Contemporary Research Analysis Journal

e-ISSN: 3050-5909
p-ISSN: 3050-5895

Open Access

Volume 03 Issue 11 November 2026
CrossRef DOI: 10.55677/CRAJ/07-2026-Vol03I11

Page no: 901-917

Random Walks, Drift, And Machine Learning: Forecasting the Philippine Peso- US Dollar Exchange Rate

Vicente Montano
University of Mindanao

ABSTRACT: Predicting the Philippine peso-US dollar exchange rate is important for monetary policy and hedging strategies. However, exchange rates are difficult to forecast due to the complexity of their time-series dynamics. This research evaluates twelve forecasting models based on the random walk principle, which include the pure random walk model, the random walk model with drift, the state-space random walk with drift using the Kalman filter method, as well as combinations of other models including autoregressive moving average, conditional volatility, neural network, support vector regression, random forest, gradient boosting, and long short-term memory. Reference rates from 1st January 2016 to 31st October 2026 were used to construct stationary log returns, which were then assessed for stationarity, unit roots, autocorrelation, heteroscedasticity, and non-linearity. None of the models outperformed the random walk in forecasting 1 to 10 days ahead. Over the 20-day horizon, random forest and gradient boosting reduced squared error by about 11 percent, and the random walk with drift by 7.6 percent. Moreover, the latter two models improved their forecasts in both halves of the sample. The Kalman filter identified a constant drift of roughly 2.8 percent per annum and yielded the best-calibrated monthly bands. These results are in line with the near-martingale behavior of exchange rates, overlaid on a gradual, fundamentals-driven depreciation trend, with machine learning gaining an advantage when the trend continues.

KEYWORDS: exchange rate forecasting, random walk with drift, Kalman filter, machine learning, Philippine peso, model confidence set

Cite the Article: *Montano, V. (2026). Random Walks, Drift, And Machine Learning: Forecasting the Philippine Peso- US Dollar Exchange Rate. Contemporary Research Analysis Journal, 3(11), 901-917.https://doi.org/10.55677/CRAJ/07-2026-Vol03I11*

License: This is an open-access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Publication Date: November 21, 2026

**Corresponding Author: Vicente Montano*

1. INTRODUCTION

The lowest exchange rate of the Philippine peso against the US dollar was recorded in October 2026, with the Philippine currency trading below P63 per dollar on 15 October, when inflation was reported at 7.2% (Villanueva, 2026), and below P64 per dollar on 27 October (Tesorero, 2026). Such fluctuations in the exchange rate are important for a small open economy. This is because they affect the cost of important fuel and food, as well as the cost of dollar debt and the volume of dollar inflows in the form of remittances. A forecast of the exchange rate is needed in such an environment.

The exchange rate is one of the most difficult economic variables to predict. Indeed, as pointed out by Meese and Rogoff (1983), structural models relying on monetary variables, output, and interest rates failed to outperform a naïve random walk out of sample, and four decades of academic efforts have only confirmed this finding (Cheung et al., 2005; Rossi, 2013). The reason such a benchmark is very difficult to beat lies in the exchange rate variable, which behaves much like an asset price and changes almost exclusively in response to unpredictable news. However, daily exchange rate changes are far from being noise. They exhibit volatile behavior, fat-tailed distributions, and non-linear dynamics, so that volatility, regimes, and interactions among previous changes provide predictive power over some forecasting horizons (Engle, 1982; Kilian & Taylor, 2003).

Machine learning approaches have reignited the appeal of these models. In particular, tree ensemble models, support vector regression, and neural network models do not require a pre-defined functional form to approximate nonlinearities, and hybrid models use a linear or a volatility model together with a neural network that captures information not captured by the first model (Kim & Won, 2018; Zhang, 2003). There are conflicting results regarding machine learning models in currency markets, but one can state that, in some cases, they perform better than the random walk, with small and inconsistent improvements (Amat et al., 2018; Rossi, 2013). Moreover, there is often an absence of the simple alternative, i.e., random walk with drift.

This study attempts to fill these gaps for the peso-dollar exchange rate. It assesses twelve models, all based on the random walk: pure random walk; random walk with drift; state space random walk with drift using a Kalman filter; and nine hybrids, which include autoregressive moving average (ARMA); generalized autoregressive conditional heteroscedasticity (GARCH); artificial neural network (ANN); support vector regression (SVR); random forest (RF); extreme gradient boosting (XGBoost); and long short-term memory (LSTM), as well as two-stage ARIMA-LSTM and GARCH-LSTM hybrids. All the models predict the stationary log return and use the random walk to convert the prediction into the level; hence, all the differences in performance between the models are clearly identified. The models are tested against a held-out sample year using point, directional, and interval measures, and tests for equal predictive ability are conducted. Finally, the models' interpretations are carried out against the background of exchange rate theories.

This study attempted to answer three research questions. First, is there a model that relies on random walks and gives a better forecast of the peso/dollar exchange rate than the random walk alone, up to 1 day and 1 month in advance? Second, does drift improve forecasts, whether this drift is constant or learned by the Kalman filter, and how does this approach perform compared to the machine learning hybrids? Third, what theories are useful to describe the working of the best-performing models? The project structure is described below.

1.1 Theoretical Framework

1.1 Efficient markets and the random walk

The efficient markets hypothesis states that asset prices are consistent with available information and, hence, price changes occur in response to news and cannot be forecast from previous prices (Fama, 1970). For the case of foreign exchange, the efficient market hypothesis implies that the expected change in the log of the exchange rate is either close to zero or close to a small risk premium. In contrast, the point forecast of the exchange rate tomorrow is the current exchange rate. This position is further structured by Engel & West (2005). In general equilibrium models, the exchange rate is the discounted present value of expected future fundamentals such as money supply, output, prices, and interest rates. If the fundamentals are unit root processes and the discount factor is close to unity, the exchange rate is almost a random walk, as it is determined entirely by fundamentals.

1.1.2 Fundamentals, parity conditions and drift

If the fundamentals themselves wander off, Engel-West's result would be an indication of a random walk with drift, which in turn is associated with two parity conditions. The relative purchasing power parity states that, over a long period of time, a currency tends to depreciate in proportion to the difference between the domestic and foreign inflation rates. Even though there are substantial deviations and lags in achieving parity, it is likely to happen in the long run (Rogoff, 1996). The uncovered interest parity suggests that the expected depreciation of the currency equals the interest differential in its favor. This condition is not valid in the short term, but it reflects the expected movement of high-interest currencies (Fama, 1984). In the case of the Philippines, where policy interest rates have been relatively higher than in the US, and inflation has also been relatively higher, both conditions suggest depreciation of the peso.

State space formulation allows testing of the above hypothesis. The local linear trend model assumes that the log exchange rate is a random walk with drift, with the drift term as a state variable that changes over time (Harvey, 1989). The Kalman filter recursively computes both the level and the drift using the data (Kalman, 1960). If the variance of the innovations in the drift is zero, it suggests a constant drift under the parity theory hypothesis; otherwise, the drift changes with changes in fundamentals/regime.

1.1.3 Volatility clustering

There is evidence that exchange rate returns exhibit volatility clustering, in which large changes are followed by large changes. Mixture-of-distribution theory states that volatility clustering in exchange rates is due to irregular news flows, such that the variance of returns is a function of the amount of news per period (Clark, 1973). ARCH and GARCH models explain this kind of dynamic as conditional variance that is a function of past shocks and variances (Bolllerslev, 1986; Engle, 1982). Volatility clustering is important primarily for forecast intervals. However, volatility itself is informative about future movements of emerging market currencies during periods of global risk aversion, when volatility coincides with a flight to the dollar (Brunnermeier et al., 2008).

1.1.4 Heterogeneous agents, managed floating and nonlinear dynamics

Various models predict nonlinear and state-contingent behavior in the exchange rate that the linear model fails to capture. In heterogeneous agent models, chartist agents extend past trends, whereas fundamentalists believe in mean reversion towards the equilibrium. The weightage of these two groups varies with the market's past profitability, allowing trends to continue temporarily before reversing (De Grauwe & Grimaldi, 2006; Frankel & Froot, 1990). The currency momentum effect, in which past winners continue to win, also corroborates this theory (Menkhoff et al., 2012). Higher transaction costs and bounded arbitrage suggest that mean reversion towards the equilibrium occurs faster if the exchange rate deviates significantly from the equilibrium (Kilian & Taylor, 2003). Central banks in emerging markets tend to lean against sudden movements rather than let their currencies float freely (Calvo & Reinhardt, 2002). This results in smoothing daily fluctuations and turning shocks into gradual changes over time. Adaptive markets theory suggests that predictability is episodic, i.e., certain patterns occur only if the market conditions permit them (Lo, 2004).

1.2 Forecasting evidence and research gap

Out-of-sample research on exchange rates has shown that the random walk is difficult to beat over short horizons (Cheung et al., 2005; Meese & Rogoff, 1983; Rossi, 2013). Benefits from fundamentals and nonlinearity are likely insignificant, specific to particular currencies, samples, and horizons, and dependent on the estimation period. Research on machine learning techniques for currency markets provides similar results: a simple learner can extract some patterns, but the gains from the random walk are insignificant and unstable (Amat et al., 2018). ARIMA-neural and GARCH-LSTM hybrid models are successful in predicting financial series, although their usefulness for forecasting exchange rates depends on the presence of patterns in the residuals of linear or volatility components (Kim & Won, 2018; Zhang, 2003).

Three gaps inspire the current research. First, empirical comparisons between the random walk with drift model and its state-space analog with machine learning ensembles are quite uncommon, even though the drift term is the most straightforward consequence of the parity relationship. Second, there is a lack of studies that test the peso-dollar exchange rate using procedures developed for nested models and multiple comparisons, and that measure the accuracy of the intervals. Third, the factors underlying the success of a specific model are unrelated to exchange rate theory.

2. MATERIALS AND METHODS

2.1 Data

The data set used in the experiment is the daily peso-US dollar exchange rate from 4 January 2016 through 30 October 2026, a total of 2,773 trading days. This rate is derived from the European Central Bank's daily euro reference rates by taking the quotient of the peso per euro and the US dollar per euro (European Central Bank, 2026). These reference rates are updated only once a day at 14:15 CET, providing a consistent daily series. They might differ a bit from the last closing rate published by the Bangko Sentral ng Pilipinas. The sample consist of three periods: the training period (from 5 January 2016 to 13 November 2024; 2,272 returns), the one-year validation period that was used for tuning the model parameters only (14 November 2024 to 6 November 2025; 250 returns), and the one-year test period that was used for the testing of the model only (7 November 2025 to 30 October 2026; 250 returns). The peso changed by 0.3% in the validation year and depreciated by 7.9% in the test year.

The analysis used the daily log return,

$$r_t = 100 \times (\ln S_t - \ln S_{t-1}) \quad (1)$$

Where S_t is the number of pesos per US dollar, so that a positive return is a depreciation of the peso, each model forecasts the cumulative return over the next h trading days and converts it into a level forecast through the random walk.

$$R_{t,h} = \Sigma_{k=1}^{h} r_{t+k}, \qquad \hat{S}_{t+h} = S_t \, exp(R_{t,h} / 100) \quad (2)$$

2.2. Stationarity and diagnostic tests

The level, the log level, the first difference, and the log return series were tested for the existence of unit roots using the Augmented Dickey-Fuller (ADF), DF-GLS, and the Phillips-Perron unit root tests; for stationarity using the KPSS test; and for a unit root in the presence of one endogenous structural break using the Zivot-Andrews test (Dickey & Fuller, 1979; Elliott et al., 1996; Kwiatkowski et al., 1992; Philipps & Perron. 1988). The tests were carried out both in the presence of a constant and in the presence of a constant and a trend. Subsequently, the returns were characterized and tested for normality (Jarque-Bera test), serial correlation (Ljung-Box Q test on returns), conditional heteroscedasticity (Ljung-Box Q test on squared returns and ARCH-LM test), nonlinear dependence (the BDS test) and the random walk hypothesis (heteroscedasticity-robust variance-ratio test) (Brock et al., 1996; Engle, 1982; Jarque & Bera, 1980; Ljung & Box, 1978; Lo & MacKinlay, 1988).

2.3 Forecasting models

All twelve models share the random walk as their anchor and differ in the forecast of the cumulative return $\hat{R}_{th}$ that they add to it. The random walk (RW) forecast is no change, $\hat{R}_{th} = 0$. The random walk with drift adds the mean daily return.

$$R_{t,h} = h \, \hat{\mu}_t, \qquad \hat{\mu}_t = (1 / t) \, \Sigma_{j=1}^{t} r_j \quad (3)$$

Where the estimation window (60, 125, 250, or 500 days, or all past data) was chosen in the validation year. The state-space random walk with drift (SS-RW-KF with drift) is the local linear trend model for $y_t = 100 \, x \ln S_t$ (Harvey, 1989);

$$y_t = \mu_t + \varepsilon_t, \qquad \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \qquad \beta_t = \beta_{t-1} + \zeta_t \quad (4)$$

Where μ_t is the level β_t the drift and ε_t, η_t, and ζ_t are independent Gaussian disturbances. The Kalman filter (Kalman, 1960) produced the filtered states, and the forecast was

$$\hat{y}_{t+h|t} = \hat{\mu}_{t|t} + h \, \beta_{t|t} \quad (5)$$

With forecast variance from the filter. The variances of ε_t and η_t were estimated by maximum likelihood. The drift variance was set to q times the level variance, with q chosen on the validation year from $\{0, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$; $q = 0$ corresponds to a constant drift learned recursively by the filter.

RW + ARIMA fitted an ARMA (p, q) model to returns, which is equivalent to an ARIMA $(p, 1, q)$ model for the log rate (Box et al., 2015).

$$r_t = c + \Sigma_{i=1}^{p} \varphi_i r_{t-i} + \varepsilon_t + \Sigma_{j=1}^{q} \theta_j \varepsilon_{t-j} \quad (6)$$

With $p, q \in \{0, \dots, 3\}$ and the constant chosen by the Akaike information criterion (AIC; Akaike, 1974). RW +GARCH used an autoregressive mean with GARCH(1,1) variance and Student-*t* errors (Bollerslev, 1986);

$$r_t = c + \Sigma_{i=1}^{k} \varphi_i r_{t-i} + \varepsilon_t, \qquad \varepsilon_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (7)$$

With autoregressive order $k \in \{0, \dots, 3\}$ chosen by AIC. Its point forecast is the conditional mean; its intervals were obtained from 2,000 simulated paths.

The four machine-learning hybrids, RW+ANN, RW+SVR, RW+RF and RW+XGBoost, learned a separate direct forecasting function for each horizon.

$$R_{t,h} = f_h(x_t) \quad (8)$$

Using 19 possible features on the close of day t: the past 10 days' returns; 5-, 10- and 20-day cumulative returns; 5-, 20- and 60-day realized volatilities; the deviation of the log rate from the 20- and 60-day moving average; and absolute return. The ANN model used a multilayer perceptron averaged across five random initializations; the SVR used a radial basis kernel (Drucker et al., 1997); the RF used an ensemble of 300 regression trees (Breiman, 2001); and XGBoost used a gradient-boosting algorithm based on gradient-boosted trees (Chen & Guestrin, 2016). In RW+LSTM, the past 20 days of three variables (return, 20-day volatility, and deviation from 20-day moving average) were input to an LSTM with one output for each horizon (Hochreiter & Schmidhuber, 1997), with averaging over five initializations.

In the two-stage hybrids, Zhang (2003) was followed. Stage one provided a statistical forecast $\hat{R}^{(1)}_{t,h}$ using the ARMA or AR-GARCH model. In contrast, in stage two, an LSTM model was trained to forecast the stage-one error using the past 20 days of stage-one residual information:

$$R_{t,h} = R_{t,h}^{(1)} + g_h(e_{t-19}, \dots, e_t) \quad (9)$$

For RW+ARIMA+LSTM, the inputs were the ARMA residuals, the 20-day volatility, and the distance from the 20-day moving average. For RW+GARCH+LSTM, the inputs were standardized GARCH residuals, the GARCH conditional volatility, and the distance from the 20-day moving average, in the spirit of Kim and Won (2018).

2.4 Forecasting design

Forecasts were generated for horizons of 1, 5, 10, and 20 trading days, spanning 6 November 2025 to 29 October 2026, across 250 origins. The model specifications for all forecasting models were selected using the mean squared error criterion on the validation year, using training data prior to the validation year: the drift window length, the drift variance window length, the ANN architecture and penalty, the SVR penalty, the RF depth and the leaf size, the XGBoost depth and tree count, and the number of LSTM units (either 16 or 32). The neural networks were trained using the Adam optimizer with mean squared error as a loss function and early stopping based on the most recent 15% of the training window. The models were re-estimated in the test year using an expanding window of 50 trading days, using only information available at the forecast origin. The models were also re-estimated using all data up to 30 October 2026 to generate a forecast for November 2026.

Each model was also associated with an 80% forecast interval. The RW interval considered the 10th and 90th percentiles of the last h-day returns over the previous 250 days, while the RW with drift did so after recentering on its drift. The ARIMA and Kalman filter intervals were analytical; the GARCH intervals were derived from simulation, while the ANN, SVR, RF, XGBoost, and RW+LSTM intervals were split conformal bands obtained from validation-year residuals (Lei et al., 2018).

2.5 Evaluation

The performance of point forecasts for exchange rate levels was measured using the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), mean absolute scaled error (MASE; Hyndman & Koehler, 2006), and Theil's U2, which is the ratio of the RMSE of a given model and the RMSE of the RW. The R^2 value was expressed as a percentage change in the error from the RW:

$$R^2_{OOS} = 100 \times (1 - \Sigma_t e^2_{m,t} / \Sigma_t e^2_{RW,t}) \quad (10)$$

The equality of forecasting accuracy across all models relative to the RW was tested using the DM test on squared errors with the Harvey-Leybourne-Newbold correction, with a t-distribution with T – 1 degrees of freedom as the alternative hypothesis (Diebold & Mariano, 1995; Harvey et al., 1997). Since each model includes the RW, the CW test, which accounts for the noise from additional parameter estimation under the null hypothesis and is one-sided, was also performed (Clark & West, 2007). Using the MCS procedure based on the range statistic and 5,000 stationary-bootstrapping replications, the models that could not be statistically distinguished from the most accurate one at the 10% significance level were identified (Hansen et al., 2011). Directional accuracy

was tested through the PT test (Pesaran & Timmermann, 1992). Forecasting efficiency was checked through the MZ regression of outcomes on forecasts (Mincer & Zarnowitz, 1969). Bias was tested using the heteroskedasticity- and autocorrelation-consistent t-statistic for the mean forecast error and the Ljung-Box test for one-day error autocorrelation. Interval forecasts were evaluated using empirical coverage and the Winkler score, which penalizes being outside the 80% interval by 10 times the interval width (Winkler, 1972).

Two robust checks were conducted. First, the validation-year mean squared error of the optimized model was compared with that of the RW, and the test year was split into two halves based on its median forecast date. Feature importance for the last RF (impurity reduction) and XGBoost (gain) models was calculated over 20 days.

2.6 Software

The estimation was performed using Python where the unit root test, ARMA and stable space model were estimated using statsmodel package (Seabold & Perktold, 2010), GARCH models by using ARCH (Sheppard, n.d.), variance ratio tests, and the model confidence set, ANN, SVR and RF by using scikit-learn package (Pedregosa et al., 2011), XGBoost (Chen & Guestrin, 2016) and LSTM by using keras with JAX backend (Chollet et al., 2015).

3. RESULTS

3.1 Properties of exchange rate

The peso weakened from P47.16 per US dollar on 4 January 2016 to P63.68 on 30 October 2026, averaging 2.8% per year, after reaching a record high of P63.99 on 27 October 2026 (Figure 1). The movement was uneven: the peso strengthened from late 2018 to mid-2021 and again in early 2025 and weakened sharply in 2022 and 2026.

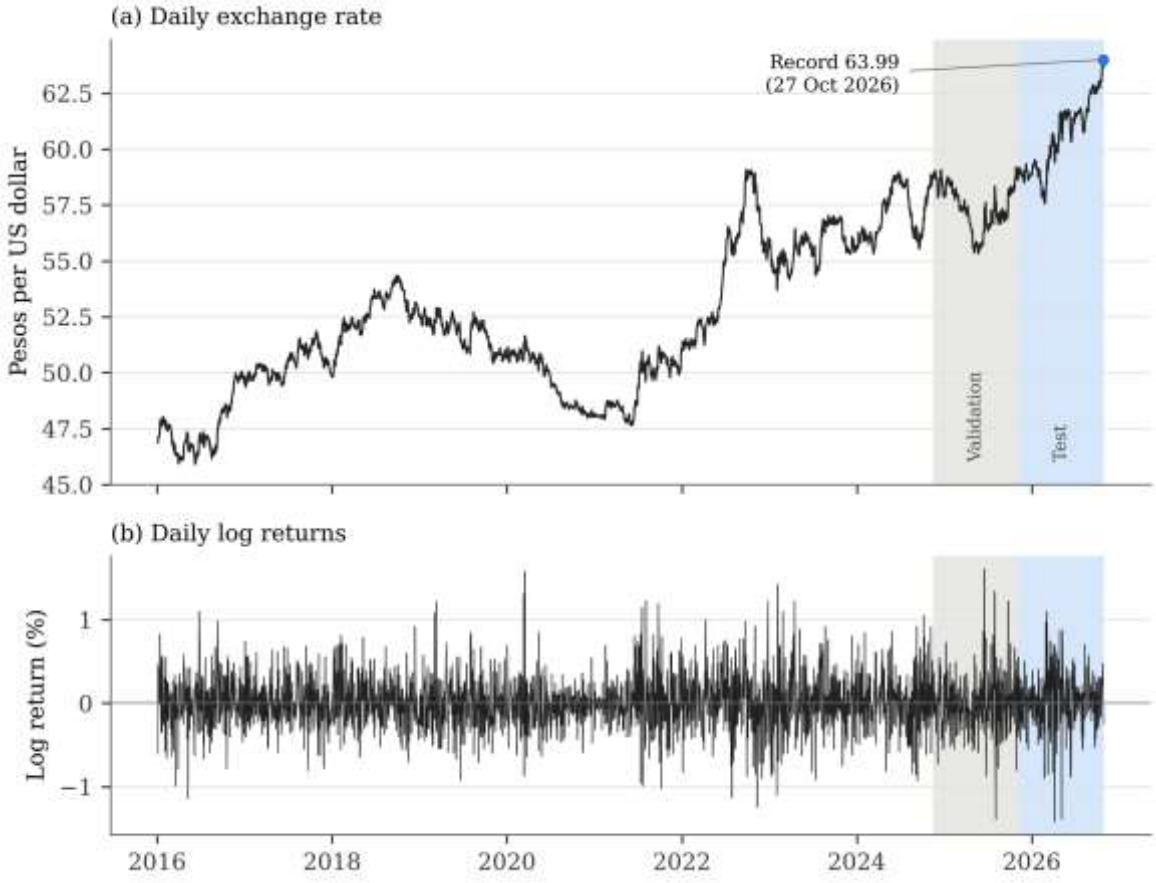

Figure 1. Daily peso-US dollar exchange rate and log returns, January 2016 – October 2026

Note. Shaded areas mark the validation year (grey) and the test year (blue). Source: author's calculations from European Central Bank reference rates.

Unit-root tests confirmed that the level and the log level were non-stationary, whereas returns were stationary (Table 1). With a constant, the log level ADF statistic was -0.34 (p=0.919) and the KPSS statistic was 6.32 (<0.01), while the ADF statistic for returns was -18.09 (p<0.001) and the KPSS statistic was 0.11 (p>0.10). The Zivot-Andrews test failed to confirm a unit root for log level with one structural break considered (p=0.693; most likely break on 15 August 2019). Thus, all models were built on returns.

Table 1. Unit-root and stationarity test, 4 January 2016 – 30 October 2026

Series	ADF	DF-GLS	Phillips–Perron	KPSS	Zivot–Andrews
Panel A. Constant					
Level (₱/US$)	−0.15 (0.944)	1.28 (0.953)	−0.27 (0.929)	6.374 (< 0.01)	−3.42 (0.728)
Log level	−0.34 (0.919)	1.16 (0.942)	−0.48 (0.897)	6.316 (< 0.01)	−3.48 (0.693)
First difference	−18.13 (< 0.001)	−3.75 (< 0.001)	−52.35 (< 0.001)	0.129 (> 0.10)	−18.27 (< 0.001)
Log return (%)	−18.09 (< 0.001)	−3.30 (0.001)	−52.35 (< 0.001)	0.107 (> 0.10)	−18.23 (< 0.001)
Panel B. Constant + trend					
Level (₱/US$)	−1.70 (0.749)	−1.85 (0.382)	−1.77 (0.719)	0.900 (< 0.01)	−3.39 (0.859)
Log level	−1.80 (0.702)	−1.92 (0.339)	−1.86 (0.673)	0.841 (< 0.01)	−3.43 (0.845)
First difference	−18.15 (< 0.001)	−7.11 (< 0.001)	−52.35 (< 0.001)	0.066 (> 0.10)	−18.32 (< 0.001)
Log return (%)	−18.10 (< 0.001)	−6.37 (< 0.001)	−52.34 (< 0.001)	0.066 (> 0.10)	−18.25 (< 0.001)

Note: Statistics with p-values in parentheses; N= 2,772 returns. ADF, DF-GLS, Phillips-Perron and Zivot-Andrews test the null of a unit root; KPSS tests the null of stationarity (p-values are interpolated and bounded at 0.01 and 0.10). ADF lag length chosen by AIC; Zivot-Andrews allows one break in the intercept.

Returns were found not to be normal, with high volatility clustering but low linear predictability (Table 2). The Jarque-Bera statistics were calculated as $\chi^2(2) = 321.97$ ($p < 0.001$), and the excess kurtosis was 1.61. The Ljung-Box test for returns was significant, $Q(10) = 28.02$ ($p = 0.002$), although the highest autocorrelation coefficient was only 0.055 at lag 6. Squared returns were highly autocorrelated: $Q(10) = 169.27$ ($p<0.001$), the ARCH-LM statistic was $\chi 2(10) = 107.31$ ($p<0.001$), and the BDS statistic rejected independence at all embedding dimensions ($p<0.001$). Variance ratios were close to one at 2, 5, 10, and 20 days (0.986 – 1.035) and did not reject the random walk hypothesis ($p\geq0.744$).

Table 2. Distribution and dependence of daily peso-dollar log returns

Statistic or test	Value	Reference distribution	*p*
Panel A. Distribution			
Mean / standard deviation (%)	0.0108 / 0.3300	—	—
Minimum / maximum (%)	−1.420 / 1.607	—	—
Skewness / excess kurtosis	0.215 / 1.613	—	—
Jarque–Bera	321.97	$\chi^2(2)$	< 0.001
Panel B. Linear and nonlinear dependence			
Ljung–Box *Q*(10), returns	28.02	$\chi^2(10)$	0.002
Ljung–Box *Q*(20), returns	41.49	$\chi^2(20)$	0.003
Ljung–Box *Q*(10), squared returns	169.27	$\chi^2(10)$	< 0.001
ARCH-LM (10 lags)	107.31	$\chi^2(10)$	< 0.001
BDS, embedding dimension 2	6.19	N(0,1)	< 0.001
BDS, embedding dimension 5	8.62	N(0,1)	< 0.001
Panel C. Lo–MacKinlay variance ratio (heteroskedasticity-robust)			
VR(2)	1.006	$z^* = 0.26$	0.792
VR(5)	0.986	$z^* = -0.28$	0.783
VR(10)	0.992	$z^* = -0.11$	0.914
VR(20)	1.035	$z^* = 0.33$	0.744

Note. Returns are 100 x the first difference of the log exchange rate, 5 January 2016 – 30 October 2026 (N = 2,772).

3.2 Model specification and estimates

Model Selection Criteria in Table 3 includes the twelve models, along with the chosen settings on the validation year. For both drift models, the more persistent setting was selected; for RW with drift, the expanding window (validation MSE relative to RW = 1.012) performed better than the 500-, 250-, 125-, and 60-day windows (validation MSE relative to RW = 1.033, 1.081, 1.137, and 1,200). In the case of the state-space model, all positive values of the drift variance increased the validation MSE at each horizon; moreover, the MLE estimated the drift to be zero, so q = 0 was chosen.

Table 3. Forecasting models and setting selected on the validation year

Model	Component added to the random walk	Inputs	Selected setting
RW	None (no-change forecast)	—	—
RW with drift	Constant drift: h × mean daily return	Past returns	All past data (from windows of 60, 125, 250, 500 days or all data)
SS-RW+KF with drift	State-space random walk with a drift state (local linear trend), Kalman filter	Log rate	Variances by maximum likelihood; drift variance 0 (grid 0 to 0.001 × level variance)
RW+ARIMA	ARMA forecast of returns (ARIMA on the log rate)	Past returns and shocks	ARMA(2,2), no constant (AIC, 32 candidates)
RW+GARCH	AR mean with GARCH(1,1) variance, Student-t errors	Past returns	AR(2) (AIC)
RW+ANN	Multilayer perceptron, 5 seeds averaged	19 features	16 hidden units; L2 penalty 1.0 (0.1 at h = 20)
RW+SVR	Support vector regression, radial basis kernel	19 features	C = 0.1 (1.0 at h = 20)
RW+RF	Random forest, 300 trees	19 features	Depth 3; minimum leaf 20 (100 at h = 1)
RW+XGBoost	Gradient-boosted trees	19 features	100 trees, depth 2, learning rate 0.03
RW+LSTM	LSTM network on 20-day windows, 5 seeds averaged	Return, 20-day volatility, distance from 20-day mean	32 units
RW+ARIMA+LSTM	ARIMA forecast plus LSTM forecast of the ARIMA error	ARIMA residuals, volatility, distance from mean	32 units
RW+GARCH+LSTM	AR-GARCH forecast plus LSTM forecast of the GARCH error	Standardized residuals, conditional volatility, distance from mean	16 units

Note. The 19 machine-learning features are returns on the last 10 days; 5-, 10- and 20-day cumulative returns;5-, 20- and 60-day volatility; distance of the log rate from its 20- and 60-day moving averages; and the absolute return. Every model forecasts the cumulative log return over h days, which is added to the last observed log rate.

Estimates are presented in Table 4. The ARMA(2,2) model contained highly significant second-order terms (AR(2) = -0.764 and MA(2) = 0.713, both p<0.001, and therefore its forecasts were very similar to those of the RW. The model residuals were not autocorrelated (Q(10) p = 0.965) but exhibited significant ARCH effects and nonlinearity. The AR(2)-GARCH(1,1)-t model had a very weak but significant AR(2) coefficient (-0.046, p = 0.023), very persistent volatility process ($\alpha + \beta = 0.994$), and heavy-tailed errors (7.0 degrees of freedom). It reduced volatility clustering substantially (ARCH LM p = 0.092), but nonlinear dependence was still present in the standardized residuals as revealed by the BDS (p = 0.034 and 0.009 for dimensions 2 and 3). In the state-space model, the irregular variance was close to zero, and the level variance was 0.108; thus, the log rate was a random walk with constant drift. The model was preferred to the alternative (stochastic drift) in likelihood (AIC 1543.1 vs 1547.1).

Table 4. Parameter estimates and residual diagnostics of the statistical models (estimation sample to 6 November 2025)

Item	ARMA(2,2)	AR(2)-GARCH(1,1)-*t*	State-space RW with drift
*Panel A. Estimates (*p*-values in parentheses)*			
AR(1) / AR(2)	−0.1348 (0.189) / −0.7643 (< 0.001)	−0.0062 (0.767) / −0.0455 (0.023)	—
MA(1) / MA(2)	0.1403 (0.203) / 0.7135 (< 0.001)	—	—
Constant	—	−0.0001 (0.980)	—
ω / α / β	—	0.0008 / 0.034 / 0.959	—
Persistence α + β; *t* degrees of freedom	—	0.994; 7.04	—
Innovation variance σ²	0.1070	—	Level 0.1076; irregular ≈ 0
Drift variance; filtered drift (% per day)	—	—	0; 0.0090
AIC	1531.7	1312.6	1543.1
*Panel B. Residual diagnostics (*p*-values)*			
Ljung–Box *Q*(10), residuals	0.965	0.133	—
Ljung–Box *Q*(10), squared residuals	< 0.001	0.084	—
ARCH-LM (10 lags)	< 0.001	0.092	—
Jarque–Bera	< 0.001	< 0.001	—
BDS, dimension 2 / 3	< 0.001 / < 0.001	0.034 / 0.009	—

Note. GARCH diagnostics use standardized residuals. The state-space model is the local linear trend model on 100 x ln S; its residual is the return less the filtered drift, so its diagnostic coincides with those of the returns in Table 2. The AIC of the state-space model is that of the fixed-drift specification which the likelihood selects over the stochastic-drift specification (AIC 1547.1).

The filtered drift value tended towards around 0.01% per day starting from 2018 (Figure 2). The filtered drift value was 0.0090% at the beginning of the year of the year of the test and 0.0111% on 30 October 2026, which is roughly 2.8% annually, and an 80% range was 0.003% to 0.019% per day. Since the drift variance was zero, the filtered drift was equal to the running average of the historical returns, so that both drift models generated identical point forecast up to four decimal places.

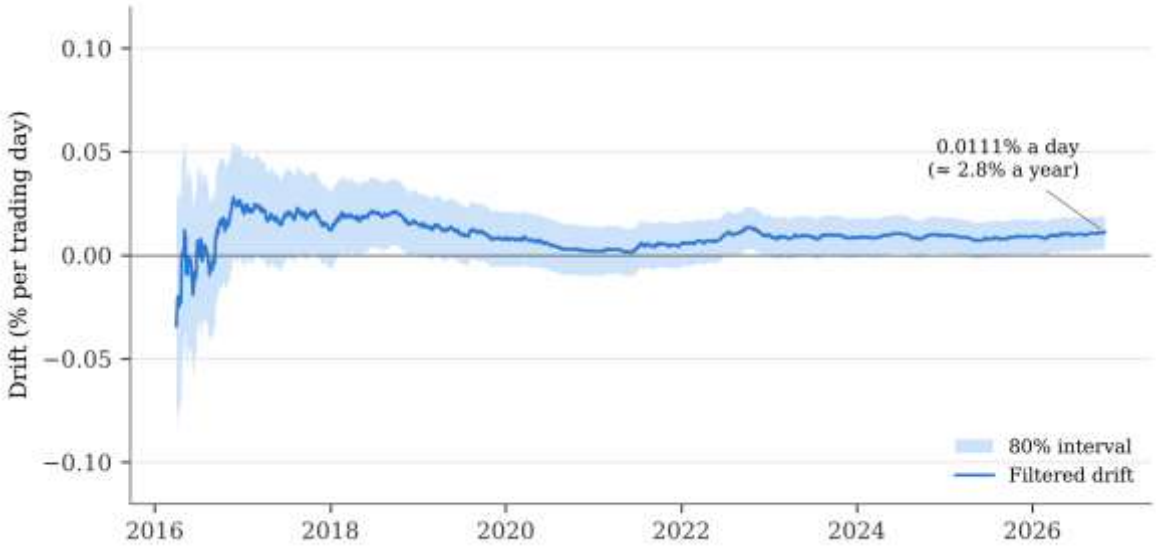

Figure 2. Kalman-filtered drift of the log-peso-dollar rate, 2016-2026

Note. Filtered drift $\hat{\beta}$ in percent per training day, with an 80% interval from the filter's state variance, using parameters estimated on all data to 30 October 2026. The first 60 days are omitted.

The LSTM-based models stopped training early: validation loss did not improve within the first few epochs, and early stopping occurred after 11 to 35 epochs (mean 13). The networks therefore learned little beyond the average return of the training window.

3.3 Point forecast accuracy

Within one to 10 days, all models were within 3% of the RW's RMSE (Table 5). The most accurate one-day model was RW+XGBoost at P0.2083, while the most accurate ten-day model was RW+RF at P0.6588, compared to RMSE of P0.2101 and P0.6785, respectively, for RW. The gap between the models increased when the horizon was increased to 20 days. Among those with the lowest RMSE were RW+XGBoost (P0.9015) and RW+RF(P0.9035), followed by RW+GARCH+LSTM (P0.9134) and the two drift models (P0.9199), while RW was at P0.9568. RW+SVR was the worst-performing model in terms of RMSE (P0.9790).

Table 5. Point forecast accuracy in the test year by horizon

Model	1 day	5 days	10 days	20 days
Panel A. Root mean squared error (₱)				
RW	0.2101	0.4999	0.6785	0.9568
RW with drift	0.2097	0.4952	0.6651	0.9199
SS-RW+KF with drift	0.2097	0.4952	0.6651	0.9199
RW+ARIMA	0.2089	0.4998	0.6781	0.9564
RW+GARCH	0.2102	0.4995	0.6773	0.9537
RW+ANN	0.2101	0.4976	0.6781	0.9474
RW+SVR	0.2104	**0.4929**	0.6670	0.9790
RW+RF	0.2085	0.4973	**0.6588**	0.9035
RW+XGBoost	**0.2083**	0.5000	0.6594	**0.9015**
RW+LSTM	0.2098	0.4986	0.6779	0.9370
RW+ARIMA+LSTM	0.2086	0.4985	0.6775	0.9368
RW+GARCH+LSTM	0.2094	0.4976	0.6860	0.9134
Panel B. Mean absolute error (₱)				
RW	0.1576	0.3825	0.5338	0.7309
RW with drift	0.1567	0.3755	0.5219	0.7028
SS-RW+KF with drift	0.1567	0.3755	0.5219	0.7028
RW+ARIMA	0.1582	0.3826	0.5340	0.7305
RW+GARCH	0.1580	0.3827	0.5341	0.7283
RW+ANN	0.1581	0.3796	0.5249	0.7415
RW+SVR	0.1582	0.3798	0.5191	0.7708
RW+RF	**0.1560**	0.3775	**0.5144**	**0.6956**
RW+XGBoost	0.1568	0.3779	0.5165	0.7047
RW+LSTM	0.1565	0.3771	0.5318	0.7177
RW+ARIMA+LSTM	0.1572	0.3772	0.5318	0.7176
RW+GARCH+LSTM	0.1574	**0.3749**	0.5421	0.7034
Panel C. Mean absolute percentage error (%)				
RW	0.261	0.632	0.881	1.203
RW with drift	0.260	0.621	0.863	1.159
SS-RW+KF with drift	0.260	0.621	0.863	1.159
RW+ARIMA	0.262	0.632	0.882	1.202
RW+GARCH	0.262	0.632	0.882	1.199
RW+ANN	0.262	0.628	0.867	1.220
RW+SVR	0.262	0.628	0.857	1.267
RW+RF	**0.258**	0.624	**0.850**	**1.145**
RW+XGBoost	0.260	0.625	0.854	1.159
RW+LSTM	0.259	0.623	0.879	1.182
RW+ARIMA+LSTM	0.260	0.624	0.879	1.182
RW+GARCH+LSTM	0.261	**0.620**	0.896	1.159

Model	1 day	5 days	10 days	20 days
Panel D. Theil's U2 (RMSE relative to the RW)				
RW	1.000	1.000	1.000	1.000
RW with drift	0.998	0.991	0.980	0.961
SS-RW+KF with drift	0.998	0.991	0.980	0.961
RW+ARIMA	0.994	1.000	0.999	1.000
RW+GARCH	1.000	0.999	0.998	0.997
RW+ANN	1.000	0.995	0.999	0.990
RW+SVR	1.001	**0.986**	0.983	1.023
RW+RF	0.992	0.995	**0.971**	0.944
RW+XGBoost	**0.991**	1.000	0.972	**0.942**
RW+LSTM	0.998	0.997	0.999	0.979
RW+ARIMA+LSTM	0.993	0.997	0.998	0.979
RW+GARCH+LSTM	0.997	0.995	1.011	0.955

Note. Test year: 7 November 2025 – 30 October 2026; 250, 246, 241 and 231 forecasts at 1, 5, 10 and 20 days. Bold = best in the column. MASE equals MAE divided by a horizon-specific constant (the in-sample MAE of the no-change forecast), so it ranks model exactly as MAE; value ranges from 1.15 to 1.32

Figure 3 shows the benefits, expressed as percentage savings in squared errors, compared to the RW. With 20 days, the savings were 11.2% for RW+XGBoost, 10.8% for RW+RF, 8.9% for RW+GARCH+STM, 7.6% for both drifts, and 4.1% for RW+LSTM+ARIMA+LSTM, while RW+ARIMA (0.1%) and RW+GARCH (0.6%) had almost the same performance as RW. None of the models managed to reduce squared error by more than 6% within 1 to 10 days. Figure 4 presents the forecast of 20 days versus actual values: all models underestimated the increase in the peso-dollar exchange rate in March and October 2026

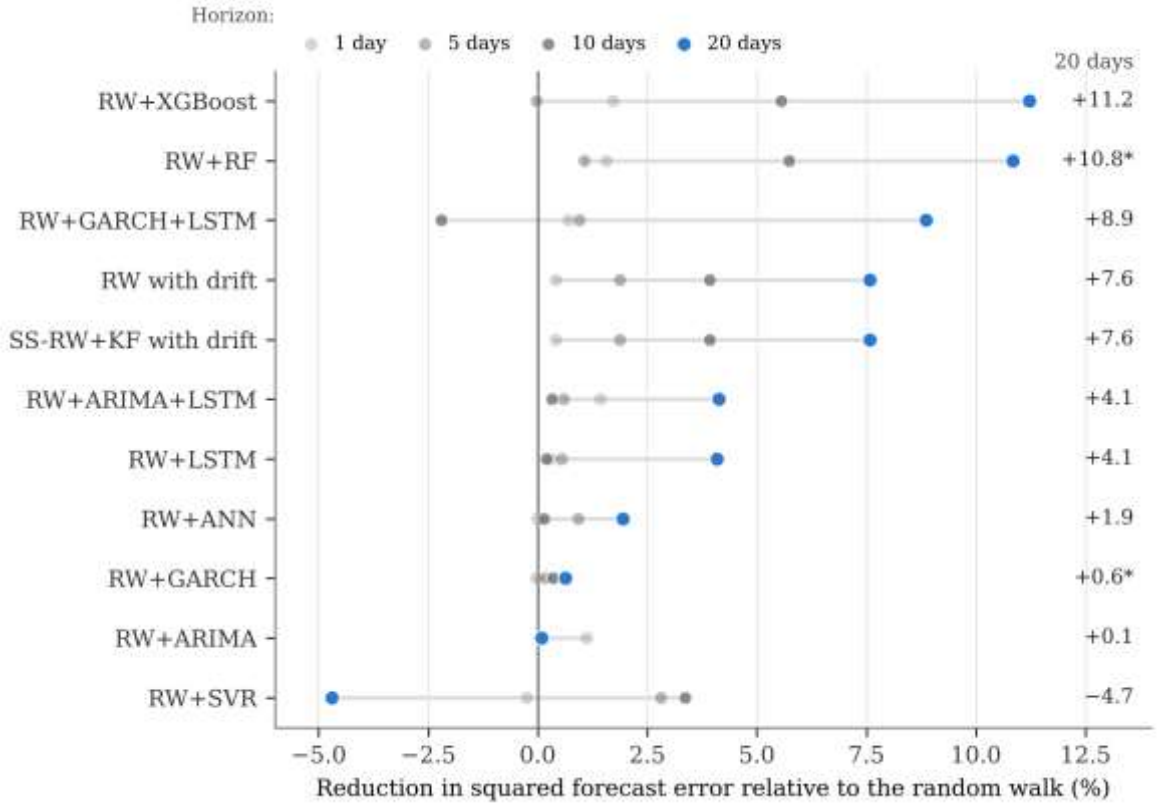

Figure 3. Reduction in squared forecast errors relative to the random walk, by model and horizon

Note. Out-of-sample R^2 (Equation 10) in the test year. Values at the right are for 20-day horizon; Diebold-Mariano p* <0.05 against the random walk.

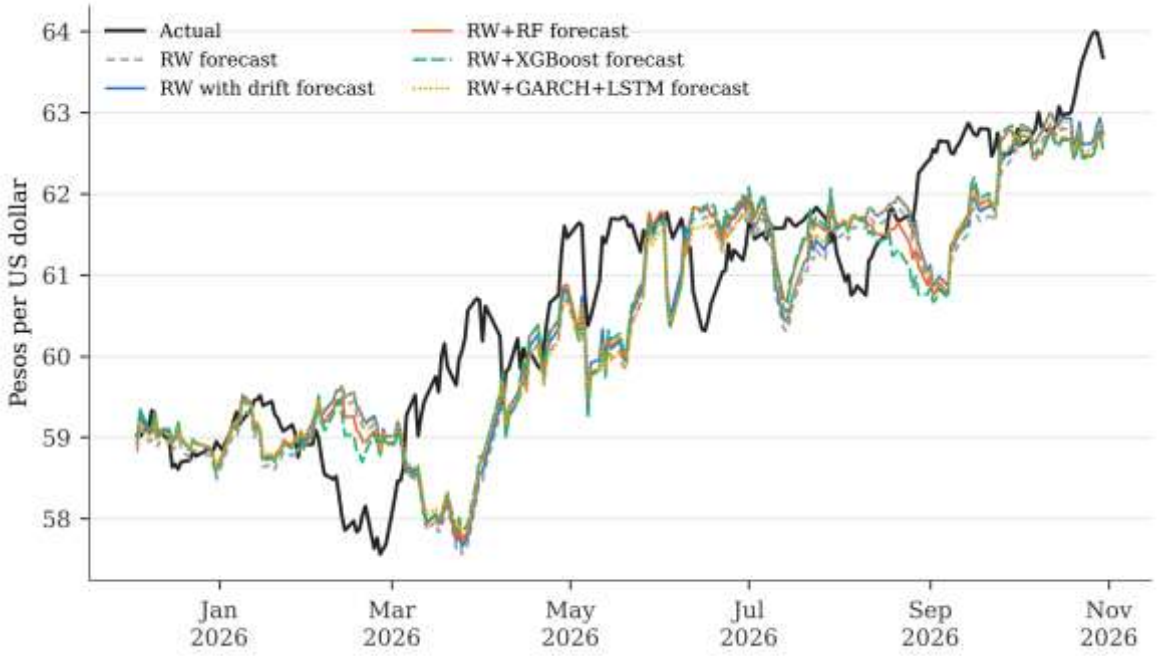

Figure 4. Twenty-day-ahead forecasts and actual exchange rates in the test year

Note. Each forecast is plotted at its target date.

3.4 Test of predictive accuracy

Only two models were significantly more accurate than the RW on the DM test, both at 20 days: RW+RF (DM = -2.07, t(230), p = 0.040) and RW+GARCH (DM = -2.29, t(230), p = 0.03) (Table 6). RW+GARCH was significant despite its small gain because its forecasts differed only slightly from the RW, so the loss differential had a small variance, and a small average gain was statistically detectable. The drift models did not reach significance on the DM test at any horizon (at 20 days, DM = -1.57, p = 0.118). Also, for the parameter noise of nested models, the CW test was significant in more cases for drift models for 5, 10, and 20 days (p=0.047, 0.037, and 0.022), RW+RF and RW+XGBoost for 1, 10, and 20 days, RW+ARIMA+LSTM for 1 day, RW+SVR for 5 days, and RW+GARCH and RW+GARCH+LSTM for 20 days. The MCS included all 12 models at 1, 5, and 10 days. At 20 days, the MCS included six: RW+XGBoost (p = 1.000), RW+RF (0.913), both drift models (0.881), RW+GARCH+LSTM (0.881), and RW+GARCH (0.149). The RW model was not included (p = 0.065).

Table 6. Tests of predictive accuracy against the random walk

Model	1 day	5 days	10 days	20 days
Panel A. Diebold–Mariano statistic, HLN-corrected (p-value); negative = more accurate than the RW				
RW with drift	−1.18 (0.239)	−1.41 (0.159)	−1.47 (0.143)	−1.57 (0.118)
SS-RW+KF with drift	−1.18 (0.239)	−1.41 (0.159)	−1.47 (0.143)	−1.57 (0.118)
RW+ARIMA	−0.86 (0.391)	−0.21 (0.831)	−0.63 (0.530)	−1.22 (0.223)
RW+GARCH	0.07 (0.946)	−0.32 (0.746)	−1.01 (0.312)	−2.29 (0.023)
RW+ANN	0.01 (0.994)	−0.38 (0.703)	−0.04 (0.970)	−0.37 (0.713)
RW+SVR	0.22 (0.822)	−1.34 (0.180)	−0.94 (0.347)	0.76 (0.447)
RW+RF	−1.75 (0.082)	−0.54 (0.588)	−1.67 (0.096)	−2.07 (0.040)
RW+XGBoost	−1.30 (0.194)	0.01 (0.993)	−1.26 (0.209)	−1.48 (0.140)
RW+LSTM	−0.44 (0.659)	−0.42 (0.674)	−0.08 (0.940)	−0.90 (0.369)
RW+ARIMA+LSTM	−1.16 (0.248)	−0.45 (0.654)	−0.13 (0.899)	−0.91 (0.364)
RW+GARCH+LSTM	−1.11 (0.269)	−0.45 (0.653)	0.60 (0.548)	−1.47 (0.144)
Panel B. Clark–West test against the RW, one-sided p-value				
RW with drift	0.087	0.047	0.037	0.022
SS-RW+KF with drift	0.087	0.047	0.037	0.022
RW+ARIMA	0.055	0.284	0.179	0.057
RW+GARCH	0.391	0.322	0.134	0.006
RW+ANN	0.138	0.144	0.268	0.122
RW+SVR	0.286	0.029	0.055	0.253

Model	1 day	5 days	10 days	20 days
RW+RF	0.018	0.187	0.017	0.003
RW+XGBoost	0.036	0.305	0.034	0.008
RW+LSTM	0.235	0.242	0.348	0.098
RW+ARIMA+LSTM	0.016	0.211	0.314	0.095
RW+GARCH+LSTM	0.063	0.183	0.584	0.037
Panel C. Model confidence set p-*value (bold = in the 90% set)*				
RW	**0.551**	**0.827**	**0.585**	0.065
RW with drift	**0.672**	**0.968**	**0.883**	**0.881**
SS-RW+KF with drift	**0.672**	**0.968**	**0.883**	**0.881**
RW+ARIMA	**0.928**	**0.869**	**0.611**	0.070
RW+GARCH	**0.551**	**0.917**	**0.635**	**0.149**
RW+ANN	**0.860**	**0.968**	**0.635**	0.068
RW+SVR	**0.672**	**1.000**	**0.883**	0.068
RW+RF	**0.968**	**0.968**	**1.000**	**0.913**
RW+XGBoost	**1.000**	**0.968**	**0.905**	**1.000**
RW+LSTM	**0.716**	**0.968**	**0.635**	0.068
RW+ARIMA+LSTM	**0.968**	**0.968**	**0.635**	0.068
RW+GARCH+LSTM	**0.860**	**0.968**	**0.635**	**0.881**

Note. Diebold-Mariano statistics use squared errors, a Newey-West variance with h-1 lags and the Harvey-Leybourne-Newbold small-sample correction, compared with t(T-1). The model confidence set uses the range statistic and a stationary bootstrap with 5,000 replications. The two drift models give identical point forecasts to four decimals

3.6 Robustness

This improved accuracy failed to spill over to the validation year (Table 9), where the peso showed hardly any change in value. In the latter, there was no method that outperformed RW at any horizon, apart from RW+SVR at 20 days (relative mean squared error 0.964); the drift model's relative errors ranged between 1.001 and 1.029. During the test period, the drift models showed improved accuracy in both halves across all horizons, from 0.2% to 10.4%. However, at the 20-day horizon, accuracy improved during both halves at all horizons, from 0.2% to 10.4%. However, at the 20-day horizon, apart from the two methods mentioned above, RW+XGBoost gained in the first half only (+18.2% versus -1.5%), while RW+SVR lost in both halves.

Table 9. Robustness: validation-year accuracy and stability across the test year

	Validation year: MSE relative to the RW				Test year, 20 days: % reduction in squared error	
Model	1 d	5 d	10 d	20 d	Nov 2025 – May 2026	May – Oct 2026
RW with drift	1.001	1.005	1.013	1.029	+6.0	+10.4
SS-RW+KF with drift	1.001	1.005	1.013	1.029	+6.0	+10.4
RW+ARIMA	—	—	—	—	+0.1	+0.1
RW+GARCH	—	—	—	—	+0.3	+1.2
RW+ANN	1.025	1.036	1.077	1.018	+3.7	−1.3
RW+SVR	1.014	1.023	1.035	**0.964**	−1.1	−11.3
RW+RF	1.007	1.027	1.058	1.066	+13.3	+6.4
RW+XGBoost	1.014	1.019	1.070	1.073	+18.2	−1.5
RW+LSTM	1.007	1.026	1.042	1.036	+4.7	+3.0
RW+ARIMA+LSTM	1.016	1.027	1.043	1.036	+4.8	+2.9
RW+GARCH+LSTM	1.004	1.037	1.042	1.054	+7.3	+11.7

Note. Validation year: 14 November 2024 – 6 November 2025 (peso +0.3%); bold = better than the RW. RW+ARIMA and RW+GARCH were not tuned on the validation year (order chosen by AIC) and have no validation entry. The test year is split at its median forecast origin; the peso weakened 2.6% in the first half and 5.5% in the second.

The tree ensembles relied mostly on volatility and trend information (Figure 5). In the 20-day RF, the 20-, 60- and 5-day volatilities accounted for 26.1%, 16.1% and 10.0% of the impurity reduction, followed by the distance from the 60-day moving average (8.6%)

and 10- and 20-day returns (7.3% and 6.8%); no single-day lag exceeded 2% XGBoost spread its importance more evenly but also ranked the 20-and 60- day volatilities first.

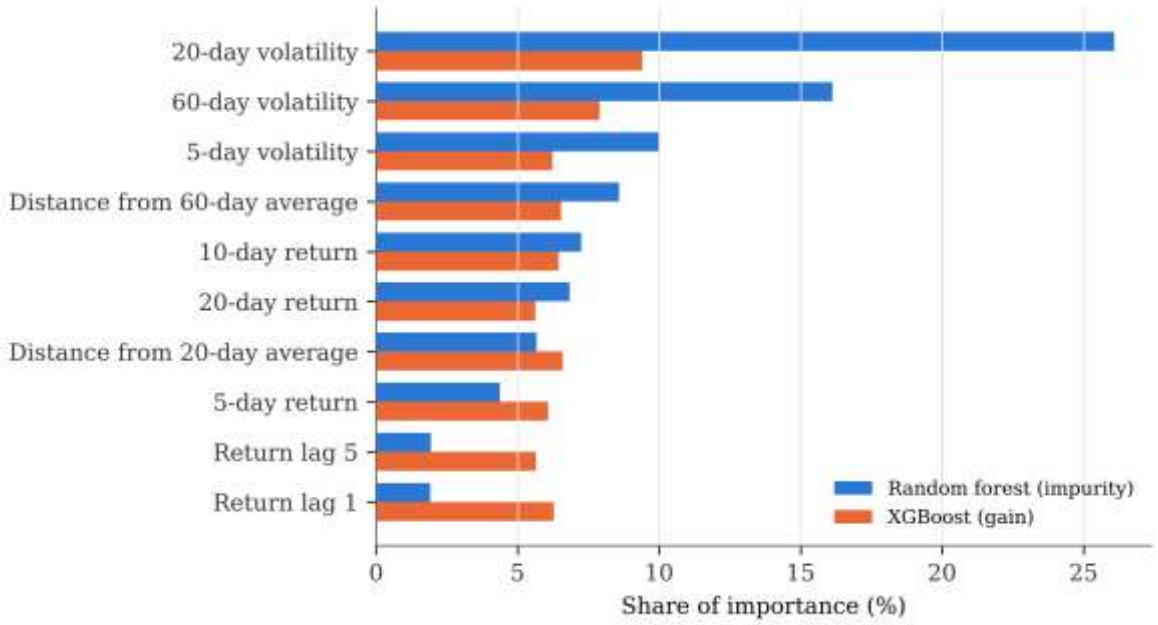

Figure 5. Feature importance of the 20-day random forest and XGBoost models

Note. Ten most important features of the final 20-day models estimated on data on 30 October 2026, as a share of total impurity reduction (random forest) or gain (XGBoost)

3.7 Forecast for November 2026

From P63.680 on 30 October 2026, the models forecast rates between P63.637 (RW+RF) and P63.914 (RW+SVR) for 27 November 2026 (Table 10). The drift models forecast P63.821, with a Kalman-filter 80% interval of P62.62 to P65.04. The spread between the models' point forecasts (P0.28) was small relative to the width of the intervals (about P2).

Table 10. Out-of-sample forecasts from 30 October 2026 (pesos per US dollar)

Model	2 Nov (1 d)	6 Nov (5 d)	13 Nov (10 d)	27 Nov (20 d)	80% interval, 27 Nov
RW	63.680	63.680	63.680	63.680	62.96 – 65.22
RW with drift	63.687	63.715	63.750	63.821	62.68 – 64.93
SS-RW+KF with drift	63.687	63.715	63.750	63.821	62.62 – 65.04
RW+ARIMA	63.685	63.684	63.680	63.681	62.52 – 64.87
RW+GARCH	63.688	63.701	63.710	63.728	62.74 – 64.63
RW+ANN	63.695	63.704	63.760	63.755	62.55 – 64.75
RW+SVR	63.710	63.709	63.683	63.914	62.71 – 64.91
RW+RF	63.688	63.698	63.692	63.637	62.49 – 64.67
RW+XGBoost	63.686	63.697	63.715	63.695	62.54 – 64.75
RW+LSTM	63.689	63.725	63.779	63.776	62.66 – 64.83
RW+ARIMA+LSTM	63.694	63.729	63.778	63.778	62.67 – 64.82
RW+GARCH+LSTM	63.697	63.776	63.854	63.780	62.66 – 64.84

Note. Last observation: P63.680 on 30 October 2026. All models were re-estimated on data 30 October 2026.

4. DISCUSSION

4.1 Why the random walk is hard to beat

The findings show that the peso-dollar exchange rate is nearly a martingale over short time spans. Near-one variance ratios, autocorrelation values lower than 0.055, and the inability of all models to significantly outperform the RW test over 1 to 10 days are consistent with predictions of the efficient markets hypothesis (Fama, 1970). The Engle-West theory is used to explain how these facts do not contradict any fundamentals, because the high persistence of the fundamental shock and a small discount factor make the exchange rate driven by far-future information (Engel & West, 2005). Near-canceling roots of the ARMA(2,2) model,

small forecasts from RW+ARIMA, and early stopping of LSTMs are evidence of a low ability to extract information from linear and non-linear data patterns relative to noise, for one- to two-week-ahead forecasts. This replicates the Meese-Rogoff phenomenon for an emerging market managed-floating currency more than four decades after its discovery (Cheung et al., 2005; Meese & Rogoff, 1983; Rossi, 2013).

4.2 The drift as an expression of fundamentals

The RW with drift is the most robust variant of the RW. This is the only model that improves over the RW in both halves of the test year and across all horizons, does not perform significantly worse than any machine learning model, and whose superiority is solely due to the RW's lower bias. The theory behind this model is straightforward. Assuming the underlying fundamentals of the Engle-West present value model exhibit a deterministic trend, the exchange rate follows an RW with drift. According to the relative purchasing power parity and uncovered interest parity, this trend is determined by inflation and interest rate differentials between the Philippines and the US, which have historically favored the steady devaluation of the peso (Fama, 1984; Rogoff, 1996). The estimated drift parameter of around 2.8% per year coincides with the average annual devaluation of the peso over the decade under consideration; using either parity, it implies a difference of 2.8 percentage points per year.

Kalman filter analysis reinforces this view. With the liberty to allow the drift process to move, maximum-likelihood estimation and validation accuracy selected a drift process. In the context of the state space model formulation, the trend of the peso is an intrinsic part of slow fundamentals rather than a series of transitory regimes (Harvey, 1989). This is in line with monetary theory, in which inflation expectations are anchored, and the central bank smooths out rather than opposes market dynamics (Calvo & Reinhart, 2002). In addition, constant drift helps explain the findings in the validation years: when the peso was constant, drift introduced a bias and slightly favored the RW, whereas when the peso showed a trend, drift favored the RW. Long history-based drift is a safe gamble on fundamentals that succeeds when fundamentals rather than transitory shocks influence the interest rate.

4.3 Tree ensembles, heterogeneous expectations and regime dependence

RW+RF and RW+XGBoost realized the highest 20-day gains, which amounted to around 3 to 4 percentage points more than the drift gain. Feature importance for them indicates what information they used. They used 20- and 60-day volatilities and the gap between the rate and its moving averages, but not yesterday's return. Thus, these models are regarded as state-dependent trend filters that place greater weight on recent movements in periods of higher volatility. Such a pattern is typical of heterogeneous-agent models, in which chartist extrapolation becomes more pronounced during periods of strong movements (De Grauwe & Grimaldi, 2006; Frankel & Froot, 1990) and is related to currency momentum (Menkhoff et al., 2012). The connection between volatility and depreciation stems from the asymmetric response of emerging markets, which depreciate sharply during periods of high volatility and heightened global risk aversion (Brunnermeier et al., 2008).

These improvements are, however, contingent. In the flat validation period, both models underperformed the RW strategy. The improvement provided by the RW+XGBoost was no longer observed in the latter half of the test period, and neither model outperformed the drift. The Adaptive Markets Hypothesis explains this phenomenon: predictability is observed only within a given regime but disappears with regime changes (Lo, 2004). Thus, tree ensembles improve on the drift regime.

4.4 Volatility models and the GARCH-LSTM hybrid

Large ARCH effects and persistence of GARCH variance ($\alpha + \beta = 0.994$) support the theory of information arrival in bunches (Clark, 1973). Modeling volatility was more beneficial for the intervals than for the point forecast, as RW+GARCH provided the most accurate one-day intervals but yielded point forecasts very similar to RW. RW+GARCH+LSTM outperformed RW+LSTM and RW+ARIMA+LSTM at 20 days and remained inside the MCS while improving both halves of the test year. Nonlinear dependence persisted after applying GARCH, and standardized residuals with conditional volatility enabled the LSTM to better recognize this pattern than raw returns; the ARIMA stage did not provide any improvement due to very small ARMA forecasts (Zhang, 2003).

4.5 Comparison with previous studies

The result supports the extensive literature on the difficulty of beating the random walk in the short term (Cheung et al., 2005; Mees & Rogoff, 1983; Rossi, 2013) and on the small and volatile gains from machine learning in currency prediction (Amat et al., 2018). In two ways, they add to that literature. First, they find that an appropriately estimated drift can account for almost all of the one-month gains from machine learning, while the Kalman filter validates the assumption of a constant drift. Thus, a comparison that ignores the drift exaggerates the contribution of machine learning. Second, they confirm the importance of the hybrid design: the GARCH-LSTM hybrid performed better than the ARIMA-LSTM hybrid, consistent with the theoretical view that the valuable structure of exchange rate return residuals lies in their volatility process rather than their linear autocorrelation.

4.6 Implications

To forecasters, the RW with drift must be used in place of the pure RW as the benchmark model for forecasting the peso-dollar exchange rate over horizons of one month, and any claim of improved performance is evaluated against it using nested-model tests and the MCS. For the Bangko Sentral ng Pilipinas and other institutional dealings with dollar-exposed banks and firms, the existing

level must serve as the one- to two-week forecast, GARCH ranges as short-term risk limits, and the Kalman filter-based ranges around the drift as the best-calibrated one-month ranges. RW+RF and RW+GARCH+LSTM are added to the set during the trend period, but their performance should be monitored and the models retested whenever the regime changes. Because one-month intervals were too small during the testing year's trending period, risk managers must increase the size of such intervals or skew them towards depreciation while the trend continues.

5. CONCLUSION

Twelve models anchored to the random walk were tested for their ability to forecast the daily Philippine peso-US dollar exchange rate from 1 to 1 month ahead. Their returns were stationary, nearly independent, and highly heteroskedastic, and no model significantly outperformed the random walk over 1-10 days ahead. At a one-month lead, the models RW+XGBoost and RW+RF decreased squared errors by 11%; the model with drift by 7.6%; and the hybrid GARCH-LSTM model by 8.9%. Six models made up the model confidence set. Drift models were the only models that performed better in both halves of the test year; the Kalman filter estimated the drift as 2.8% per year, and the intervals produced by the Kalman filter were the best-calibrated ones at one month ahead. In theoretical terms, the Philippine peso exhibits the behavior of a near-martingale that drifts slowly downward according to the trend implied by the difference between inflation and interest rates under parity conditions, and whose volatility clusters due to clustered news. Thus, the episodic and regime-dependent momentum can be exploited by tree ensembles. The random walk with drift should be taken as a benchmark for peso forecasting, and the gains from machine learning should be treated as conditional on the market regime.

REFERENCES

1. Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705
2. Amat, C., Michalski, T., & Stoltz, G. (2018). Fundamentals and exchange rate forecastability with simple machine learning methods. *Journal of International Money and Finance*, *88*, 1-24. https://doi.org/10.1016/j.jimonfin.2018.06.003
3. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1
4. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
5. Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324
6. Broock, W. A., Scheinkman, J. A., Dechert, W. D., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric reviews*, *15*(3), 197-235. https://doi.org/10.1080/07474939608800353
7. Brunnermeier, M. K., Nagel, S., & Pedersen, L. H. (2008). Carry trades and currency crashes. *NBER macroeconomics annual*, *23*(1), 313-348. https://doi.org/10.1086/593088
8. Calvo, G. A., & Reinhart, C. M. (2002). Fear of floating. *The Quarterly journal of economics*, *117*(2), 379-408. https://doi.org/10.1162/003355302753650274
9. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785
10. Cheung, Y. W., Chinn, M. D., & Pascual, A. G. (2005). Empirical exchange rate models of the nineties: Are any fit to survive?. *Journal of international money and finance*, *24*(7), 1150-1175. https://doi.org/10.1016/j.jimonfin.2005.08.002
11. Chollet, F., et al. (2015). *Keras* [Computer software]. https://keras.io
12. Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, 135-155. https://doi.org/10.2307/1913889
13. Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, *138*(1), 291-311. https://doi.org/10.1016/j.jeconom.2006.05.023
14. De Grauwe, P., & Grimaldi, M. (2006). Exchange rate puzzles: a tale of switching attractors. *European Economic Review*, *50*(1), 1-33. https://doi.org/10.1016/j.euroecorev.2004.07.005
15. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427-431. https://doi.org/10.1080/01621459.1979.10482531
16. Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), 134-144. https://doi.org/10.1198/073500102753410444
17. Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, *9*.
18. Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, *64*(4), 813–836. https://doi.org/10.2307/2171846

19. Engel, C., & West, K. D. (2005). Exchange rates and fundamentals. *Journal of Political Economy*, *113*(3), 485–517. https://doi.org/10.1086/429137
20. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007. https://doi.org/10.2307/1912773
21. European Central Bank. (2026). *Euro foreign exchange reference rates* [Data set]. https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/index.en.html
22. Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486
23. Fama, E. F. (1984). Forward and spot exchange rates. *Journal of Monetary Economics*, *14*(3), 319–338. https://doi.org/10.1016/0304-3932(84)90046-1
24. Frankel, J. A., & Froot, K. A. (1990). Chartists, fundamentalists, and trading in the foreign exchange market. *The American Economic Review*, *80*(2), 181-185. https://www.jstor.org/stable/2006566
25. Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453-497. https://doi.org/10.3982/ECTA5771
26. Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press. https://doi.org/10.1017/CBO9781107049994
27. Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, *13*(2), 281-291. https://doi.org/10.1016/S0169-2070(96)00719-4
28. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
29. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001
30. Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255–259. https://doi.org/10.1016/0165-1765(80)90024-5
31. Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552
32. Kilian, L., & Taylor, M. P. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, *60*(1), 85–107. https://doi.org/10.1016/S0022-1996(02)00060-0
33. Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, *103*, 25-37. https://doi.org/10.1016/j.eswa.2018.03.002
34. Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of econometrics*, *54*(1-3), 159-178. https://doi.org/10.1016/0304-4076(92)90104-Y
35. Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, *113*(523), 1094-1111. https://doi.org/10.1080/01621459.2017.1307116
36. Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303. https://doi.org/10.1093/biomet/65.2.297
37. Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, *30*(5), 15–29. https://doi.org/10.3905/jpm.2004.442611
38. Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, *1*(1), 41–66. https://doi.org/10.1093/rfs/1.1.41
39. Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1–2), 3–24. https://doi.org/10.1016/0022-1996(83)90017-X
40. Menkhoff, L., Sarno, L., Schmeling, M., & Schrimpf, A. (2012). Currency momentum strategies. *Journal of financial economics*, *106*(3), 660-684. https://doi.org/10.1016/j.jfineco.2012.06.009
41. Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3-46). NBER.
42. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.
43. Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, *10*(4), 461-465. https://doi.org/10.1080/07350015.1992.10509922
44. Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *biometrika*, *75*(2), 335-346. https://doi.org/10.1093/biomet/75.2.335
45. Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic literature*, *34*(2), 647-668.

https://www.jstor.org/stable/2729217
46. Rossi, B. (2013). Exchange rate predictability. *Journal of economic literature*, *51*(4), 1063-1119. https://doi.org/10.1257/jel.51.4.1063
47. Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (pp. 92–96). https://doi.org/10.25080/Majora-92bf1922-011
48. Sheppard, K. (n.d.). *arch* (Version 8.0.0) [Computer software]. https://github.com/bashtage/arch
49. Tesorero, A. (2026, October 27). Philippine peso sinks past Php64 vs US dollar, lowest value in history. *Khaleej Times*. https://www.khaleejtimes.com/world/asia/philippine-peso-sinks-past-php64-vs-us-dollar-lowest-value-in-history
50. Villanueva, R. (2026, October 15). Philippine peso hits record low of P63.15 against US dollar as inflation surges. *Khaleej Times*. https://www.khaleejtimes.com/world/asia/philippine-peso-hits-record-low-of-p6315-against-us-dollar-as-inflation-surges
51. Winkler, R. L. (1972). A decision-theoretic approach to interval estimation. *Journal of the American Statistical Association, 67*(337), 187–191. https://doi.org/10.1080/01621459.1972.10481224
52. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0