Predictability and Specification in Models of Exchange Rate Determination

Esfandiar Maasoumi* & Levent Bulut†

January 17, 2012

Abstract

We examine a class of popular structural models of exchange rate determination and compare them to a random walk with and without drift. Given almost any set of conditioning variables, we find parametric specifications fail. Our findings are based on a broad entropy functional of the whole distribution of variables and forecasts. We also find significant evidence of nonlinearity and/or "higher moment" influences which seriously questions the habit of forecast and model evaluation based on mean-variance criteria. Taylor rule factors may improve out of sample "forecasts" for some models and exchanges, but do not offer similar improvement for in-sample (historical) fit. We estimate models of exchange rate determination nonparametrically so as to avoid functional form issues. Taylor rule and some other variables are smoothed out, being statistically irrelevant in sample. The metric entropy tests suggest significant differences between the observed densities and their in- and out- of sample forecasts and fitted values. Much like the Diebold-Mariano approach, we are able to report statistical significance of the differences with our more general measures of forecast performance.

*Arts & Sciences Distinguished Professor of Economics, Emory University, Atlanta, GA, 30322 Tel: (404) 727-9817, E-mail: esfandiar.maasoumi@emory.edu

†Visiting Assistant Professor of Economics, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, 30303 Tel: (404) 413-6680, E-mail: bulut@gsu.edu

1 Introduction

Rational expectation hypothesis is at the core of the modern exchange rate determination models: assuming the 'true' structural model of the economy is known, forecasts are unbiased, uncorrelated, and efficient. In out of sample forecastabality of the *change* in spot rates, popular models seemingly fail to systematically beat a random walk based on *standard point forecast criteria.* In an influential paper, Meese & Rogoff (1983) examined this "puzzle", and based on data from 1970s, they found that the random walk model performs as well as any estimated structural or various time series models. In other words, knowing the past, current and even one period ahead true values of exchange rate fundamentals such as income, money supply growth rates, inflation rates, output gaps, and interest rates for two countries seemingly fails to produce better forecasts of the change in the nominal exchange rate, than simply using the current nominal exchange rate. This uncomfortable finding is known as the "Exchange rate disconnect puzzle" or the "Meese-Rogoff Puzzle".

Meese & Rogoff (1983) findings have been re-examined in the literature. One line of research re-examines different currency pairs, different time periods, real time versus revised macro data, and different linear structural models to assess the evidence based on the first and second moments of the distribution of the forecast errors. In this line of research, until recently, different attempts had failed to show forecasting power in the short run. A few, for example Mark (1995) and Engel et al.(2007), find better performing structural models only at horizons two-to-four years out! Cheung et al.(2005) applied a wide range of structural models and adopted different test statistics, yet failed to show that structural models consistently outperform the random walk model. Some researchers, instead of revised data, examined real time data since it was the only available information at the time of decision making. Faust et al.(2003) found better performing structural models in out-of-sample by using fully revised data, yet they failed to outperform random walk model. Other studies extend the set of structural models by including further macro variables such as the Taylor-rule principles. Molodtsova & Papell (2009) test the out-of-sample predictive power of Taylor-rule based exchange rate models and find that those models significantly outperform the random walk model in the short horizon. Rogoff & Stavrakeva (2008), on the other hand, are critical of the findings in Molodtsova & Papell (2009), Gourinchas & Rey (2007), and Engel et al.(2007), arguing that they are not robust to alternative time windows and alternative tests. They notably speculate that this lack of robustness to alternative time windows may be due to potential non-linearities and/or structural breaks. We agree, and would add inadequate conditioning sets and other mispecifications to the list of potential culprits. A very recent literature pioneered by Evans & Lyons (2002, 2005) incorporates the micro determinants of exchange rates (micro-structured models) into macro exchange rate models to form a 'hybrid' model. In this approach, the order flows (the detail on the size, direction and initiator of the transactions of exchange rates) are considered as signals of heterogeneous investor expectations and it is suggested that the order flows contain information on (and changing expectations of) exchange rate fundamentals that are more timely than the data releases. While there is some evidence of exchange rate predictability and forecastability of the 'hybrid' model with the very high frequency data (where data on exchange rate fundamentals do not exist), at higher forecast horizons (one month or longer), the evidence based on the conventional point forecast accuracy criteria is mixed(see Berger et al.(2008) and Chinn & Moore (2011)).[1]

A second related line of research questions model specification and seeks alternative, mostly non-linear/nonparametric, specifications to overcome the Meese & Rogoff (1983) findings. Diebold & Nason (1990) use univariate non-parametric time series methods to forecast the *conditional mean* of the spot exchange rates but fail to find better prediction performance over the random walk. Meese & Rose (1991) look for non-linear relationship between the exchange rates and their fundamentals to minimize the potential mis-specification in the linear models. They use non-parametric techniques and conclude that the poorer explanatory power of the structural models over random walk cannot be attributed to nonlinearities. This is suggestive of inadequate conditioning variables which we confirm in this paper.

Finally, an emergent third strand in the literature questions the conventional point forecast accuracy criteria and offers some alternatives, such as the density forecast approach. In this approach, the model-based forecast distributions are compared with the true (data-driven) distribution of the actual change in exchange rate series. Were these densities to be fully characterized by second moments, as in the case of the linear/Gaussian processes,

[1] It might be interesting to evaluate the performance of the 'hybrid' model by using the metric entropy criterion, but we leave that to future studies.

this approach would find the same results obtained with Diebold & Mariano (1995) and West (1996) type second moment tests. As for the evaluation of density forecasts, several different methods have been proposed. Diebold et al.(1998) propose to first estimate the forecast density of the model and transform it by the probability integral at each observed value over the forecast period. They suggest testing the implied i.i.d.'ness of these uniformly distributed transformed series. Clements & Smith (2000) test for the implied uniform distribution using the Kolmogorov-Smirnov statistic. In particular, Berkowitz (2001) proposes to transform Diebold et al.(1998) transformed statistics to a normal distribution to avoid the difficulties of testing for a uniform null. Even though Berkowitz (2001) methodology helps understanding how well a model's predictive density approximates the predictive density of the data, it doesn't allow for model comparison. To solve that problem, Corradi & Swanson (2006), in the spirit of Diebold & Mariano (1995), test for equal point forecast accuracy by proposing a testing strategy which tests the null of equal density forecast accuracy of two competing models. Along this line of research, and more encouragingly, Wang & Wu (2010) estimate the semi-parametric interval distribution of change in exchange rates to compare their forecast interval range with random walk model. They find supporting evidence of better forecast performance of Taylor-rule based structural models over the random walk model.

The approach in this paper encompasses the second and third strands of research highlighted above. We examine the same set of popular structural models of exchange rate determination, and compare them to a random walk with and without drift. Our criterion for assessment of closeness between distributions is a general entropy functional of such distributions. This reflects our critical view of mean squared type performance criteria as inadequate and/or inappropriate. Given almost any set of conditioning variables, we find parametric specifications fail by our criterion. We find significant evidence of nonlinearity and/or "higher moment" influences which seriously questions the habit of forecast and model evaluation based on mean-variance criteria. Conditioning variables, such as Taylor rule factors may improve out-of-sample "forecasts" for some models and exchanges, but do not offer uniform improvement for in-sample (historical) fit. Our findings benefit from nonparametric estimation. This is consistent with findings of nonlinear effects of fundamental and Taylor rule variables, as in this paper, Wang & Wu (2010), and Rogoff & Stavrakeva (2008). Such findings suggest the "Meese-Rogoff puzzle" may be an artifact of the parametric specifications of the traditional models, as well as due to inordinate focus on the first two moments of the forecast distributions. The tightness of the distribution of the forecast errors reported by Wang & Wu (2010) may be accompanied by other differences in asymmetry, kurtosis, and higher order moments. This is important information in risk management. All of the potential additional effects can be picked up by broader "distribution metrics", such as "entropy" that go well beyond the variance for non-Gaussian/nonlinear processes.

The metric entropy criterion was suggested in Maasoumi & Racine (2002), and Granger et al.(2004). It is capable of contrasting whole distributions of (conditional) predictions of parametric and nonparametric models, as well as that of random walk. It serves equally well as a measure of in-sample and out-of-sample fit or model adequacy, and it can assess the (nonlinear) affinity between the actual series and their predictions obtained from various models. This same metric serves as a measure of generic "dependence" in time series, as demonstrated in Granger et al.(2004).

Our findings, being robust to functional form misspecification, forecast criteria limitations, and a large set of popular explanatory variables, indicate the parametric forms are misspecified, and no current theory provides uniformly good forecast of the distribution of the observed changes in exchange rates.

The nonparametric approach suggests Taylor rule and some other variables are smoothed out, being statistically irrelevant in sample. The metric entropy tests suggests significant difference between forecast and fitted densities, on the one hand, and densities of the corresponding observed series. We are able to report statistical significance of differences for our more general measures of forecast performance. A by-product of our work is a more complete characterization of the gains over traditional specifications obtained from nonparametric implementations and additional fundamental variables and Taylor rule effects.

1.1 Entropy Measure of Dependence for Forecast Performance

- **As a Measure of Association**

Diebold & Mariono (1995) employed a quadratic loss function and computed the squared prediction errors. They tested the null of equal predictive accuracy by estimating whether the population mean of the difference between loss differential functions of two contestant models is zero. Such testing strategy is focused on the first and second moments of functions of predictions errors, and not designed to capture differences in higher moments. Given a set of conditioning variables, conditional mean is known as the best predictor under mean squared error criterion. Here we adopt the nonparametric entropy measure of dependence, as suggested in Granger et al.(2004), to compare forecasting power of several different models. It has been documented that this metric entropy has good power in detecting dependence structure of various linear and nonlinear models. The measure is similar to the Kullback-Leibler information criteria, but unlike the latter, it satisfies the triangularity property of metrics, and is a normalized Bhattacharya-Matusita-Hellinger measure as follows:

$$S\rho_1 = \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(f_1^{1/2} - f_2^{1/2})^2 da db \tag{1}$$

where $f_1 = f(a, b)$ is a joint density and $f_2 = g(a)h(b)$ is the product of the marginal densities of the random variables a and b. Two random variables are independent if and only if $f_1 = f_2$ which implies $S\rho_1 = 0$. For continuous variables, the upper limit of the function is normalized to unity, but for discrete variables the upper limit depends on the underlying distributions (see Giannerinni et al.(2011) for details). A statistically significant positive value of $S\rho_1$ implies existence of possibly nonlinear dependence/association. Theoretical properties of the cross validated, nonparametric estimation of $S\rho_1$ are examined in a number of papers by Skaug & Tjøstheim (1996), Su & White (2007), Granger et al.(2004), and lately by Giannerini et al.(2011) under more demanding sample dependence assumptions for time series. Giannerini et al.(2011) consider both sorugate method and bootstrap resampling approaches. Bootstrap is adequate to the task at hand in this paper, and is widely known to be a major improvement over the asymptotic approximate theory for these entropy measures.

In this paper, we utilize this entropy based measure of dependence for two traditionally distinct purposes: as an in-sample *goodness-of-fit* measurement, and as a test of predictive performance of alternative models. To measure the in-sample goodness-of-fit of a model i, we set $a = y_t$ and $b = \hat{y}_t^i$ $(t = 1, 2, ..., T)$, where y_t refers to observed change in nominal exchange rate, while $\hat{y}_t^i$ indicates the **fitted value** for y_t generated by model i. Based on existing empirical evidence, we assume that the change in nominal exchange rate is a stationary continuous random variable with a marginal density $g(y_t)$. An adequate model is expected to produce strong relationship between the actual and the fitted values of exchange rate change series. Therefore, *higher* values of $S\rho_1$ imply well-performing models with higher predictive ability, much as R^2 as a goodness of fit measure for linear models. This metric reduces to a monotonically increasing function of the correlation coefficient for linear Gaussian processes. Nothing is lost by employing this metric when correlation analysis may suffice!

Similarly, when measuring and testing *predictive performance* of model i, we set $a = y_t$ and $b = \hat{y}_t^i$ $(t = R+1, R+2, ..., T)$, where y_t refers to actual out- of-sample change in nominal exchange rate, whereas $\hat{y}_t^i$ indicates the non-recursive **forecast** of y_t generated by model i based on the estimates obtained from the training sample of size R.[2] The *higher* $S\rho_1$ the better the fit and forecast performance of a model.

- **As an Alternative Measure of Density**

We also use the same entropy measure to test the equality of densities for two univariate random variable a and b as suggested in Maasoumi & Racine (2002). We use the nonparametric kernel estimates of the following metric entropy statistics:

$$S\rho_2 = \frac{1}{2} \int (f_1^{1/2} - f_2^{1/2})^2 da \tag{2}$$

but, now $f_1 = f(a)$ and $f_2 = f(b)$ are both marginal densities of the random variables a and b. As before, $a = y_t$ and $b = \hat{y}_t^i$ may indicate either the in-sample fitted values or the non-recursive **forecast** of y_t generated by model i. Specifically random variables y_t and $\hat{y}_t^i$ are identically distributed if and only if the marginal densities are equal. Under the null of equality, $S\rho_2 = 0$. Here the *lower* values of $S\rho_2$ indicate *better* predictive performance, or a poorer fit.

In the following sections, we will first discuss the exchange rate forecasting methodology, then we will briefly summarize the data and the structural models considered in this paper. Finally, we will discuss our estimation results and compare them with the traditional point forecast accuracy criterion.

[2] The detailed information about the structural models and the out-of-sample forecasting methodology will be provided in the following section.

2 Out-of-Sample Exchange Rate Forecasting

For each model, we construct the first one-month ahead forecast with the first estimation window, then repeat the process by rolling the window one period ahead in the sample until all the observations are exhausted. Let s_t denote the natural logarithm of the nominal exchange rate at time t, and the change in the logged nominal exchange rate as $y_{t+1} = s_{t+1} - s_t$. If X_t is a vector of "fundamentals" at time t, a typical parametric model in our set may be represented as follows:

$$y_{t+1} = \alpha + \beta' X_t + \varepsilon_{t+1} \tag{3}$$

In this regression equation (3), the change in the natural logarithm of exchange rates is determined by some fundamentals and unexpected shocks. Under rational expectations, the unobservable expectation of one period ahead exchange rate will be the conditional expectation implied by the structural model, assuming uncorrelated and mean-zero error terms. In rolling regressions with a sample of size N, a sub-sample of size R is designated as the training or estimation sample to produce P forecasts where $N = R + P$. For example, the first R observations may be used to obtain estimates $\widehat{\alpha}$ and $\widehat{\beta}$; then the realized values of economic fundamentals at time $(t+1)$ are employed to produce an out-of-sample forecast as follows:

$$\hat{y}_{t+1} = \widehat{\alpha} + \widehat{\beta}' X_{t+1} \quad . \tag{4}$$

The process is repeated by rolling the window by one period ahead: the first observation is dropped from the sample and $(R+1)^{th}$ observation is added to the sample, leaving the sample size constant. This produces the new estimates of $\widehat{\alpha}$ and $\widehat{\beta}$. Eventually, P number of forecasts are produced for each model to be used for forecast comparison.

2.1 Data and the "Structural" Models

In the selection of countries, data coverage and model selection, we follow the predominant choices in the recent literature. The data is taken from Molodtsova & Papell (2009) who employed monthly data from March 1973 through December 1998 for Euro area countries (France, Germany, Italy, Netherlands, and Portugal), and through June 2006 for the remaining OECD countries (Australia, Canada, Denmark, Japan, Sweden, Switzerland, and the United Kingdom).[3] The US dollar is treated as the "home currency", and exchange rate is defined as the US Dollar price of one unit of foreign currency. An increase in the exchange rate indicates depreciation of the dollar.

The conditioning variables are total income $y_t(y_t^*)$, inflation rate $\Pi_t(\Pi_t^*)$, output gap $y_{gap}(y_{gap}^*)$, the real exchange rate $q_t(q_t^*)$, money supply $m_t(m_t^*)$, price level $p(p^*)$ and short term interest rates $i_t(i_t^*)$.[4] Nominal interest rates are defined in percentages, while all other variables are transformed by taking the natural logarithm multiplied by 100. In the original database the seasonally adjusted industrial production index is used as proxy for total in-

[3] Meese & Rogoff (1983) also uses monthly data in their paper, while some studies such as Engel & West (2005), Cheung et al.(2005) and Gourinchas & Rey (2007) use quarterly data.

[4] Variables in parentheses denote the foreign country counterparts.

come. Inflation rate is measured by the annual growth rate of monthly CPI index p_t. Output gap is measured (by using the quasi-real data) as the percentage deviation of industrial production from its quadratic trending level. Real exchange rate for the foreign country is calculated as $q_t = s_t + p_t^* - p_t$. As for the money supply, for the majority of countries with available data, M1 is used as a proxy for the quantity of money, and M2 for a few.

We estimate nine exchange rate models, Models 1-7 are the structural models extant in the literature, Model 8 is the driftless random walk model, and Model 9 is the random walk with drift. These models are not nested. An "encompassing model" is sometimes a convenient statistical construct, and may be stated as follows:

$$\Delta y_{t+1} = \alpha + \beta_1 \widetilde{y_t}^{gap} + \beta_2 \widetilde{\Pi}_t + \beta_3 q_t + \beta_4 i_{t-1} + \beta_5 i_{t-1}^* + \beta_6 \widetilde{i_t} + \beta_7 m_t^* + \varepsilon_{t+1} \quad (5)$$

For any variable x, we denote by $\widetilde{x}$ the fundamental variable x in the home country (United States), minus the fundamental variable x^* in the foreign country (such that $\widetilde{x} = x - x^*$). $m_t^* = (m_t - m_t^*) - (y_t - y_t^*) - s_t$ refers to the predictor in the monetary model. Each parametric model may be viewed as a restricted form of the this artificial comprehensive form.

Table (8) in the appendix summarizes these models. Model 1 is the Taylor rule model examined in Wang and Wu (2010) as their benchmark. It is asymmetric with no smoothing. Models 2-4 are also Taylor rule models studied in Molodtsova & Papell (2009). Model 2 is the constrained (symmetric) Taylor rule that assumes PPP, Model 3 is the smoothing Taylor rule, and Model 4 is the constrained (symmetric) smoothing Taylor rule model where the lagged value of interest rate is included to control for the potential interest rate smoothing affect. Model 5 is the PPP model with a single variable q_t. Model 6 is the uncovered interest parity model, and Model 7 is the monetary model.[5]

2.2 Evaluating Point Forecasts

The literature does the out-of-sample forecast comparison by comparing the prediction error implied by the structural model with the one implied by the benchmark model; here in our paper, we will use both the driftless random walk (Model 8) and random walk with drift (Model 9) for model comparison. 2-state markov-switching model is commonly used in the literature to control for long periods of appreciations and depreciations in nominal exchange rates. Instead, we follow a strategy which can be characterized as a P-period markov-switching model when the model produces P out-of-sample forecasts. In other words, we define the drift in each estimation window as the mean of the first differences of the actual exchange rates in the training sample.

Following the methodology of Diebold & Mariono (1995) and West (1996), we first evaluate the out-of-sample performance of the models based on the mean-squared prediction error (MSPE) comparison. In this approach, the quadratic loss functions for the structural model i and the benchmark model

[5] See Molodtsova & Papell (2009) and Wang & Wu (2010) for the derivation of the models. A specification search approach to these models may be a worthy topic of research. The appropriate approach in that setting would be the data snooping techniques proposed by White (2000) in which no model may be correctly specified. This realism is an enduring aspect of techniques developed by Hal White. The object of inference in such settings would be the "pseudo parameters" which are afforded a compelling and clear definition based on entropy concepts such as the ones employed in this paper.

b are defined as follows:

$$L(y_t^i) = (y_t - \hat{y}_t^i)^2, \qquad L(y_t^b) = (y_t - \hat{y}_t^b)^2 \tag{6}$$

where y_t is the actual series and $\hat{y}_t^i$ and $\hat{y}_t^b$ are the forecasts obtained from the structural model i and the benchmark model b, respectively. The forecast accuracy testing is based on whether the population mean of the loss differential series d_t is zero where:

$$d_t = L(y_t^b) - L(y_t^i) = (y_t - \hat{y}_t^b)^2 - (y_t - \hat{y}_t^i)^2. \tag{7}$$

In Diebold & Mariono (1995) and West (1996), the null of equal predictive accuracy is:

$$H_0 : E[d_t] = E[L(y_t^b) - L(y_t^i)] = MSPE^b - MSPE^i = 0 \tag{8}$$

Clark & McCracken (2001) show that when comparing nested models, Diebold & Mariono (1995) test statistics will be non-normal and the use of standard critical values results in poorly sized tests. Accordingly, Clark & West (2006) propose a corrected Diebold-Mariano statistic which takes into account the fact that under the null, MSPE of the structural model and the benchmark model are not the same. If the null is true, estimation of the structural model produces a noisy estimate of the parameter, supposed to be zero in population, increasing the MSPE in the sample. They suggest an adjusted MSPE for the alternative model which is adjusted downwards to have equal MSPEs under the null. Accordingly, the loss differential function can be adjusted as follows:

$$d-adj_t = L(y_t^b) - L(y_t^i) - adj = (y_t - \hat{y}_t^b)^2 - (y_t - \hat{y}_t^i)^2 - (\hat{y}_t^b - \hat{y}_t^i)^2. \quad (9)$$

Clark and West (2006) test if the population mean of the adjusted series $d-adj_t$ is zero, based on the following statistic:

$$CW = \frac{\widetilde{d}}{(\widehat{avar}(\widetilde{d}))^{1/2}} = \frac{MSPE^b - MSPE^i_{adj}}{(\widehat{avar}(MSPE^b - MSPE^i_{adj}))^{1/2}} \quad (10)$$

where $\widetilde{d}$ refers to mean of $d-adj_t$, and $MSPE^i_{adj}$ refers to MSPE for the structural model adjusted for the bias. If the CW test statistics is significantly positive, one may conclude that the structural model outperforms the random walk model. Clark & West (2006) suggest standard normal critical values for inference in comparing these nested models.[6]

Note that, these procedures assume an encompassing form that correctly nests the competing models. Misspecifications of functional form and/or omitted variables are not accommodated. We find evidence for both types of misspecification. To see this, consider Table (1) in which the parametric models are subjected to the nonparametric specification test proposed by Hsiao et al.(2007). Note that for each model, this test takes the conditioning variables as given. But cross validation in NP estimation is indeed capable of identifying irrelevant variables (see Table (9) in the appendix on full sample smoothing of the largest model, Model 3, through least-square cross

[6] Rogoff & Stavrakeva (2008) argue that CW test statistics cannot be used to evaluate forecasting performance as it is not testing the null of equal predictive accuracy, hence they suggest to use bootstrapped critical values. There is less evidence in favor of Taylor-rule based models when CW test statistics with bootstrapped critical values are used.

Table 1: Kernel Consistent Model Specification Test Results

Models	1	2	3	4	5	6	7
AUS	0.032**	0.452	0.068*	0.288	0.540	0.136	0.126
CAN	0.566	0.564	0.106	0.202	0.540	0.268	0.044**
DEN	0.362	0.544	0.030**	0.000***	0.536	0.002***	0.548
FRA	0.014**	0.008***	0.002***	0.012**	0.546	0.040**	0.008***
GER	0.810	0.456	0.048**	0.006***	0.596	0.002***	0.010**
ITL	0.074*	0.092*	0.000***	0.000***	0.582	0.020**	0.004***
JPN	0.000***	0.004***	0.000***	0.014**	0.556	0.004***	0.002***
NTH	0.556	0.118	0.002***	0.004***	0.552	0.000***	0.152
POR	0.032**	0.144	0.052*	0.294	0.546	0.526	0.032**
SWE	0.020**	0.022**	0.006***	0.000***	0.626	0.000***	0.046**
SWI	0.202	0.482	0.060*	0.058*	0.542	0.314	0.532
U.K.	0.000***	0.000***	0.022**	0.004***	0.592	0.066*	0.592

*Notes: The table shows the test results for correct specification of parametric regression models as described in Hsiao et al.(2007). The numbers show the p-values at which we reject the null of correctly specified parametric model. The distribution of the test statistics is derived with 500 IID bootstrapped replications. ***, **, and * denote significance at 1 percent, 5 percent and 10 percent, respectively.*

validated bandwidth selection).

As can be seen from Table (1) p-values, most parametric model-currency combinations are rejected. Only Model 5, with a single variable q, is generally parametrically (linearly) well specified! Additional variables appear to have nonlinear effects, to various degrees, for different exchange rates. In relative terms, the Symmetric Taylor rule models with no smoothing (Model 2) do better in sample. Addition of linear interest rate smoothing variables (in Models 3 and 4) tends to be rejected.

A shed further light on the possibility of irrelevant explanatory variables, we examine more closely a nonparametric estimation of the largest model above, Model 3 (Asymmetric Taylor Rule with no smoothing), which contains the majority of explanatory variables in any of the 7 models. Smoothing with cross validation in kernel estimation is known to be able to smooth out irrelevant regressors; see Li & Racine (2007, Chapter 4). Table (9) in the appendix supports the following inferences: Taylor rule variables appear to be insignificant in the full sample, when all the conditioning variables are considered together! Output gap differences, inflation differences and real exchange rates are smoothed out in at least 5 currencies for the full sample (not-shown in the appendix) and in at least half of the currencies for the rolling regressions. While irrelevant variables may not generally induce bias or inconsistency in estimation, they do increase uncertainty, be it through MSPEs (as observed in the CW tests), and as will be seen in our metric entropy examination of the whole forecast distribution. Combined with the parametric tests in Table (1), it would seem that many of these models suffer from parametric misspecifications, as well as inappropriate set of conditioning variables. In this setting one has to seriously question the propriety of the conditional mean forecasting paradigm based on the mean squared error assessments.

With above caveat in mind, in Table (2), we report the CW test statistics and p-values for each model and currency when the benchmark model is the driftless random walk. This is done to provide a benchmark for what can be learned from our broader distribution metrics. For one-month ahead forecasts of exchange rates changes, the following observations are indicated:

(a) Neither OLS nor NP forecasts provide enough evidence in favor of the exchange rate models 1 (Asymmetric Taylor Rule with no smoothing) and 5 (PPP); (b) With NP out-of-sample forecasting, there is some evidence in favor of models for five currencies in Models 2 and 3, and for four currencies in Models 4 and 7 where OLS estimates did not do well; (c) For Model 3, only NP forecasts provide favorable evidence for five currencies (Australian Dollar, British Pound, Deutsche Mark, Dutch Guilder, and Swiss Franc), all of which exhibit specification problems (see Table (1)). For Model 4, with NP forecasts, there is evidence in favor for four currencies, three of which involve specification issues. A similar pattern is present in the rest of the models except Models 1 and 2; (d) Out of 15 currency–model pairs where both OLS and NP forecasts produce favorable results, nine cases have parametric specification problems. This does not support the idea that NP estimation is a panacea for prediction from misspecified models. Out of 20 currency-model pairs where we have evidence in favor of some NP empirical exchange rate models, we reject the null of parametric specification in 12 cases. So, NP forecasts tend to produce better results in favor of the exchange rate models relative to a driftless random walk; (e) Only for the Japanese yen, we see evidence in favor of the random walk with drift against driftless random walk. This suggests a constant growth in that exchange series.

In Table (3) the CW test results are given with the null of a random walk with drift. In CW test of equal predictability of two nested models, under the null the series follows a martingale difference against the alternative that the series is linearly predictable. However, Clark & West (2006) argue that the same asymptotic distribution critical values can be applied even for nonlinear and Markov Switching models. Nikolsko-Rzhevskyy & Prodan (2011) also show that their simulation results imply properly sized CW test in the case of the nonlinear models. Therefore, we use the asymptotic normal critical values. We allow drift to change for every forecast window. In the NP forecasts, we compute the least-squares cross-validated bandwidths for the local linear estimators. For each currency model pair, the first entry shows the CW statistics while the one below it is the p-value. The first columns are parametric (OLS), the second columns are NP values. Our results show that random walk with drift model outperforms the driftless random walk model in 3 out of 12 currencies.

According to the results in Table (3), when the null is a random walk with a drift, parametric and NP forecasts reach the same conclusion (in favor of the empirical models) more than the case where the null is a driftless random walk. One noteworthy finding is that, similar to the results in Table (2), out of 21 currency-model pairs where we have evidence in favor of the NP models, we reject the null of parametric models in 16 cases. So, at times where we have a model specification problem, NP forecasts tend to produce better results in favor of exchange rate models against the null of random walk with drift.

Table 2: P-values for the CW Test under the Null of Driftless Random Walk

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 9
	Lin	NP	Lin	NP	Lin	NP	Lin	NP	Lin	NP	Lin	NP	Lin	NP	Lin
AUS	-0.524	0.619	0.299	**1.653**	1.055	**2.115**	**1.691**	**2.434**	-0.801	0.082	**1.551**	**1.992**	0.199	-0.810	0.317
	0.700	0.268	0.383	**0.050**	0.146	**0.018**	**0.046**	**0.008**	0.788	0.467	**0.061**	**0.024**	0.421	0.791	0.376
CAN	**2.006**	-0.145	**2.122**	**2.501**	**1.295**	-0.964	**1.657**	**1.431**	-0.837	-2.233	**1.886**	**1.306**	**1.898**	**1.364**	0.393
	0.023	0.558	**0.017**	**0.006**	**0.098**	0.832	**0.049**	**0.077**	0.798	0.987	**0.030**	**0.096**	**0.029**	**0.087**	0.347
DEN	0.018	0.203	0.096	-0.515	0.119	0.489	0.589	1.185	-1.034	-0.992	-0.259	**1.285**	0.855	0.247	-0.920
	0.493	0.420	0.462	0.696	0.453	0.312	0.278	0.119	0.849	0.839	0.602	**0.100**	0.197	0.402	0.821
FRA	0.594	0.083	0.133	0.707	**1.568**	1.249	**1.521**	1.073	-0.315	-0.352	-1.194	-0.216	-0.067	**2.005**	-0.758
	0.277	0.467	0.447	0.240	**0.059**	0.107	**0.065**	0.142	0.624	0.638	0.883	0.585	0.527	**0.023**	0.775
GER	-0.317	1.281	0.222	**2.147**	0.168	**2.650**	1.210	**3.370**	-0.150	-0.196	0.659	0.281	-0.487	**2.364**	0.209
	0.624	0.101	0.412	**0.017**	0.433	**0.004**	0.114	**0.000**	0.560	0.578	0.255	0.389	0.687	**0.010**	0.417
ITL	**2.568**	**1.580**	**1.714**	**1.487**	**2.773**	**1.883**	**2.741**	**2.288**	-0.544	-0.817	0.354	0.862	-0.926	0.573	-0.433
	0.006	**0.058**	**0.044**	**0.069**	**0.003**	**0.031**	**0.003**	**0.012**	0.706	0.793	0.362	0.195	0.822	0.284	0.667
JPN	0.860	1.104	0.849	0.653	**2.536**	**2.387**	**2.859**	**2.169**	1.268	-0.437	**2.537**	**2.762**	0.967	**1.677**	**1.346**
	0.195	0.135	0.198	0.257	**0.006**	**0.009**	**0.002**	**0.015**	0.103	0.669	**0.006**	**0.003**	0.167	**0.047**	**0.090**
NTH	-0.915	1.002	-0.496	**1.751**	-0.101	**1.801**	0.460	**2.383**	-0.065	-0.051	0.947	**2.816**	-0.116	0.330	0.018
	0.819	0.159	0.690	**0.041**	0.540	**0.037**	0.323	**0.009**	0.526	0.520	0.172	**0.003**	0.546	0.371	0.493
POR	-0.622	-0.176	0.308	**1.452**	-1.319	0.966	0.389	**1.631**	-2.298	-2.119	0.609	0.679	0.200	-0.169	-1.298
	0.733	0.570	0.379	**0.074**	0.904	0.169	0.349	**0.054**	0.989	0.982	0.272	0.250	0.421	0.567	0.902
SWE	0.539	0.226	0.745	1.173	-0.981	0.256	-0.790	**2.301**	-0.561	-1.811	-1.116	0.644	0.284	0.607	-1.168
	0.295	0.411	0.228	0.121	0.836	0.399	0.785	**0.011**	0.712	0.964	0.867	0.260	0.388	0.272	0.878
SWI	-0.666	**1.344**	-0.053	**1.910**	0.992	**2.125**	**1.342**	**2.579**	-0.782	-1.066	**1.996**	**3.143**	0.893	**2.475**	0.382
	0.747	**0.090**	0.521	**0.029**	0.161	**0.017**	**0.090**	**0.005**	0.783	0.856	**0.024**	**0.001**	0.186	**0.007**	0.351
UK	0.885	1.171	0.751	1.031	0.490	**2.109**	0.370	0.610	-0.081	-0.148	0.403	0.819	-0.282	-0.453	-0.370
	0.188	0.121	0.227	0.152	0.312	**0.018**	0.356	0.271	0.532	0.559	0.344	0.207	0.611	0.674	0.644

Notes: The Table shows the Clark and West (CW) test results (p-values are in the second rows) for the OLS and NP out-of-sample forecasts for alternative exchange rate models under the null of driftless random walk. We use rolling regression method with a window of 120 observations. In the NP forecasts, we compute the least-squares cross-validated bandwidths for the local linear estimators. For each currency model pair, the first data shows the CW test statistics while the second data shows the p-values. CW assumes two models compared are nested, under the null the exchange rate follows a driftless RW. The limiting distribution of the CW under the null is standard normal.

Table 3: P-values for the CW Test under the Null of Random Walk with a Drift

	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**		**Model 6**		**Model 7**		**Model 8**
	Lin	NP	Lin	NP	Lin	NP	Lin	NP	Lin	NP	Lin	NP	Lin	NP	Lin
AUS	-0.527	0.363	-0.014	**1.394**	1.225	**2.078**	**1.684**	**2.446**	-1.631	-0.133	**1.480**	**2.039**	-0.343	-0.900	**1.404**
	0.701	0.359	0.505	**0.082**	0.111	**0.019**	**0.047**	**0.008**	0.948	0.553	**0.070**	**0.021**	0.634	0.815	**0.081**
CAN	**2.407**	0.134	**2.335**	**2.697**	**1.957**	-0.932	**1.939**	**1.453**	-0.281	-1.737	**2.353**	**1.545**	**1.812**	**1.282**	**1.804**
	0.008	0.447	**0.010**	**0.004**	**0.026**	0.824	**0.027**	**0.074**	0.610	0.958	**0.010**	**0.062**	**0.036**	**0.100**	**0.036**
DEN	0.492	0.686	0.892	0.123	0.619	0.943	**1.322**	**1.402**	-0.521	-0.711	-0.278	**2.324**	**1.290**	0.517	**2.518**
	0.312	0.246	0.187	0.451	0.268	0.173	**0.094**	**0.081**	0.699	0.761	0.610	**0.010**	**0.099**	0.303	**0.006**
FRA	**1.770**	1.038	**1.383**	**1.506**	**2.359**	**1.325**	**2.184**	1.145	0.111	0.082	-0.934	0.550	0.420	**2.138**	**2.304**
	0.039	0.150	**0.084**	**0.067**	**0.010**	**0.093**	**0.015**	0.127	0.456	0.468	0.824	0.291	0.338	**0.017**	**0.011**
GER	0.259	1.217	0.349	**2.221**	0.281	**2.876**	1.095	**3.393**	0.543	0.628	0.338	0.215	0.324	**2.356**	1.178
	0.398	0.113	0.364	**0.014**	0.390	**0.002**	0.137	**0.000**	0.294	0.265	0.368	0.415	0.373	**0.010**	0.120
ITL	**3.264**	**1.982**	**2.641**	**1.846**	**3.293**	**2.054**	**3.549**	**2.630**	-0.370	-0.800	1.025	**1.352**	0.821	**1.532**	**2.575**
	0.001	**0.024**	**0.004**	**0.033**	**0.001**	**0.021**	**0.000**	**0.005**	0.644	0.788	0.153	**0.089**	0.207	**0.064**	**0.005**
JPN	-0.033	0.543	-0.388	-0.229	**2.013**	**2.136**	**2.505**	**1.849**	0.630	-0.739	**1.880**	**2.437**	0.201	1.042	0.709
	0.513	0.294	0.651	0.591	**0.023**	**0.017**	**0.006**	**0.033**	0.264	0.770	**0.031**	**0.008**	0.420	0.149	0.239
NTH	-0.236	1.164	-0.767	**1.693**	-0.281	**1.656**	-0.005	**2.465**	0.570	0.653	0.642	**2.872**	0.336	0.548	**1.358**
	0.593	0.123	0.778	**0.046**	0.611	**0.050**	0.502	**0.007**	0.285	0.257	0.261	**0.002**	0.369	0.292	**0.088**
POR	**2.352**	**2.265**	**2.490**	**3.586**	-1.103	0.701	-0.338	1.247	-0.526	-0.383	-0.287	-0.155	-0.133	-0.482	**4.634**
	0.010	**0.012**	**0.007**	**0.000**	0.863	0.243	0.632	0.108	0.700	0.649	0.613	0.561	0.553	0.684	**0.000**
SWE	**1.486**	0.637	**1.489**	1.271	-0.868	0.792	-0.522	**2.685**	-0.269	-1.355	-0.699	0.860	1.081	1.275	**3.246**
	0.069	0.262	**0.069**	0.102	0.807	0.215	0.699	**0.004**	0.606	0.912	0.758	0.195	0.140	0.102	**0.001**
SWI	-0.578	1.129	-0.426	**1.818**	0.758	**1.916**	0.857	**2.516**	0.004	-0.201	**1.537**	**2.905**	0.458	**2.286**	1.040
	0.718	0.130	0.665	**0.035**	0.225	**0.028**	0.196	**0.006**	0.498	0.580	**0.063**	**0.002**	0.324	**0.012**	0.150
UK	**1.291**	1.235	1.144	1.034	0.944	**2.279**	0.948	0.751	0.296	0.044	0.924	**1.350**	0.146	-0.236	**1.391**
	0.099	0.109	0.127	0.151	0.173	**0.012**	0.172	0.227	0.384	0.482	0.178	**0.089**	0.442	0.593	**0.083**

Notes: The Table shows the Clark and West (CW) test results for the OLS and NP out-of-sample forecasts for alternative exchange rate models under the null of random walk with a drift model (we allow drift to change in every forecast window). We use rolling regression method with a window of 120 observations. In the NP forecasts, we compute the least-squares cross-validated bandwidths for the local linear estimators. For each currency model pair, the first row shows the CW test statistics while the second row shows the p-values. The limiting distribution of the CW under the null is standard normal.

3 Constructing and Evaluating Density Forecasts

In estimating $S\rho_1$, we use Guassian kernel density estimates of the marginal density functions for the actual series $g(y_t)$, the in-sample fitted values for each structural model $h(\hat{y}_t^i)(i = 1, 2, .., 7)$, and the bivariate density of the actual and fitted values $f(y_{t,}\hat{y}_t^i)$. Least-squares cross-validation is employed to select the optimal bandwidth. Results are shown in Table (4).[7] As stated earlier, this is a "goodness of fit" indicator of general dependence, comparable to "R^2" assessments of "association" in linear models. The higher the values of the $S\rho_1$ the better is the in sample and out-of-sample performance of a model. The critical levels under the null were generated by bootstrap methods as described in the NP package in R (see Hayfield & Racine (2008)).

Only Models 3 and 4 perform consistently well across different currencies, and only Australian Dollar and Japanese Yen can be consistently predicted well by all models. We do find the fitted values are statistically significantly "related" with the actual series in more than half of the currencies. Specifically, Model 3, Constrained (Asymmetric) smoothing Taylor rule model, is the best performing as it produces significant relation with the actual series for all currencies, having the highest dependence in 8 out of 12 currencies.

To consider the performance of the random walk models with or without drift, we note that these models have unique values for $S\rho_1$. A driftless random walk (RW) model suggests $y_{t+1} = \varepsilon_{t+1}$, predicting a zero change. On the other hand, the RW model with drift, Model 9, is $y_{t+1} = c + \varepsilon_{t+1}$ where

[7] Metric entropy measurements are done in R by using the np package (Hayfield & Racine (2008)).

Table 4: S-rho 1 ($S\rho_1$) Measure of Goodness of Fit:
In-sample NP fits versus the Actual Series

Models	1	2	3	4	5	6	7
AUS	**0.019**	**0.011**	**0.181**	**0.024**	**0.026**	**0.014**	**0.018**
	0.000	**0.016**	**0.000**	**0.000**	**0.000**	**0.006**	**0.094**
CAN	**0.021**	**0.007**	**0.034**	**0.014**	**0.011**	0.011	**0.010**
	0.000	**0.024**	**0.000**	**0.000**	**0.068**	0.116	**0.000**
DEN	**0.015**	0.005	**0.024**	**0.016**	0.010	**0.011**	0.008
	0.000	0.546	**0.000**	**0.000**	0.264	**0.000**	0.572
FRA	**0.016**	**0.018**	**0.034**	**0.016**	0.009	0.012	**0.017**
	0.000	**0.096**	**0.000**	**0.000**	0.182	0.472	**0.000**
GER	**0.019**	**0.018**	**0.042**	**0.018**	0.016	0.012	**0.019**
	0.000	**0.000**	**0.000**	**0.000**	0.386	0.158	**0.000**
ITL	**0.024**	**0.019**	**0.084**	**0.049**	**0.010**	0.016	**0.018**
	0.000	**0.000**	**0.000**	**0.000**	**0.006**	0.280	**0.000**
JPN	**0.010**	**0.007**	**0.034**	**0.018**	**0.011**	**0.021**	**0.047**
	0.000	**0.002**	**0.000**	**0.000**	**0.026**	**0.000**	**0.006**
NTH	0.012	**0.015**	**0.034**	**0.038**	**0.012**	**0.011**	0.011
	0.140	**0.000**	**0.000**	**0.000**	**0.070**	**0.000**	0.448
POR	**0.023**	**0.014**	**0.108**	**0.041**	0.024	0.011	0.035
	0.000	**0.000**	**0.000**	**0.000**	0.268	0.174	0.752
SWE	**0.015**	**0.013**	**0.021**	**0.042**	0.010	**0.007**	**0.000**
	0.000	**0.000**	**0.000**	**0.000**	0.546	**0.000**	**0.000**
SWI	**0.030**	**0.034**	**0.032**	**0.060**	**0.010**	**0.014**	0.010
	0.000	**0.000**	**0.000**	**0.000**	**0.052**	**0.000**	0.520
UK	**0.030**	**0.022**	**0.088**	**0.036**	**0.010**	**0.016**	0.014
	0.000	**0.000**	**0.000**	**0.000**	**0.002**	**0.000**	0.124

Notes: The first row for each country gives the integral version of the nonparametric metric entropy $S\rho_1$ for testing pairwise nonlinear dependence between the densities of the actual series and the NP fits in-sample. The second row shows the p-values generated with 500 bootstrap replications. Under the null, actual series and the in-sample NP fits are independent, hence the integrated value of the dependence matrix $S\rho_1$ (Maasoumi & Racine, 2002) takes the value of zero.

c is a constant. For each period, it will predict c as its forecast. The marginal density functions for RW models are thus degenerate. Accordingly, the bivariate density of the actual values and the forecasts from RW models will be the marginal density function for the actual series. Therefore $S\rho_1 = 0$ for RW models. Consequently, the rejection of the null hypothesis of "independence" with the entropy metric constitutes a rejection of the random walk hypothesis, supporting the inference that the models with statistically significant non-zero $S\rho_1$ perform better than the random walk model. And the findings indicate that, quite a few models for each currency do a better job than the random walk model in terms of our general in-sample goodness of fit criteria, especially so when the models are estimated nonparametrically.

Table (5) shows the out-of-sample predictive performance of alternative models. While there are a few model-currency pairs which suggest some predictability, these results indicate a generally poor in and out-of-sample association between the actual series and the forecast values from these models. Very few of these values are significantly larger than zero. The fact that some of these distributions may have lower second moments, when suggested by the CW type tests, is not comforting, given significant evidence of higher order differences between the series and its forecasts.

In Table (5) where the higher moment effects are considered, we find that the performance of the structural models against the driftless random walk model improves significantly. We find that for those models where CW test has failed to show a better performance, metric entropy $S\rho_1$ values show that, out of 12 countries, structural models outperform the random walk model in three currencies for Models 2, 5, and 6, and in two currencies for Models 3, 4, and 7.

By including higher moment effects, our entropy based nonparametric $S\rho_1$ statistic reveals that the RW with drift does even better against the linear models compared with the assessments based on the traditional second moment tests, see Table (5): for Italy, RW with drift has higher pairwise relation with the actual series than five out of six well-performing models.

Table 5: S-rho 1 ($S\rho_1$) Measure of Predictability:
Out-of-sample NP forecasts versus the Actual Series

Models	1	2	3	4	5	6	7	9
AUS	0.011	0.005	0.008	0.005	0.015	**0.020**	**0.010**	0.016
	0.888	0.906	0.640	0.380	0.650	**0.000**	**0.094**	0.126
CAN	0.007	0.005	**0.008**	0.010	0.002	**0.013**	**0.007**	0.012
	0.342	0.164	**0.036**	0.390	0.258	**0.030**	**0.014**	0.174
DEN	0.008	0.006	0.005	0.003	0.008	0.006	0.004	0.011
	0.110	0.146	0.942	0.882	0.328	0.504	0.652	0.768
FRA	0.005	0.005	**0.005**	0.007	0.006	0.004	0.008	0.018
	0.654	0.632	**0.022**	0.460	0.704	0.916	0.134	0.344
GER	0.005	**0.010**	**0.011**	**0.016**	0.011	0.004	**0.013**	0.008
	0.342	**0.048**	**0.088**	**0.012**	0.978	0.312	**0.006**	0.852
ITL	**0.010**	**0.006**	**0.009**	**0.010**	**0.011**	**0.014**	0.008	**0.012**
	0.026	**0.020**	**0.008**	**0.006**	**0.100**	**0.000**	0.172	**0.090**
JPN	0.009	0.004	0.004	0.008	0.003	**0.007**	0.006	0.017
	0.162	0.892	0.274	0.230	0.172	**0.024**	0.314	0.166
NTH	0.005	0.006	0.006	0.006	0.009	0.007	0.004	0.007
	0.498	0.400	0.736	0.178	0.930	0.340	0.876	0.786
POR	0.007	0.007	0.010	0.007	**0.021**	0.027	0.014	**0.023**
	0.196	0.290	0.764	0.864	**0.018**	0.412	0.664	**0.018**
SWE	0.006	**0.009**	**0.011**	**0.008**	0.011	**0.006**	0.004	0.012
	0.138	**0.010**	**0.034**	**0.016**	0.102	**0.006**	0.910	0.262
SWI	0.008	0.009	0.008	0.003	0.006	**0.011**	0.012	0.014
	0.898	0.356	0.110	0.336	0.882	**0.096**	0.302	0.118
UK	0.004	**0.006**	0.006	0.010	**0.009**	**0.014**	0.002	**0.015**
	0.636	**0.018**	0.366	0.164	**0.020**	**0.014**	0.448	**0.032**

Notes: The first row for each country gives the integral version of the nonparametric metric entropy $S\rho_1$ for testing pairwise nonlinear dependence between the densities of the actual series and the NP out-of-sample forecasts. The second row shows the p-values generated with 500 bootstrap replications. Under the null, actual series and the out-of-sample NP forecasts are independent, hence the integrated value of the dependence matrix $S\rho_1$ (Maasoumi & Racine, 2002) takes the value of zero.

For Portugal, RW with drift produces higher pairwise relation with the actual series than the single well-performing structural model. Finally, for the UK, RW with drift has the highest pairwise relation with the observed series over the three well-performing models.

3.1 Density of Forecasts

Table (6) and Panel-A of Table (7) show the nonparametric estimation results of the metric $S\rho_2$ tests. Large values of this statistic provide evidence against the structural models. The results are rather emphatic with these consistent and powerful entropy tests. These models generally fail to produce forecasts close in distributions to observed series. There is good reason why they do not forecast well, "on average". Broadly speaking Taylor rule based models produce smaller $S\rho_2$ values both in-and out-of-sample. We also calculate the $S\rho_2$ value for the RW with drift for out-of-sample forecasting. Therefore, we can calculate the relative $S\rho_2$ values for model comparison. As shown in Panel-B of Table (7), except for two currency/model pairs, the relative $S\rho_2$ value is less than 1, indicating that structural models produce densities of exchange rate forecasts out-of-sample that are closer than the densities implied by the RW with drift model.

Table 6: S-rho 2 ($S\rho_2$): Metric Entropy Density Equality Test Results for the Fitted (in-sample) Values

Models	1	2	3	4	5	6	7
AUS	0.257*	0.549*	0.023*	0.188*	0.552*	0.453*	0.607*
CAN	0.273*	0.517*	0.182*	0.344*	0.635*	0.571*	0.668*
DEN	0.380*	0.703*	0.241*	0.360*	0.721*	0.551*	0.788*
FRA	0.325*	0.581*	0.177*	0.347*	0.685*	0.714*	0.458*
GER	0.336*	0.353*	0.144*	0.364*	0.671*	0.638*	0.433*
ITL	0.269*	0.273*	0.096*	0.169*	0.780*	0.712*	0.452*
JPN	0.380*	0.451*	0.144*	0.289*	0.712*	0.468*	0.541*
NTH	0.662*	0.397*	0.224*	0.268*	0.667*	0.451*	0.830*
POR	0.192*	0.291*	0.039*	0.185*	0.625*	0.584*	0.627*
SWE	0.408*	0.407*	0.246*	0.128*	0.659*	0.588*	0.632*
SWI	0.253*	0.243*	0.206*	0.138*	0.590*	0.534*	0.788*
UK	0.195*	0.299*	0.082*	0.218*	0.680*	0.533*	0.742*

*Notes:The table shows the consistent univariate density difference metric entropy test statistics for the NP fitted values (in-sample) and the actual series. Under the null of equality of densities, * denote significance at 1 percent. The null distribution is obtained with 500 bootstrap replications.*

4 Conclusion

Whether structural models of exchange rate movements are predictive or not is not well suited to mean squared prediction error criteria, and the underlying series appear to have distributions with significant higher order moments characteristics. Conditioning variables such as have been proposed so far appear to have nonlinear effects, at best, which are more robustly examined with nonparametric estimation. Comparison with forecasts from random walk models is somewhat misleading, as this may indicate "good relative performance" for models that have very poor forecasting ability, as clearly demonstrated with our entropy distance metrics. Our measures are metric and allow a ranking of closeness of forecasts to realized values.

5 References

Berger, David W. and Chaboud, Alain P. and Chernenko, Sergey V. and Howorka, Edward and Jonathan H. Wright (2008). "Order Flow and Exchange Rate Dynamics in Electronic Brokerage System Data," *Journal of International Economics*, Elsevier, vol. 75(1), pages 93-109, May.

Berkowitz, Jeremy (2001). "Testing Density Forecasts with Applications to Risk Management," *Journal of Business and Economic Statistics*, 19, 465-474.

Cheung, Yin-Wong & Chinn, Menzie D. & Antonio G. Pascual (2005). "Empirical Exchange Rate Models of the Nineties: Are any Fit to Survive?" *Journal of International Money and Finance*, Elsevier, vol. 24(7), pages 1150-1175, November.

Chinn, Menzie D. & Michael J. Moore (2011). "Order Flow and the Monetary Model of Exchange Rates: Evidence from a Novel Data Set," Forthcoming in *Journal of Money, Credit and Banking*.

Clark, Todd E. & Michael W. McCracken (2001). "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, Elsevier, vol. 105(1), pages 85-110, November.

Clark, Todd E. & Kenneth D. West (2006). "Using out-of-sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics*, Elsevier, vol. 135(1-2), pages 155-186.

Clements, Michael P. & Jeremy Smith (2000). "Evaluating the Forecast Densities of Linear and Non-Linear Models: Applications to Output Growth and Unemployment," *Journal of Forecasting*, 19, 255-276.

Corradi, Valentina & Norman R. Swanson (2006). "Predictive Density Evaluation", in: *Handbook of Economic Forecasting*, eds. Clive W.J. Granger, Graham Elliot and Allan Timmerman, Elsevier, Amsterdam, pp. 197-284.

Diebold, Francis X & Gunther, Todd A. & Anthony S. Tay (1998). "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, vol. 39(4), pages 863-83, November.

Diebold, Francis X & Roberto S. Mariano (1995). "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, American Statistical Association, vol. 13(3), pages 253-63, July.

Diebold, Francis X & James A. Nason (1990). "Nonparametric Exchange Rate Prediction?" *Journal of International Economics*, Elsevier, vol. 28(3-4), pages 315-332, May.

Engel, Charles & Mark, Nelson C. & Kenneth D. West (2007). "Exchange Rate Models are not as Bad as You Think," NBER Chapters, in: *NBER Macroeconomics Annual 2007*, Volume 22, pages 381-441 National Bureau of Economic Research, Inc.

Engel, Charles & Kenneth D. West (2005). "Exchange Rates and Fundamentals," *Journal of Political Economy*, vol. 113(3), pages 485-517, June.

Evans, M.D.D. & Richard K. Lyons (2002). "Order Flow and Exchange Rate Dynamics," *Journal of Political Economy*, vol. 110(1), 170-180.

Evans, M.D.D. & Richard K. Lyons (2005). "Meese and Rogoff Redux: Micro-Based Exchange Rate Forecasting," *American Economic Review*, vol. 95(2), pages 405-414, May.

Faust, Jon & Rogers, John H. & Jonathan H. Wright (2003). "Exchange Rate Forecasting: The Errors We've Really Made," *Journal of International Economics*, Elsevier, vol. 60(1), pages 35-59, May.

Giannnerini, Simone & Dagum, Estela B. & Esfandiar Maasoumi (2011). "A Powerful Entropy Test for Linearity Against Nonlinearity in Time Series," *Working Paper Series.*

Gourinchas, Pierre-Olivier & Helene Rey (2007). "International Financial Adjustment," *Journal of Political Economy*, vol. 115(4), pages 665-703.

Granger, Clive W. J. & Maasoumi, Esfandiar & Jeff Racine (2004). "A Dependence Metric For Possibly Nonlinear Processes," *Journal of Time Series Analysis*, 25, Issue 5, p. 649-669.

Hayfield, Tristen & Jeffrey S. Racine (2008). "Nonparametric Econometrics: The np Package," *Journal of Statistical Software*, Volume 27 (5).

Hsiao, Cheng & Li, Qi & Jeffrey S. Racine (2007). "A Consistent Model Specification Test with Mixed Discrete and Continuous Data," *Journal of Econometrics*, Elsevier, vol. 140(2), pages 802-826, October.

Li, Qi & Jeffrey S. Racine (2007). *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, ISBN: 0691121613, 768 Pages.

Maasoumi, Esfandiar & Jeff Racine (2002). "Entropy and Predictability of Stock Market Returns," *Journal of Econometrics*, Elsevier, vol. 107(1-2), pages 291-312, March.

Mark, Nelson C. (1995). "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review*, American Economic Association, vol. 85(1), pages 201-18, March.

Meese, Richard A. & Kenneth Rogoff (1983). "Empirical Exchange Rate Models of the Seventies : Do They Fit out of Sample?" *Journal of International Economics*, Elsevier, vol. 14(1-2), pages 3-24, February.

Meese, Richard A & Andrew K. Rose (1991). "An Empirical Assessment of Non-linearities in Models of Exchange Rate Determination," *Review of Economic Studies*, Wiley Blackwell, vol. 58(3), pages 603-19, May.

Molodtsova, Tanya & Nikolsko-Rzhevskyy, Alex & David H. Papell (2008). "Taylor Rules with Real-time Data: A Tale of Two Countries and One Exchange Rate," *Journal of Monetary Economics*, Elsevier, vol. 55(Supplement), pages S63-S79, October.

Molodtsova, Tanya & David H. Papell (2009). "Out-of-sample Exchange Rate Predictability with Taylor Rule Fundamentals," *Journal of International Economics*, Elsevier, vol. 77(2), pages 167-180, April.

Nikolsko-Rzhevskyy, Alex & Ruxandra Prodan (2011). "Markov Switching and Exchange Rate Predictability," Forthcoming in *International Journal of Forecasting.*

Rogoff, Kenneth S. & Vania Stavrakeva (2008). "The Continuing Puzzle of Short Horizon Exchange Rate Forecasting," *NBER Working Papers* 14071, National Bureau of Economic Research, Inc.

Skaug, H. & Dag Tjøstheim (1996). "Testing for Serial Independence Using Measures of Distance Between Densities," in P. Robinson & M. Rosenblatt, eds, *Athens Conference on Applied Probability and Time Series*, Springer Lecture Notes in Statistics, Springer.

Su, Liangjun & Halbert White (2007). "A Consistent Characteristic Function-Based Test for Conditional Dependence," *Journal of Econometrics,* vol. 141(2), pages 807-834, December.

Wang, Jian & Jason J. Wu (2010). "The Taylor Rule and Forecast Intervals for Exchange Rates," Forthcoming in *Journal of Money, Credit and Banking.*

West, Kenneth D. (1996). "Asymptotic Inference about Predictive Ability," *Econometrica*, Econometric Society, vol. 64(5), pages 1067-84, September.

White, Halbert (2000). "A Reality Check for Data Snooping," *Econometrica*, Econometric Society, vol. 68(5), pages 1097-1126, September.

Table 7: S-rho 2 ($S\rho_2$): Metric Entropy Density Equality Test Results for Forecasted (out-of-sample) Values

PANEL-A: Metric Entropy Density Equality Test Statistics (NP Forecasts vs. Actual Series)

Models	1	2	3	4	5	6	7
AUS	0.063*	0.206*	0.025*	0.040*	0.507*	0.279*	0.200*
CAN	0.188*	0.284*	0.086*	0.143*	0.523*	0.338*	0.287*
DEN	0.166*	0.350*	0.080*	0.119*	0.448*	0.425*	0.218*
FRA	0.215*	0.270*	0.103*	0.115*	0.424*	0.335*	0.127*
GER	0.201*	0.220*	0.075*	0.071*	0.506*	0.241*	0.161*
ITL	0.195*	0.202*	0.090*	0.114*	0.477*	0.490*	0.235*
JPN	0.131*	0.151*	0.021*	0.027*	0.575*	0.193*	0.178*
NTH	0.223*	0.206*	0.113*	0.120*	0.483*	0.239*	0.310*
POR	0.163*	0.157*	0.083*	0.082*	0.405*	0.569*	0.257*
SWE	0.147*	0.173*	0.066*	0.078*	0.480*	0.289*	0.309*
SWI	0.057*	0.144*	0.023*	0.046*	0.518*	0.296*	0.193*
UK	0.168*	0.338*	0.078*	0.125*	0.489*	0.327*	0.277*

PANEL-B: Relative $S\rho_2$ (against random walk with a drift null) in out-of-sample Forecasting

Models	1	2	3	4	5	6	7
AUS	**0.097**	**0.318**	**0.039**	**0.062**	**0.784**	**0.431**	**0.308**
CAN	**0.302**	**0.457**	**0.138**	**0.230**	**0.841**	**0.543**	**0.461**
DEN	**0.278**	**0.586**	**0.134**	**0.199**	**0.750**	**0.712**	**0.364**
FRA	**0.394**	**0.495**	**0.189**	**0.211**	**0.777**	**0.614**	**0.233**
GER	**0.312**	**0.342**	**0.116**	**0.110**	**0.786**	**0.374**	**0.251**
ITL	**0.329**	**0.341**	**0.152**	**0.193**	**0.806**	**0.828**	**0.396**
JPN	**0.231**	**0.266**	**0.037**	**0.048**	1.014	**0.340**	**0.314**
NTH	**0.356**	**0.329**	**0.180**	**0.191**	**0.770**	**0.381**	**0.494**
POR	**0.331**	**0.319**	**0.169**	**0.167**	**0.823**	1.157	**0.523**
SWE	**0.236**	**0.277**	**0.106**	**0.125**	**0.769**	**0.463**	**0.495**
SWI	**0.087**	**0.219**	**0.035**	**0.070**	**0.788**	**0.451**	**0.294**
UK	**0.271**	**0.546**	**0.126**	**0.202**	**0.790**	**0.528**	**0.448**

*Notes: Panel-A shows the consistent univariate density difference metric entropy test statistics for the NP forecasted values (out-of-sample) and the actual series. Under the null of equality of densities ($S\rho_2 = 0$), * denote significance at 1 percent. The null distribution is obtained with 500 bootstrap replications. Panel-B shows, in out-of-sample forecasting, the ratio of the integrated value of the metric entropy measure of univariate density differences, Srho-2 ($S\rho_2$), for each model to the ($S\rho_2$) for the benchmark (random walk drift) model. Higher Srho-2 measures imply lower predictive powers. Therefore, a ratio less than 1 implies that structural model out-predicts the null (random walk with a drift) model.*

Table 8: Model Description

	Model Name	List of explanatory variables
Model 1:	Asymmetric Taylor Rule with no smoothing	$(y_{gap} - y^*_{gap}), (\pi - \pi^*), q$ and *a constant*
Model 2:	Symmetric Taylor Rule with no smoothing	$(y_{gap} - y^*_{gap}), (\pi - \pi^*)$, and *a constant*
Model 3:	Asymmetric Taylor Rule with smoothing	$(y_{gap} - y^*_{gap}), (\pi - \pi^*), q, i_{t-1}, i^*_{t-1}$, and *a constant*
Model 4:	Symmetric Taylor Rule with smoothing	$(y_{gap} - y^*_{gap}), (\pi - \pi^*), i_{t-1}, i^*_{t-1}$ and *a constant*
Model 5:	Purchasing Power Parity Model	q and *a constant*
Model 6:	Interest Parity Model	$(i - i*)$ and *a constant*
Model 7:	Monetary Model	$(m - m^*) - (y - y^*) - s$
Model 8:	Driftless Random Walk	*none*
Model 9:	Random Walk with a drift	*a constant*

Variable Definitions

y_{gap}	Quasi-real quadratic trending based measure of the output gap in US.
y^*_{gap}	Quasi-real quadratic trending based measure of the output gap in the foreign country.
π	Inflation rate in US.
π^*	Inflation rate in the foreign country.
p	Price level in US.
p^*	Price level in the foreign country.
q	Real exchange rate in the foreign country. It is calculated as: $q = p - p^* - s$.
i	Monthly nominal interest rate in U.S.
i_{t-1}	One-month lagged monthly interest rate in U.S.
i^*	Monthly nominal interest rate in the foreign country.
i^*_{t-1}	One-month lagged monthly interest rate in the foreign country.
m	Natural logarithm of the money supply in U.S.
m^*	Natural logarithm of the money supply in the foreign country.
y	Natural logarithm of the real GDP in U.S.
y^*	Natural logarithm of the real GDP in the foreign country.
s	Natural logarithm of the dollar price of foreign currency.
Δs	Percentage change in the Natural logarithm of the nominal spot exchange rate.

Notes: In outputgap measurement, we follow Molodtsova & Papell (2009) and calculate the potential output (quadratic trending output) by using the "quasi-real-time" data: we still use ex-post revised output data at time $t-1$ (not the future observations) to calculate the potential output at time t, and for the following each year, we update the sample and calculate potential output at that year accordingly.

Table 9: Cross-validated Bandwidths in out-of-sample Forecasts

		Cross-validated bandwidths in each rolling window			
		Bandwidths			
		median	**10th percentile**	**90th percentile**	**st. dev.**
AUS	$\widetilde{y_t}^{gap}$	35260.74	0.02	218779.24	0.04
	$\widetilde{\Pi}_t$	1.59	0.80	5877020.38	2.54
	q_t	0.04	0.03	9925.49	0.11
	i_{t-1}	1.95	0.66	2508301.63	2.42
	i^*_{t-1}	0.88	0.44	4973632.00	2.74
CAN	$\widetilde{y_t}^{gap}$	0.04	0.02	366035.50	0.03
	$\widetilde{\Pi}_t$	4973061.80	1.21	26125346.44	1.36
	q_t	0.28	0.02	703888.32	0.08
	i_{t-1}	228159.48	0.54	17470521.79	2.28
	i^*_{t-1}	185809.19	1.36	310479024.80	2.58
DEN	$\widetilde{y_t}^{gap}$	241385.39	0.11	1201050.51	0.06
	$\widetilde{\Pi}_t$	0.91	0.45	8097023.70	1.53
	q_t	101566.07	0.08	1946581.91	0.15
	i_{t-1}	4.42	0.73	25549599.61	2.42
	i^*_{t-1}	4.71	1.62	101520853.43	2.88
FRA	$\widetilde{y_t}^{gap}$	0.03	0.02	271520.01	0.04
	$\widetilde{\Pi}_t$	228337.87	0.74	11918011.74	1.86
	q_t	215631.05	0.25	1050485.47	0.16
	i_{t-1}	2897456.10	1.12	46750812.45	2.84
	i^*_{t-1}	12.50	0.82	31944044.52	2.50
GER	$\widetilde{y_t}^{gap}$	82145.73	0.03	410423.70	0.04
	$\widetilde{\Pi}_t$	14.81	0.98	9947322.06	1.99
	q_t	64493.60	0.07	1434618.94	0.16
	i_{t-1}	2.55	0.83	11222742.85	2.84
	i^*_{t-1}	0.96	0.80	3867907.12	2.34
ITL	$\widetilde{y_t}^{gap}$	39073.22	0.02	548697.67	0.04
	$\widetilde{\Pi}_t$	890113.88	0.98	13898521.08	2.49
	q_t	0.26	0.03	926787.13	0.15
	i_{t-1}	2.79	1.07	17138191.42	2.84
	i^*_{t-1}	1.93	0.67	22555640.87	2.92
JPN	$\widetilde{y_t}^{gap}$	0.03	0.02	148294.40	0.05
	$\widetilde{\Pi}_t$	0.62	0.41	2.09	1.68
	q_t	23719.70	0.12	853226.44	0.15
	i_{t-1}	1.61	0.67	1295104.14	2.42
	i^*_{t-1}	1.01	0.64	1.76	1.81
NTH	$\widetilde{y_t}^{gap}$	0.04	0.02	22294.20	0.04
	$\widetilde{\Pi}_t$	1657121.24	2.41	12788187.32	2.06
	q_t	148929.61	0.24	1138899.70	0.16
	i_{t-1}	10.29	0.64	20898299.67	2.84
	i^*_{t-1}	2.48	1.06	7561953.49	2.32
POR	$\widetilde{y_t}^{gap}$	150939.94	0.07	632244.74	0.06
	$\widetilde{\Pi}_t$	1.97	1.49	3485564.84	3.46
	q_t	218204.39	0.23	1160393.73	0.15
	i_{t-1}	0.53	0.43	0.68	2.04
	i^*_{t-1}	8548160.95	29.25	63600482.45	3.50
SWE	$\widetilde{y_t}^{gap}$	0.04	0.02	120508.07	0.05
	$\widetilde{\Pi}_t$	348965.98	0.91	13828866.69	2.34
	q_t	37012.03	0.19	1427051.78	0.20
	i_{t-1}	6.13	0.85	12613983.53	3.18
	i^*_{t-1}	7.14	1.72	9869856.03	3.09
SWI	$\widetilde{y_t}^{gap}$	0.04	0.02	228780.81	0.04
	$\widetilde{\Pi}_t$	652660.54	0.40	18467288.02	1.55
	q_t	338518.86	0.17	1588662.72	0.15
	i_{t-1}	2.00	0.39	8179778.01	2.23
	i^*_{t-1}	3.95	0.70	22462155.95	2.00
UK	$\widetilde{y_t}^{gap}$	16759.44	0.02	297811.95	0.03
	$\widetilde{\Pi}_t$	2411331.57	1.61	13253603.52	1.76
	q_t	29224.76	0.07	818628.17	0.12
	i_{t-1}	2.22	0.67	9449292.76	2.42
	i^*_{t-1}	1.72	0.60	7189358.43	2.77

Notes: The table shows the summary statistics of the least squared cross-validated bandwidth selections for each explanatory variable of Model 3 (the largest model) in each rolling window. The last column shows the standard deviation of the regressor.

Table 10: S-rho 1 ($S\rho_1$) Measure of Goodness of Fit: In-sample OLS fits versus the Actual Series

Models	1	2	3	4	5	6	7
AUS	0.016	0.017	0.011	0.010	0.026	0.013	0.025
	0.000	0.020	0.002	0.008	0.000	0.002	0.016
CAN	0.008	0.007	0.008	0.009	0.011	0.011	0.019
	0.014	0.024	0.000	0.000	0.068	0.084	0.002
DEN	0.007	0.005	0.007	0.008	0.010	0.014	0.008
	0.102	0.546	0.000	0.000	0.264	0.008	0.572
FRA	0.010	0.011	0.013	0.018	0.009	0.015	0.010
	0.066	0.302	0.012	0.002	0.182	0.266	0.332
GER	0.008	0.008	0.007	0.010	0.016	0.010	0.014
	0.070	0.022	0.002	0.000	0.386	0.008	0.004
ITL	0.010	0.015	0.010	0.012	0.010	0.006	0.021
	0.002	0.014	0.002	0.046	0.006	0.334	0.000
JPN	0.009	0.010	0.014	0.010	0.011	0.018	0.012
	0.034	0.636	0.000	0.000	0.026	0.000	0.068
NTH	0.012	0.009	0.008	0.008	0.012	0.013	0.009
	0.142	0.132	0.050	0.090	0.070	0.000	0.660
POR	0.009	0.010	0.012	0.013	0.024	0.011	0.037
	0.002	0.004	0.058	0.206	0.268	0.174	0.366
SWE	0.010	0.006	0.004	0.004	0.010	0.004	0.014
	0.058	0.142	0.000	0.000	0.546	0.000	0.032
SWI	0.009	0.008	0.009	0.015	0.010	0.016	0.010
	0.062	0.364	0.014	0.000	0.052	0.034	0.520
UK	0.014	0.012	0.007	0.010	0.010	0.013	0.013
	0.006	0.008	0.002	0.000	0.002	0.000	0.120

Notes: The first row for each country gives the integral version of the non-parametric metric entropy $S\rho_1$ for testing pairwise nonlinear dependence between the densities of the actual series and the OLS fits in-sample. The second row shows the p-values generated with 500 bootstrap replications. Under the null, actual series and the in-sample OLS fits are independent, hence the integrated value of the dependence matrix $S\rho_1$ (Maasoumi & Racine, 2002) takes the value of zero.

Table 11: S-rho 1 ($S\rho_1$) Measure of Predictability:
Out-of-sample OLS forecasts versus the Actual Series

Models	1	2	3	4	5	6	7	9
AUS	0.009	0.019	0.007	**0.012**	0.011	**0.013**	**0.014**	0.016
	0.256	0.800	0.262	**0.098**	0.336	**0.016**	**0.014**	0.126
CAN	**0.011**	0.012	**0.009**	**0.009**	0.008	**0.014**	**0.011**	0.012
	0.026	0.108	**0.080**	**0.076**	0.346	**0.024**	**0.030**	0.174
DEN	0.005	0.009	0.005	**0.005**	0.009	0.009	**0.007**	0.011
	0.696	0.476	0.126	**0.062**	0.446	0.446	**0.072**	0.768
FRA	0.005	0.011	0.008	**0.010**	0.015	**0.015**	0.007	0.018
	0.478	0.174	0.338	**0.086**	0.830	**0.008**	0.832	0.344
GER	0.006	0.008	0.007	0.007	0.012	0.011	0.008	0.008
	0.506	0.360	0.140	0.120	0.986	0.226	0.146	0.852
ITL	**0.011**	**0.014**	**0.010**	**0.012**	**0.011**	**0.031**	**0.012**	**0.012**
	0.030	**0.070**	**0.000**	**0.000**	**0.024**	**0.000**	**0.106**	**0.090**
JPN	0.007	0.010	**0.010**	**0.011**	**0.011**	**0.012**	**0.011**	0.017
	0.346	0.114	**0.018**	**0.006**	**0.004**	**0.000**	**0.018**	0.166
NTH	0.008	0.016	0.005	0.010	0.011	**0.013**	0.010	0.007
	0.258	0.534	0.434	0.222	0.970	**0.006**	0.230	0.786
POR	**0.014**	**0.015**	0.016	0.022	**0.020**	0.024	0.023	**0.023**
	0.100	**0.012**	0.264	0.350	**0.008**	0.278	0.428	**0.018**
SWE	**0.011**	**0.007**	**0.008**	**0.009**	0.015	0.003	0.013	0.012
	0.022	**0.054**	**0.020**	**0.024**	0.138	0.184	0.134	0.262
SWI	0.007	0.007	**0.007**	0.014	0.015	0.013	0.012	0.014
	0.848	0.920	**0.080**	0.106	0.802	0.120	0.750	0.118
UK	0.007	0.008	**0.008**	0.015	**0.012**	**0.013**	**0.017**	**0.015**
	0.150	0.362	**0.030**	0.150	**0.000**	**0.002**	**0.010**	**0.032**

Notes: The first row for each country gives the integral version of the nonparametric metric entropy $S\rho_1$ for testing pairwise nonlinear dependence between the densities of the actual series and the OLS out-of-sample forecasts. The second row shows the p-values generated with 500 bootstrap replications. Under the null, actual series and the out-of-sample OLS forecasts are independent, hence the integrated value of the dependence matrix $S\rho_1$ (Maasoumi & Racine, 2002) takes the value of zero.

Table 12: S-rho 2 ($S\rho_2$): Metric Entropy Density Equality Test Results for the OLS Fitted (in-sample) and Forecasted (out-of-sample) Values

PANEL-A: S-rho 2 ($S\rho_2$): Metric Entropy Density Equality Test Results for the OLS Fitted (in-sample) Values

Models	1	2	3	4	5	6	7
AUS	0.464*	0.547*	0.374*	0.377*	0.552*	0.570*	0.622*
CAN	0.507*	0.517*	0.420*	0.445*	0.635*	0.497*	0.560*
DEN	0.689*	0.703*	0.437*	0.464*	0.721*	0.602*	0.788*
FRA	0.479*	0.602*	0.441*	0.498*	0.685*	0.649*	0.960*
GER	0.579*	0.577*	0.420*	0.466*	0.671*	0.686*	0.665*
ITL	0.440*	0.559*	0.365*	0.510*	0.780*	0.718*	0.622*
JPN	0.563*	0.649*	0.341*	0.386*	0.712*	0.436*	0.706*
NTH	0.662*	0.786*	0.469*	0.569*	0.667*	0.546*	0.930*
POR	0.353*	0.375*	0.396*	0.483*	0.625*	0.584*	0.622*
SWE	0.652*	0.718*	0.399*	0.396*	0.659*	0.741*	0.853*
SWI	0.580*	0.674*	0.431*	0.506*	0.590*	0.527*	0.788*
UK	0.510*	0.553*	0.390*	0.443*	0.680*	0.537*	0.742*

PANEL-B: S-rho 2 ($S\rho_2$): Metric Entropy Density Equality Test Results for the OLS Forecasted (out-of-sample) Values

Models	1	2	3	4	5	6	7	9
AUS	0.320*	0.448*	0.230*	0.278*	0.550*	0.358*	0.487*	0.647*
CAN	0.355*	0.392*	0.314*	0.284*	0.553*	0.476*	0.391*	0.622*
DEN	0.339*	0.439*	0.305*	0.334*	0.484*	0.485*	0.338*	0.597*
FRA	0.335*	0.439*	0.328*	0.299*	0.441*	0.526*	0.450*	0.546*
GER	0.325*	0.426*	0.316*	0.293*	0.527*	0.399*	0.320*	0.644*
ITL	0.327*	0.306*	0.245*	0.272*	0.479*	0.426*	0.349*	0.592*
JPN	0.282*	0.276*	0.220*	0.168*	0.611*	0.339*	0.272*	0.567*
NTH	0.411*	0.454*	0.332*	0.305*	0.492*	0.331*	0.501*	0.627*
POR	0.339*	0.328*	0.438*	0.375*	0.397*	0.571*	0.500*	0.492*
SWE	0.334*	0.309*	0.279*	0.275*	0.496*	0.418*	0.421*	0.624*
SWI	0.364*	0.412*	0.268*	0.235*	0.544*	0.419*	0.332*	0.657*
UK	0.264*	0.376*	0.221*	0.275*	0.493*	0.370*	0.432*	0.619*

*Notes:Panel-A shows the consistent univariate density difference metric entropy test statistics for the OLS fitted values (in-sample) and the actual series. Panel-B shows the consistent univariate density difference metric entropy test statistics for the OLS forecasted values (out-of-sample) and the actual series. Under the null of equality of densities ($S\rho_2 = 0$), * denote significance at 1 percent. The null distribution is obtained with 500 bootstrap replications.*

Table 13: Relative $S\rho_1$ and $S\rho_2$ values (against random walk with a drift null) in out-of-sample Forecasting

PANEL-A: Relative $S\rho_1$ (against random walk with a drift null) in out-of-sample Forecasting

	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**		**Model 6**		**Model 7**	
	Lin	**NP**	**Lin**	**NP**	**Lin**	**NP**	**Lin**	**NP**	**Lin**	**NP**	**Lin**	**NP**	**Lin**	**NP**
AUS	0.563	0.688	**1.188**	0.313	0.438	0.500	0.750	0.313	0.688	0.938	0.813	**1.250**	0.851	0.642
CAN	0.917	0.583	**1.000**	0.417	0.750	0.667	0.750	0.833	0.667	0.167	**1.167**	**1.083**	0.892	0.598
DEN	0.455	0.727	0.818	0.545	0.455	0.455	0.455	0.273	0.818	0.727	0.818	0.545	0.674	0.389
FRA	0.278	0.278	0.611	0.278	0.444	0.278	0.556	0.389	0.833	0.333	0.833	0.222	0.396	0.447
GER	0.750	0.625	**1.000**	**1.250**	0.875	**1.375**	0.875	**2.000**	**1.500**	**1.375**	**1.375**	0.500	0.985	**1.651**
ITL	0.917	0.833	**1.167**	0.500	0.833	0.750	**1.000**	0.833	0.917	0.917	**2.583**	**1.167**	0.997	0.626
JPN	0.412	0.529	0.588	0.235	0.588	0.235	0.647	0.471	0.647	0.176	0.706	0.412	0.668	0.345
NTH	**1.143**	0.714	**2.286**	0.857	0.714	0.857	**1.429**	0.857	**1.571**	**1.286**	**1.857**	**1.000**	**1.435**	0.590
POR	0.609	0.304	0.652	0.304	0.696	0.435	0.957	0.304	0.870	0.913	**1.043**	**1.174**	**1.020**	0.594
SWE	0.917	0.500	0.583	0.750	0.667	0.917	0.750	0.667	**1.250**	0.917	0.250	0.500	**1.109**	0.315
SWI	0.500	0.571	0.500	0.643	0.500	0.571	**1.000**	0.214	**1.071**	0.429	0.929	0.786	0.877	0.859
UK	0.467	0.267	0.533	0.400	0.533	0.400	**1.000**	0.667	0.800	0.600	0.867	0.933	**1.161**	0.128

PANEL-B: Relative $S\rho_2$ (against random walk with a drift null) in out-of-sample Forecasting

	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**		**Model 6**		**Model 7**	
	Lin	**NP**	**Lin**	**NP**	**Lin**	**NP**	**Lin**	**NP**	**Lin**	**NP**	**Lin**	**NP**	**Lin**	**NP**
AUS	**0.495**	**0.097**	**0.692**	**0.318**	**0.355**	**0.039**	**0.430**	**0.062**	**0.850**	**0.784**	**0.553**	**0.431**	**0.753**	**0.308**
CAN	**0.571**	**0.302**	**0.630**	**0.457**	**0.505**	**0.138**	**0.457**	**0.230**	**0.889**	**0.841**	**0.765**	**0.543**	**0.629**	**0.461**
DEN	**0.568**	**0.278**	**0.735**	**0.586**	**0.511**	**0.134**	**0.559**	**0.199**	**0.811**	**0.750**	**0.812**	**0.712**	**0.567**	**0.364**
FRA	**0.614**	**0.394**	**0.804**	**0.495**	**0.601**	**0.189**	**0.548**	**0.211**	**0.808**	**0.777**	**0.963**	**0.614**	**0.824**	**0.233**
GER	**0.505**	**0.312**	**0.661**	**0.342**	**0.491**	**0.116**	**0.455**	**0.110**	**0.818**	**0.786**	**0.620**	**0.374**	**0.497**	**0.251**
ITL	**0.552**	**0.329**	**0.517**	**0.341**	**0.414**	**0.152**	**0.459**	**0.193**	**0.809**	**0.806**	**0.720**	**0.828**	**0.590**	**0.396**
JPN	**0.497**	**0.231**	**0.487**	**0.266**	**0.388**	**0.037**	**0.296**	**0.048**	1.078	1.014	**0.598**	**0.340**	**0.480**	**0.314**
NTH	**0.656**	**0.356**	**0.724**	**0.329**	**0.530**	**0.180**	**0.486**	**0.191**	**0.785**	**0.770**	**0.528**	**0.381**	**0.800**	**0.494**
POR	**0.689**	**0.331**	**0.667**	**0.319**	**0.890**	**0.169**	**0.762**	**0.167**	**0.807**	**0.823**	1.161	1.157	1.017	**0.523**
SWE	**0.535**	**0.236**	**0.495**	**0.277**	**0.447**	**0.106**	**0.441**	**0.125**	**0.795**	**0.769**	**0.670**	**0.463**	**0.674**	**0.495**
SWI	**0.554**	**0.087**	**0.627**	**0.219**	**0.408**	**0.035**	**0.358**	**0.070**	**0.828**	**0.788**	**0.638**	**0.451**	**0.506**	**0.294**
UK	**0.426**	**0.271**	**0.607**	**0.546**	**0.357**	**0.126**	**0.444**	**0.202**	**0.796**	**0.790**	**0.598**	**0.528**	**0.698**	**0.448**

Notes: Panel-A shows, in out-of-sample forecasting, the ratio of the integrated value of the dependence metric for each model (estimated in OLS and NP, respectively) to the S-rho for the random walk drift model. Higher Srho-1 values imply better forecast. Hence, a ratio higher than 1 implies that structural model out-predicts the null of random walk with a drift. Benchmark is Random walk with drift. Since S-rho is a metric, the relative Srho shows the relative performance of the structural models against random walk drift model. The numbers in bold indicate the well-performing structural models against the random walk with drift.
Panel-B shows, in out-of-sample forecasting, the ratio of the integrated value of the metric entropy measure of univariate density differences, Srho-2 ($S\rho_2$), for each model to the ($S\rho_2$) for the benchmark (random walk drift) model. Higher Srho-2 measures imply lower predictive powers. Therefore, a ratio less than 1 implies that structural model out-predicts the null of random walk with a drift. The numbers in bold indicate the well-performing structural models against the random walk with drift.