous section regarding herding are not explained by a dispersion measure over observed forecast disagreement in each currency. If we expand the measure of volatility and use the global volatility measure of Menkoff et al. (2012), defined in equation (11), or if we include the range of currency-specific forecast uncertainty measures defined by Rossi and Sekhposyan (2015), we observe the same result. That the inclusion of these terms in the regressions does not explain the conclusions regarding herding, as the proportion of rejections remains very close to when they are not included, and hence are robust in this regard.

When uncertainty in the FX market is high, for example due to geopolitical events, economic crises, policy changes, investors might find it challenging to predict future exchange rate movements. Potentially, this uncertainty can lead market participants to look for cues from other traders, analysts, or market sentiment to form their own expectations. Hence, in times of increased uncertainty, individual investors might rely more on the collective actions of others, assuming that others possess better or more timely information. Hence, we also condition on, one at a time, three measures of policy uncertainty. The US-specific policy uncertainty, as all our currencies are relative to the dollar, global policy uncertainty, and trade policy uncertainty given it's association with foreign exchange flows. The outcome remains the same as with the other conditioning variables, namely it makes little difference to the proportions of null hypothesis rejected.

[37] Furthermore, reported in table A2, we test the joint hypothesis of a zero constant, unit parameter in the expectations term, and zero coefficients in the additional variables. The overriding result is to reject the null in nearly all cases, across currencies, forecast horizons, and the full range of additional variables.

Macro-economic and financial uncertainty can be a cause of herding behaviour. Events such as unexpected changes in interest rates, inflation, or political instability can lead to uncertainty. During such periods, herding behaviour might become more evident as investors move in a unified manner toward safer currencies (such as the US dollar) or away from riskier ones. Hence we include (separately) the JLN 1-month, 3-month, and 1-year ahead measures of macro and financial uncertainty, an innovations-based version of the financial uncertainty, two equity based volatility measures, an overall measure and one specific to the macro environment, and finally a Geo-political Index. Note that the table reports results for JLN 3-month indices only, as the 1-month and 1-year versions yield essentially the same outcomes, and that we tried a macro innovation version also, and this gave similar results to the financial version. The overriding result across this wide range of uncertainty measures is that the proportion of rejections remains unaffected. This is despite some of these variables appearing to have predictive power for exchange rate changes.

When uncertainty limits the availability of reliable information, market participants may default to the consensus view, reinforcing herding behaviour. In such environments, if a handful of major institutions or prominent analysts take a strong position—such as forecasting a currency depreciation—others frequently follow. The assumption is that these influential players possess superior insight and this imitation leads investors to adopt similar positions or strategies in the FX market. The resulting convergence magnifies the underlying market movements. To account for this dynamic, we incorporate the three standard foreign exchange predictors employed by Dahlquist and Söderlind (2024): the forward premium, the real exchange rate, and the past depreciation, representing the carry, value, and momentum, respectively. These predictors are specific to each currency and forecast horizon. However, consistent with the other controls we evaluate, their inclusion has only a minimal impact on the proportion of rejections in our forecast revision and forecast error tests, relative to regressions that omit conditioning variables altogether.

We extend the analysis by adding inflation expectations from the Michigan survey, four-quarter-ahead mean SPF forecasts for output growth and inflation, and the first three principal components derived from the broader combined set of variables described earlier.[38] As in all the previous specifications, these additions do not alter the original herding results estimated without conditioning variables, and as such they remain robust.

Finally, in additional regressions not reported in detail, we incorporate a wide set of conditioning variables using a mean-group estimator that accounts for cross-sectional dependence. The results closely mirror those of the individual-currency regressions. The p-values associated with both the forecast revision and the forecast error tests—each evaluating the null of no herding under different conditioning environments—remain nearly unchanged. This stability holds at the aggregate level, across individual currencies, over alternative forecast horizons, and even when both current and lagged consensus forecasts are included.

Taken together, the evidence strongly indicates that the herding findings are not an artefact of omitted variables. Instead, they persist in an extensive range of controls and estimation approaches, underscoring the robustness of the baseline herding results.

[38] We report one PCA representation in the table, as the other two yield very similar outcomes.

Table 9: Proportion of individual forecaster regressions rejecting the herding nulls (Current vs Lagged Consensus)

	Current consensus			Lagged consensus			Current consensus		
Conditioning Variable(s)	$\phi_1 < 0$	$\phi_1 > 0$	Neither	$\phi_1 < 0$	$\phi_1 > 0$	Neither	$\rho_1 < 0$	$\rho_1 > 0$	Neither
Panel A: 1-month horizon									
NO CONDITIONING	**0.921**	**0.001**	**0.078**	**0.443**	**0.022**	**0.535**	**0.331**	**0.014**	**0.655**
Cross-sectional std.	0.919	0.000	0.081	0.533	0.008	0.459	0.433	0.015	0.553
Global FX Volatility	0.931	0.000	0.069	0.458	0.026	0.516	0.360	0.009	0.631
R&S Uncertainty	0.902	0.001	0.096	0.438	0.027	0.535	0.248	0.150	0.602
Econ Policy Uncertainty : US	0.926	0.001	0.073	0.466	0.022	0.512	0.408	0.013	0.579
Global Econ Policy Uncertainty	0.885	0.004	0.111	0.446	0.021	0.532	0.233	0.177	0.590
Econ Policy Uncertainty : Trade	0.920	0.000	0.080	0.442	0.021	0.537	0.357	0.012	0.631
JLN 3-month ahead macro uncertainty	0.916	0.001	0.083	0.483	0.023	0.494	0.374	0.014	0.613
JLN 3-month ahead financial uncertainty	0.922	0.001	0.077	0.466	0.025	0.509	0.369	0.014	0.617
JLN 1-month innovation financial uncertainty	0.938	0.000	0.062	0.550	0.022	0.428	0.449	0.010	0.541
Equity Market Volatility: overall	0.932	0.000	0.068	0.466	0.022	0.512	0.330	0.015	0.656
Equity Market Volatility: macro	0.936	0.000	0.064	0.461	0.019	0.519	0.347	0.018	0.635
Geopolitical Index	0.920	0.001	0.079	0.482	0.029	0.489	0.377	0.015	0.609
FX Predictors	0.927	0.000	0.073	0.470	0.017	0.513	0.4244	0.0190	0.557
Michigan Inflation Expectations	0.929	0.006	0.071	0.540	0.030	0.430	0.416	0.017	0.567
1-year Inflation Expectations	0.940	0.000	0.060	0.515	0.030	0.455	0.404	0.018	0.578
4-quarter SPF Output growth forecast	0.903	0.004	0.093	0.459	0.018	0.523	0.348	0.014	0.638
4-quarter SPF Inflation (PGDP) forecast	0.920	0.000	0.080	0.532	0.023	0.445	0.396	0.009	0.596
PCA – All Variables	0.909	0.001	0.090	0.496	0.029	0.476	0.419	0.015	0.566
Panel B: 3-month horizon									
NO CONDITIONING	**0.881**	**0.003**	**0.116**	**0.469**	**0.010**	**0.521**	**0.533**	**0.009**	**0.458**
Cross-sectional std.	0.901	0.004	0.095	0.536	0.006	0.458	0.653	0.004	0.343
Global FX Volatility	0.883	0.004	0.114	0.540	0.005	0.455	0.558	0.013	0.429
R&S Uncertainty	0.866	0.004	0.130	0.493	0.009	0.498	0.419	0.206	0.375
Econ Policy Uncertainty : US	0.926	0.001	0.073	0.563	0.010	0.426	0.599	0.003	0.398
Global Econ Policy Uncertainty	0.872	0.004	0.124	0.494	0.008	0.498	0.353	0.190	0.457
Econ Policy Uncertainty : Trade	0.883	0.004	0.113	0.477	0.012	0.511	0.550	0.007	0.444
JLN 3-month ahead macro uncertainty	0.883	0.005	0.112	0.507	0.008	0.485	0.561	0.011	0.428
JLN 3-month ahead financial uncertainty	0.885	0.005	0.110	0.508	0.011	0.482	0.590	0.012	0.398
JLN 1-month innovation financial uncertainty	0.894	0.003	0.104	0.489	0.009	0.502	0.601	0.005	0.394
Equity Market Volatility: overall	0.882	0.002	0.116	0.563	0.010	0.426	0.566	0.009	0.425
Equity Market Volatility: macro	0.874	0.002	0.124	0.455	0.012	0.534	0.542	0.007	0.451
Geopolitical Index	0.884	0.004	0.113	0.500	0.009	0.491	0.585	0.007	0.408
FX Predictors	0.915	0.006	0.078	0.578	0.004	0.418	0.583	0.003	0.414
Michigan Inflation Expectations	0.897	0.003	0.101	0.540	0.011	0.450	0.631	0.005	0.359
1-year Inflation Expectations	0.893	0.003	0.104	0.542	0.009	0.449	0.615	0.005	0.380
4-quarter SPF Output growth forecast	0.870	0.003	0.128	0.518	0.006	0.476	0.565	0.013	0.422
4-quarter SPF Inflation (PGDP) forecast	0.895	0.003	0.102	0.552	0.010	0.437	0.677	0.012	0.320
PCA – All Variables	0.879	0.003	0.119	0.556	0.008	0.436	0.608	0.002	0.390
Panel C: 6-month horizon									
NO CONDITIONING	**0.823**	**0.007**	**0.171**	–	–	–	**0.594**	**0.017**	**0.390**
Cross-sectional std.	0.841	0.005	0.154	–	–	–	0.711	0.011	0.277
Global FX Volatility	0.823	0.011	0.166	–	–	–	0.691	0.010	0.300
R&S Uncertainty	0.801	0.010	0.189	–	–	–	0.419	0.206	0.375
Econ Policy Uncertainty : US	0.825	0.012	0.162	–	–	–	0.627	0.006	0.367
Global Econ Policy Uncertainty	0.798	0.019	0.183	–	–	–	0.428	0.210	0.361
Econ Policy Uncertainty : Trade	0.829	0.007	0.165	–	–	–	0.567	0.017	0.417
JLN 3-month ahead macro uncertainty	0.818	0.011	0.171	–	–	–	0.626	0.016	0.358
JLN 3-month ahead financial uncertainty	0.825	0.011	0.164	–	–	–	0.645	0.018	0.337
JLN 1-month innovation financial uncertainty	0.843	0.011	0.146	–	–	–	0.584	0.009	0.407
Equity Market Volatility: overall	0.824	0.011	0.165	–	–	–	0.609	0.017	0.374
Equity Market Volatility: macro	0.826	0.007	0.167	–	–	–	0.610	0.017	0.373
Geopolitical Index	0.821	0.012	0.167	–	–	–	0.400	0.208	0.392
FX Predictors	0.833	0.015	0.152	–	–	–	0.708	0.007	0.285
Michigan Inflation Expectations	0.837	0.007	0.157	–	–	–	0.671	0.012	0.317
1-year Inflation Expectations	0.838	0.007	0.155	–	–	–	0.656	0.011	0.332
4-quarter SPF Output growth forecast	0.824	0.008	0.168	–	–	–	0.608	0.013	0.379
4-quarter SPF Inflation (PGDP) forecast	0.838	0.012	0.150	–	–	–	0.667	0.012	0.320
PCA – All Variables	0.816	0.009	0.175	–	–	–	0.615	0.018	0.367

Notes: The table presents the average proportion (of the nine currencies) of the individual forecasters who reject the null, in favor of the stated alternative hypothesis, for the tests involving revisions and forecast errors. The first row in each panel, in bold in the original tables, acts as a benchmark reporting the proportions where no conditioning variables are included (as in the empirical implementation section). The remaining rows report proportions of rejections when the stated conditioning variables are included in the regressions.

7 Conclusions

This paper has investigated herding and anti-herding in foreign exchange forecasts using a range of complementary tests applied to a large panel of individual professional forecasts covering nine currencies over a maximum sample from January 1994 to December 2024, with around 40–50 forecasters per currency on average. Overall, the evidence does not support a robust conclusion of herding in FX markets. While revision-based tests suggest herding in 89% of cases when current consensus forecasts are used, this evidence weakens substantially when lagged information is employed, with only 46% of cases rejecting the null. This sensitivity is important for interpretation, since it suggests that stronger evidence of herding in some specifications may partly reflect the informational assumptions imposed by the tests.

By contrast, the forecast-error based tests suggest that in 51% of cases there is neither herding nor anti-herding, and, where the null is rejected, the evidence more often favours anti-herding. In the pooled over-reaction regressions in section 5.3, only one of the nine currencies shows a significant coefficient at the 1-month horizon, but this rises to five of the nine currencies at the 3-month horizon and all nine currencies at the 6-month horizon, with coefficients mostly negative. This indicates that forecast revisions tend to overshoot rather than adjust gradually, especially at longer horizons, and therefore reinforces the anti-herding interpretation.

The broader properties of the forecasts are also consistent with this interpretation. Exchange rate forecasts are mostly unconditionally unbiased, but they are conditionally inefficient, forecast errors are persistent at short horizons, and forecast accuracy is generally weak relative to a random walk benchmark. At the same time, disagreement across forecasters is substantial, persistent, and economically meaningful. These features are difficult to reconcile with a simple account in which forecasters converge on a common view. Instead, they point to a setting in which expectations remain heterogeneous and where noise and idiosyncratic error play an important role.

The analysis of alternative explanations for expectation formation strengthens this conclusion. A wide range of measures intended to capture uncertainty, volatility, and policy-related information frictions has little effect on the overall pattern of results. This is notable, given that such variables are often thought to provide a natural environment for herding to emerge. In our data, however, they do not systematically shift the evidence toward herding. Standard FX predictors, such as the forward premium, the real exchange rate, and past depreciation, have somewhat more influence, but their effects remain modest and do not alter the main message.

Overall, our interpretation is that there is mixed evidence of robust herding in professional FX forecasts. Where the null of no herding is rejected, the balance of the evidence more often favours anti-herding or over-reaction. Nonetheless, the dominant evidence of anti-herding can have several important implications for FX markets and for the interpretation of survey expectations. First, it suggests that forecast dispersion reflects not only macroeconomic uncertainty but also strategic behaviour among forecasters. Consequently, disagreement measures may overstate the degree of underlying economic uncertainty if part of the dispersion is driven by deliberate forecast differentiation. This also highlights the importance of looking beyond the mean or consensus forecast when analysing exchange-rate expectations. A sole focus on the average forecast may conceal substantial heterogeneity in beliefs and strategic positioning across agents, particularly during episodes of market stress or elevated uncertainty. Examining the cross-sectional distribution of expectations can therefore provide additional information about market sentiment, informational frictions, and behavioural dynamics that are not captured by the consensus alone. More broadly, evidence of anti-herding challenges the notion that survey forecasts represent efficient information aggregation. Instead, expectations may embody a combination of private information, behavioural incentives, and strategic signalling.

References

[1] Bachmann, R. , Topa G. and W. van der Klaauw, *eds.* (2022), *Handbook of Economic Expectations*, Elsevier.

[2] Bernhardt, D., Campello, M. and E. Kutsoati (2006), "Who Herds?", *Journal of Financial Economics*, 80, 657-675.

[3] Born, B., Enders, Z., Müller, G.J. and K. Niemann (2022). "Firm expectations about production and prices: facts, determinants, and effects", *Handbook of Economic Expectations*. Ed. by R.Bachmann, Gi. Topa, and W. van der Klaauw. Elsevier. Chap. 12, 355–383.

[4] Born, B., Enders, Z., Müller, G.J. (2025). "On FIRE, News, and Expectations", *The Routledge Handbook of Economic Expectations in Historical Perspective*. Ed. by L. Lenel, A. Nutzenadel, J. Streb and I. Köhler. Routledge. Chap. 10, 178–194.

[5] Bordalo, P., Gennaioli, N., Ma, Y. and A. Shleifer (2020), "Overreaction in Macroeconomic Expectations", *American Economic Review*, 110, 2748-2782.

[6] Caldara, D., and M. Iacoviello (2022). "Measuring Geopolitical Risk", *American Economic Review*, 112(4), 1194–1225.

[7] Coibion, O. and Y. Gorodnichenko (2015), "Information Rigidity and the Expectations Formation Process: A simple Framework and New Facts", *American Economic Review*, 109, 2644-2678. 657-675.

[8] Clements, M.P., (2018), "Do Macro-Forecasters Herd?", *Journal of Money Credit and Banking*, 50, 265-292.

[9] Dahlquist, M. and Soderland, P. (2023), "Individual Forecasts of Exchange Rates", SSRN.

[10] Dahlquist, M. and Ibert, M. (2024), "Institutions Return Expectations across Assets and Time ", SSRN.

[11] Diebold, F.X, Mora, A., and M. Shin (2024), "On the Wisdom of Crowds (of Economists) ", *Mimeo*.

[12] Farmer, L.E., Nakamura, E. and J. Steinsson (2024), "Learning about the long run ", *Journal of Political Economy*, 132 (10), 3334-3377.

[13] Frenkel, M., Mauch, M. and J.C. Rulke (2020), "Do forecasters of major exchange rates herd? ", *Economic Modelling* , 84, 214-221.

[14] Fritsche, U., Pierdzioch, C., Rulke, J.C. and G. Stadtmann (2015), "Forecasting the Brazilian real and the Mexican peso: Asymmetric loss, forecast rationality, and forecaster herding?", *International Journal of Forecasting*, 31, 130-139.

[15] Gallo, G.M., Granger, C.W.J., and Y. Jeon (2002), "Copycats and Common Swings: The Impact of the Use of Forecasts in Information Sets", *IMF Staff Papers*, 49, No.1, 4-21.

[16] Giacomini, R., and H. White (2006), "Tests of conditional predictive ability", *Econometrica*, 74, 6, 1545-1578.

[17] Ince, O. and T. Molodtsova (2017), "Rationality and Forecasting Accuracy of Exchange Rate Expectations: Evidence from Survey-Based Forecasts", *Journal of International Financial Markets, Institutions and Money*, 47, 131-151.

[18] Jongen, R., Verschoor, W.F.C., Wolff, C.C.P., and R.C.J. Zwinkels (2012), "Explaining dispersion in Foreign exchange expectations: A heterogenous agent approach", *Journal of Economic Dynamics and Control* , 36, 719-735.

[19] Jurado, K., Ludvigson, S. C. and S. Ng (2015), "Measuring Uncertainty", *American Economic Review*, 105(3): 1177–1216.

[20] Lahari, K. and X. Sheng (2008), "Evolution of forecast disagreement in a bayesian learning model", *Journal of Econometrics*, 144, 325–340.

[21] Lahari, K. and X. Sheng (2010), "Measuring Forecast Uncertainty by Disagreement: The Missing Link", *Journal of Applied Econometrics*, 25, 514–538.

[22] Laster, D., Bennet, P. and S. Geoum (1999), "Rational Bias in Macroeconomic Forecasts", *Quarterly Journal of Economics*, 114(1), 293-318.

[23] Menkoff, L., Sarno, L., Schmelling, M., and A. Schrimf (2012), "Carry Trades and Global Foreign Exchange Volatility", *Journal of Finance*, LXVII, 2, 681-718.

[24] Ottaviani, M. and P.N. Sorenson (2006), "The Strategy of professional forecasting", *Journal of Financial Economics*, 81, 441-466.

[25] Pesaran, M.H. (2004), "General Diagnostic Tests for Cross-sectional Dependence in Panels", CESifo Working Paper No. 1229.

[26] Pesaran, M.H. (2015), "Testing Weak Cross-sectional Dependence in Large Panels", *Econometric Reviews*, 34, 1089-1117.

[27] Pesaran, M.H. (2015), *Time Series and Panel Data Econometrics* , Oxford, Oxford University Press.

[28] Pesaran, M.H. and T. Yamagata (2008), "Testing Slope Homogeneity in Large Panels", *Journal of Econometrics*, 142, 50-93.

[29] Pesaran, M.H. and R.P. Smith (1995), "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels", *Journal of Econometrics*, 68, 79-113.

[30] Rossi, B. and T. Sekhposyan (2015), "Macroeconomic Uncertainty Indices Based on Nowcast and Forecast Error Distributions", *American Economic Review: Papers and Proceedings*, 105, 650-655.

[31] Yang, F., Calvo-Pardo, H.F. and J. Olmo (2025), "Hedging economic uncertainty from the cross section of stock returns", Available at SSRN: https://ssrn.com/abstract=5786573 or http://dx.doi.org/10.2139/ssrn.5786573.

Appendix

Table A1: Consensus Economics Average Forecasts Bias and Accuracy

Panel A: 1-month horizon								
	Forecast error			RMSE Ratio		Unbiasedness		
	Mean	t-stat	Std. deviation	Ratio	GW test	$\hat{\beta}_0$	$\hat{\beta}_1$	p-test
UK	0.48	1.68	3.46	1.431	0.000	0.000	0.015	0.000
Australia	0.87	1.63	5.14	1.413	0.000	0.000	-0.008	0.000
Canada	0.23	0.85	2.96	1.480	0.000	0.000	-0.407	0.000
Switzerland	0.84	2.45	3.74	1.326	0.000	0.003	0.160	0.000
Euro	0.20	0.51	4.06	1.478	0.000	0.000	-0.030	0.000
Japan	-0.04	-0.11	4.11	1.421	0.000	0.000	-0.030	0.000
New Zealand	1.15	2.23	4.60	1.352	0.000	0.002	0.144	0.000
Norway	-1.06	-1.59	4.63	1.464	0.000	-0.004	0.009	0.000
Sweden	-0.44	-0.60	4.07	1.531	0.000	-0.003	-0.025	0.000
Panel B: 3-month horizon								
	Forecast error			RMSE Ratio		Unbiasedness		
	Mean	t-stat	Std. deviation	Ratio	GW test	$\hat{\beta}_0$	$\hat{\beta}_1$	p-test
UK	0.53	0.98	5.17	1.212	0.0000	-0.002	-0.109	0.000
Australia	1.46	1.51	7.67	1.197	0.0006	-0.002	-0.090	0.000
Canada	0.54	0.97	4.58	1.101	0.0026	0.002	-0.064	0.000
Switzerland	1.49	2.47	5.04	1.200	0.0045	0.007	0.194	0.000
Euro	0.14	0.18	6.00	1.239	0.0000	0.001	-0.052	0.000
Japan	0.28	0.38	6.62	1.225	0.0000	-0.001	-0.139	0.000
New Zealand	1.83	2.16	5.80	1.183	0.0137	0.005	0.243	0.000
Norway	-2.27	-1.87	6.08	1.139	0.0781	-0.015	0.256	0.000
Sweden	-1.22	-0.96	5.69	1.199	0.0038	-0.010	0.034	0.000
Panel D: 12-month horizon								
	Forecast error			RMSE Ratio		Unbiasedness		
	Mean	t-stat	Std. deviation	Ratio	GW test	$\hat{\beta}_0$	$\hat{\beta}_1$	p-test
UK	-0.06	-0.05	8.52	0.9765	0.0000	-0.002	0.638	0.564
Australia	2.87	1.17	12.24	0.9726	0.0001	0.029	0.998	0.505
Canada	0.66	0.42	7.98	0.9269	0.0006	0.007	1.469	0.542
Switzerland	3.83	2.77	7.24	1.0276	0.1176	0.032	0.676	0.004
Euro	-0.17	-0.09	10.40	1.0411	0.0000	0.003	0.344	0.145
Japan	0.35	0.17	11.89	1.1345	0.0000	-0.003	-0.037	0.000
New Zealand	3.03	1.41	8.79	1.0334	0.0000	0.018	0.605	0.163
Norway	-8.49	-2.46	11.04	1.1375	0.0000	-0.072	0.645	0.013
Sweden	-6.16	-1.88	10.69	1.1119	0.0001	-0.046	0.449	0.040

Notes: The table presents the mean of the forecast errors, defined as $s_{t+h} - \bar{s}_{t+h|t}$, Newey-West t-statistic testing significant differences from zero, and standard deviation; Root Mean Square Error (RMSE) ratio relative to a random walk and accompanying Giacomini and White (GW) (2006) test of significance; the estimated coefficients from the Mincer-Zarnowitz regression, $(s_{t+h} - s_t) = \beta_0 + \beta_1(\bar{s}_{t+h|t} - s_t) + u_{t+h}$, and the Newey-West adjusted p-value of the F-test of the the joint hypothesis $\beta_0 = 0$ and $\beta = 1$.

Table A2: Tests of Forecast Efficiency Using Mincer-Zarnowitz Regression

Proportion of Individual Forecasters Rejecting Efficiency Null: $\gamma = 0$				
Conditioning Variable	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Cross-sectional Standard Deviation	0.107	0.223	0.384	0.581
Global FX Volatility	0.260	0.680	0.823	0.951
Econ Policy Uncertainty: US & Trade	0.187	0.498	0.648	0.672
JLN Macro Uncertainty: 1 month, 3-month, 1 year	0.226	0.439	0.604	0.662
JLN Financial Uncertainty: 1 month, 3-month, 1 year	0.131	0.594	0.904	0.915
Equity Market Volatility: Overall and Macro	0.236	0.411	0.456	0.537
Geopolitical Index	0.104	0.299	0.399	0.374
FX Predictors	0.260	0.680	0.823	0.951
Michigan and 1-year Inflation Expectations	0.292	0.620	0.715	0.687
SPF Output Growth and Inflation 4-quarter Forecast	0.130	0.549	0.856	0.768
Proportion of Individual Forecasters Rejecting Joint Null: $\beta_0 = 0$, $\beta_1 = 1$ **and** $\gamma = 0$				
Conditioning Variable	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Cross-sectional Standard Deviation	0.993	0.989	0.978	0.978
Global FX Volatility	0.989	0.994	0.994	0.995
Econ Policy Uncertainty: US & Trade	0.992	0.988	0.975	0.970
JLN Macro Uncertainty: 1 month, 3-month, 1 year	0.993	0.992	0.978	0.970
JLN Financial Uncertainty: 1 month, 3-month, 1 year	0.993	0.996	0.989	0.991
Equity Market Volatility: Overall and Macro	0.996	0.991	0.975	0.973
Geopolitical Index	0.996	0.992	0.975	0.962
FX Predictors	0.989	0.994	0.994	0.995
Michigan and 1-year Inflation Expectations	0.993	0.997	0.994	0.986
SPF Output Growth and Inflation 4-quarter Forecast	0.996	0.993	0.989	0.981

Notes: In the first panel, the table presents the proportion of individual forecasts rejecting the null hypothesis $\gamma = 0$, from the augmented Mincer-Zarnowitz regression: $(s_{t+h} - s_t) = \beta_0 + \beta_1(\bar{s}_{t+h|t} - s_t) + \gamma X_t + u_{t+h}$, where X_t contains the conditioning variables, at the 10% significance level, using Newey-West adjusted p-values. The second panel reports the proportion rejecting the joint null hypothesis: $\beta_0 = 0$, $\beta_1 = 1$ and $\gamma = 0$, using a Wald test at the 10% level of significance.

Acknowledgements

We thank Michael Clements, Cesare Robotti, Rory Mullen, Kevin Lee, Kalvinder Shields, Robin Lumsdaine, Magnus Dahlquist, Hector Carlo-Pardo, Arie Gozluklu, and participants in the 56th Annual Conference of the Money, Macro and Finance Society and the IAAE Annual Conference 2025 for helpful discussions and comments.
The views expressed here are those of the authors and do not necessarily reflect those of the ECB or the Eurosystem.

Anastasia Allayioti
European Central Bank, Frankfurt am Main, Germany; email: anastasia.allayioti@ecb.europa.eu

Anthony Garratt
University of Warwick, Coventry, United Kingdom; email: anthony.garratt@wbs.ac.uk

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-7900-9 ISSN 1725-2806 doi:10.2866/5290020 QB-01-26-144-EN-N