Technological and Economic Development of Economy
ISSN: 2029-4913 / eISSN: 2029-4921
2023 Volume 29 Issue 4: 1216–1238
https://doi.org/10.3846/tede.2023.19294

TIME SERIES FORECASTING WITH THE CIR# MODEL: FROM HECTIC MARKETS SENTIMENTS TO REGULAR SEASONAL TOURISM

Giuseppe ORLANDO [1*], Michele BUFALO [2]

[1] *Department of Economics and Finance, Università degli Studi di Bari Aldo Moro, Via C. Rosalba 53, Bari, 70124 Italy*
[2] *Department of Methods and Models for Economics, Territory and Finance, Università degli Studi di Roma "La Sapienza", Via del Castro Laurenziano 9, Roma, 00185 Italy*

Received 13 April 2022; accepted 12 April 2023

Abstract. This research aims to propose the so-called CIR#, which takes its cue from the well-known Cox-Ingersoll-Ross (CIR) model originally devised for pricing, as a general econometric model. To this end, we present the results on two very different time series such as Polish interest rates (subject to market sentiments) and seasonal tourism (subject to pandemic lock-down measures). For interest rates, as reference models, we consider an improved version of the CIR model (denoted CIR_{adj}), the Hull and White model, the exponentially weighted moving average (EWMA) which is often adopted whenever no structure is assumed in the data and a popular machine learning model such as the short-term memory network (LSTM). For tourism, as a benchmark, we consider seasonal autoregressive integrated moving average (SARIMA) complemented by the generalized autoregressive conditional heteroskedasticity (GARCH) for modelling the variance, the classic Holt-Winters model and the aforementioned LSTM. Results support the claim that the CIR# performs better than the other models in all considered cases being able to deal with erratic behaviour in data.

Keywords: tourism demand prediction, interest rate forecasting, cluster volatility and jumps fitting, SARIMA, CIR model, Hull and White model.

JEL Classification: G12, E43, E47, Z31.

Introduction

Interest rates represent the cost of the loan for those wishing to borrow. Like any price, the value of money is determined in the market by supply and demand. Economic agents base their actions on rational considerations and irrational feelings. The latter were called by J. M. Keynes "animal spirits" to summarize "instincts, proclivities and emotions that osten-

*Corresponding author. E-mail: *giuseppe.orlando@uniba.it*

sibly influence and guide human behavior" (Keynes, 1936). This work set off from precedent research by Orlando et al. (2018, 2019a, 2019b), Orlando and Bufalo (2021) in which the authors have suggested a new framework, that they have called CIR#, to predict interest rates. This of paramount importance because interest rates are key for setting the exchange rate i.e. the terms of trade between countries (e.g. Bjørnland & Hungnes, 2006), the default probability of the issuer (e.g. Gruppe et al., 2017), the conditions at which borrowers compete for credit on the market (e.g. Thakur et al., 2018), real estate evaluation (e.g. Akimov et al., 2019), availability of funds, market liquidity, etc. In addition, "interest rates are very difficult to predict" (Schwarzbach et al., 2012).

The idea behind the CIR # model is to maintain the analytical tractability and simplicity of the original single-factor model but, at the same time, to capture significant changes in time series. To this end, by performing a suitable sampling of the dataset, it is possible to modelling cluster volatility or jumps (see Orlando et al., 2020). Moreover, as the original CIR framework by construction was unable to deal with near-to-zero or even negative interest rates, data are appropriately shifted similarly to what Brigo and Mercurio (2000) did.

Finally, a key aspect of the suggested model is the calibration algorithm. Differently from existing literature that relies on standard Brownian motion, in the CIR# model the random part comes from Gaussian residuals of an "optimal" Autoregressive Integrated Moving Average (ARIMA) model. The ARIMA is used for filtering raw data so that the residuals allow to get an exact trajectory of CIR fitted values. This substitutes the standard Monte Carlo approach, to significantly reduce the computational cost of the algorithm.

While in literature there is a large number of models, for the reasons above exposed, in this work, we do not consider multi-factor models. Thus, we compare the CIR# model with the Exponentially Weighted Moving Average (EWMA), the CIR_{adj} which is an improved (through data partitioning, shift and calibration) version of the CIR model suggested by us, the single-factor Hull and White model and the short-term memory network (LSTM) used for machine learning. The performance of all those models is scrutinized on Polish interest rates (which include both turmoil and calm periods) where we test the forecasting capability regarding the directionality of interest rates (i.e. success of prediction) and accuracy (i.e. quality of prediction).

In addition to interest rates forecasting for which the CIR# has been developed, we asked ourselves about the performance of the suggested model when dealing with completely different time series such as tourism demand. There are three reasons for this choice: i) importance of the sector, ii) completely different behaviour of the historical series on tourism demand from those on interest rates (e.g. seasonal), iii) high dependence on external factors such as crime, environment, politics and health problems.

Tourism is a crucial industry in modern economies, with a direct contribution to the world GDP of approximately 4.7 trillion U.S. dollars in 2020, according to Lock (2022). In 2020, the United States' travel and tourism industry alone contributed 1.1 trillion U.S. dollars. Seasonality, defined as a temporal imbalance in tourist flows, is a characteristic feature of tourism (Baum & Lundtorp, 2001), involving the concentration of tourists in relatively short periods of the year. This has led to questions about the sustainability management of tourism (Grundey, 2008). Additionally, the sensitivity of the tourism industry to external factors means that it was severely impacted by the global COVID-19 pandemic that began in early 2020. Given these reasons, we believe that predicting tourism demand presents a challenging test case for the CIR#.

The novelty of this work lies in the proposal of the so-called CIR#, as a general econometric model suitable not only for forecasting interest rates exposed to volatile market sentiments but also for data such as those of the tourism industry with very different, more regular patterns but characterized by a high cyclicality. For these data, established models such as Holt-Winters have proven their worth, however the COVID-19 pandemic has severely hampered the industry and destroyed the above mentioned regularity. For this reason we claim that a model like CIR#, specially designed for such irregularities could be of valid alternative. To demonstrate our claim, apart from the four reference models mentioned for interest rates, for tourism we use: a) the CIR_{adj}), b) the EWMA, c) the seasonal autoregressive integrated moving average (SARIMA) complemented by the generalized autoregressive conditional heteroskedasticity (GARCH) for modeling the variance, d) the classic Holt-Winters model and, e) the LSTM.

The organization of the article is as follows. Section 1 deals with the literature on the CIR model, its evolution and the prediction problem in the tourism sector. Section 2 explains the proposed model starting with its reasons. Section 3 presents the results together with an evaluation of the modelling capabilities of the time series examined. The last Section concludes.

1. Literature review

Cox, Ingersoll and Ross suggested a term structure model (CIR) 1985, for modelling, under no-arbitrage condition, the price of discount zero-coupon bonds. The CIR model was perceived as an improvement over the Vasicek model (1977) because does not allow negative interest rates thanks to the mean-reverting mechanic as the random perturbation dampens when rates get low. Furthermore, the CIR model made it possible to consider the case of non-constant volatility with an underlying short-term interest rate thought as a diffusion process, that is a continuous Markov process. Furthermore, the CIR model made it possible to consider the case of non-constant volatility with an underlying short-term interest rate, conceived as a diffusion process, namely a continuous Markov process. Both, the analytical tractability and simplicity of the CIR model made it very popular among practitioners and scholars when dealing with short-term rates. Expansions to stochastic volatility modelling for option pricing or default intensities in credit risk include Heston (1993), Duffie (2005), Mininni et al. (2020).

Even though the same authors warned about potential limitations of their model because it can "produce only normal, inverse or humped shapes" (1985) practical issues in terms of calibrations were reported, for example, by Brigo and Mercurio (2001) and Carmona and Tehranchi (2006). In fact, the zero coupon curve is difficult to accurately replicate, even with various parameter values chosen in the model's dynamics. Additionally, certain shapes, such as an inverted yield curve, may not be captured by the model, further contributing to inaccuracies in reproducing the curve (Brigo & Mercurio, 2001). Adjusting the model's parameters can result in yield curves with a single peak or trough, but it is challenging or even impossible to accurately calibrate these parameters to achieve the desired location of the peak or trough. Additionally, it should be noted that the model's parameters may not be enough to completely explain all the observed characteristics that are present in market prices (Carmona & Tehranchi, 2006). Keller-Ressel and Steiner (2008) state that any time-homogeneous, affine one-factor model can only produce yield-to-maturity curves that fall into one of three categories: normal, humped, or inverse. This provides some insight into the limitations of such models in capturing the complexity of real-world yield curves.

This inability of the CIR model, in its original version, to fit to the observed yield curve and to sudden jumps in interest rates, sparkled a quest on alternative solutions and improvements such as Hull and White (1990) who introduced non-autonomous coefficients; Chen (1996) with a three-factor model; Brigo and Mercurio (2006) and Brigo and El-Bachir (2006) who added the jump diffusion JCIR model and the JCIR++, respectively; d) Brigo and Mercurio (2006) the CIR2 and CIR2++ two-factor models. More recently, Zhu (2014) came out with a CIR process with Hawkes "clustering effect", Moreno and Platania (2015) suggested an harmonic oscillation for the long-run mean, Najafi and Mehrdoust (2017) studied the CIR model in the context of mixed fractional Brownian motion and Ascione et al. (2023) suggested a combination of the Heston and CIR model under Levy process framework for pricing FX options.

As a result of the financial crisis that began in 2007 and the subsequent implementation of quantitative easing policies, the notion that interest rates should always maintain a positive value, which was a fundamental characteristic of the original CIR model, was rendered invalid. Between December 2014 and the end of May 2015, an estimated $2 trillion worth of long-term sovereign debt was traded globally, with a significant portion of it being issued by euro area sovereigns, at negative yields. As mentioned by Engelen (2015) and Bank for International Settlements [BIS] (2015), the existence of such yields is unprecedented. Furthermore, policy rates are even lower now than they were at the peak of the Great Financial Crisis, both in nominal and real terms. In fact, real interest rates have been negative for a longer period than they were during the Great Inflation of the 1970s. Despite the exceptional nature of this situation, many anticipate that it will persist. The remarkably low interest rates are just one of several notable indications of a larger issue plaguing the global economy. The current economic expansion is characterized by an imbalance, with excessive debt burdens and financial risks still prevalent, while productivity growth remains insufficient, and the scope for macroeconomic policy manoeuvring is limited. As a result, actions that would have been unthinkable in the past are now becoming normalized and perceived as the "new normal".

Within this new normal there is a need to model regimes' changes, shocks and negative rates. For this reason, with the idea of maintaining the analytical tractability of the original CIR model, we have developed our methodology inside the said design Orlando et al. (2018, 2019a, 2019b, 2020), Orlando and Bufalo (2021) which fits well into the interest rate structure.

Regarding tourism, seminal works on seasonality are Allcock (1989), Butler (1998), Baum and Lundtorp (2001) while a recent overview can be found in Corluka (2019). In terms of forecasting, the use of time series models is simple and effective such as ARIMA is largely documented. For example, Chang and Liao (2010) adopt a seasonal ARIMA model for forecasting monthly outbound tourism departures of major destinations from Taiwan (i.e. Hong Kong, Japan and the USA). Choden and Unhapipat (2018) predict the monthly number of international visitors in Bumthang (Bhutan) with 91% accuracy. Zong and Wang (2018) found that an ARIMA model (2,1,2) is able to predict the urban residents' travel rate in China from 2016 to 2020 was predicted. More in general, Li et al. (2021), when providing a review of articles on tourism forecasting research with Internet data published in academic journals from 2012 to 2019, found that AR, ARMA, and ARIMA are among the most popularly used time series models, accounting for 44% of the reviewed articles.

About applications of GARCH models, Dutta et al. (2021) found that shocks such as Brexit uncertainty impact on agents' utility function and claim strong evidence of long-run persistence in volatility in tourist arrival. Shanika and Jahufer (2021) employ a GARCH model for analysing the volatility of international tourist arrivals to Sri Lanka. Ampountolas (2021) provides a comparison between SARIMAX, Neural Networks, and GARCH models when studying "daily demand observations from a hotel in a US metropolitan city from 2015 to 2019 and a set of exogenous social and environmental features such as temperature, holidays, and hotel competitive set ranking". The outcome is that the SARIMAX outperforms the other models, in every horizon except one out of seven forecast horizons. Finally, Santamaria and Filis (2019), by the means of a Dynamic Conditional Correlation GARCH model, investigate the relationship between the Spanish term structure of interest rates and tourism-expected economic growth.

Another model considered for tourism demand forecast is the classic Holt-Winters. Lim and McAleer (2001) test a range of different exponential smoothing models "over the period 1975–1999 to forecast quarterly tourist arrivals to Australia from Hong Kong, Malaysia, and Singapore". According to the root mean squared error used as a criterion for measuring the forecast accuracy, the best results are provided by the Holt-Winters additive and multiplicative seasonal models.

Last but not least, as machine learning methods enjoy great momentum among academics and practitioners, we have added to the considered models the short-term memory network (LSTM) which is a recurrent neural network (RNN) quite popular in time series forecasting (e.g. Kudo et al., 1999; Zhao et al., 2017; Qadeer et al., 2020). In finance, Yıldırım et al. (2021) uses LSTM deep learning to make direction predictions in Forex with macroeconomic (fundamental) data and a technical indicator used by traders in technical analysis. Lu et al. (2020) suggest LSTM for predicting the stock price based on features extracted by convolutional neural networks (CNN) from "opening price, highest price, lowest price, closing price, volume, turnover, ups and downs, and change". Li and Cao (2018) employ LSTM finding it more efficient in the short-term than other models "such as ARMA, ARIMA which are mainly linear models and cannot describe the stochastic and non-linear nature of tourism flow". Polyzos et al. (2021) use LSTM neural network to investigate the effect of the COVID-19 outbreak to arrivals of Chinese tourists to the USA and Australia. Claveria et al. (2017) suggest data pre-processing for neural network-based forecasting. Finally, we mention He et al. (2021) who introduce a SARIMA–CNN–LSTM model to forecast tourist demand data at a daily frequency.

2. Methods and material

2.1. Data

In this section, we describe the data and their statistical characteristics. The main issues are that interest rates are very different from tourism data and both have a unit root. These two types of problems make it difficult not only to predict but, also, to find a suitable model for both time series.

2.1.1. Description of data

The market data used in this study consists of interest rates denominated in Polish zloty (PLN) from January 2, 1995, to February 5, 2021. The data is sampled on a weekly basis and covers six distinct tenors: Overnight, 1 Week, 1 Month, 3 Months, 6 Months, and 1 Year as collected from Trading Economics. Figure 1 shows all six time series of fairly similar shapes. However, note that the overnight and the one week tenor are more erratic as they are more exposed to market sentiments.

Data on foreign tourism to Italy (arrivals) consist in hotels, holiday and other short-stay accommodation, camping grounds, recreational vehicle parks and trailer parks. Monthly data span from Jan-90 to May-21 and have been retrieved from Eurostat. Analyses are carried out on log changes of data.

From Figure 2 it can be seen how the relatively regular and oscillatory behaviour of the time series due to the seasonal component is significantly perturbed starting from the 300th month because of the COVID-19 pandemic.

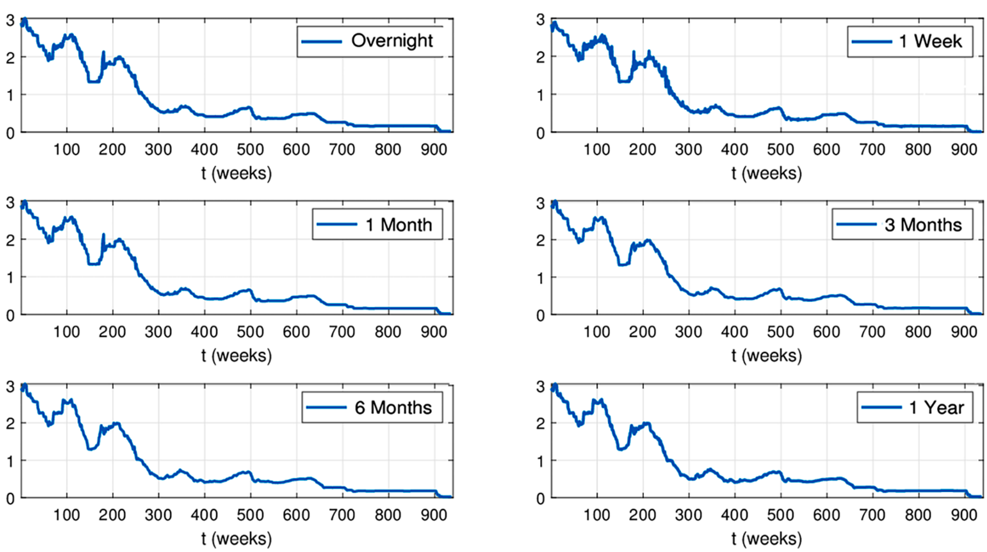


Figure 1. Market data – Polish zloty (PLN) interest rates. The downward trend in interest rates was temporarily interrupted by sudden spikes due to market sentiments of the moment (source: Trading Economics, 2023)

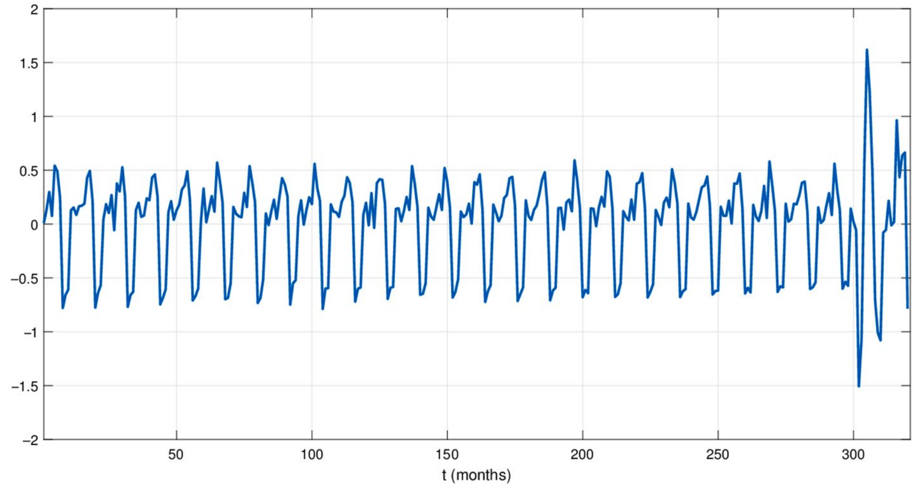


Figure 2. Changes in foreign tourism in Italy (arrivals) from Jan-90 to May-21. The periodic pattern due to seasonality is quite regular until the COVID-19 pandemic around the 300th month (source: Eurostat, 2022)

2.1.2. Statistical characteristics of data

Table 1 and Figure 3 highlight the dissimilarities between the PLN interest rates (changes) and tourism data in terms of moments and distributions, respectively.

Regarding the other statistical properties of the studied time series, Lim and McAleer (2001) underpin the need to check whether data is stationary (i.e. mean and variance constant over time). This is because, in the case of nonstationary (i.e. when there is a unit root), "it is difficult to estimate the mean with any degree of precision because it does not converge to some constant value as the number of observations increases. Hence, the estimated mean will be unreliable and will tend to provide poor forecasts. Moreover, the variance of the forecast error will increase with additional observations" (Lim & McAleer, 2001). Table 2 and Table 3 show that all the considered time series, despite having very different behaviours and statistical moments, share the fact that they have a unit root, thus making the forecast more challenging.

Table 1. First four central moments for (changes) in Polish zloty interest rates and foreign tourism time series

Time series	Mean	Standard Deviation	Skewness	Kurtosis
Polish zloty (Overnight)	8.2888	8.0051	1.1505	2.9170
Polish zloty (1 Week)	8.2069	7.9411	1.1180	2.8073
Polish zloty (1 Month)	8.2888	8.0051	1.1505	2.9170
Polish zloty (3 Months)	8.3611	7.9731	1.1592	2.9587
Polish zloty (6 Months)	8.3404	7.9098	1.1812	3.0319
Polish zloty (1 Year)	8.3651	7.8709	1.1898	3.0658
Foreign tourism	0.0029	0.4310	–0.5565	3.2476

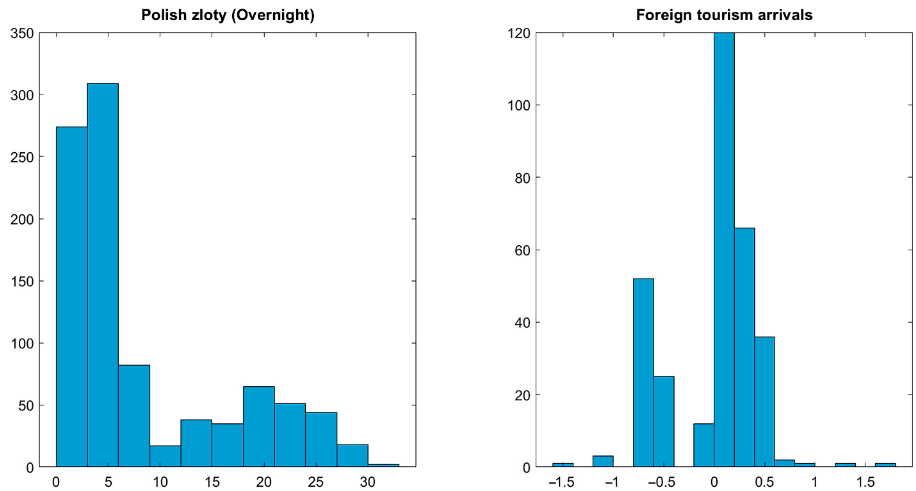


Figure 3. Histograms of changes in Polish zloty and foreign tourism arrival

Table 2. Augmented Dickey-Fuller test on first differences of Polish zloty interest rates and foreign tourism time series. h denotes the test response

Augmented Dicky-Fuller test on first differences							
Stat.\TS	Overnight	1 Week	1 Month	3 Months	6 Months	1 Year	Tourism
h	1	1	1	1	1	1	1
p-Value	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Table 3. Philips Perron test for the presence of unit root on first differences of Polish zloty interest rates and foreign tourism time series. h denotes the test response

Time series	h	p-value	statistic
Polish zloty (Overnight)	1	0.0010	3.8620
Polish zloty (1 Week)	1	0.0035	2.9825
Polish zloty (1 Month)	1	0.0010	3.8620
Polish zloty (3 Months)	1	0.0010	4.8627
Polish zloty (6 Months)	1	0.0011	4.9262
Polish zloty (1 Year)	1	0.0030	4.9023
Foreign tourism	1	0.0010	9.4441

Finally, Table 4 shows that, even when the data is first differentiated, in all cases except 2, there is still autocorrelation which highlights the presence of some periodic pattern obscured by noise. Once again, note that autocorrelation is common to both interest rates and tourism time series.

Table 4. Ljung-Box Q-test test for the presence of unit root on first differences of Polish zloty interest rates and foreign tourism time series. h denotes the test response

Ljung-Box Q-test on first differences (lags from 1 to 5)							
Stat.\TS	Overnight	1 Week	1 Month	3 Months	6 Months	1 Year	Tourism
h	0	1	0	1	1	1	1
p-Value	0.6556	0	0.6556	0	0	0	0

2.2. Models

This section starts with a summary of some key features of the CIR framework and, then, the CIR_{adj}) and the CIR# models are described. Next, we briefly present the reference models already mentioned: EMWA, SARIMA-GARCH, Hull-White, and LSTM.

2.2.1. The CIR model

The Cox-Ingersoll-Ross (CIR) interest rate model (1985) is one of the most popular for pricing discount zero-coupon bonds under no-arbitrage condition. Concerning the Vasicek model (1977), the CIR model permitted non-constant volatility in the stochastic differential equation (SDE) i.e.

$$dr(t) = k\left[\theta - r(t)\right]dt + \sigma\sqrt{r(t)}dW(t), \tag{1}$$

where $r(t)$ is the short-rate at time t, $r(0) = r_0 > 0$ is the initial condition. $\left(W(t)\right)_{t\geq 0}$ denotes a standard Brownian motion, $\sigma > 0$ the volatility of the instantaneous short rate, and $k\left[\theta - r(t)\right]$, is the "mean reverting" drift to make sure that the rate $r(t)$ is dragged to a long-run mean value $\theta > 0$ with velocity $k > 0$. The standard deviation $\sigma\sqrt{r(t)}$ acts as a scaling factor for the random component $W(t)$.

To make a fair comparison, we introduce an improved version of the original CIR model that we call CIR adjusted (CIR_{adj}). In this version, the calibration is performed on an appropriately shifted and partitioned time series resulting from the application of our algorithm on the original data. Concerning the calibration, we apply the estimating function for ergodic diffusion models suggested by Bibby et al. (2010). Orlando et al. (2019a) have shown that the said method provides optimal estimators for the parameters of discretely sampled diffusion-type models. In fact, as the likelihood function is frequently not explicitly known, Bibby et al. (2010) method may suit better than other methods based on the maximum likelihood (ML) estimation (e.g. Kladıvko, 2007).

2.2.2. The CIR# model

In this Section we recap the CIR# model, for more details, a reader can refer to Orlando et al. (2019b). In terms of notation let us say that we use the subscript j for the constants and the superscript j for the vectors, so the shifted weekly interest rate is

$$r_{shift} = \{r_{real,h} + \alpha \,|\, h = 1,\ldots,n\},$$

with a suitable partition, for $j = 1, \ldots, J$

$$r_{shift}^{(j)} = \{r_{shift,h}^{(j)} \mid h = n_{j-1}+1,\dots,n_j\} \ (n_0 = 0),$$

where $\sum_{j=1}^{J} n_j = n$ and $J > 1$ represents the partitioned sub-group (for details on the partitioning see Orlando et al. (2020).

Denote $r_{shift}^{(j)}$ by $r^{(j)}$. The fitted rates $\hat{r}^{(j)} = \{\hat{r}_h^{(j)} \mid h = n_{j-1}+1,\dots,n_j\}$ are given by the Milstein discretization scheme applied to the SDE Eq. (1), for $j = 1, \dots, J$

$$\hat{r}_{h+1}^{(j)} = \hat{r}_h^{(j)} + \hat{k}_j\left(\hat{\theta}_j - \hat{r}_h^{(j)}\right)\Delta + \hat{\sigma}_j\sqrt{\hat{r}_h^{(j)}\Delta}\,\hat{Z}_h^{(j)} + \frac{\left(\hat{\sigma}_j\right)^2}{4}\left[\left(\sqrt{\Delta}\,\hat{Z}_h^{(j)}\right)^2 - \Delta\right], \tag{2}$$

where $\Delta = 1/30$,

$$\hat{\theta}_j = \frac{1}{n_j}\sum_{h=n_{j-1}+1,}^{n_j} r_h^{(j)}, \ \hat{\sigma}_j = \sqrt{\frac{\sum_{h=n_{j-1}+1,}^{n_j}\left(r_h^{(j)} - \hat{\theta}_j\right)^2}{n_j - 1}},$$

$$\hat{k}_j = \min_{k>0} S_j(k) = \min_{k>0}\sqrt{\frac{\sum_{h=n_{j-1}+1,}^{n_j}\left(u_h^{(j)}(k)\right)^2}{n_j - 1}},$$

with $u^{(j)}(k) = \left\{\mathfrak{r}_h^{(j)}(k) - r_h^{(j)} \,\middle|\, h = n_{j-1}+1,\dots,n_j,\ k > 0\right\}$ and $\mathfrak{r}_h^{(j)} : \mathbb{R} \to \mathbb{R}$ such that

$$\mathfrak{r}_{h+1}^{(j)}(k) = \mathfrak{r}_h^{(j)}(k) + k\left(\hat{\theta}_{(j)} - \mathfrak{r}_h^{(j)}\right)\Delta + \hat{\sigma}_j\sqrt{\mathfrak{r}_h^{(j)}\Delta}\,\hat{Z}_h^{(j)} + \frac{\left(\hat{\sigma}_j\right)^2}{4}\left[\left(\sqrt{\Delta}\,\hat{Z}_h^{(j)}\right)^2 - \Delta\right]. \tag{3}$$

Specify that the quantities $\hat{Z}_h^{(j)}$'s in (2) and (3) are Gaussian standardized residuals obtained by an "optimal" ARIMA model chosen as follows. For $j = 1, \dots, J$, consider the set

$$\{Z_h^{(j)} = f\left(\left(r_{h+1}^{(j)} - \hat{r}_{ARIMA,h+1}^{(j)}\left(p_j, i_j, q_j\right) - \mu_j\right)/\eta_j\right) \mid h = n_{j-1}+1,\dots,n_j, \left(p_j, i_j, q_j\right) \in \mathcal{I}_{AC}\},$$

where $f : \mathbb{R} \to \mathbb{R}$ denotes the Johnson transformation (1949), $\hat{r}_{ARIMA,h+1}^{(j)}\left(p_j, i_j, q_j\right)$ is the estimate of $r_{h+1}^{(j)}$ through an ARIMA$\left(p_j, i_j, q_j\right)$ from a set $\mathcal{I}_{AC}$ of suitable models satisfying certain conditions (see Orlando et al. (2019b, Section 4.4.1)), and μ_j, η_j are the mean and the standard deviation of $\{r_h^{(j)} - \hat{r}_{ARIMA,h}^{(j)}\left(p_j, i_j, q_j\right) \mid h = n_{j-1}+1,\dots,n_j\}$, respectively. Then, the "optimal" ARIMA $\left(\hat{p}_j, \hat{i}_j, \hat{q}_j\right)$ model for the *jth*-subgroup is chosen in the set $\mathcal{I}_{AC}$ minimizing the quantity

$$\min_{\hat{r}^{(j)}} \varepsilon_j = \min_{\hat{r}^{(j)}} \sqrt{\frac{1}{n_j}\sum_{h=n_{j-1}+1}^{n_j}\left(r_h^{(j)} - \hat{r}_h^{(j)}\right)^2} \tag{4}$$

with respect to all the samples $\hat{r}^{(j)} = \{\hat{r}_h^{(j)} \mid h = n_{j-1}+1,\dots,n_j\}$ provided by Eq. (2). Hence, the $\hat{Z}_h^j$'s are the residuals of the ARIMA $\left(\hat{p}_j, \hat{i}_j, \hat{q}_j\right)$. These residuals replace the standard Brownian motion, to obtain an exact trajectory of the fitted CIR values instead of a curve obtained with an average of 100,000 simulated trajectories, thus greatly reducing the computational time.

In order to forecast the next rates, firstly, we calibrate the model, i.e. the six parameters (k,θ,σ,p,i,q) through a rolling window w of length m of historical data, say $w=\{r_h,\dots,r_{h+m-1}\}, h\geq 1$, as detailed above, then, the future interest rates $r^F_{h+m+s}, s\geq 0$, may be computed with the procedure described in Orlando et al. (2019b).

2.2.3. The EWMA

The Exponentially Weighted Moving Average (EWMA) (Perry, 2010) is a weighting scheme to predict future values averaging on historical data. The EWMA weights decrease exponentially at a fixed rate λ the observations that are far in the past, i.e.,

$$r_t^F=\lambda r_t+(1-\lambda)r_{t-1}.$$

Generally, the EWMA is used when no model is assumed. The EWMA, also, could be suitable for predictions over a short horizon as it is able to capture the changes in volatility.

2.2.4. The Hull-White model

The single-factor, non-autonomous model by Hull-White (HW) (1990) is among the best known in finance. In the HW model, the dynamics of the short interest rate $r(t)$ are given by

$$dr(t)=\big(\theta(t)-\alpha r(t)\big)dt+\sigma dW(t)\quad (r_0>0). \tag{5}$$

In Eq. (5), α, σ and θ represent, respectively, the strength of the mean reversion, the volatility and the long-run mean, that is a function of time. $\theta(t)$ has to match the term structure of zero-coupon bond prices (or, equivalently, the yields). Since θ depends on time, the Hull-White model may be considered a generalization of the Vasicek model.

The solution of Eq. (5) is given by

$$r(t)=e^{-\alpha t}r(0)+\int_0^t e^{\alpha(u-t)}\theta(u)du+\sigma e^{-\alpha t}\int_0^t e^{\alpha u}\,dW(u), \tag{6}$$

in particular, $r(t)$ is normally distributed, with

$$\mathbb{E}\big[r(t)\big]=e^{-\alpha t}r(0)+\int_0^t e^{\alpha(u-t)}\theta(u)du,$$

and

$$\mathrm{Var}\big(r(T)\big)=\frac{\sigma^2}{2\alpha}\big(1-e^{2\alpha t}\big).$$

In accordance with its economic meaning, $\theta(u)$ is chosen like the EWMA time series. Then, the integral in Eq. (6) is approximated by the trapezoidal rule, i.e.,

$$\int_0^t e^{\alpha(u-t)}\theta(u)du\simeq\frac{1}{2}\sum_{i=1}^{N-1}\Big(e^{\alpha(u_i-t)}\theta(u_i)+e^{\alpha(u_{i+1}-t)}\theta(u_{i+1})\Big)(u_{i+1}-u_i),$$

where $0=u_1\leq u_2\leq\ldots\leq u_N=t$. The unknown term $\theta(u_N)$ can be obtained through the EWMA prediction at time u_N. The remaining parameters α, σ are computed by solving the optimization problem

$$\min_{(\alpha,\sigma)}\sum_i(r_i-r_i^{HW}(\alpha,\sigma))^2,$$

where r_i, r_i^{HW} represent the market rates and the Hull-White simulated rates (defined in Eq. (6)), respectively.

Lastly, we predict the future interest rate $r(t)$ by the conditioned expectation

$$\mathbb{E}\left[r(t) \mid r(s)\right] = e^{-\alpha(t-s)} r(s) + \int_s^t e^{\alpha(u-t)} \theta(u) du, \;\; 0 \le s < t.$$

2.2.5. The SARIMA-GARCH model

Autoregressive integrated moving average (ARIMA) models are widely employed in statistics and econometrics for time series analysis. ARIMA models fit well non-stationary (in mean), seasonal data. The AR part of the ARIMA model indicates a regression on its own lagged values while the MA part indicates the regression error in terms of a linear combination of error terms. The I stands for the differencing step (i.e. the "integrated" part of the model). The ARIMA models can be estimated following the Box–Jenkins approach (see Asteriou & Hall, 2011). SARIMA models are a variant for seasonal time series such as tourism data.

Generalized autoregressive conditional heteroskedasticity (GARCH) model are widely employed in modelling time series that exhibit volatility clustering and time- varying volatility (see Bollerslev, 2008).

For the proposed analysis, a combined SARIMA-GARCH was used to exploit the characteristics of both models.

2.2.6. Holt-Winters model

Holt (1957)[1] and then Winters (1960) introduced the so-called Holt-Winters model to describe seasonality in data. The model is composed of the three smoothing equations for the levels the trend and the seasonal component. Depending on the nature of the seasonal component there are two different specifications of the Holt-Winters model. The additive specification is when "the seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation the series is seasonally adjusted by subtracting the seasonal component" (Hyndman & Athanasopoulos, 2018). The multiplicative version of the model, is when "the seasonal component is expressed in relative terms (percentages), and the series is seasonally adjusted by dividing through by the seasonal component".

The additive method is preferred when "the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportionally to the level of the series" (Hyndman & Athanasopoulos 2018). As part of this study, since data is almost constant, it was found that the most performing method is the additive one.

2.2.7. Machine learning (LSTM network)

To compare the aforementioned models with the currently widespread machine learning methods, we considered the short-term memory network (LSTM) which updates the state of the network cyclically over time (see Kudo et al., 1999; Qadeer et al., 2020). The LSTM is a recurrent neural network (RNN) and the algorithm is such that, at all times, the network

[1] Reprinted in Holt (2004).

status contains information stored on all previous time passages. Then LSTM predicts the next values of a given time series or data sequence. Learning an LSTM network consists of a regression, in which at each time step of the input sequence, the LSTM network optimizes the predicted value of the next time step.

There are two forecasting methods: open loop and closed loop. The open loop forecast uses only the input data for forecasting and updating the network. The closed-loop forecast, on the other hand, uses the previous forecasts as input. In our case, we implemented the former in Matlab (2022) as it provides more accurate predictions. The set-up we adopted is 1/3 of data for training and 2/3 for forecasting.

2.3. Forecasting accuracy

To test if the forecasts closely fit real-world data, we use the normalized root mean square error version (NRMSE) and the percentage of mean absolute error (MAPE) to measure the accuracy of the prediction. The directionality, if correctly anticipated, is tested with the "success" criterion (IDX).

2.3.1. Root mean square error (RMSE)

The RMSE is the square root of the mean square error (MSE), i.e.,

$$RMSE = \sqrt{\frac{1}{n}\sum_{h=1}^{n} e_h^2},$$

where e_h denotes the residuals between the observations and their predictions, over n times. Hence, a value of 0 indicates a perfect fit of the data.

The RMSE is sensitive to outliers because it depends on the scale of the observed data so, for the sake of comparison, we take the normalized root mean square error (NRMSE):

$$NRMSE = \frac{RMSE}{r_{\max} - r_{\min}},$$

where $r_{\max}$ and $r_{\min}$ correspond to the maximum and the minimum value of the sample.

2.3.2. Mean absolute percentage error (MAPE)

The mean absolute percentage error (MAPE) measures the accuracy of the forecasting model and takes the form:

$$MAPE = \frac{}{n}\sum_{h=}^{n} \left|\frac{e_h}{r_h}\right|.$$

2.3.3. Directionality of forecasting (IDX)

As mentioned, we are not only interested in forecasting error but, also, in understanding whether the CIR# can correctly predict an increase or decrease in the time series. The index of directionality (IDX) is aimed at exploring that. By r_t we denote the value at time t and by r_t^F its corresponding forecast. With $\alpha_{t+1} := r_{t+1} - r_t$ we define the difference between two consecutive interest rates, and with $\beta_{t+1} := r_{t+1}^{F} - r_t$ difference between the forecast at time $t + 1$ and the actual interest rates at time t. Then, the boolean variable $H(t + 1)$ corresponds to

$$H(t+1) = 1 \text{ if } \operatorname{sgn}(\alpha_{t+1}) = \operatorname{sgn}(\beta_{t+1}),$$
$$H(t+1) = 0 \text{ if } \operatorname{sgn}(\alpha_{t+1}) \neq \operatorname{sgn}(\beta_{t+1}).$$

$H(t + 1) = 1$ when the forecast coincides (in sign) with the actual occurrence. Thus, the forecasting "success" index (IDX) represents the mean of the $H(t + 1)$ values on the number of forecasts over a time series of length T. In other terms is

$$\text{IDX} := \frac{1}{T-1}\sum_{t=1}^{T-1} H(t+1),$$

where IDX expresses the percentage of correct directional predictions.

3. Results

Unless otherwise stated, forecasts are calculated based on the frequency of the data. This means that for interest rates (weekly frequency) the forecast is the next data point (i.e. the true value one week ahead). For tourism (monthly frequency) the forecast corresponds to the value of the following month.

In this section we show the out-of sample results (forecasts) for PLN interest rates and tourism data. Regarding the first dataset, we compare the CIR# versus the baseline models providing both graphical and statistical evidence that the CIR# model betters the others by any metric (i.e. NMRSE, MAPE, and IDX) and tenor. This holds true, also, for different forecast horizons (i.e. 1 week, 4 and 12 weeks).

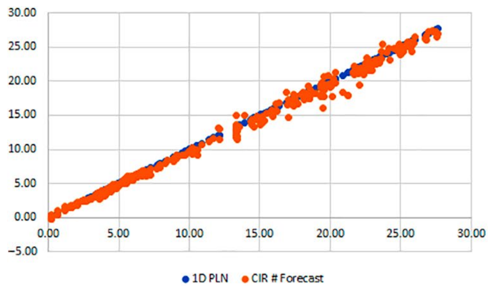
Regarding tourism forecasts (arrivals), among the reference considered models we have included the Holt-Winters model specifically designed for this type of periodic data. However, as the COVID-19 pandemic has severely disrupted the industry, the model may be affected by a lack of accuracy. In our simulations, the MAPE is between 6% and 8.5% which is not very far from the result obtained by other authors in the literature. However, we found out that the CIR# substantially overperforms the Holt-Winters as the MAPE goes from 0.7% (1 month ahead forecast) to 2% (6 months ahead forecast).

3.1. Interest rates forecasting

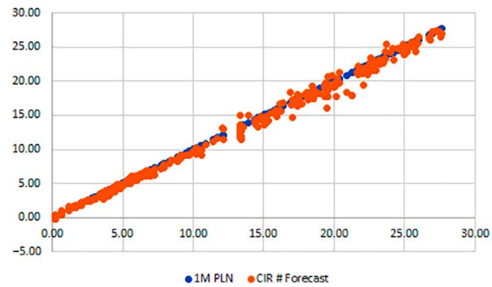
We start by graphing the actual data against their forecast for the overnight (Figure 4a), 1 month (Figure 4b), 3 months (Figure 4c), and 1-year tenors (Figure 4d). Figure 4 shows that the predictions closely follow the realizations for all tenors and that they are not scattered.

Next, as all tenors display a similar path, we focus our attention on the overnight because it is the most erratic and difficult to predict. Figure 5 compares the forecasts for all considered models on the overnight tenor. By visual inspection, the CIR# is the most accurate. To confirm that, Table 5 compares the CIR# versus the considered baseline models i.e. CIR_{adj}, EWMA, Hull-White (HW) and LSTM. The measures adopted are: MAPE for forecasting error, NRMSE for accuracy error, and IDX for directionality error. As demonstrated, for all of the time series considered, the CIR# model outperforms the benchmark models.

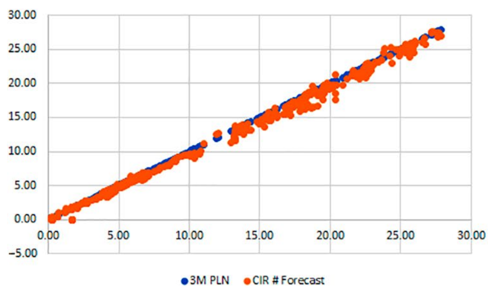
a) PLN overnight. Actual rates versus (abscissa) forecasted values (ordinate)



b) 1 month PLN interest rate. Actual rates versus (abscissa) forecasted values



c) 3 months PLN interest rate. Actual rates versus (abscissa) forecasted values (ordinate)



d) 1 year PLN interest rate. Actual rates versus (abscissa) forecasted values (ordinate)

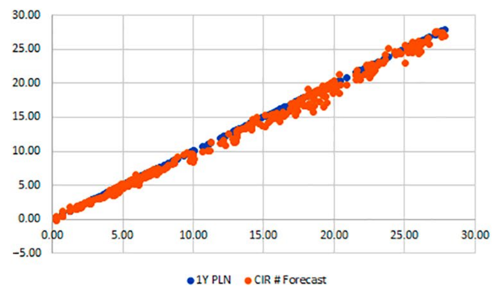


Figure 4. Comparisons for Polish PLN interest rates between different tenors against their corresponding forecasts. Period: 1995-01-02 to 2021-02-05

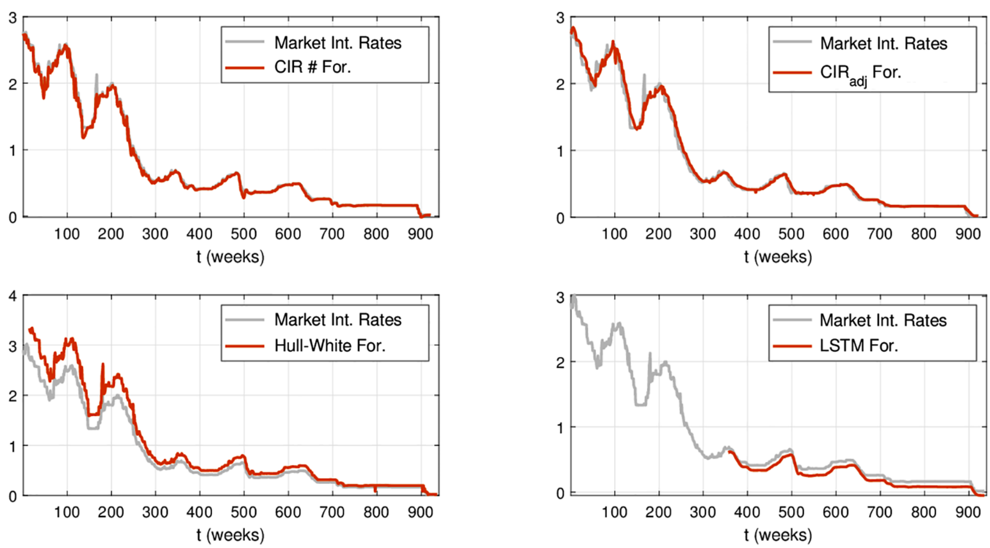


Figure 5. Overnight – Polish zloty (PLN). Real data versus forecasts obtained by the CIR# and the baseline models. Notice that LSTM forecasts start after training (based on 1/3 of the dataset. Period: 1995-01-02 to 2021-02-05)

The Tukey's Honestly Significant Difference procedure utilized in the analysis is most appropriate for samples of equal size and balanced one-way ANOVA. However, when samples are of different sizes, it becomes a conservative one-way ANOVA. The reliability of the Tukey's Honestly Significant Difference procedure in situations where the variables being compared are correlated has not been proven. However, the Tukey-Kramer conjecture suggests that this procedure may still be applicable in such scenarios, as noted by Hochberg and Tamhane (1989). The results of the comparisons between the models being considered and the actual data are presented in Table 6, displaying the obtained p-values. The table shows that, although there is no indication that the mean responses for the CIR# and CIR_{adj} differ significantly from the real data, the same cannot be said for the Hull-White model and the LSTM network.

Finally, to appreciate how accuracy can vary with the forecast horizon, in Table 7 we have taken one week tenor and predicted the next value in one week, four weeks and twelve weeks. Given the simplicity of the EWMA, for reasons of space, we do not show the result for the model as well as we do not show the results for the other tenors because they are similar.

Table 5. Forecasting accuracy: CIR# vs. CIR_{adj}, EWMA, Hull-White and LSTM. Period: 1995-01-02 to 2021-02-05

Measure	Tenor	CIR#	CIR_{adj}	EWMA	Hull-White	LSTM
MAPE	Overnight	0.0409	0.0757	0.2172	0.0856	0.1344
NRMSE		0.0168	0.0266	0.0856	0.0281	0.0273
IDX		72.14%	56.95%	54.69%	59.63%	63.44%
MAPE	1 Week	0.0519	0.0920	0.2187	0.1053	0.1618
NRMSE		0.0205	0.0280	0.0872	0.0284	0.0278
IDX		67.68%	58.97%	55.93%	57.01%	62.61%
MAPE	1 Month	0.0409	0.0757	0.2172	0.0856	0.1320
NRMSE		0.0168	0.0266	0.0856	0.0281	0.0273
IDX		72.14%	59.63%	54.95%	56.69%	66.35%
MAPE	3 Months	0.0401	0.0777	0.2139	0.0820	0.1357
NRMSE		0.0288	0.0331	0.0847	0.0275	0.0284
IDX		73.99%	59.63%	57.12%	58.10%	60.70%
MAPE	6 Months	0.0405	0.0700	0.2140	0.0801	0.1279
NRMSE		0.0169	0.0254	0.0843	0.0279	0.0298
IDX		79.32%	67.57%	56.58%	59.84%	68.19%
MAPE	1 Year	0.0415	0.0708	0.2143	0.0812	0.140
NRMSE		0.0172	0.0256	0.0843	0.0281	0.0314
IDX		78.99%	68.00%	56.47%	59.95%	60.80%

Table 6. Tukey's Honestly Significant Difference: CIR# CIR_{adj} , Hull-White and LSTM vs. real data (interest rates). Period: 1995-01-02 to 2021-02-05

p-values of 1-week ahead predictions

Tenor	CIR#	CIR_{adj}	Hull-White	LSTM
Overnight	0.60432	0.62882	0.00003	0.00083
1 Week	0.70499	0.62817	0.00003	0.00006
1 Month	0.60432	0.62882	0.00003	0
3 Months	0.90816	0.57571	0.00002	0.00884
6 Months	0.50936	0.61708	0.00002	0
1 Year	0.49922	0.62167	0.00002	0.00025

Table 7. Forecasting accuracy: CIR# vs. CIR_{adj}, EWMA, Hull-White and LSTM. Period: 1995-01-02 to 2021-02-05

Err. Measure	Tenor	CIR#	CIR_{adj}	EWMA	Hull-White	LSTM
MAPE	Overnight	0.0409	0.0757	0.2172	0.0856	0.1344
NRMSE		0.0168	0.0266	0.0856	0.0281	0.0273
IDX		72.14%	56.95%	54.69%	59.63%	63.44%
MAPE	1 Week	0.0519	0.0920	0.2187	0.1053	0.1618
NRMSE		0.0205	0.0280	0.0872	0.0284	0.0278
IDX		67.68%	58.97%	55.93%	57.01%	62.61%
MAPE	1 Month	0.0409	0.0757	0.2172	0.0856	0.1320
NRMSE		0.0168	0.0266	0.0856	0.0281	0.0273
IDX		72.14%	59.63%	54.95%	56.69%	66.35%
MAPE	3 Months	0.0401	0.0777	0.2139	0.0820	0.1357
NRMSE		0.0288	0.0331	0.0847	0.0275	0.0284
IDX		73.99%	59.63%	57.12%	58.10%	60.70%
MAPE	6 Months	0.0405	0.0700	0.2140	0.0801	0.1279
NRMSE		0.0169	0.0254	0.0843	0.0279	0.0298
IDX		79.32%	67.57%	56.58%	59.84%	68.19%
MAPE	1 Year	0.0415	0.0708	0.2143	0.0812	0.140
NRMSE		0.0172	0.0256	0.0843	0.0281	0.0314
IDX		78.99%	68.00%	56.47%	59.95%	60.80%

3.2. Tourism demand forecasting

As mentioned, we intend to test the performance of the CIR# model for fore- casting tourism demand. Figure 6 and Table 8 demonstrate that the CIR# model not only outperforms the other models but, also, copes well with the shock caused by the COVID-19 pandemic. For example, the minimum MAPE for a number of countries (from France, to Italy, from Japan to the USA, etc.) reported by Claveria et al. (2017) is 4%. Similarly, Supriatna et al. (2019), in predicting the number of international tourists arriving in Indonesia, from 2013 to 2017 (thus excluding the period of the COVID-19 pandemic) had a MAPE of 4.7098%. The results obtained on the complete dataset exhibit inferior performance as they incorporate the impact of the pandemic-induced disruption on the tourism industry. The (one month ahead) MAPE for the Holt-Winters model is 7% versus 0.7% of the CIR# model.

Table 9 confirms the statistical robustness of the correspondence between data and forecasts and, regarding the period of the COVID-19 pandemic, Table 10 provides a zoom-in for the accuracy of all considered models. Once again, the CIR# appears to be the best.

Table 8. Forecasting accuracy: CIR# vs. CIR_{adj}, EWMA, SARIMA-GARCH, Holt-Winters models and LSTM neural network. Period: Jan-90 to May-21

	Forecast Horizon	MAPE	NRMSE	IDX		MAPE	NRMSE	IDX
CIR#	1 Month	0.0068	0.0064	96.71%	CIR_{adj}	0.0604	0.0956	78.61%
	3 Months	0.0037	0.0070	99.91%		0.0179	0.0452	87.91%
	6 Months	0.0216	0.0166	51.35%		0.0532	0.0384	51.31%
EWMA	1 Month	0.1022	0.1217	68.42%	SARIMA-GARCH	0.1253	0.0733	81.90%
	3 Months	0.0818	0.1714	70.32%		0.1899	0.0756	75.82%
	6 Months	0.1757	0.1641	50.00%		0.1284	0.2303	27.63%
Holt-Winters	1 Month	0.0737	0.0616	87.50%	LSTM	0.0982	0.1698	80.28%
	3 Months	0.0844	0.1093	92.30%		0.0868	0.1912	76.74%
	6 Months	0.0594	0.1705	44.73%		0.0950	0.1923	72.25%

Table 9. Tukey's Honestly Significant Difference: CIR# vs. CIR_{adj}, EWMA, SARIMA-GARCH, Holt-Winters models and LSTM vs. real data (tourism). Period: Jan-90 to May-21

p-values of 1-month ahead predictions (period Jan-90 to May-21)					
CIR#	CIR_{adj}	EWMA	SARIMA-GARCH	Holt-Winters	LSTM
0.9527	0.7084	0.8898	0.7084	0.9813	0.8471

Table 10. Forecasting accuracy: CIR# vs. CIR_{adj}, EWMA, SARIMA-GARCH, Holt-Winters models and LSTM neural network. Period: Jan-19 to May-21

	Forecast Horizon	MAPE	NRMSE	IDX		MAPE	NRMSE	IDX
CIR#	1 Month	0.0096	0.0076	95.91 %	CIR_{adj}	0.0528	0.1100	75.55%
	3 Months	0.0046	0.0107	90.02%		0.0193	0.0406	93.75%
	6 Months	0.0471	0.0222	100.00%		0.0428	0.0247	100.00%
EWMA	1 Month	0.1218	0.1642	67.34%	SARIMA-GARCH	0.1494	0.0977	83.67%
	3 Months	0.1120	0.2145	75.01%		0.1258	0.0797	75.04%
	6 Months	0.4273	0.2678	100.00%		0.1963	0.2297	42.85%
Holt-Winters	1 Month	0.0819	0.1281	81.63%	LSTM	0.1608	0.1720	83.33%
	3 Months	0.1013	0.1825	90.00%		0.5671	0.2776	73.33%
	6 Months	0.0850	0.2472	100.00%		0.3655	0.4232	66.70%

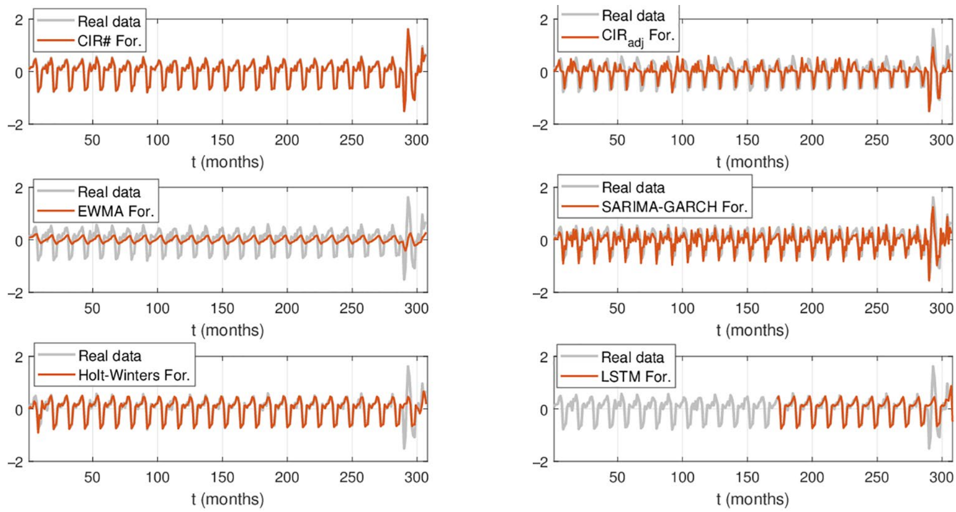


Figure 6. CIR# forecasts of foreign tourism versus baseline models. Period: Jan-90 to May-21. Notice that LSTM forecasts start after training (based on 1/3 of the dataset)

Conclusions

Over time, numerous extensions of the original CIR model have been proposed to address its limitations. These extensions encompass a broad range of variations, from one-factor models with time-varying coefficients to jump diffusions and multi-factor models. In most cases, these models maintain the positivity of interest rates, but some of them may not retain the analytical tractability of the CIR model. The first contribution to the literature is that we have shown how the CIR# model, while staying within the original framework by Cox, Ingersoll and Ross, it can deal with cluster volatility, jumps and low to-negative interest rates in time series. To confirm that, the model was tested on different tenors during both turbulent calmer periods on both accounts: forecast errors and directionality of interest rates. These findings on the PLN confirm previous research made on currencies from developed economies (from money market to one year rates) thus corroborating the ability of the CIR# model to tackle a wide range of interest rate curves.

Similarly, the CIR# model, when tested for forecasting tourism demand, outperformed all reference models. Furthermore, we have given evidence of the ability of the CIR# model to cope well with the shock to the tourism industry caused by the COVID-19 pandemic where established models may lose accuracy. This result provides the second contribution to the literature showing the forecasting power of the proposed model for fields outside of interest rates for which the CIR# model was originally conceived.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors have no conflicts of interest to declare.

References

Akimov, A., Lee, C. L., & Stevenson, S. (2019). Interest rate sensitivity in European public real estate markets. *Journal of Real Estate Portfolio Management*, *25*(2), 138–150. https://doi.org/10.1080/10835547.2020.1803694

Allcock, J. B. (1989). Seasonality. In Witt, S. F. & Moutinho, L. (Eds.), *Tourism marketing and management handbook* (pp. 387–392). Prentice Hall.

Ampountolas, A. (2021). Modeling and forecasting daily hotel demand: A comparison based on SARIMAX, neural networks, and GARCH models. *Forecasting*, *3*(3), 580–595. https://doi.org/10.3390/forecast3030037

Ascione, G., Mehrdoust, F., Orlando, G., & Samimi, O. (2023). Foreign exchange options on Heston-CIR model under Levy process framework. *Applied Mathematics and Computation*, *446*, 1–31. https://doi.org/10.1016/j.amc.2023.127851

Asteriou, D., & Hall, S. G. (2011). ARIMA models and the Box–Jenkins methodology. *Applied Econometrics*, *2*(2), 265–286.

Bank for International Settlements. (2015). *Is the unthinkable becoming routine*? Technical report. BIS.

Baum, T., & Lundtorp, S. (2001). Seasonality in tourism: Issues and implications. In *Seasonality in Tourism* (pp. 13–30). Routledge. https://doi.org/10.4324/9780080516806-6

Bibby, B. M., Jacobsen, M., & Sørensen, M. (2010). Estimating functions for discretely sampled diffusion-type models. In Y. Ait-Sahalia & L. P. Hansen (Eds.), *Handbook of financial econometrics: Tools and techniques* (vol. 1, pp. 203–268). North-Holland, Oxford. https://doi.org/10.1016/B978-0-444-50897-3.50007-9

Bjørnland, H. C., & Hungnes, H. (2006). The importance of interest rates for forecasting the exchange rate. *Journal of Forecasting*, *25*(3), 209–221. https://doi.org/10.1002/for.983

Bollerslev, T. (2008). *Glossary to ARCH* (*GARCH*). CREATES Research paper 2008-49. SSRN. https://doi.org/10.2139/ssrn.1263250

Brigo, D., & El-Bachir, N. (2006). *Credit derivatives pricing with a smile-extended jump stochastic intensity model* (ICMA Centre Discussion Papers in Finance DP2006-13). SSRN. https://doi.org/10.2139/ssrn.950208

Brigo, D., & Mercurio, F. (2000). The CIR++ model and other deterministic- shift extensions of short rate models. In *Proceedings of the 4th Columbia-JAFEE Conference for Mathematical Finance and Financial Engineering* (pp. 563–584). https://doi.org/10.2139/ssrn.292060

Brigo, D., & Mercurio, F. (2001). A deterministic-shift extension of analytically-tractable and time-homogeneous short rate models. *Finance and Stochastics*, *5*, 369–387. https://doi.org/10.1007/PL00013541

Brigo, D., & Mercurio, F. (2006). *Interest rate models – Theory and practice: With smile, inflation and credit* (2nd ed.). Springer-Verlag.

Butler, R. (1998). Seasonality in tourism: Issues and implications. *The Tourist Review*, *53*(3), 18–24. https://doi.org/10.1108/eb058278

Carmona, R. A., & Tehranchi, M. R. (2006). *Interest rate models: An infinite dimensional stochastic analysis perspective*. Springer-Verlag.

Chang, Y.-W., & Liao, M.-Y. (2010). A seasonal ARIMA model of tourism forecasting: The case of Taiwan. *Asia Pacific Journal of Tourism Research*, *15*(2), 215–221. https://doi.org/10.1080/10941661003630001

Chen, L. (1996). *Stochastic mean and stochastic volatility: A three-factor model of the term structure of interest rates and its applications and its applications in derivatives pricing and risk management*. Blackwell Publishers. https://doi.org/10.1007/978-3-642-46825-4_1

Choden, & Unhapipat, S. (2018). ARIMA model to forecast international tourist visit in Bumthang, Bhutan. *Journal of Physics: Conference Series*, *1039*, 012023. https://doi.org/10.1088/1742-6596/1039/1/012023

Claveria, O., Monte, E., & Torra, S. (2017). Data pre-processing for neural network-based forecasting: Does it really matter? *Technological and Economic Development of Economy*, *23*(5), 709–725. https://doi.org/10.3846/20294913.2015.1070772

Corluka, G. (2019). Tourism seasonality – an overview. *Journal of Business Paradigms*, *4*(1), 21–43.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, *53*(2), 385–407. https://doi.org/10.2307/1911242

Duffie, D. (2005). Credit risk modeling with affine processes. *Journal of Banking & Finance*, *29*(11), 2751–2802. https://doi.org/10.1016/j.jbankfin.2005.02.006

Dutta, A., Mishra, T., Uddin, G. S., & Yang, Y. (2021). Brexit uncertainty and volatility persistence in tourism demand. *Current Issues in Tourism*, *24*(16), 2225–2232. https://doi.org/10.1080/13683500.2020.1822300

Engelen, K. C. (2015). The unthinkable as the new normal. *The International Economy*, *29*(3), 30.

Eurostat. (2022). *Database – Tourism*. https://ec.europa.eu/eurostat/web/tourism/data/database

Grundey, D. (2008). Managing sustainable tourism in Lithuania: Dream or reality? *Technological and Economic Development of Economy*, *14*(2), 118–129. https://doi.org/10.3846/1392-8619.2008.14.118-129

Gruppe, M., Basse, T., Friedrich, M., & Lange, C. (2017). Interest rate convergence, sovereign credit risk and the European debt crisis: A survey. *Journal of Risk Finance*, *18*(4), 432–442. https://doi.org/10.1108/JRF-01-2017-0013

He, K., Ji, L., Wu, C. W. D., & Tso, K. F. G. (2021). Using SARIMA–CNN–LSTM approach to forecast daily tourism demand. *Journal of Hospitality and Tourism Management*, *49*, 25–33. https://doi.org/10.1016/j.jhtm.2021.08.022

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, *6*(2), 327–343. https://doi.org/10.1093/rfs/6.2.327

Hochberg, Y., & Tamhane, A. C. (1989). *Multiple comparison procedures*. John Wiley & Sons.

Holt, C. C. (1957). *Forecasting seasonals and trends by exponentially weighted moving averages* (Office of Naval Research Memorandum, vol. 52). Carnegie Institute of Technology.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Hull, J., & White, A. (1990). Pricing interest-rate-derivative securities. *The Review of Financial Studies*, *3*(4), 573–592. https://doi.org/10.1093/rfs/3.4.573

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, *36*(1/2), 149–176. https://doi.org/10.1093/biomet/36.1-2.149

Keller-Ressel, M., & Steiner, T. (2008). Yield curve shapes and the asymptotic short rate distribution in affine one-factor models. *Finance and Stochastics*, *12*(2), 149–172. https://doi.org/10.1007/s00780-007-0059-z

Keynes, J. M. (1936). The general theory of the rate of interest. In *The General theory of employment, interest, and money* (pp. 145–153). Springer. https://doi.org/10.1007/978-3-319-70344-2_13

Kladıvko, K. (2007). *Maximum likelihood estimation of the Cox-Ingersoll- Ross process: The Matlab implementation*. https://it.mathworks.com/matlabcentral/fileexchange/37297-maximum-likelihood-estimation\-of-the-cox-ingersoll-ross-process-the-matlab-implementation

Kudo, M., Toyama, J., & Shimbo, M. (1999). Multidimensional curve classification using passing-through regions. *Pattern Recognition Letters*, *20*(11), 1103–1111. https://doi.org/10.1016/S0167-8655(99)00077-X

Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. *Tourism Management*, *83*, 104245. https://doi.org/10.1016/j.tourman.2020.104245

Li, Y., & Cao, H. (2018). Prediction for tourism flow based on LSTM neural network. *Procedia Computer Science*, *129*, 277–283. https://doi.org/10.1016/j.procs.2018.03.076

Lim, C., & McAleer, M. (2001). Forecasting tourist arrivals. *Annals of Tourism Research*, *28*(4), 965–977. https://doi.org/10.1016/S0160-7383(01)00006-8

Lock, S. (2022). *Global tourism industry – statistics & facts*. https://www.statista.com/topics/962/global-tourism/#dossierContents__outerWrapper

Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. *Complexity*, *2020*, 6622927. https://doi.org/10.1155/2020/6622927

MATLAB. (2022). *MathWorks*. https://www.mathworks.com/products/matlab.html

Mininni, M., Orlando, G., & Taglialatela, G. (2020). Challenges in approximating the black and scholes call formula with hyperbolic tangents. *Decisions in Economics and Finance*, 1–28. https://doi.org/10.1007/s10203-020-00305-8

Moreno, M., & Platania, F. (2015). A cyclical square-root model for the term structure of interest rates. *European Journal of Operational Research*, *241*(1), 109–121. https://doi.org/10.1016/j.ejor.2014.08.010

Najafi, A. R., & Mehrdoust, F. (2017). Bond pricing under mixed generalized CIR model with mixed Wishart volatility process. *Journal of Computational and Applied Mathematics*, *319*(C), 108–116. https://doi.org/10.1016/j.cam.2016.12.039

Orlando, G., & Bufalo, M. (2021). Interest rates forecasting: Between Hull and White and the CIR#. How to make a single factor model work. *Journal of Forecasting*, *40*(8), 1566–1580. https://doi.org/10.1002/for.2783

Orlando, G., Mininni, R. M., & Bufalo, M. (2018). A new approach to CIR short-term rates modelling. In Mili, M., Samaniego Medina, R., & di Pietro, F. (Eds.), *New methods in fixed income modeling – fixed income modeling* (pp. 35–44). Springer International. https://doi.org/10.1007/978-3-319-95285-7_2

Orlando, G., Mininni, R. M., & Bufalo, M. (2019a). Interest rates calibration with a CIR model. *The Journal of Risk Finance*, *20*(4), 370–387. https://doi.org/10.1108/JRF-05-2019-0080

Orlando, G., Mininni, R. M., & Bufalo, M. (2019b). A new approach to forecast market interest rates through the CIR model. *Studies in Economics and Finance*, *37*(2), 267–292. https://doi.org/10.1108/SEF-03-2019-0116

Orlando, G., Mininni, R. M., & Bufalo, M. (2020). Forecasting interest rates through Vasicek and CIR models: A partitioning approach. *Journal of Forecasting*, *39*(4), 569–579. https://doi.org/10.1002/for.2642

Perry, M. B. (2010). *The exponentially weighted moving average*. Wiley Encyclopedia of Operations Research and Management Science. https://doi.org/10.1002/9780470400531.eorms0314

Polyzos, S., Samitas, A., & Spyridou, A. Ef. (2021). Tourism demand and the COVID-19 pandemic: An LSTM approach. *Tourism Recreation Research*, *46*(2), 175–187. https://doi.org/10.1080/02508281.2020.1777053

Qadeer, K., Rehman, W. U., Sheri, A. M., Park, I., Kim, H. K., & Jeon, M. (2020). A Long Short-Term Memory (LSTM) network for hourly estimation of $PM_{2.5}$ concentration in two cities of South Korea. *Applied Sciences*, *10*(11), 3984. https://doi.org/10.3390/app10113984

Santamaria, D., & Filis, G. (2019). Tourism demand and economic growth in Spain: New insights based on the yield curve. *Tourism Management*, *75*, 447–459. https://doi.org/10.1016/j.tourman.2019.06.008

Schwarzbach, C., Kunze, F., Rudschuck, N., & Windels, T. (2012). Asset management in the German insurance industry: The quality of interest rate forecasts. *Zeitschrift für die gesamte Versicherungswissenschaft*, *101*(5), 693–703. https://doi.org/10.1007/s12297-012-0218-y

Shanika, A., & Jahufer, A. (2021). *Volatility analysis of international tourist arrivals to Sri Lanka using GARCH models*. Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sammanthurai.

Supriatna, A., Lesmana, E., Aridin, L., Sukono, & Napitupulu, H. (2019). Comparison between multiplicative Holt Winter and decomposition method in predicting the number of incoming international tourists to Indonesia. *IOP Conference Series: Materials Science and Engineering*, *567*(1), 012047. https://doi.org/10.1088/1757-899X/567/1/012047

Thakur, B. P. S., Kannadhasan, M., & Goyal, V. (2018). Determinants of corporate credit spread: Evidence from India. *Decision*, *45*(1), 59–73. https://doi.org/10.1007/s40622-018-0179-7

Trading Economics. (2023). *Polish Zloty - 2023 Data – 1993–2022 Historical - Quote*. https://tradingeconomics.com/poland/currency

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, *5*(2), 177–188. https://doi.org/10.1016/0304-405X(77)90016-2

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*(3), 324–342. https://doi.org/10.1287/mnsc.6.3.324

Yıldırım, D. C., Toroslu, I. H., & Fiore, U. (2021). Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. *Financial Innovation*, *7*(1), 1–36. https://doi.org/10.1186/s40854-020-00220-2

Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). LSTM network: A deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, *11*(2), 68–75. https://doi.org/10.1049/iet-its.2016.0208

Zhu, L. (2014). Limit theorems for a Cox-Ingersoll-Ross process with Hawkes jumps. *Journal of Applied Probability*, *51*(3), 699–712. https://doi.org/10.1239/jap/1409932668

Zong, C.-L., & Wang, L. (2018). Prediction of urban residents' travel rate in China based on ARIMA models. *Journal of Interdisciplinary Mathematics*, *21*(5), 1285–1290. https://doi.org/10.1080/09720502.2018.1497999