Stochastic Models of Short-term Interest Rates

Vazgen Badalyan under supervision of Boris Demeshev
National Research University - Higher School of Economics

June 2012

Abstract

In this paper I decided to investigate some of the basic stochastic one-factor short-term interest rate models, including those of Vasicek (1977), Cox, Ingersoll & Ross (1985) and Dothan (1978), which were assessed in terms of how well they fit actual data (LIBOR and EURIBOR datasets) and forecast future values (EURIBOR dataset).

1 Introduction

Many renowned academics have been trying to apply the techniques of stochastic calculus to various seemingly random events that people face in their lives for a considerable amount of time. For those interested in finance the possible application of such tools to the prediction of the behaviour of different financial indicators, such as key interest rates, is still very intriguing, especially taking into account the recent events surrounding the European sovereign debt crisis and the extra attention of the global investors to the borrowing costs of the governments and inter-bank interest rates in the European economies. The ability to construct a consistent model that would enable one to predict the likely directions and magnitudes of change in the inter-bank interest rates with reasonable precision is very likely to shape an absolutely new approach to the entire field of finance, so realizing the high relevance of the mentioned topic, in this paper I decided to investigate some of the basic one-factor short-term interest rate models to try to grasp a general idea of how mathematical and statistical techniques can be applied to real financial data.

Short-term interest rate datasets were chosen for this investigation (1-day LIBOR and 1-week EURIBOR) since in the real world the longer the time horizon of interest rates examined, the more likely it is that factors like expectations of future interest rates or other ones will influence the current values, so such a choice allows us to concentrate on a relatively more random cluster of data available.

The second section introduces the models considered in this paper, which comprise the Vasicek model (1977) and those of Cox, Ingersoll & Ross (1985) and Dothan (1978), which, as it has already been mentioned, are all one-factor models, i.e. the predicted magnitude of change in the interest rates considered only depends on a single explanatory factor.

Using the discrete version of the equations based on the original differential equations, which are the core components of the models considered, it is possible to obtain an estimate of the linear regression equation for each model, which could then be tested as to how well it fits the actual data. This estimation, as well as an example of calibration, are provided in section 3, which are followed by a comparative analysis of the properties of the regression models.

Afterwards, some further analysis is conducted for the larger set of EURIBOR values in section 4 in order to assess not only the significance of the explanatory factors or the regression models in whole, but the forecasting possibility of each model. In order to achieve this aim, the regression models designed to forecast the future change in the interest rates are developed in a way that would concentrate on predicting the value of the future interest rate itself. In addition, the stability of the coefficients in the models estimated is investigated via a cycle of similar regressions to analyse the consistency of the regression models. Finally, the errors in forecasting produced by each model are investigated to compare the models and the possible advantages of some model over the others in several aspects.

The objective of section 5 is the assessment of the justification behind the 2 basic assumptions underlying the linear regression models analysed in this paper, i.e. those of homoscedasticity and normality of residuals.

Finally, the general inference concerning the comparison of the three one-factor models is drawn in the conclusion, where the summarized results of the investigation are outlined.

2 The three models tested

General description of the models [1, p.58 – 68]:

1. The Vasicek Model

   Vasicek assumed that the relationship between the change in the interest rate and its value is linear and over time the latter converges to some mean value

$$dr(t) = k[\theta - r(t)]dt + \sigma dW(t), r(0) = r_0 \tag{1}$$

   It is evident that when $r(t) < \theta$, the interest rate is expected to rise as $dr > 0$, unless some large negative random error occurs, and for $r(t) > \theta$ the opposite holds, so the existence of a long-term limit of the interest rate equal to θ is revealed. Scholars claim that the major drawback of the Vasicek model is the fact that the underlying assumption of a normal distribution of the interest rate makes negative values of it possible. Although considering such a scenario as likely as the case of positive values is not, to my mind, coherent in the real financial world, in some cases negative values of the *real* interest rate, which is the difference between the nominal one and the inflation rate, are indeed possible, e.g. in Russia prior to the financial crisis of 2008 with rates of inflation over 10% annually and nominal interest rates under that value. Therefore, the nature of the rates considered should be accounted for.

   The estimation of the regression model that would enable us to test the Vasicek model requires a bit more vigorous work. I would like to describe the stages that need to be completed to obtain the desired estimate [2, p.266 – 270].

   Firstly, as outlined by Back, having gone through all the necessary stochastic transformations, one arrives at the following solution for the interest rate estimated by the Vasicek model

$$r(b) = \theta - e^{-k(b-a)}[\theta - r(a)] + \sigma \int_a^b e^{-k(b-q)}\, dB(q) \tag{2}$$

   Where B is a Geometric Brownian Motion. Now, if, say, we are currently at the moment of time denoted as t and our objective is to forecast the value of the interest rate at the point $t+1$, asssuming the value of the interest rate for the current moment is known and using the previous equation, we get the following estimate

$$r(t+1) = \theta - e^{-k\Delta t}[\theta - r(t)] + \sigma \int_t^{t+1} e^{-k(t+1-q)}\, dB(q) \tag{3}$$

   Where $\Delta t = t + 1 - t = 1$. Finally, we need to subtract $r(t)$ from both sides of the equation to get $\Delta r(t+1) = r(t+1) - r(t)$ on the left-hand side. After some minor transformations we arrive at the following estimation rule for $\Delta r(t+1)$

$$\Delta r(t+1) = \theta(1 - e^{-k}) - (1 - e^{-k})r(t) + \sigma \int_t^{t+1} e^{-k(t+1-q)}\, dB(q) \tag{4}$$

   Consequently, the model tells us that to predict the difference between the next day's interest rate and the current one, we need to use the current rate, which is known, on the right-hand side of the equation, so the last one can be rewritten as

$$r(t+1) - r(t) = a + br(t) + \epsilon \tag{5}$$

   And this equation is equivalent to setting up the following regression model

$$r(t) - r(t-1) = a + br(t-1) \tag{6}$$

Obtaining the estimates of the 2 coefficients in the last equation allows one to derive the values of the original coefficients of the model — namely, σ, k and θ. On the basis of the last few equations, it is evident that the following is true, keeping in mind that $\Delta t = 1$ in our case

$$a = (1 - e^{-k})\theta \tag{7}$$

$$b = -(1 - e^{-k}) \tag{8}$$

$$var(\epsilon) = \frac{\sigma^2(1 - e^{-2k})}{2k} \tag{9}$$

Therefore, one can easily estimate the initial coefficients, which are equal to

$$k = -ln(b + 1) \tag{10}$$

$$\theta = -\frac{a}{b} \tag{11}$$

$$\sigma = \sqrt{\frac{var(\epsilon)(-2ln(b+1))}{1 - (b+1)^2}} \tag{12}$$

This technique can be applied to the remaining models by analogy. An example of the calibration based on real data is provided in the next section.

2. The Cox, Ingersoll & Ross (CIR) Model

   The CIR model was introduced as an effort to eliminate the assumption of the possibility of negative values of the interest rate in the Vasicek model. This aim was achieved by introducing the square root of the interest rate in the error term that represents a random walk away from the estimated value, following Geometric Brownian Motion, so the new model were to be described as

$$dr(t) = k[\theta - r(t)]dt + \sigma\sqrt{r(t)}dW(t), r(0) = r_0 \tag{13}$$

   The equation differs from the Vasicek one only in the error term. So by analogy with the Vasicek case, constructing a linear regression model according to the assumptions of the CIR model will involve using $r(t-1)$ terms on the right-hand side to estimate $dr(t)$. It enables us to test the CIR model using past data, assuming every prediction of future interest rates or the changes in them is possible if we know the current ones. However, to transform the above equation into a regression model comparable to the Vasicek one, we need to make the error terms in the respective regression models follow the same distribution. Therefore, taking into account both of the above ideas, we should divide both sides of the above equation by $\sqrt{r(t-1)}$ to obtain the regression equation corresponding to the CIR model

$$\frac{\Delta r(t)}{\sqrt{r(t-1)}} = \frac{a}{\sqrt{r(t-1)}} + b\sqrt{r(t-1)}, r(0) = r_0 \tag{14}$$

3. The Dothan Model

   Dothan also tried to eliminate the problem of negative interest rates from his model, so he developed it in the following way

$$dr(t) = ar(t)dt + \sigma r(t)dW(t) \tag{15}$$

Under this condition, $r(t)$ is lognormally distributed, so the possibility of negative interest rates is eliminated. As shown by Brigo in his textbook, the interest rate in the Dothan model follows the path below

$$r(t+1) = r(t)e^{(a-\frac{1}{2}\sigma^2)\Delta t+\sigma(W(t+1)-W(t))} \tag{16}$$

Where $\Delta t = t+1-t = 1$. Subtracting $r(t)$ from both sides, we get

$$r(t+1) - r(t) = r(t)(e^{(a-\frac{1}{2}\sigma^2)\Delta t+\sigma(W(t+1)-W(t))} - 1) \tag{17}$$

This is equivalent to the following equation (m is a constant)

$$\Delta r(t) = mr(t-1) \tag{18}$$

So like the 2 previous cases, here the known interest rate is used to estimate the future change. Also, like the case with the CIR Model, here again we need to modify the equation to obtain a comparable one by getting rid of the excess term in the diffusion element. Dividing by $r(t-1)$, we obtain the final version of the regression model to be used

$$\frac{\Delta r(t)}{r(t-1)} = m \tag{19}$$

Which means that the LS predicted value of the dependent variable here is just the average over the dataset available.

3 The calibration of models for the LIBOR dataset

Below is an example of calibration for the three models, namely, the Vasicek, the CIR and the Dothan ones based on the actual LIBOR rates from 01.07.2011 till 30.12.2011. The values are provided in Appendix 1 and are reflected in the graph below. The variables considered were denoted as

$r\_t1 = r(t-1)$ — interest rate on the previous business day

$d\_r = \Delta r(t)$ — difference between the current and the previous day's interest rates

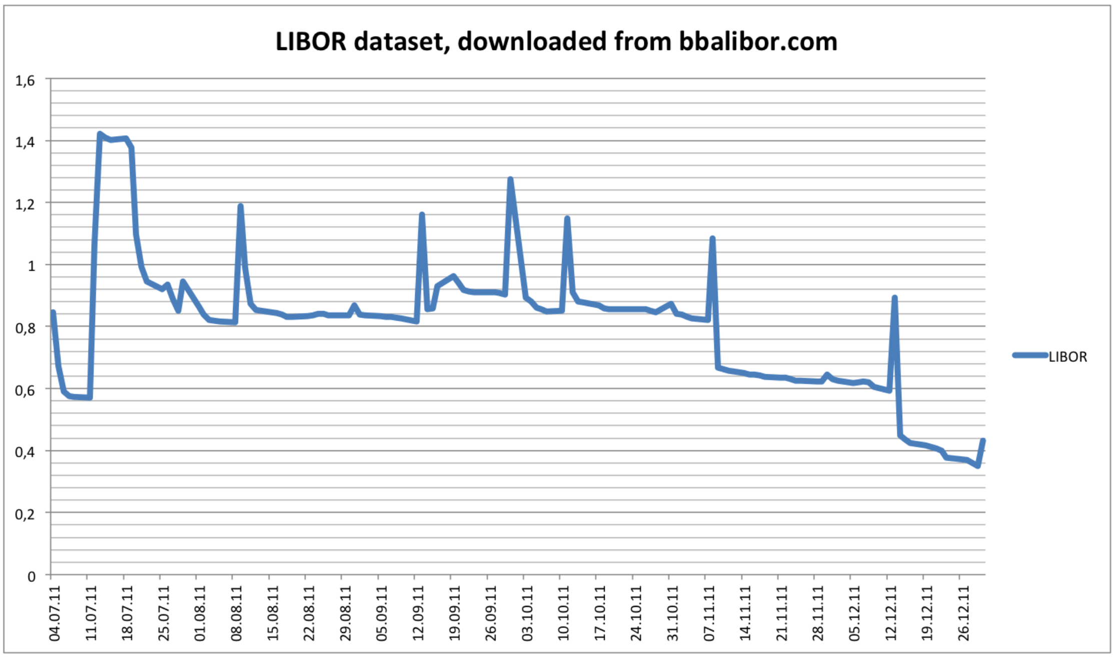
$const$ — constant coefficient

$dr\_oversqrt\_rt1 = \frac{\Delta r(t)}{\sqrt{r(t-1)}}$

$sqrt\_rt1 = \sqrt{r(t-1)}$

$oneoversqrt\_rt1 = \frac{1}{\sqrt{r(t-1)}}$

$dr\_over\_rt1 \ = \ \frac{\Delta r(t)}{r(t-1)}$



Using the regression models derived in the previous section we arrive at the following results (summary regression statistics are provided in Appendix 2)

Vasicek model

$$\Delta \hat{r}(t) = 0.12 - 0.153 r(t-1) \tag{20}$$

CIR model

$$\frac{\Delta \hat{r}(t)}{\sqrt{r(t-1)}} = \frac{0.0826}{\sqrt{r(t-1)}} - 0.1066\sqrt{r(t-1)} \tag{21}$$

Dothan model

$$\frac{\Delta \hat{r}(t)}{r(t-1)} = 0.004 \tag{22}$$

Using the regression coefficients obtained and the material in the previous section, we can now derive the estimated coefficients of the initial equations. For instance, if we consider the Vasicek equation, then equation (20) tells us that $a = 0.12$ and $b = -0.153$. So the findings of the previous section produce the following results

$$\hat{k} = 0.166, \hat{\theta} = 0.784, \hat{\sigma} = 0.0165 \tag{23}$$

Putting these into the equation outlined by Vasicek, we get the estimated model for the interest rate below. Although we considered the discrete case, since the original equation is for a continuous scenario, then, assuming our loss of precision due to discretization is not crucial, we get the following equation

$$dr(t) = 0.166[0.784 - r(t)]dt + 0.0165dW(t) \tag{24}$$

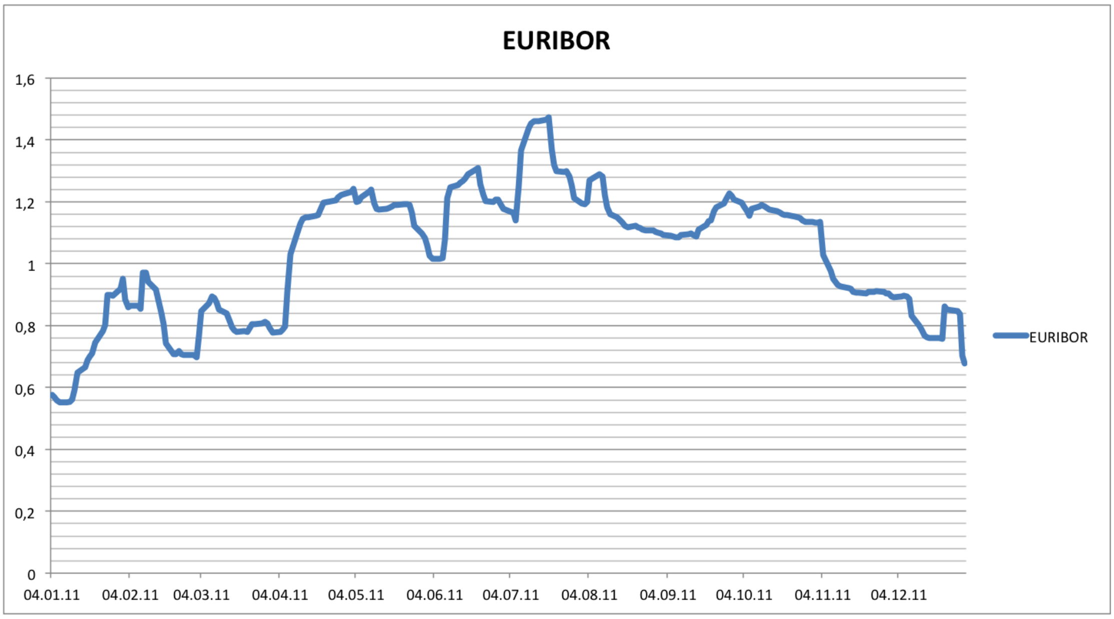
Such a transition to the original coefficients can be conducted by analogy for the other models as well.

As to the regression statistics, they exhibit the following patterns: firstly, concerning the Vasicek and CIR models, although the coefficient of determination is less than 10% in both cases (which is, in fact, very low; however, one should keep in mind that the process being investigated is rather random since the fluctuations in short-term LIBOR rates are difficult to predict consistently, especially using just one explanatory factor), the independent variables are significant. The Vasicek model incorporates a single one, namely $r(t-1)$, which is significant at all reasonable significance levels as the P-value is approximately 0.2%. The same, though to a smaller degree, is true for the CIR model, where $\sqrt{r(t-1)}$ and $\frac{1}{\sqrt{r(t-1)}}$ are both significant at the 5% and the 3% significance levels. These arguments are supported by the F-statistics for the respective regressions in whole, which reveal the fact the regression derived from the Vasicek model is highly significant (the F-statistic coincides with the t-statistic here), whereas the one for the CIR model is relatively less significant with $P - value = 0.0664$, but still not crucially.

On the other hand, the assessment of the Dothan model seems a bit problematic in this case since the regression model that we established only incorporates a constant on the right-hand side, so the coefficient of determination is identically 0. In addition, one could evaluate the relative advantages of the models by comparing the corresponding sums of squares of errors. However, keeping in mind that each regression model, the way it was established in the previous section, is designed to estimate a different variable and the left-hand sides are unique for each equation, such a comparison does not seem fair enough. Therefore, comparing the three models in a more or less objective manner will require some additional tools as shown in the next section.

4 The calibration of models and further analysis for the EURIBOR dataset

The aim of this section is to check whether the results of the previous one still hold for a similar dataset (EURIBOR is the European analogue for LIBOR) but with a greater number of observations - 9 months of 2011, the values are provided in Appendix 3 and are reflected in the graph below. Afterwards, I would like to try to assess the forecasting efficiency of each model considered by estimating the respective regression equations based on the 9-month period, and then evaluating the quality of each model's prediction over the remaining 3 months of 2011 by comparing with the actual figures, which are as well provided in Appendix 3 and the graph below. Finally, I consider it interesting to assess the errors produced by each model more thoroughly and compare some aspects of error terms across the 3 models.



Using the same denomination of variables as in the previous section of LIBOR data analysis and again incorporating the regression equations derived from the respective models, we obtain the following equations (summary statistics are provided in Appendix 4)

Vasicek model

$$\Delta \hat{r}(t) = 0.0243 - 0.0204 r(t-1) \tag{25}$$

CIR model

$$\frac{\Delta \hat{r}(t)}{\sqrt{r(t-1)}} = \frac{0.0223}{\sqrt{r(t-1)}} - 0.0184\sqrt{r(t-1)} \tag{26}$$

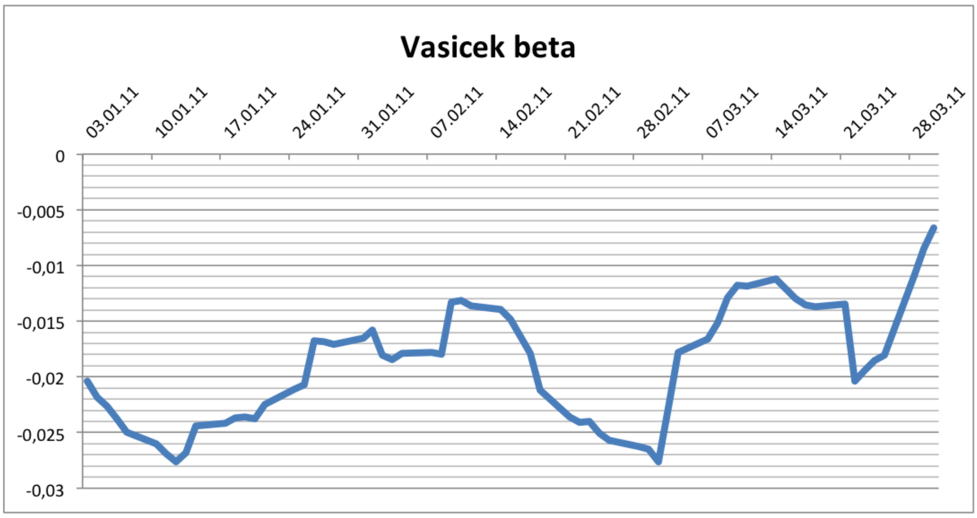
Dothan model

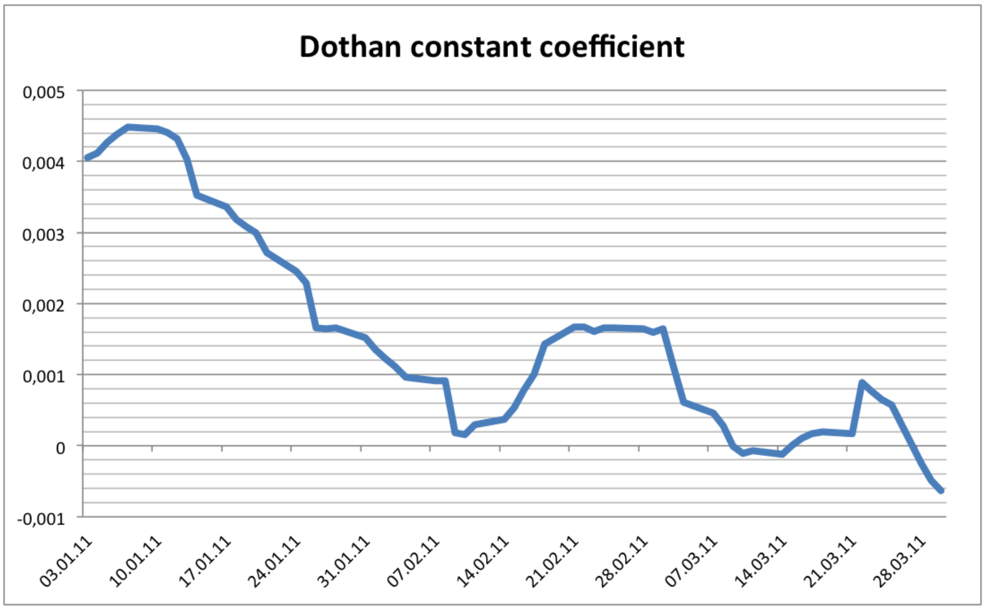
$$\frac{\Delta \hat{r}(t)}{r(t-1)} = 0.004 \tag{27}$$

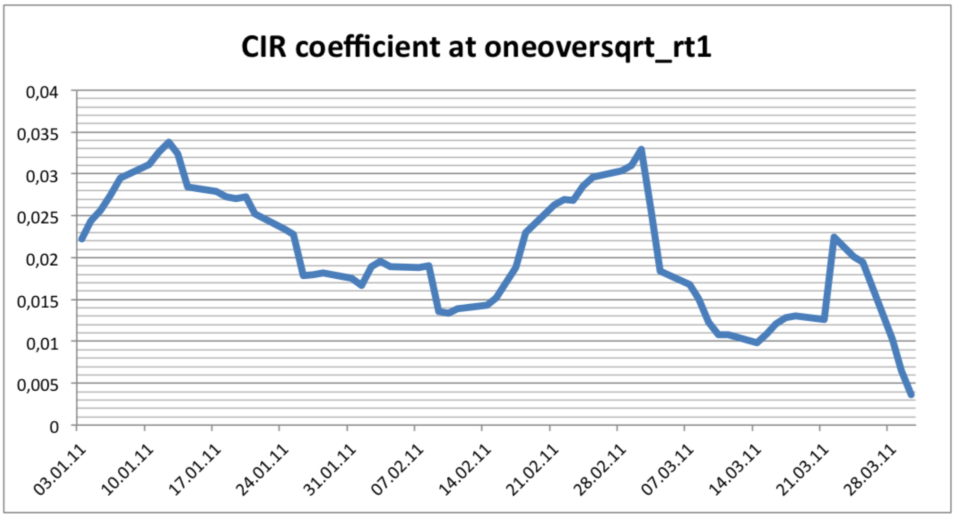
The statistics show that the Vasicek and CIR model regressions are significant at the 5% significance level as P-values of both respective F-statistics are less than 5%. Moreover, for the CIR, which incorporates 2 variables on the right-hand side of the equation, the P-values of the corresponding t-statistics are also quite small - less than 6% for the $\sqrt{r(t-1)}$ and less than 3% for the $\frac{1}{\sqrt{r(t-1)}}$. Although the R^2 coefficients are low for both regressions (again less than 10%) due to the same reasons as described in the previous section, the 2 regressions considered are of rather high quality since both are significant at the 5% level.

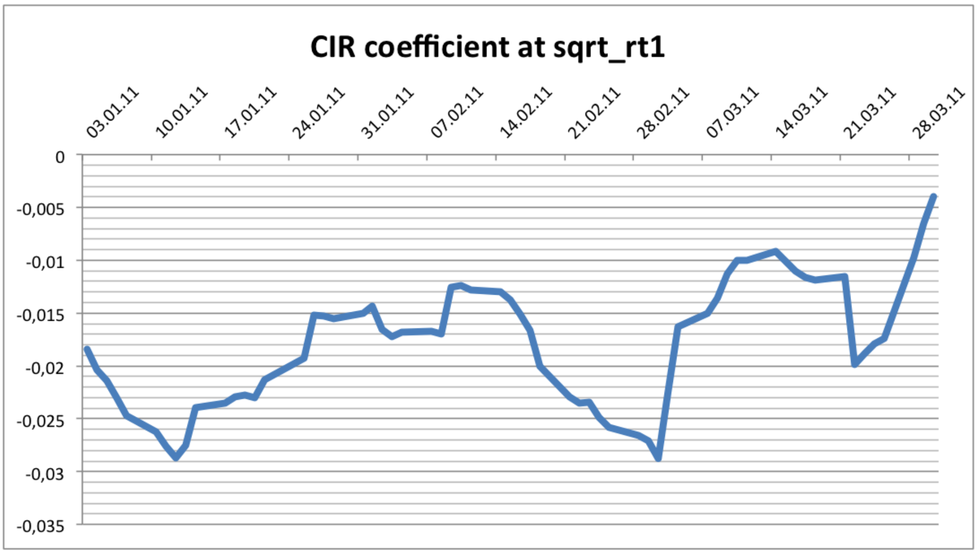
The problems outlined in the previous section with respect to the assessment of the Dothan model still hold, but I find it useful to evaluate the model even in these circumstances by using some additional techniques to compare it with the Vasicek and CIR models.

A point worth considering is the stability of the coefficients estimated by the regression equations above. To assess this aspect, a special script was constructed in the GRETL package to generate 3 cycles of regressions (corresponding to the three models). Each regression in these cycles was designed to replicate the original regressions given above by incorporating the same number of observations (9 months of data), the only difference was that each consecutive one among the 63 generated was based on values of the interest rate starting and ending one day later than the previous one. So, obviously, the first regression in each cycle produces the same coefficients as those provided above and it can be checked in Appendix 6, where the actual coefficients are provided for the three cycles. The graphs below represent the fluctuations in the coefficients estimated, what allows one to evaluate their relative stability and, thus, see whether the regression models are consistent.









The table in Appendix 6 reveals the fact that the Dothan model's constant coefficient is the one with the least standard deviation among all. However, the relative difference between the mean value of the coefficient over the entire cycle and the one obtained in the original regression, which is the first in the cycle, is much larger for the Dothan model compared to the corresponding values for the Vaiscek and the CIR models. One way to calculate such a relative difference is, for example, to subtract the average value from the actual one and divide the result by the average value. Therefore, the advantages in some aspects seem outweighed by disadvantages and we can infer that generally there is no evidence on this stage of our investigation that some model is more consistent than the others, so further consideration is required.

Consequently, having obtained the estimates of the 3 regression lines, we should now try to use them to see which of the models is best capable of forecasting the rates for the last 3 months of 2011. But we need to account for the fact that except for the Vasicek model, the dependent variables are not merely $\Delta\hat{r}(t)$ and simply adding the predicted value of the dependent variable to the known $r(t-1)$ will not produce the correct estimate of $r(t)$. Therefore, what we need to do is to modify the equations to get $\Delta\hat{r}(t)$ on the left-hand side, which means multiplying both sides of the CIR equation by $\sqrt{r(t-1)}$ and the Dothan equation by $r(t-1)$. These operations yield the following forecasting rules for $\Delta r(t)$ using the 2 models:

CIR model

$$\Delta\hat{r}(t) = 0.0223 - 0.0184r(t-1) \tag{28}$$

Dothan model

$$\Delta\hat{r}(t) = 0.004r(t-1) \tag{29}$$

Now, knowing $r(t-1)$ and having derived an estimate for the $\Delta r(t)$, summing these 2 terms will produce the predicted value of $r(t)$. After rearranging some terms where needed, we arrive at the following prediction rules for the three models

Vasicek model

$$\hat{r(t)} = 0.9796r(t-1) + 0.0243 \tag{30}$$

CIR model

$$\hat{r(t)} = 0.9816r(t-1) + 0.0222 \tag{31}$$

Dothan model

$$\hat{r(t)} = 1.004r(t-1) \tag{32}$$

Then, using these prediction schemes, not only can we find out which equation best fits the actual data for the 9-month period considered, but also we can use them to estimate the values of $r(t)$ over the next three months of 2011 and again compare with the actual data for that period. I assumed such a prediction for the 3 months ahead could be conducted in a similar way, i.e. each time $r(t-1)$ was known and the model based on the 9-month period was used.

One way to evaluate the comparative advantages of the three models is to calculate the sum of squares of errors, which, unlike the case with the LIBOR data, is consistent in the current case because the same dependent variable $\hat{r(t)}$ stands on the left-hand side of the equations. The following relationship holds for the entire year 2011

$$\sum e^2(CIR) = 0.2564 < \sum e^2(Vasicek) = 0.2569 < \sum e^2(Dothan) = 0.2596 \tag{33}$$

As for the errors in the last three months, the results are the following:

$$\sum e^2(Dothan) = 0.0564 < \sum e^2(CIR) = 0.0569 < \sum e^2(Vasicek) = 0.0572 \tag{34}$$

So we can infer that although a model may seem not quite comparable initially, it may happen that its predictions are rather on target, so the decision to keep the Dothan model in our consideration seems justified now. I think it is interesting to try to understand what may have caused such an unexpected outcome for the Dothan model. To my mind, the reason for what we observe could have been the fact that the Dothan model is the only one to use a prediction rule for $r(t)$ without a constant. Since the general trend for $r(t)$ in October–December 2011 was a decrease in the interest rate, (although the coefficient in the Dothan model of $1.004 > 1$ implies an increasing general trend in the interest rates) it could happen that the constants in the Vasicek and the CIR models based on the first 9 months of 2011 were too "large" for the next three months, what caused an overestimate in rates, and the impact of that attribute could easily outweigh the impact of the implied increase in interest rates in the equation corresponding to the Dothan model. This argument seems supported by the graph provided below this paragraph, where it is illustrated that starting in mid-November after the rates plummeted and onwards, the Dothan model constantly "strikes" closer to the actual value than the other 2 models do (the table in Appendix 5 reveals that the CIR and the Vasicek models provide almost coinciding estimates over the 3-month period and it is why the line representing the Vasicek model's estimates is hard to trace in the mentioned graph).

An alternative way to evaluate the relative sizes of errors is via the sums of absolute values, not the squared ones. In this case, the following relationship is true (with respect to the entire 2011 period):

$$\sum |e(Dothan)| = 4.9589 < \sum |e(CIR)| = 5.0514 < \sum |e(Vasicek)| = 5.0799 \tag{35}$$

As for the last three months of 2011, then

$$\sum |e(Dothan)| = 1.0346 < \sum |e(CIR)| = 1.0537 < \sum |e(Vasicek)| = 1.0619 \tag{36}$$

These 2 findings again illustrate that the Dothan model seems the most efficient predictor of the interest rates in our case.

An interesting area to investigate is the patterns in errors that the models produce. Firstly, except for several cases each model yields only positive errors, i.e. the actual rates appear to be lower than those forecast as shown in Appendix 5. Again, in my opinion, this could be a result of the general downward sloping trend line for the actual numbers, which could be due to some economic instability, some policy measures or other factors. Such a direction of change could make all the models on average "strike" too high.

Also, on average, the following relationship holds for the absolute values of errors:

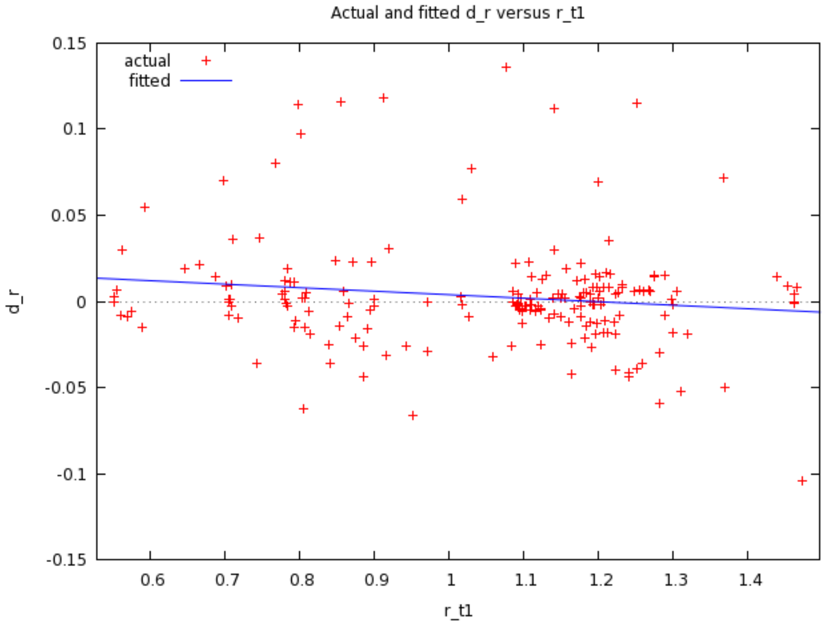
$$|e(Vasicek)| > |e(CIR)|, |e(CIR)| > |e(Dothan)|, |e(Vasicek)| > |e(Dothan)| \tag{37}$$

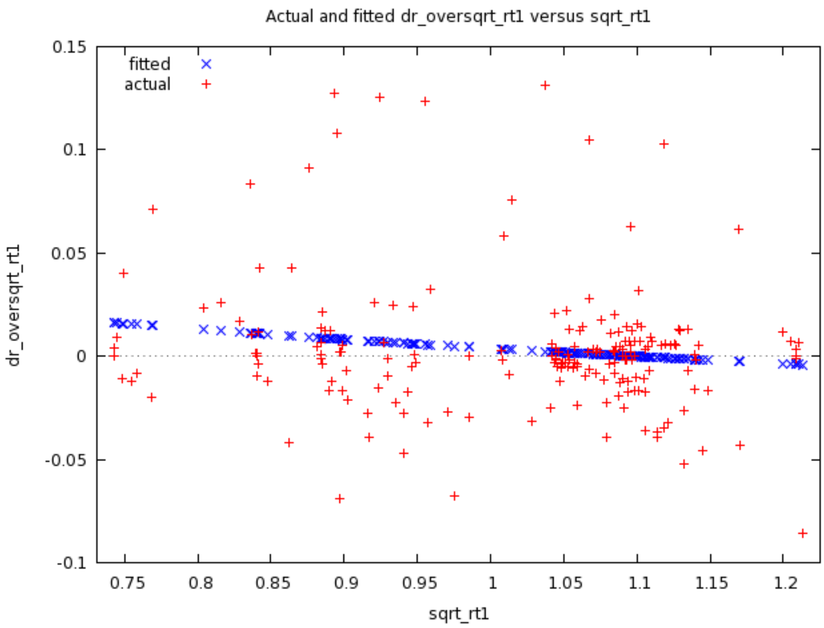
This is based on the calculation of the average value of the set, the terms of which take the following values: "1" if, for instance, $|e(Vasicek)| > |e(CIR)|$ is true and "0" otherwise (again, please check the table in Appendix 5). The resulting average value is greater than 0.5, so the given conclusion is drawn, and the 2 other cases were implemented by analogy.

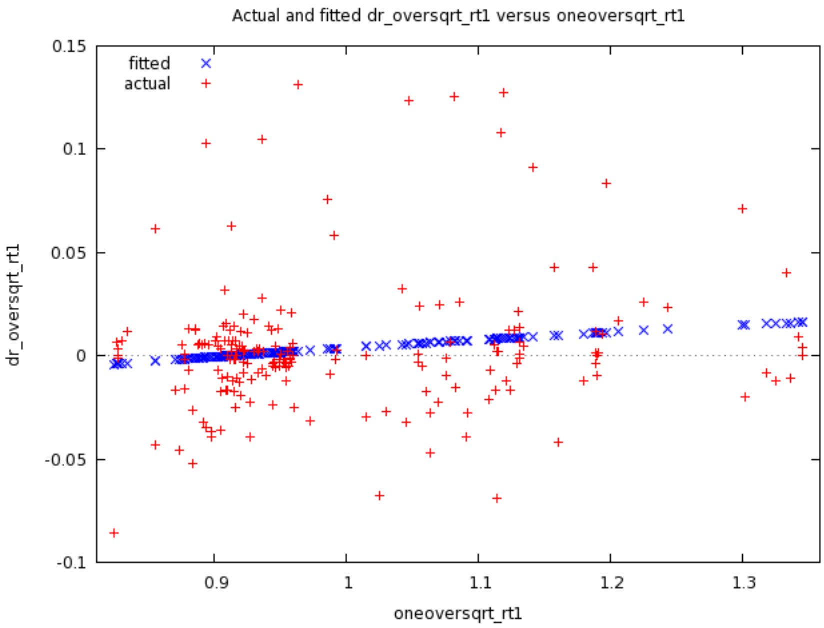
All in all, different evaluations of the forecasting quality of the three models described in this section exhibit the relative superiority of the Dothan model over the other 2 models. Concentrating on the period of the last three months over which the predictions, constructed according to the rules outlined by the models on the basis of the prior 9 months, were tested, we have found out that, in fact, all of the aspects of comparison (sum of errors squared or in absolute values and pairwise comparison of absolute values of errors) reveal a slight advantage of the Dothan model over the others. However, the actual numbers tell us that such a superiority is merely a slight one, as the differences with what other models show are quite small. Therefore, generally, there are no serious grounds to claim that the models differ significantly as concerns their forecasting power over the EURIBOR dataset under the conditions analysed in this section. Some issues concerning these conditions are considered in the next section.

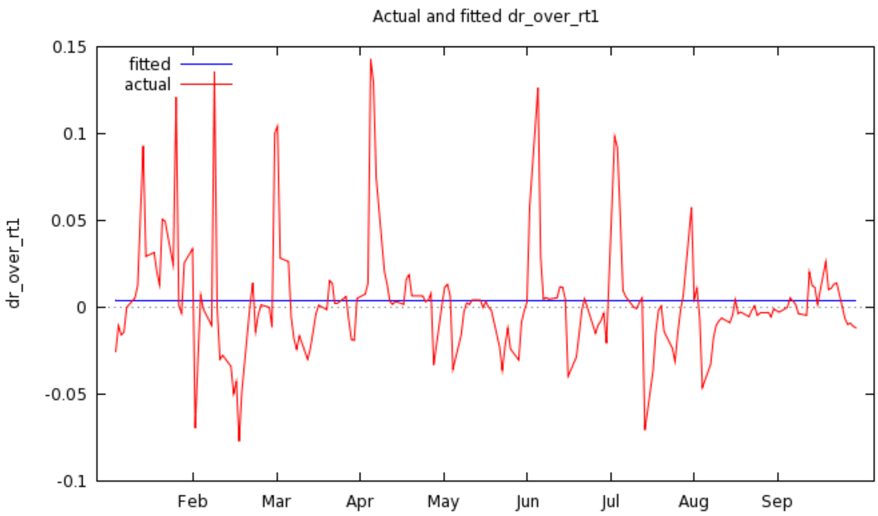
5 Homoscedasticity analysis and normality of residuals test for the EURIBOR dataset

Since interest rates are one of the crucial indicators of the general condition within the economy, they are influenced by numerous factors, such as investors' expectations, macroeconomic stability or specific policy measures. Therefore, when the aim is to explain the fluctuations in the interest rates by using a unique factor, some specific issues may arise which need to be analysed. Firstly, I would like to check how justified the assumption of homoscedasticity, underlying the 9-month regression models developed in the previous section, seems in the given circumstances. Such analysis is convenient to start with examining the graphs of fitted and actual values in the independent variable-dependent variable coordinate system, so our approach will involve the use of a single graph for the Vasicek case and 2 diagrams for the CIR case, which incorporates 2 variables on the right hand side. As to the Dothan model, since the right-hand side of the original regression model in the last section consisted of only a constant, the actual and fitted values are presented in the time vs dependent variable coordinate system. All the mentioned graphs are provided below.





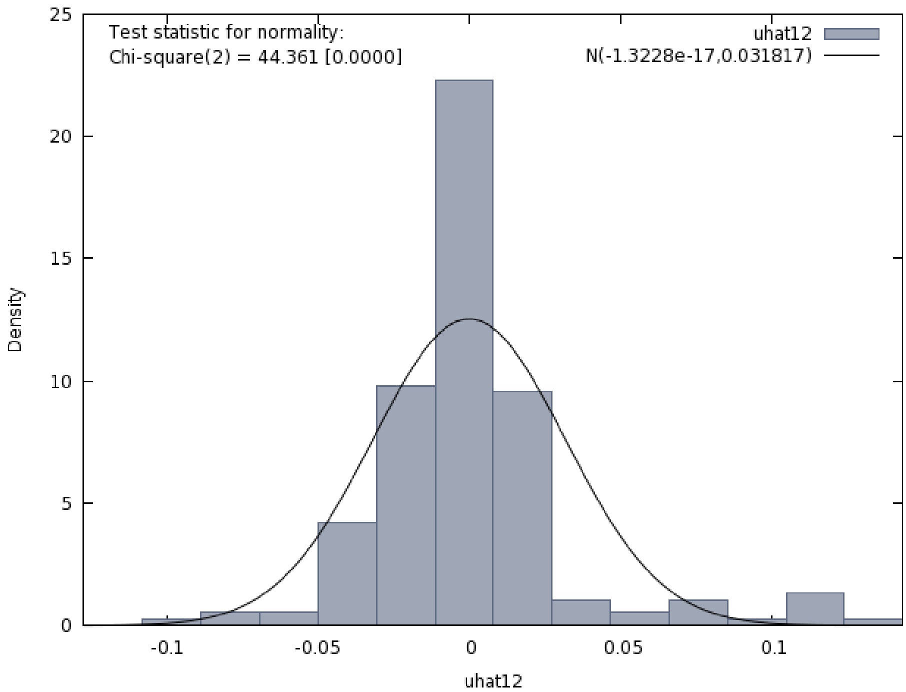




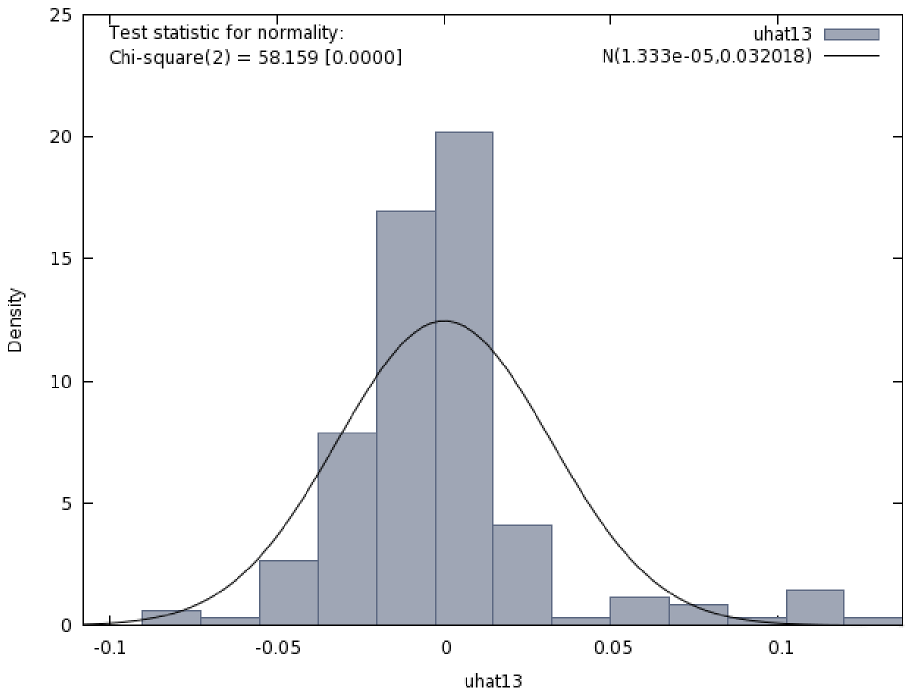
None of the diagrams reveals explicit patterns that could be associated with heteroscedasticity, as the overall spread of the actual values around the fitted ones seem pretty much the same over the entire set of the variables on the horizontal axes. This intuitive inference is, in fact, supported by the results of White's test for the Vasicek and CIR regressions. As illustrated in Appendix 7, the P-values of both test statistics do not lead to the rejection of the null hypothesss (no heteroscedasticity present) at the 5% significance level. However, it is worth mentioning that although the Vasicek regression, in fact, provides a P-value so high that the null hypothesis cannot be rejected at any reasonable significance level, the CIR case is much more "vulnerable" since H_0 can be rejected at greater levels. So in this aspect, the assumption on which both regression models are based, seems more justified for the Vasicek model than for the CIR one.

Another crucial assumption on which the linear regression models of this paper were based is the normality of residuals. To assess its level of justification in the EURIBOR dataset regressions, it is useful to construct histograms of residuals produced by each model and check how well each histogram corresponds to the bell-shaped curve representing normal distribution. If a histogram is rather symmetric, does not have "heavy" tails and is not centered far away from 0, then there is a high chance that the distribution of errors is approximately normal. For further analysis, one can use a chi-squared goodness of fit test, whereby the P-value of the test statistic will provide a more objective measure of how well the given assumption is justified. Below are the mentioned histograms for each model, which also contain the corresponding chi-squared test statistics and the respective P-values in square brackets.

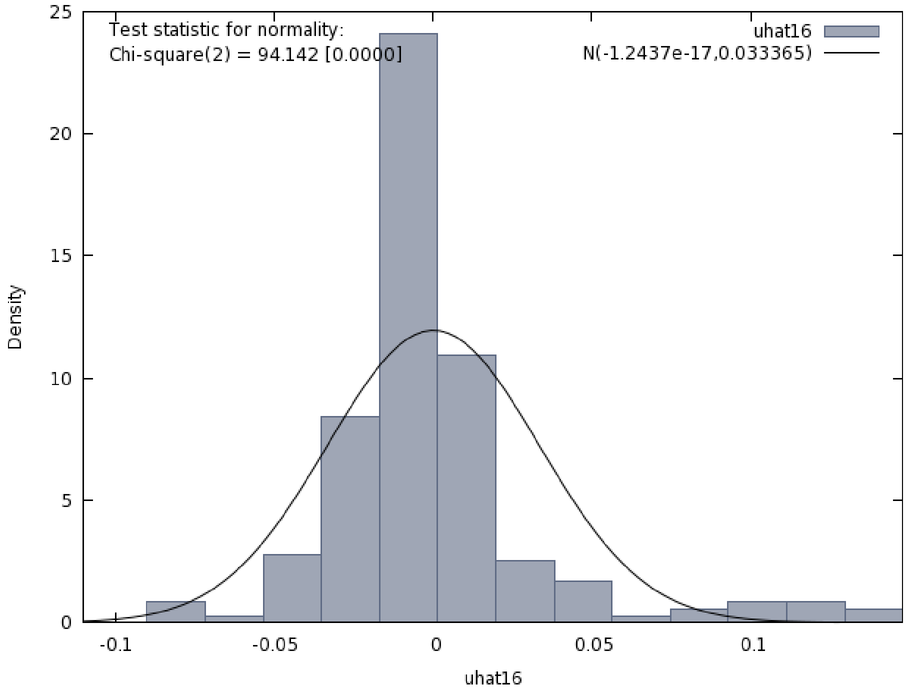
Vasicek model



CIR model



Dothan model



To begin with, all the chi-squared test statistics make us infer that the null hypothesis of normal distribution of residuals is rejected for all three models at all reasonable significance levels. However, one should still keep in mind that this inference is only valid for the particular cases discussed in this paper and, for instance, the use of a larger time period in all regressions can change the results. Moreover, the value of the test statistic is highly dependent on the way the entire span of residuals is divided into intervals, for which the relative frequencies are computed afterwards and represented in the histograms provided, so, generally, one cannot be absolutely sure that the distributions of residuals provided by the 3 regression models are not normal.

Although the P-values of the three test statistics are identical up to the 4th decimal, the Vasicek statistic is the least of all. Indeed, the corresponding histogram is quite symmetric, but the same is true about the histogram for the Dothan model, which produced the largest chi-squared value, even larger than the CIR one. To my mind, the reason for this is the presence of some large positive outliers in the Dothan model histogram and although other histograms, including the most non-symmetric CIR one, also contain some of these, the magnitudes are smaller than in the Dothan case.

All in all, the 2 fields of analysis incorporated in this section do not make any of the models stand out. Although both the homoscedastcity and normality points reveal some advantage of the Vasicek model over the CIR one in our case, the effect is merely a slight one, since the differences in the aspects considered are not crucial across the three models.

6 Conclusion

Each section of this paper includes some interesting insights, which I would like to summarize in order to arrive at some overall inference.

Firstly, a couple of times we have witnessed a comparative advantage of the properties of the Vasicek model compared to the CIR one. Namely, over the LIBOR dataset the former one results in a quite more significant regression and at the same time the EURIBOR dataset reveals the fact that the assumption of constant variance of residuals, i.e. homoscedasticity, is much less justified when dealing with the latter model than with the Vasicek one. On the other hand, the same EURIBOR dataset showed that, on average, the magnitudes of errors are smaller in the CIR case, so, in whole, I find it coherent to claim that there is not enough evidence that any of the 2 models consistently outperforms the other one.

As to the Dothan model, although initially some problems arose concerning the comparability of that model due to the properties of the regression equation in the LIBOR section, the EURIBOR dataset analysis exhibited an interesting pattern: on average, the Dothan model produced the least errors in the given circumstances. However, the difference was not reasonably significant.

Therefore, our entire analysis illustrates that the one-factor models analysed do not produce considerably different results. Each aspect may reveal a comparative advantage, but no sustainable pattern appears. Such an inference is supported by the fact that all the models produced relatively similar outcomes in the coefficients stability checking cycle, meaning that the consistency of the models is pretty much the same. However, one should keep in mind that these conclusions are based on the datasets considered and further analysis involving larger datasets and more sophisticated techniques can show considerably different patterns.

Appendix

1. LIBOR Dataset from 01.07.2011 till 30.12.2011, official data downloaded from http://bbalibor.com

   Entire data are available at http://goo.gl/nLUI3

2. LIBOR Regression Models

   (a) Vasicek: OLS, using observations 2011/07/04–2011/12/30 ($T = 129$)
   Dependent variable: d_r

	Coefficient	Std. Error	t-ratio	p-value
const	0.120177	0.0409992	2.9312	0.0040
r_t1	−0.153099	0.0490362	−3.1222	0.0022

Mean dependent var	−0.003627	S.D. dependent var	0.122312
Sum squared resid	1.778408	S.E. of regression	0.118335
R^2	0.071284	Adjusted R^2	0.063971
$F(1,127)$	9.747950	P-value(F)	0.002223
Log-likelihood	93.28097	Akaike criterion	−182.5619
Schwarz criterion	−176.8423	Hannan–Quinn	−180.2379
$\hat{\rho}$	−0.099865	Durbin–Watson	2.198832

   (b) CIR: OLS, using observations 2011/07/04–2011/12/30 ($T = 129$)
   Dependent variable: dr_oversqrt_rt1

	Coefficient	Std. Error	t-ratio	p-value
sqrt_rt1	−0.106603	0.0456646	−2.3345	0.0211
oneoversqrt_rt1	0.0825780	0.0354777	2.3276	0.0215

Mean dependent var	−0.000067	S.D. dependent var	0.131651
Sum squared resid	2.125723	S.E. of regression	0.129375
R^2	0.041821	Adjusted R^2	0.034276
$F(2,127)$	2.771551	P-value(F)	0.066353
Log-likelihood	81.77462	Akaike criterion	−159.5492
Schwarz criterion	−153.8296	Hannan–Quinn	−157.2252
$\hat{\rho}$	−0.105561	Durbin–Watson	2.208056

   (c) Dothan: OLS, using observations 2011/07/04–2011/12/30 ($T = 129$)
   Dependent variable: dr_over_rt1

	Coefficient	Std. Error	t-ratio	p-value
const	0.00402767	0.0129557	0.3109	0.7564

Mean dependent var	0.004028	S.D. dependent var	0.147148
Sum squared resid	2.771533	S.E. of regression	0.147148
R^2	0.000000	Adjusted R^2	0.000000
Log-likelihood	64.66349	Akaike criterion	−127.3270
Schwarz criterion	−124.4672	Hannan–Quinn	−126.1650
$\hat{\rho}$	−0.143527	Durbin–Watson	2.260881

3. EURIBOR Dataset from 04.01.2011 till 30.12.2011, official data downloaded from http://www.euribor-ebf.eu

   Entire data are available at http://goo.gl/oXJak

4. EURIBOR Regression Models

   (a) Vasicek: OLS, using observations 2011/01/03–2011/09/30 ($T = 195$)
   Dependent variable: d_r

	Coefficient	Std. Error	t-ratio	p-value
const	0.0242851	0.0109380	2.2202	0.0276
r_t1	−0.0203573	0.0102194	−1.9920	0.0478

Mean dependent var	0.002974	S.D. dependent var	0.032059
Sum squared resid	0.195376	S.E. of regression	0.031817
R^2	0.020146	Adjusted R^2	0.015069
$F(1, 193)$	3.968127	P-value(F)	0.047779
Log-likelihood	396.6254	Akaike criterion	−789.2507
Schwarz criterion	−782.7047	Hannan–Quinn	−786.6003
$\hat{\rho}$	0.533085	Durbin–Watson	0.930090

   (b) CIR: OLS, using observations 2011/01/03–2011/09/30 ($T = 195$)
   Dependent variable: dr_oversqrt_rt1

	Coefficient	Std. Error	t-ratio	p-value
sqrt_rt1	−0.0184328	0.00966856	−1.9065	0.0581
oneoversqrt_rt1	0.0222705	0.00984577	2.2619	0.0248

Mean dependent var	0.003477	S.D. dependent var	0.032300
Sum squared resid	0.197858	S.E. of regression	0.032018
R^2	0.033666	Adjusted R^2	0.028659
$F(2, 193)$	3.361909	P-value(F)	0.036711
Log-likelihood	395.3945	Akaike criterion	−786.7890
Schwarz criterion	−780.2430	Hannan–Quinn	−784.1386
$\hat{\rho}$	0.526935	Durbin–Watson	0.940211

   (c) Dothan: OLS, using observations 2011/01/03–2011/09/30 ($T = 195$)
   Dependent variable: dr_over_rt1

	Coefficient	Std. Error	t-ratio	p-value
const	0.00404908	0.00238931	1.6947	0.0917

Mean dependent var	0.004049	S.D. dependent var	0.033365
Sum squared resid	0.215963	S.E. of regression	0.033365
R^2	0.000000	Adjusted R^2	0.000000
Log-likelihood	386.8575	Akaike criterion	−771.7150
Schwarz criterion	−768.4420	Hannan–Quinn	−770.3898
$\hat{\rho}$	0.525947	Durbin–Watson	0.944143

5. Predicted interest rates and information on errors

date	r_t	rt_vas	rt_cir	rt_dth	e_vas	e_cir	e_dth	e_vas>0	e_cir>0	e_dth>0	|e_vas|> |e_cir|	|e_cir|> |e_dth|	|e_vas|> |e_dth|
03.10.11	1.197	1.208	1.208	1.213	0.011	0.011	0.016	1	1	1	0	0	0
04.10.11	1.184	1.197	1.197	1.202	0.013	0.013	0.018	1	1	1	0	0	0
05.10.11	1.170	1.184	1.184	1.189	0.014	0.014	0.019	1	1	1	0	0	0
06.10.11	1.155	1.170	1.171	1.175	0.015	0.016	0.020	1	1	1	0	0	0
07.10.11	1.178	1.156	1.156	1.160	-0.022	-0.022	-0.018	0	0	0	1	1	1
10.10.11	1.185	1.178	1.179	1.183	-0.007	-0.006	-0.002	0	0	0	1	1	1
11.10.11	1.191	1.185	1.185	1.190	-0.006	-0.006	-0.001	0	0	0	1	1	1
12.10.11	1.186	1.191	1.191	1.196	0.005	0.005	0.010	1	1	1	0	0	0
13.10.11	1.181	1.186	1.186	1.191	0.005	0.005	0.010	1	1	1	0	0	0
14.10.11	1.174	1.181	1.182	1.186	0.007	0.008	0.012	1	1	1	0	0	0
17.10.11	1.171	1.174	1.175	1.179	0.003	0.004	0.008	1	1	1	0	0	0
18.10.11	1.166	1.171	1.172	1.176	0.005	0.006	0.010	1	1	1	0	0	0
19.10.11	1.161	1.167	1.167	1.171	0.006	0.006	0.010	1	1	1	0	0	0
20.10.11	1.158	1.162	1.162	1.166	0.004	0.004	0.008	1	1	1	0	0	0
21.10.11	1.158	1.159	1.159	1.163	0.001	0.001	0.005	1	1	1	0	0	0
24.10.11	1.153	1.159	1.159	1.163	0.006	0.006	0.010	1	1	1	0	0	0
25.10.11	1.151	1.154	1.154	1.158	0.003	0.003	0.007	1	1	1	0	0	0
26.10.11	1.147	1.152	1.152	1.156	0.005	0.005	0.009	1	1	1	0	0	0
27.10.11	1.140	1.148	1.148	1.152	0.008	0.008	0.012	1	1	1	0	0	0
28.10.11	1.136	1.141	1.141	1.145	0.005	0.005	0.009	1	1	1	0	0	0
31.10.11	1.136	1.137	1.137	1.141	0.001	0.001	0.005	1	1	1	0	0	0
01.11.11	1.134	1.137	1.137	1.141	0.003	0.003	0.007	1	1	1	0	0	0
02.11.11	1.134	1.135	1.135	1.139	0.001	0.001	0.005	1	1	1	0	0	0
03.11.11	1.136	1.135	1.135	1.139	-0.001	-0.001	0.003	0	0	1	1	0	0
04.11.11	1.028	1.137	1.137	1.141	0.109	0.109	0.113	1	1	1	0	0	0
07.11.11	0.976	1.031	1.031	1.032	0.055	0.055	0.056	1	1	1	1	0	0
08.11.11	0.952	0.980	0.980	0.980	0.028	0.028	0.028	1	1	1	1	1	1
09.11.11	0.942	0.957	0.957	0.956	0.015	0.015	0.014	1	1	1	1	1	1
10.11.11	0.932	0.947	0.947	0.946	0.015	0.015	0.014	1	1	1	1	1	1
11.11.11	0.926	0.937	0.937	0.936	0.011	0.011	0.010	1	1	1	1	1	1
14.11.11	0.921	0.931	0.931	0.930	0.010	0.010	0.009	1	1	1	1	1	1
15.11.11	0.919	0.927	0.926	0.925	0.008	0.007	0.006	1	1	1	1	1	1

date	r_t	rt_vas	rt_cir	rt_dth	e_vas	e_cir	e_dth	e_vas>0	e_cir>0	e_dth>0	|e_vas|> |e_cir|	|e_cir|> |e_dth|	|e_vas|> |e_dth|
16.11.11	0.908	0.925	0.924	0.923	0.017	0.016	0.015	1	1	1	1	1	1
17.11.11	0.906	0.914	0.914	0.912	0.008	0.008	0.006	1	1	1	1	1	1
18.11.11	0.906	0.912	0.912	0.910	0.006	0.006	0.004	1	1	1	1	1	1
21.11.11	0.905	0.912	0.912	0.910	0.007	0.007	0.005	1	1	1	1	1	1
22.11.11	0.910	0.911	0.911	0.909	0.001	0.001	-0.001	1	1	0	1	0	0
23.11.11	0.909	0.916	0.915	0.914	0.007	0.006	0.005	1	1	1	1	1	1
24.11.11	0.909	0.915	0.915	0.913	0.006	0.006	0.004	1	1	1	1	1	1
25.11.11	0.911	0.915	0.915	0.913	0.004	0.004	0.002	1	1	1	1	1	1
28.11.11	0.909	0.917	0.916	0.915	0.008	0.007	0.006	1	1	1	1	1	1
29.11.11	0.905	0.915	0.915	0.913	0.010	0.010	0.008	1	1	1	1	1	1
30.11.11	0.904	0.911	0.911	0.909	0.007	0.007	0.005	1	1	1	1	1	1
01.12.11	0.894	0.910	0.910	0.908	0.016	0.016	0.014	1	1	1	1	1	1
02.12.11	0.892	0.900	0.900	0.898	0.008	0.008	0.006	1	1	1	1	1	1
05.12.11	0.895	0.898	0.898	0.896	0.003	0.003	0.001	1	1	1	1	1	1
06.12.11	0.896	0.901	0.901	0.899	0.005	0.005	0.003	1	1	1	1	1	1
07.12.11	0.895	0.902	0.902	0.900	0.007	0.007	0.005	1	1	1	1	1	1
08.12.11	0.887	0.901	0.901	0.899	0.014	0.014	0.012	1	1	1	1	1	1
09.12.11	0.831	0.893	0.893	0.891	0.062	0.062	0.060	1	1	1	1	1	1
12.12.11	0.800	0.838	0.838	0.834	0.038	0.038	0.034	1	1	1	1	1	1
13.12.11	0.787	0.808	0.808	0.803	0.021	0.021	0.016	1	1	1	1	1	1
14.12.11	0.767	0.795	0.795	0.790	0.028	0.028	0.023	1	1	1	1	1	1
15.12.11	0.762	0.776	0.775	0.770	0.014	0.013	0.008	1	1	1	1	1	1
16.12.11	0.760	0.771	0.770	0.765	0.011	0.010	0.005	1	1	1	1	1	1
19.12.11	0.759	0.769	0.768	0.763	0.010	0.009	0.004	1	1	1	1	1	1
20.12.11	0.760	0.768	0.767	0.762	0.008	0.007	0.002	1	1	1	1	1	1
21.12.11	0.757	0.769	0.768	0.763	0.012	0.011	0.006	1	1	1	1	1	1
22.12.11	0.861	0.766	0.765	0.760	-0.095	-0.096	-0.101	0	0	0	0	0	0
23.12.11	0.853	0.868	0.867	0.864	0.015	0.014	0.011	1	1	1	1	1	1
27.12.11	0.846	0.860	0.860	0.856	0.014	0.014	0.010	1	1	1	1	1	1
28.12.11	0.836	0.853	0.853	0.849	0.017	0.017	0.013	1	1	1	1	1	1
29.12.11	0.704	0.843	0.843	0.839	0.139	0.139	0.135	1	1	1	1	1	1
30.12.11	0.677	0.714	0.713	0.707	0.037	0.036	0.030	1	1	1	1	1	1

6. Regression cycle coefficients

date	beta_DTH	beta_VAS	beta_CIR_sqrt_rt1	beta_CIR_oneoversqrt_rt1
03.01.11	0.004049	-0.02036	-0.01843	0.02227
04.01.11	0.004114	-0.02187	-0.02034	0.02433
05.01.11	0.004269	-0.02262	-0.02137	0.02562
06.01.11	0.004381	-0.02377	-0.02297	0.02746
07.01.11	0.00448	-0.02499	-0.02473	0.02947
10.01.11	0.004459	-0.02601	-0.02622	0.0311
11.01.11	0.004409	-0.02692	-0.02756	0.03258
12.01.11	0.004314	-0.02765	-0.02867	0.03378
13.01.11	0.004027	-0.02685	-0.02749	0.03244
14.01.11	0.003529	-0.02444	-0.02395	0.02841
17.01.11	0.003356	-0.02417	-0.02356	0.02793
18.01.11	0.003181	-0.02372	-0.02296	0.02722
19.01.11	0.003077	-0.0236	-0.02279	0.02702
20.01.11	0.002989	-0.02379	-0.023	0.02723
21.01.11	0.00272	-0.02245	-0.02135	0.02529
24.01.11	0.002448	-0.0211	-0.01978	0.02341
25.01.11	0.002292	-0.02067	-0.01931	0.02281
26.01.11	0.001654	-0.01675	-0.01519	0.01782
27.01.11	0.001648	-0.01686	-0.01528	0.01793
28.01.11	0.001656	-0.0171	-0.01549	0.01818
31.01.11	0.001525	-0.01653	-0.01499	0.01754
01.02.11	0.001361	-0.01578	-0.01435	0.0167
02.02.11	0.001229	-0.01801	-0.01658	0.01895
03.02.11	0.001121	-0.01843	-0.01721	0.01952
04.02.11	0.0009586	-0.01787	-0.0168	0.01892
07.02.11	0.0009107	-0.01777	-0.01675	0.01884
08.02.11	0.0009096	-0.01795	-0.01697	0.01907
09.02.11	0.0001809	-0.0133	-0.01256	0.0136
10.02.11	0.0001532	-0.01313	-0.01242	0.01341
11.02.11	0.0002952	-0.01365	-0.01278	0.01394
14.02.11	0.0003754	-0.01397	-0.01301	0.01428
15.02.11	0.0005376	-0.01481	-0.01372	0.01521
16.02.11	0.0007926	-0.01635	-0.01511	0.01696
17.02.11	0.001006	-0.0179	-0.01664	0.01883
18.02.11	0.00143	-0.02118	-0.02006	0.02295
21.02.11	0.001673	-0.02362	-0.02292	0.0263
22.02.11	0.001665	-0.02406	-0.02351	0.02698
23.02.11	0.001604	-0.02397	-0.02344	0.02687
24.02.11	0.001664	-0.02508	-0.02488	0.02854
25.02.11	0.001663	-0.02573	-0.02582	0.02959
28.02.11	0.001651	-0.02622	-0.02656	0.03042
01.03.11	0.001594	-0.02646	-0.0271	0.03101
02.03.11	0.00164	-0.02761	-0.02876	0.0329
03.03.11	0.001143	-0.02282	-0.0224	0.02552
04.03.11	0.000615	-0.01778	-0.01633	0.01839
07.03.11	0.0004641	-0.01657	-0.01504	0.01683
08.03.11	0.000283	-0.01517	-0.0136	0.01508
09.03.11	-1.21E-05	-0.01288	-0.01129	0.01225
10.03.11	-0.0001112	-0.01179	-0.01003	0.01076
11.03.11	-7.13E-05	-0.01188	-0.01002	0.01079
14.03.11	-0.0001175	-0.01124	-0.00919	0.009835
15.03.11	1.88E-06	-0.01213	-0.01009	0.01094
16.03.11	0.0001081	-0.01299	-0.01104	0.01207
17.03.11	0.0001723	-0.01353	-0.01165	0.01279
18.03.11	0.0001987	-0.01374	-0.01189	0.01308
21.03.11	0.0001719	-0.01344	-0.01153	0.01266
22.03.11	0.000883	-0.02038	-0.0199	0.02248
23.03.11	0.0007565	-0.01943	-0.01888	0.02126
24.03.11	0.0006433	-0.01855	-0.01793	0.02011
25.03.11	0.00057	-0.01801	-0.01737	0.01943
28.03.11	-0.0002525	-0.01096	-0.009771	0.01031
29.03.11	-0.0004809	-0.008442	-0.006514	0.006531
30.03.11	-0.0006324	-0.006616	-0.00398	0.003619
mean	0.001544835	-0.018943143	-0.017965476	0.020608968
st dev	0.001470151	0.005290681	0.00595309	0.007313835

7. White's test of heteroscedasticity for Vasicek and CIR regressions

```
White's test for heteroskedasticity
OLS, using observations 2011/01/03-2011/09/30 (T = 195)
Dependent variable: uhat^2

             coefficient    std. error   t-ratio   p-value
  --------------------------------------------------------
  const       0.000481877   0.00347569    0.1386   0.8899
  r_t1        0.00139000    0.00719328    0.1932   0.8470
  sq_r_t1    −0.000816231   0.00359679   −0.2269   0.8207

  Unadjusted R-squared = 0.000668

Test statistic: TR^2 = 0.130281,
with p-value = P(Chi-square(2) > 0.130281) = 0.936936
```

```
White's test for heteroskedasticity
OLS, using observations 2011/01/03-2011/09/30 (T = 195)
Dependent variable: uhat^2

                    coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------------
  sqrt_rt1           −5.28867      3.13964     −1.684    0.0937  *
  oneoversqrt_rt1    −4.59550      2.88438     −1.593    0.1128
  sq_sqrt_rt1         1.39344      0.808727     1.723    0.0865  *
  X1_X2               7.44015      4.53346      1.641    0.1024
  sq_oneoversqr       1.05213      0.682069     1.543    0.1246

  Unadjusted R-squared = 0.044631

Test statistic: TR^2 = 8.703135,
with p-value = P(Chi-square(4) > 8.703135) = 0.068963
```

References

[1] Brigo, D. & Mercurio, F., *Interest Rate Models — Theory and Practice*. Springer, 2006.

[2] Back, K., *A Course in Derivative Securities*. Springer, 2005.

Other materials used concerning LaTeX

[3] Lvovskiy, S.M., *Typing and Coding in the LaTeX System (in Russian)*. MCNMO, 2003.

[4] Vorontsov, K.V., *LaTeX 2ε in Examples (in Russian)*, 2005.