

Indirect Robust Estimation of the Short-term Interest Rate Process

Veronika Czellar¹

Dept. of Econometrics, University of Geneva, Switzerland

G. Andrew Karolyi

Fisher College of Business, Ohio State University

Elvezio Ronchetti

Dept. of Econometrics, University of Geneva, Switzerland

February 17, 2005

Abstract

We introduce Indirect Robust Generalized Method of Moments (IRGMM), a new simulation-based estimation methodology, to model short-term interest rate processes. The primary advantage of IRGMM relative to classical estimators of the continuous-time short-rate diffusion processes is that it corrects both the errors due to discretization and the errors due to model misspecification. We apply this new approach to various monthly and weekly Eurocurrency interest rate series and show its good predictive performance.

Keywords: GMM and RGMM estimators, CKLS one factor model, indirect inference.

¹Much of this work was completed while Czellar was visiting the Department of Finance of the Fisher College of Business at Ohio State University whose hospitality is gratefully acknowledged. She would like to thank also Lombard Odier Darier Hentsch&Cie (Geneva) for the financial support of this visit. Also helpful comments by seminar participants at Ohio State University and the University of Geneva are gratefully acknowledged.

1. Introduction

Understanding the dynamics of the short-term interest rate is of fundamental importance for many financial applications. Although these data have been subjected to extensive analysis, some basic issues remain unresolved. This may stem in part from the difficulties associated with the statistical analysis of continuous-time processes. For example, in complex statistical models, like diffusion models described by stochastic differential equations (SDE) of the form,

$$dy_t = \eta(y_t)dt + \sigma(y_t)dW_t, \quad (1)$$

where $\eta(\cdot)$ and $\sigma(\cdot)$ are the drift and the volatility and W_t a Wiener process, it is often difficult or impossible to carry out standard likelihood based estimation and inference. Such SDEs play an important role when modeling the short-term interest rate; see, for example, Vasicek (1977), Dothan (1978), Brennan and Schwartz (1977), Cox, Ingersoll and Ross (1981, 1985), Chan, Karolyi, Longstaff and Sanders (1992), Brenner, Harjes and Kroner (1996), Ahn and Gao (1999) among others.

A great deal of progress has been made recently in developing efficient tools for estimating and testing continuous-time models of the short-rate process; for instance, the Efficient Method of Moments (Gallant and Tauchen, 1996; Andersen and Lund, 1997), weighted least squares estimation (Chapman and Pearson, 1999), simulated maximum likelihood estimation (Durham and Gallant, 2002; Durham, 2003), a Gibbs sampling-based Markov Chain Monte Carlo algorithm (Kalimipalli and Susmel, 2004), and other innovative techniques as in Conley, Hansen, Luttmer and Scheinkman (1997), Chapman, Long and Pearson (1999).

It is well known that estimators based on a discretized version of (1) are biased, see for instance Gouriéroux, Monfort and Renault (1993). To tackle the problem arising from discretization several approaches have been proposed. Among others, non-parametric techniques (Aït-Sahalia, 1996; Stanton, 1997; Pritsker, 1998; Hong and Li, 2005; Johannes, 2004), pseudo-likelihood estimations (Aït-Sahalia, 1999; Aït-Sahalia, 2002) and indirect techniques (Broze, Scaillet and Zakoïan, 1995) were used to estimate the short-term interest rate process. An additional problem arising when estimating the short rate process is the possible model misspecification which can lead to biased estimators and misleading testing results; see a simple example in Section 3. The theory of robust statistics can be used to avoid this problem. Specifically, Dell'Aquila, Ronchetti and Trojani (2003) used robust techniques to estimate and compare discrete-time interest rate processes. Finally, Genton and Ronchetti (2003) combined robust techniques with indirect inference and its robust indirect inference.

Typically the estimation of (1) is performed by means of an auxiliary model which is a discretized version of the SDE. The resulting indirect estimation (Gouriéroux *et al.*, 1993) is based on the following idea. Given a sample of observations generated from a probability model $F(\theta)$, $\theta \in \mathbb{R}^p$, define an auxiliary model $\tilde{F}(\mu)$ where the parameter $\mu \in \mathbb{R}^r$ is easier to estimate than θ . For instance, $F(\theta)$ could be the diffusion model (1) or a fine discretization of (1) and $\tilde{F}(\mu)$ a crude discretization of (1). The main steps of the indirect estimation are the following:

- (i) the auxiliary estimator $\tilde{\mu}$ is calculated with the original sample;

- (ii) pseudo-observations are simulated from the true model $F(\theta)$ generating S samples of pseudo-data

$$\{y_t^1(\theta)\}_{t=1,\dots,n}, \dots, \{y_t^S(\theta)\}_{t=1,\dots,n};$$

- (iii) the auxiliary estimators $\tilde{\mu}_s(\theta)$ are calculated with the simulated pseudo-observations;
- (iv) finally the indirect estimator $\hat{\theta}_I$ of θ is obtained by minimizing the distance between the auxiliary estimators $\tilde{\mu}$ and $\overline{\tilde{\mu}_S}(\theta) = \frac{1}{S} \sum_{s=1}^S \tilde{\mu}_s(\theta)$.

Gouriéroux and Monfort (1996) proved that under certain conditions the indirect estimator $\hat{\theta}_I$ is consistent for θ and asymptotically normal. Moreover, if the bias of the auxiliary estimator is of order $\mathcal{O}(1)$, then indirect estimation reduces this bias, *i.e.* the bias of $\hat{\theta}_I$ is of order $\mathcal{O}(n^{-1})$.

This procedure relies on the assumption that the underlying model $F(\theta)$ is exact, *i.e.* it has generated the data. If the underlying model $F(\theta)$ is misspecified, Genton and Ronchetti (2003) showed that even the indirect estimators are biased. To eliminate this bias they developed *robust indirect estimation* which is based on a robust estimator of the auxiliary parameter. They proved that if the auxiliary estimator is a consistent and robust estimator of the parameter, then the associated indirect estimator is also robust. Applying indirect robust estimation to SDEs of the form (1) reduces both the bias due to discretization and to contamination.

In this paper we combine and extend these results and we apply them explicitly to the statistical analysis of models described by (1). In particular, we define the *indirect robust GMM (IRGMM)*, a new simulation-based estimation of SDEs and apply it to various monthly and weekly Eurocurrency interest rate series. We compare empirical results obtained with IRGMM to those obtained with classical GMM, robust GMM (RGMM) and indirect GMM (IGMM). For each data we considered, indirect estimators, and in most of cases the IRGMM, have the higher predictive performance.

The paper is organized as follows. Section 2 describes the IGMM estimation applied to the well-known CKLS models. In section 3 we propose IRGMM to estimate interest rates and explain the advantages of this technique with respect to classical estimation methods. Section 4 presents the results of IRGMM estimation of Eurocurrency for US data and compare our results with classical estimators. Moreover, we propose a methodology for forecasting using IRGMM estimators. Finally section 5 concludes the article with some open problems and suggestions for further work.

2. IGMM : correction of the bias due to the discretization

Let y_0, \dots, y_n be observations generated from the CKLS diffusion model (Chan *et al.*, 1992):

$$dy_t = (\alpha + \beta y_t)dt + \sigma y_t^\gamma dW_t, \quad (2)$$

where α represents the long term drift, β the mean reversion parameter, σ the unconditional average volatility and γ the elasticity parameter. We consider the data as

realizations of the diffusion model (2) at discrete times $t = 0, 1, \dots, n$. In general, it is impossible to determine the form of the distribution of $\{y_t\}_{t=0,1,\dots,n}$ from the continuous model (2). To estimate $\theta = (\alpha, \beta, \sigma, \gamma)$ let us consider a discretization of equation (2) with auxiliary parameters $\mu = (\alpha^*, \beta^*, \sigma^*, \gamma^*)$:

$$y_t = y_{t-1} + \alpha^* + \beta^* y_{t-1} + \sigma^* y_{t-1}^{\gamma^*} \epsilon_t \quad (3)$$

where $\{\epsilon_t\}_{t=1,\dots,n}$ are independent identically distributed standard normal variables. Hereafter we will refer to this equation as "crude discretization". For instance, Chan *et al.* (1992) estimated (3) for US treasury bill data by GMM. The CKLS orthogonality function is given by:

$$h(\{y_t\}, \mu) = \begin{pmatrix} \nu_t \\ \nu_t y_{t-1} \\ \nu_t^2 - \sigma^{*2} y_{t-1}^{2\gamma^*} \\ (\nu_t^2 - \sigma^{*2} y_{t-1}^{2\gamma^*}) y_{t-1} \end{pmatrix}, \quad (4)$$

where $\nu_t = y_t - (1 - \beta^*) y_{t-1} - \alpha^*$. In this situation we have the fully identified case, i.e. 4 parameters to estimate with 4 orthogonality conditions and the GMM estimator associated with the orthogonality function (4) is $\tilde{\mu}$ such that

$$\frac{1}{n} \sum_{t=1}^n h(\{y_t\}, \tilde{\mu}) = 0. \quad (5)$$

The GMM estimator $\tilde{\mu}$ is a consistent estimator of $\mu = (\alpha^*, \beta^*, \sigma^*, \gamma^*)$ in the model with crude discretization (3) but is inconsistent for $\theta = (\alpha, \beta, \sigma, \gamma)$ in the diffusion model (2). To correct the inconsistency under the diffusion model due to discretization we can use the *indirect GMM (IGMM)* estimator according to the following steps:

1) We simulate *pseudo-observations* from a fine discretization of (2): we divide the time interval $\Delta t = 1$ into m subintervals of length $\delta = \frac{1}{m}$. The Euler approximation corresponding to the time interval δ is the process $\{y_{k\delta}^{(\delta)}\}_{k=0,1,\dots,mn}$ defined by

$$y_{(k+1)\delta}^{(\delta)} = y_{k\delta}^{(\delta)} + \delta(\alpha + \beta y_{k\delta}^{(\delta)}) + \sigma y_{k\delta}^{(\delta)\gamma} \sqrt{\delta} \epsilon_k^{(\delta)}, \quad (6)$$

where the $\epsilon_k^{(\delta)}$ are standard normal variables. The process can be simulated² for every θ . Selecting the data at time $k\delta \in \mathbb{N}$, we obtain pseudo-data $\{y_t^{(1)}(\theta)\}_{t=0,1,\dots,n}$. Then we simulate S pseudo-datasets from this model:

$$\{y_t^s(\theta)\}_{t=0,1,\dots,n}, \quad s = 1, \dots, S, \quad S \geq 1. \quad (7)$$

2) For every s , we construct the *auxiliary* GMM estimator defined in (5) with the pseudo-data $\{y_t^s(\theta)\}_{t=0,1,\dots,n}$ and obtain estimators $\tilde{\mu}_s(\theta)$, $s = 1, \dots, S$ which are functions of the parameter θ . Let us denote by $\overline{\tilde{\mu}}_S(\theta) = \frac{1}{S} \sum_{s=1}^S \tilde{\mu}_s(\theta)$ the mean of these estimators.

²If $y_{k\delta}^{(\delta)} < 0$ we replace $y_{k\delta}^{(\delta)\gamma}$ by $|y_{k\delta}^{(\delta)}|^\gamma$.

3) The IGMM estimator $\hat{\theta}_{IGMM}$ is the one which minimizes the distance between the auxiliary estimator $\tilde{\mu}$ (computed on real data) and the mean $\overline{\tilde{\mu}_S}(\theta)$:

$$\hat{\theta}_{IGMM} = \arg \min_{\theta} (\tilde{\mu} - \overline{\tilde{\mu}_S}(\theta))^T \Omega (\tilde{\mu} - \overline{\tilde{\mu}_S}(\theta)), \quad (8)$$

where Ω is a positive definite symmetric matrix.

3. IRGMM: consistency under misspecifications of the diffusion model

One of the hypotheses assuring the consistency of the indirect estimator is that the real observations $\{y_t\}$ are generated from the diffusion model (2). In reality, the presence of jumps and high kurtosis in the increments on real data shows that the diffusion model may be *misspecified*. As an illustration, let us consider the diffusion model (2) where the increments come from an ε -neighborhood of a standard normal distribution:

$$W_t - W_{t-1} = \epsilon_t \sim (1 - \varepsilon)\mathcal{N}(0, 1) + \varepsilon G, \quad (9)$$

with $0 \leq \varepsilon < 1$ and G an unknown distribution. In Figure 3.1 we represent simulated data of size $n = 300$ generated from a contaminated diffusion model (2) with (9), where $G = \mathcal{N}(0, 5^2)$, $\varepsilon = 0.05$, $\alpha = 0$, $\beta = -0.001$, $\sigma = 0.3$, $\gamma = 1$.³

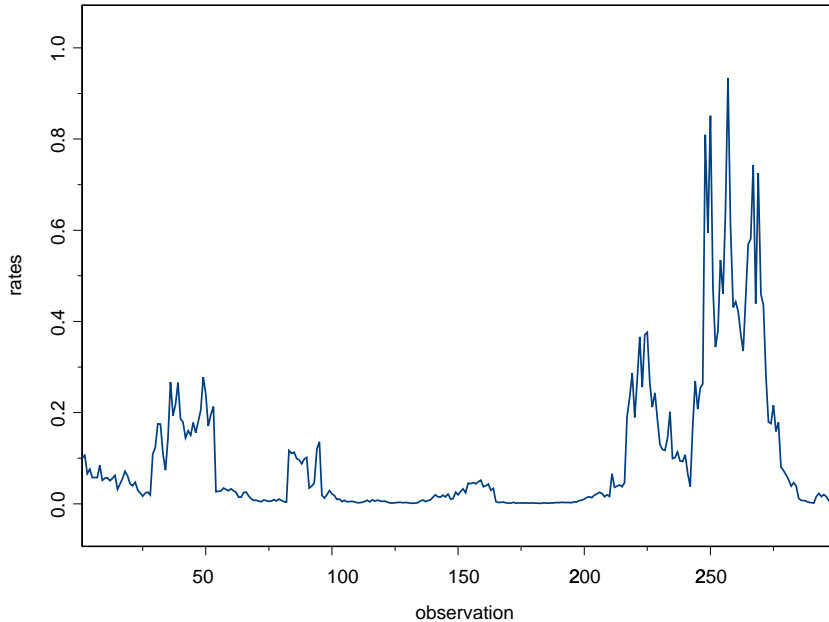


Figure 3.1: Simulated data from a diffusion model with 5% contamination.

³This case corresponds to the geometric Brownian motion with drift, which has an "exact discretization": $\log\left(\frac{y_t}{y_{t-1}}\right) = \beta - \frac{\sigma^2}{2} + \sigma\epsilon_t$.

Figure 3.1 shows that a contaminated diffusion model can generate jumps in the data. The kurtosis of the increments in a contaminated diffusion model with $G = \mathcal{N}(0, \tau^2)$ is $\kappa = 3(1 - \varepsilon + \varepsilon\tau^2)^2$ which is around 14.5 in the example given above. To illustrate how this kind of contamination affects the classical and indirect estimation procedures, let us consider the CKLS model (2) with $\alpha = \beta = 0$ and $\gamma = 1$. We obtain the model of Dothan:

$$dy_t = \sigma y_t dW_t. \quad (10)$$

Let $(n + 1)$ observations of y_t , $t = 0, 1, \dots, n$ to be given. To estimate σ^2 , we proceed to the crude discretization (3) with restrictions: $\alpha = \beta = 0$ and $\gamma = 1$. We obtain

$$y_t = y_{t-1} + \sigma^* y_{t-1} \epsilon_t, \quad (11)$$

where the ϵ_t are i.i.d. standard normal variables. The maximum likelihood estimator of σ^{*2} is

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (r_t - 1)^2, \quad (12)$$

where $r_t = y_t/y_{t-1}$.

Figure 3.2 compares the bias of the auxiliary and the asymptotic bias of the indirect estimators of σ^2 under 1%, 3%, 5% contamination as a function of τ . The real parameter was set to $\sigma = 0.5$. For the computation of the bias of auxiliary and indirect estimators see Appendix A.1-A.3.

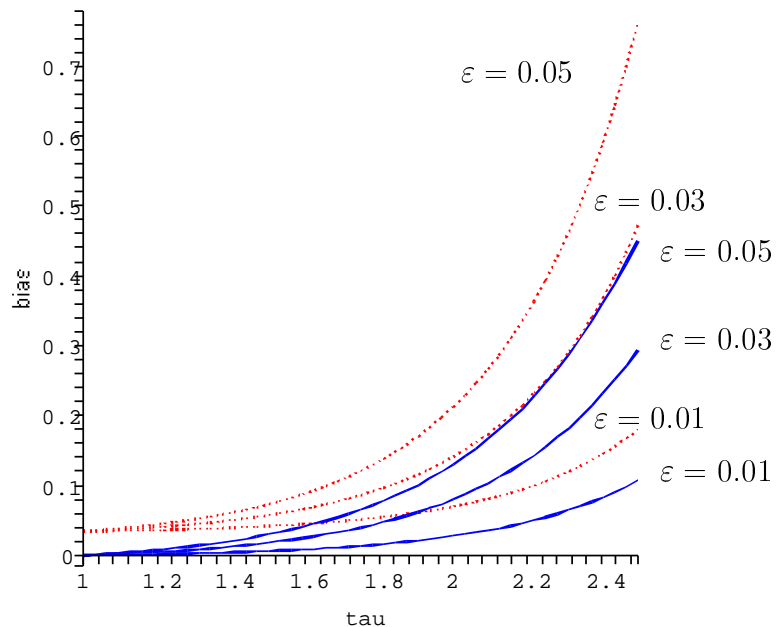


Figure 3.2: Bias of the auxiliary (\dots) and indirect ($—$) estimators of σ^2 .

Figure 3.2 shows that auxiliary estimator is biased even under the model without contamination. Under the model (without contamination) the indirect estimator corrects the bias due to discretization. But neither the auxiliary nor the indirect estimators can correct the bias due to contamination when $\tau > 2$. Indeed, the bias increases exponentially with τ^2 .

Our goal is to construct an indirect estimator which is robust against misspecifications of the underlying stochastic structure of the model, i.e. a robust estimator of the diffusion model (2) which represents the structure of the majority of the data. Genton and Ronchetti (2003) have shown that indirect estimators are robust if the auxiliary estimator is a robust and consistent estimator of the parameter of the auxiliary model. In our case, the auxiliary model being the crude discretization (3), we need a robust estimator of the parameter μ in (3). The robust version of the GMM estimator (see Ronchetti and Trojani, 2001 and Dell’Aquila, Ronchetti and Trojani, 2003) is defined by (5), where the classical orthogonality condition h is replaced by a truncated orthogonality function:

$$h_c^{A,\tau}(\{y_t\}, \mu) = A[h(\{y_t\}, \mu) - \tau] \min \left(1, \frac{c}{\|A[h(\{y_t\}, \mu) - \tau]\|} \right), \quad (13)$$

where c is a tuning constant. The non-singular matrix $A \in \mathbb{R}^4 \times \mathbb{R}^4$ and the vector $\tau \in \mathbb{R}^4$ are determined by the implicit equations (20) and (21) defined in Ronchetti and Trojani (2001).

Let us denote the robust GMM (RGMM) estimator associated with the truncated orthogonality function (13) by $\tilde{\mu}_{RGMM}$. The *indirect robust GMM (IRGMM)* estimator $\hat{\theta}_{IRGMM}$ is an indirect estimator constructed with RGMM auxiliary estimators following the steps described in Section 2. The IRGMM estimation procedure is summarized in Figure 3.3⁴.

Since the RGMM estimator is a consistent and robust estimator of the auxiliary parameter μ , the IRGMM is a consistent and robust estimator of the diffusion parameter θ (for consistency and robustness of RGMM estimators see Ronchetti and Trojani, 2001; for consistency of indirect estimators see Gouriéroux *et al.*, 1993; for robustness of indirect estimators see Genton and Ronchetti, 2003). The IRGMM estimator corrects both the errors due to discretization and the errors due to contamination of the underlying diffusion model.

⁴For numerical computation of the IRGMM, our C program can be downloaded from the webpage:

<http://www.unige.ch/ses/metri/assistants/czellar/Research.htm>

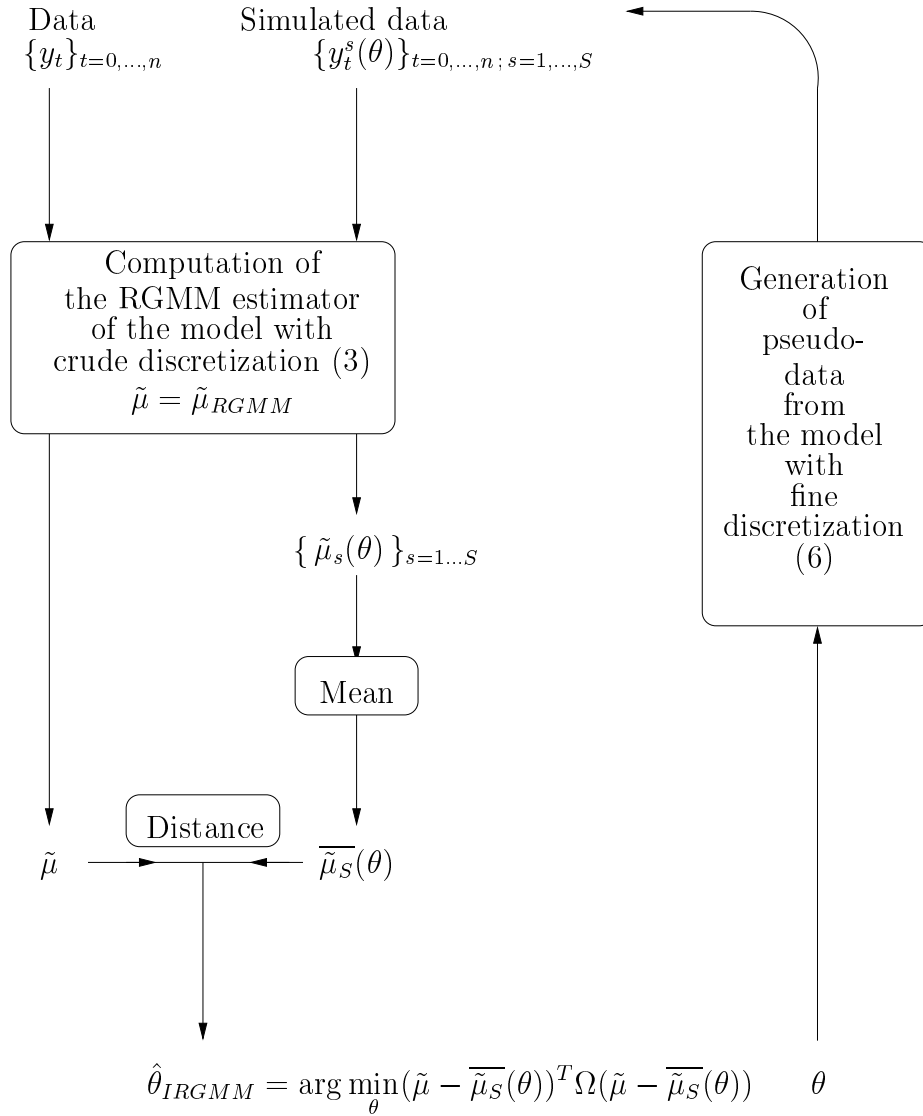


Figure 3.3 Construction of the IRGMM estimator $\hat{\theta}_{IRGMM}$

4. Estimation and forecast

4.1 Estimation of models for US Eurocurrency rates

We consider the monthly Eurocurrency rates for US covering the period from February 28, 1975 to December 31, 2002. The data is plotted in Figure 4.1.

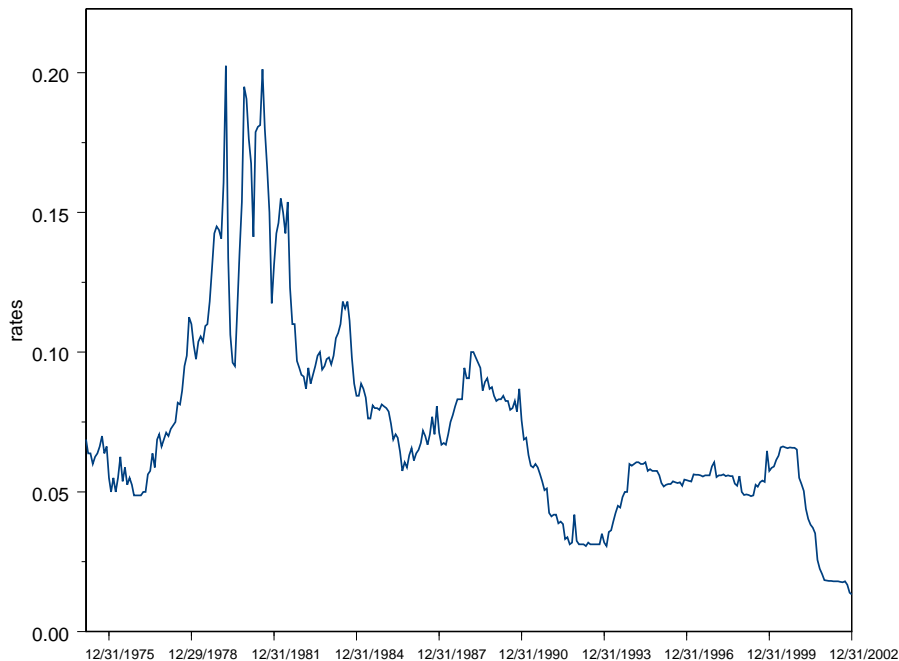


Figure 4.1: US Eurocurrency rates from February 28, 1975 to December 31, 2002.

Table 4.1 represents the means, standard deviations, skewness, kurtosis and correlations of the monthly interest rates.

	<i>Data</i>	<i>N</i>	<i>Mean</i>	<i>StDev</i>	<i>Sk</i>	<i>Ku</i>	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
<i>Monthly</i>	y_t	335	0.07319	0.03537	1.2507	1.9307	0.9691	0.9347	0.9024	0.8692	0.8533
	Δy_t	334	-0.00016	0.00816	-1.1001	20.9559	0.0673	-0.0331	0.0185	-0.3243	-0.0081

Table 4.1: US Eurocurrency rates statistics.

The classic and indirect GMM estimators are given in Table 4.2 with t-statistics⁵ in parentheses. For indirect estimators, we chose $\Omega = \text{diag}\left(\frac{1}{\tilde{\mu}_i^2}\right)_{i=1,\dots,4}$ with $\tilde{\mu}$ the auxiliary estimator and the constants were set to $\delta = 1/22$ and $S = 25$. The choice of these parameters is explained in Appendix A.4. For RGMM the value of the tuning constant was $c = 5.85$. IRGMM estimators were computed with auxiliary RGMM estimators defined by (13) where $\tau = 0$ ⁶.

⁵For IGMM and IRGMM we used the asymptotic distribution calculated in A.2 of the Appendix with $\varepsilon = 0$.

⁶The consistency parameter τ can be dropped because indirect inference corrects the inconsistency of the auxiliary estimator.

In order to compare the classical and indirect estimations, we perform a Monte Carlo simulation based on the estimated parameters to forecast the monthly Eurocurrency rates for 2003.

4.2 Forecast of US Eurocurrency rates

When forecasting with classical GMM, a natural way is to simulate from the model with crude discretization (3) with parameters from Table 4.2 and residuals ϵ_t generated from a standard normal distribution. Figure 4.2 represents a path of simulated future observations for monthly values in 2003 using the last observed interest rate in December 2002 and GMM parameter estimates from Table 4.2.

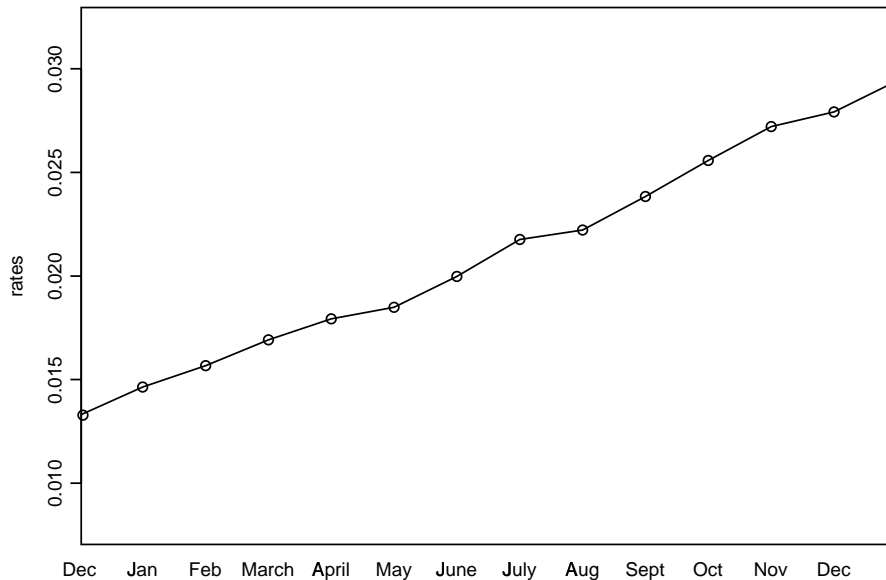


Figure 4.2: Simulated path for future US Eurocurrency rates for 2003.

We simulate 1000 of such paths and Figure 4.3 represents the simulated future values by means of boxplots. The plots represent the real values that occurred in 2003. These values are far away from the forecasted intervals given by GMM estimation and so the forecast with GMM doesn't perform well. Figure 4.4 represents the forecast for 2003 with RGMM parameter estimates from Table 4.2. The forecast based on the RGMM estimator performs better than the forecast based on the GMM estimator but most of the observed values are still outside the forecasted intervals in the second half of the year.

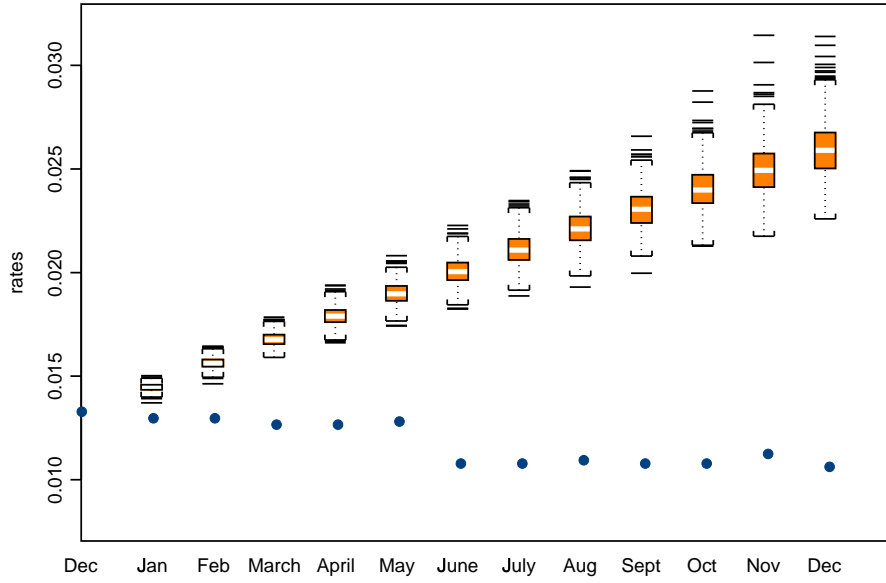


Figure 4.3: Forecasted intervals for US Eurocurrency rates for 2003 using GMM.

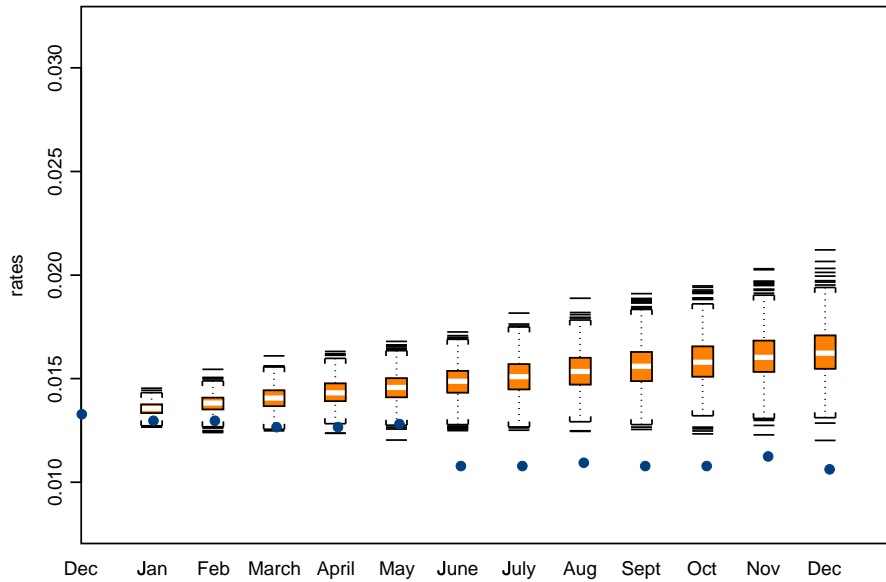


Figure 4.4: Forecasted intervals for US Eurocurrency rates for 2003 using RGMM.

For the forecast with IGMM and IRGMM, we generate data from the finer discretization defined in (6) with $\delta = 1/22$ and with the estimated parameters in Table 4.2. In this way we simulate daily data and by collecting every 22nd value we obtain simulated monthly rates. Figure 4.5 represents a simulated path for monthly interests rates in 2003 with the indirect method and Figure 4.6 represents 1000 such paths.

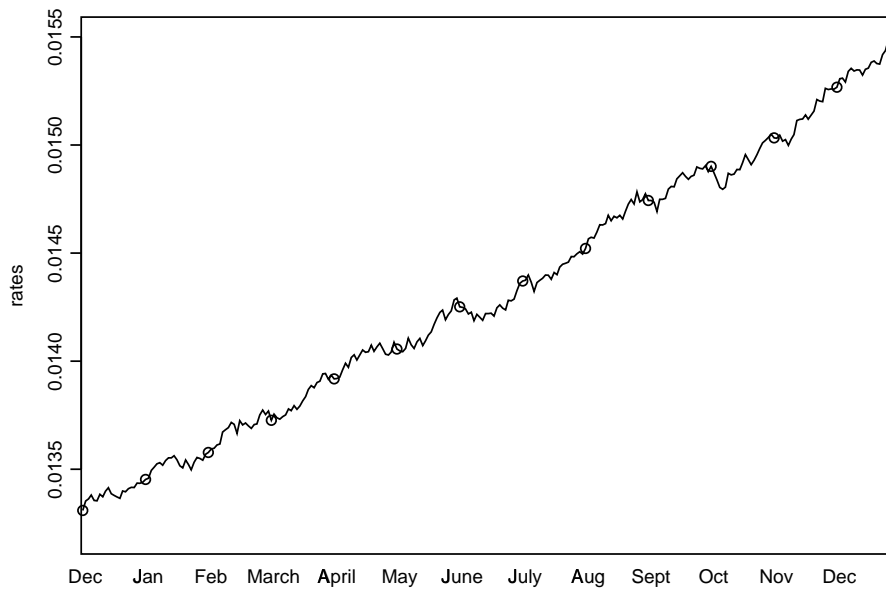


Figure 4.5: Simulated path for future US Eurocurrency rates for 2003 using IGMM.

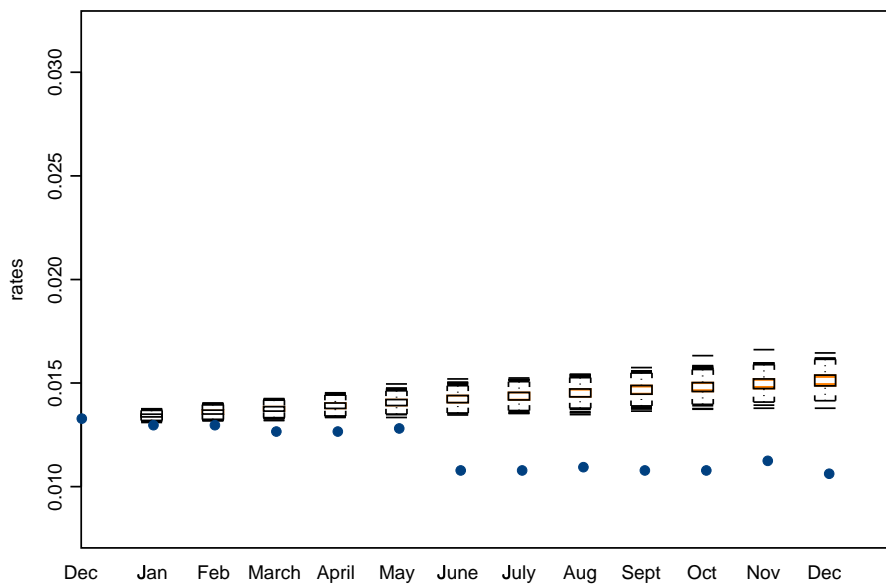


Figure 4.6: Forecasted intervals for US Eurocurrency rates for 2003 using IGMM.

Finally, Figure 4.7 shows the forecast based on IRGMM estimation.

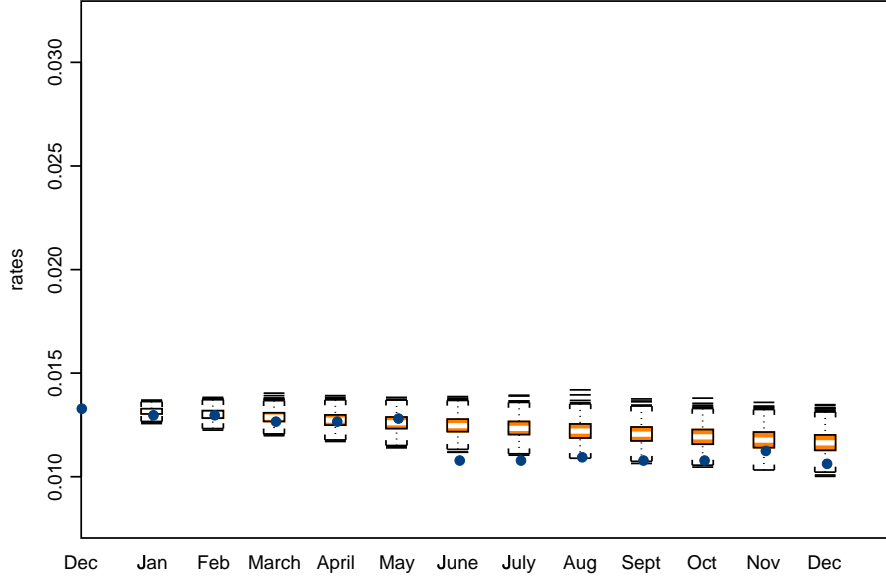


Figure 4.7: Forecasted intervals for US Eurocurrency rates for 2003 using IRGMM.

Clearly the forecast with IRGMM estimators performs better than those computed with GMM, RGMM and IGMM estimators.

In order to compare numerically the goodness-of-fit of different forecasting techniques we define the following measures:

$$\widehat{RAMSE} = \left(\frac{1}{n} \sum_{t=1}^n \frac{1}{m} \sum_{s=1}^m (y_t^{(s)} - y_t)^2 \right)^{\frac{1}{2}}, \quad (14)$$

$$\widehat{AMAD} = \frac{1}{n} \sum_{t=1}^n \text{median}_s (|y_t^{(s)} - \text{median}_l(y_t^{(l)})|), \quad (15)$$

$$\widehat{AMEDBIAS} = \frac{1}{n} \sum_{t=1}^n (|\text{median}_s(y_t^{(s)}) - y_t|), \quad (16)$$

where $\{y_t^{(s)}\}_{t=1, \dots, n, s=1, \dots, m}$ denote m simulated values for n number of periods, and $\{y_t\}_{t=1, \dots, n}$ are the real observations. $\widehat{AMEDBIAS}$ measures the bias of the the forecast, \widehat{AMAD} measures the variability of the forecast and \widehat{RAMSE} is the root mean squared error of the forecast, a combination of bias and variability.

In Table 4.2 we report the \widehat{RAMSE} , \widehat{AMAD} and $\widehat{AMEDBIAS}$ for each technique with $m = 1000$ and $n = 12$. For each goodness-of-fit measure, the smallest values in each column are underlined.

	α	β	σ	γ	\widehat{RAMSE}	\widehat{AMAD}	$\widehat{A.M.B.}$	$\cdot 10^{-3}$
GMM	0.00148 (0.80)	-0.02246 (-0.75)	0.73331 (1.50)	1.93049 (6.73)	9.28	0.42	7.92	
IGMM	0.00020 (0.11)	-0.00349 (-0.11)	0.82777 (0.91)	2.09933 (4.31)	2.87	<u>0.16</u>	2.39	
RGMM	0.00034 (0.29)	-0.00583 (-0.30)	0.31669 (1.99)	1.61110 (7.79)	3.60	0.49	2.93	
IRGMM	-0.00016 (-0.10)	0.00230 (0.08)	0.52401 (1.64)	1.83978 (7.25)	<u>1.02</u>	0.26	<u>0.70</u>	

Table 4.2: Parameter estimates and goodness-of-fit test for US Eurocurrency rates.

Table 4.2 shows that the smallest bias of the forecast is obtained using the IRGMM technique. The smallest variability of the forecast is obtained using the IGMM estimator but the gain is not so important compared to the variability of the forecasts computed by means of other techniques. Finally the accuracy of the forecast is the best with the IRGMM estimator which provides the smallest value of \widehat{RAMSE} . These values confirm the results obtained in Figures 4.3-4.7.

4.3 Forecast of international rates

We consider monthly Eurocurrency rates for Japan during the period 1/31/1990-8/30/2002:

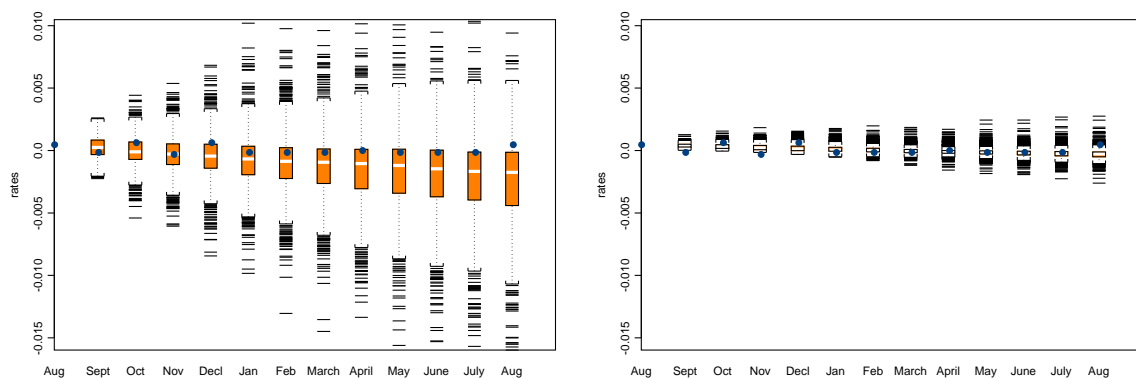


Figure 4.14: Forecasts for monthly Eurocurrency rates of Japan for the period 8/30/2002-08/29/2003 using RGMM and IRGMM.

	α	β	σ	γ	\widehat{RAMSE}	\widehat{AMAD}	$\widehat{A.M.B.}$	$\cdot 10^{-3}$
GMM	-0.00019 (-0.91)	-0.01132 (-1.08)	0.00788 (3.48)	0.28067 (3.70)	2.99	1.33	0.70	
IGMM	-0.00040 (-1.88)	-0.00421 (-0.40)	0.00882 (2.95)	0.32863 (3.42)	3.44	1.23	1.74	
RGMM	-0.00024 (-2.94)	-0.01030 (-2.53)	0.00786 (8.93)	0.29101 (9.87)	3.02	1.25	0.84	
IRGMM	-0.00006 (-0.77)	-0.01204 (-2.96)	0.01020 (11.60)	0.52590 (17.84)	<u>0.51</u>	<u>0.14</u>	<u>0.27</u>	

Table 4.3: Parameter estimates and goodness-of-fit test for monthly Eurocurrency rates of Japan.

For Japan the 3 goodness-of-fit measures are the smallest when using the IRGMM technique.

Estimation of monthly Eurocurrency rates for Switzerland during the period 1/31/1990-12/31/2002:

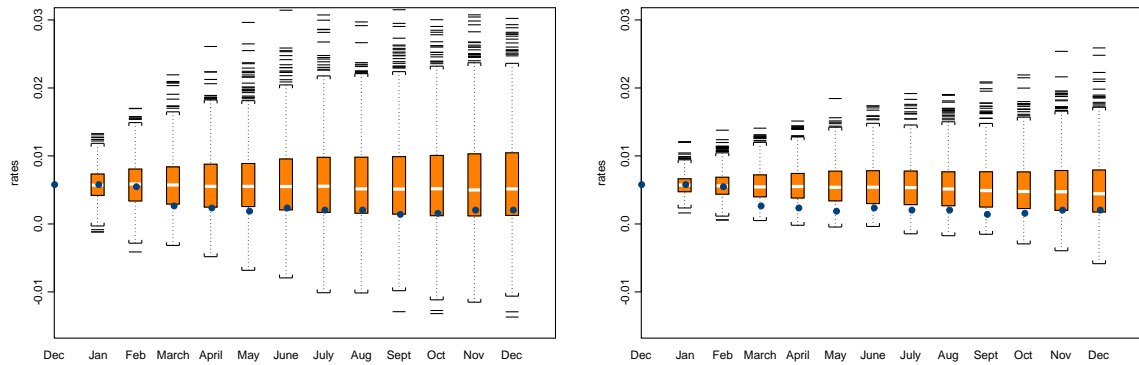


Figure 4.15: Forecasts for monthly Eurocurrency rates of Switzerland for the 2003 using RGMM and IRGMM.

	α	β	σ	γ	\widehat{RAMSE}	\widehat{AMAD}	$\widehat{A.M.B.}$	$\cdot 10^{-3}$
GMM	0.00014 (0.29)	-0.01881 (-1.53)	0.00769 (0.02)	0.22646 (2.34)	6.66	3.54	2.60	
IGMM	-0.00003 (-0.07)	-0.01091 (-0.87)	0.00749 (2.65)	0.23395 (2.11)	5.81	3.25	<u>1.82</u>	
RGMM	0.00018 (0.98)	-0.02130 (-4.40)	0.00699 (6.96)	0.20595 (5.41)	6.88	3.67	2.89	
IRGMM	0.00005 (0.26)	-0.01466 (-3.03)	0.00747 (7.44)	0.31878 (8.37)	<u>4.70</u>	<u>2.15</u>	2.58	

Table 4.4: Parameter estimates and goodness-of-fit test for monthly Eurocurrency rates of Switzerland.

For Switzerland the smallest value of the $\widehat{AMEDBIAS}$ is obtained by the IGMM but the variability of the forecast is large. The IRGMM gives the smallest value for \widehat{AMAD} and also for \widehat{RAMSE} which means that the forecast with IRGMM is the most accurate.

Estimation of monthly UK Eurocurrency rates during the period 1/31/1990-12/31/2002:

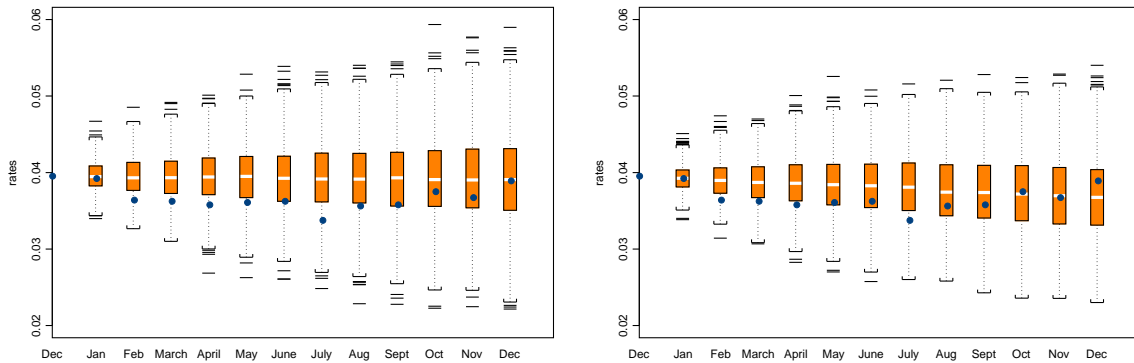


Figure 4.16: Forecasts for monthly UK Eurocurrency rates for the 2003 using RGMM and IRGMM.

	α	β	σ	γ	\widehat{RAMSE}	\widehat{AMAD}	$\widehat{A.M.B.}$	$\cdot 10^{-3}$
GMM	0.00094 (1.56)	-0.02282 (-2.55)	0.01129 (2.12)	0.55218 (3.12)	5.69	2.91	2.98	
IGMM	0.00073 (1.96)	-0.02075 (-2.28)	0.01222 (1.92)	0.60581 (3.06)	4.93	<u>2.55</u>	2.33	
RGMM	0.00078 (3.28)	-0.02042 (-5.80)	0.01015 (4.85)	0.51808 (7.43)	5.53	3.00	2.71	
IRGMM	0.00029 (1.23)	-0.01382 (-3.93)	0.01147 (5.48)	0.58527 (8.40)	<u>4.73</u>	2.74	<u>1.93</u>	

Table 4.5: Parameter estimates and goodness-of-fit test for monthly UK Eurocurrency rates.

For UK similar conclusions apply.

Tables 4.2-4.5 show that for every dataset we considered, IRGMM provides the smallest value of \widehat{RAMSE} and has overall the best performance.

Now we turn to the estimation and forecast using shorter time periods.

4.4. Forecast with 5-year periods of weekly data

Here we present the forecasts with 5-year periods of weekly US Eurocurrency data. We subdivided the time interval of 02/28/75-12/31/02 into periods of 5 years and using weekly data we forecast the 3 first months of the following (out-of-sample) period. The general pattern and the conclusions are similar to those in the previous case.

		α	β	σ	γ	\widehat{RAMSE}	\widehat{AMAD}	$\widehat{A.M.B.}$	$\cdot 10^{-3}$
75 – 79 :	GMM	-0.00037 (-0.50)	0.00945 (0.81)	0.12213 (1.16)	1.53424 (4.57)	20.18	9.46	9.26	
	IGMM	-0.00067 (-0.88)	0.01217 (1.02)	0.22844 (0.57)	1.95989 (2.73)	22.00	7.70	<u>9.00</u>	
	RGMM	-0.00048 (-1.14)	0.00730 (1.19)	0.00997 (2.04)	0.61084 (3.26)	15.74	4.58	9.80	
	IRGMM	-0.00031 (-0.41)	0.00732 (0.63)	0.03558 (0.05)	1.68691 (0.22)	<u>13.48</u>	<u>2.11</u>	9.35	
80 – 84 :	GMM	0.00257 (1.56)	-0.02241 (-1.48)	0.08206 (3.22)	1.16842 (7.39)	11.14	6.69	2.11	
	IGMM	0.00075 (0.44)	-0.00636 (-0.41)	0.09736 (2.91)	1.27609 (7.11)	10.63	6.28	2.06	
	RGMM	0.00257 (1.20)	-0.02147 (-1.12)	0.11705 (3.50)	1.35374 (10.24)	10.55	6.24	2.70	
	IRGMM	0.00094 (0.54)	-0.00786 (-0.49)	0.14145 (2.62)	1.49654 (7.31)	<u>8.84</u>	<u>5.27</u>	<u>1.61</u>	
85 – 89 :	GMM	0.00977 (2.26)	-0.12686 (-2.21)	1.03427 (0.15)	2.07052 (0.78)	11.16	6.58	3.56	
	IGMM	0.00878 (1.97)	-0.11301 (-1.91)	1.47481 (0.09)	2.31423 (0.49)	8.61	5.07	2.08	
	RGMM	0.00120 (1.60)	-0.01566 (-1.63)	0.00381 (0.95)	0.25763 (0.63)	5.20	3.09	1.96	
	IRGMM	0.00256 (0.97)	-0.03388 (-0.96)	0.22061 (0.02)	1.93304 (0.11)	<u>4.42</u>	<u>2.73</u>	<u>1.43</u>	
90 – 94 :	GMM	0.00150 (1.17)	-0.03219 (-1.08)	5.75370 (2.04)	2.59929 (2.33)	9.39	4.64	4.92	
	IGMM	0.00037 (0.28)	-0.00927 (-0.31)	14.2944 (0.86)	3.13175 (2.76)	6.20	2.69	3.41	
	RGMM	0.00039 (2.13)	-0.01101 (-2.95)	0.00327 (2.29)	0.30182 (2.13)	<u>5.07</u>	<u>2.15</u>	3.45	
	IRGMM	-0.00003 (-0.05)	-0.00208 (-0.16)	0.15822 (0.16)	1.53222 (0.68)	6.10	3.18	<u>3.35</u>	
95 – 00 :	GMM	0.01137 (8.84)	-0.20885 (-7.01)	24502.1 (1507.53)	5.63413 (5.06)	4.92	2.02	3.33	
	IGMM	0.01147 (8.74)	-0.20320 (-6.69)	189797 (11450.8)	6.56324 (5.78)	<u>2.89</u>	<u>1.17</u>	1.87	
	RGMM	0.00014 (0.16)	-0.00469 (-0.31)	0.00092 (0.31)	0.01259 (0.01)	3.34	1.46	1.98	
	IRGMM	0.00212 (0.55)	-0.03543 (-0.49)	16.3467 (0.05)	3.33949 (0.53)	3.08	1.69	<u>1.14</u>	

Table 4.6: Parameter estimates and goodness-of-fit test for weekly US Eurocurrency rates using 5-year period data.

		α	β	σ	γ	\widehat{RAMSE}	\widehat{AMAD}	$\widehat{A.M.B.}$	$\cdot 10^{-3}$
00 – 02 :	GMM	0.00031 (-1.74)	0.00101 (0.20)	0.00545 (1.67)	0.42439 (2.35)	4.30	1.48	3.39	
	IGMM	0.00000 (0.00)	0.06883 (4.73)	2.07202 (3.04)	0.91987 (9.04)	12.81	<u>0.02</u>	12.75	
	RGMM	-0.00048 (-2.89)	0.00464 (1.02)	0.00422 (2.19)	0.37601 (2.77)	2.89	1.33	1.63	
	IRGMM	-0.00038 (-0.71)	0.00164 (0.24)	0.00127 (0.09)	0.43122 (0.10)	<u>1.60</u>	0.31	<u>1.24</u>	

Table 4.7: Parameter estimates and goodness-of-fit test for weekly US Eurocurrency rates using 3-year period data.

5. Conclusion

In this paper we presented an empirical comparison of four estimation techniques of the diffusion model (1). We checked the performance of these techniques by comparing their predictive power on monthly Eurocurrency rates for US, UK, Japan and Switzerland and weekly rates for US. With each dataset, indirect estimators and particularly the IRGMM estimator provided the most accurate forecasts.

Two main research directions could be considered in the future. From a computational point of view, it would be useful to develop more efficient algorithms for the computation of the IRGMM estimator which would reduce the computational complexity of the present algorithm for the minimization of (8). This would speed up the computation of IRGMM. Secondly, in this paper we applied the new technique in the standard CKLS single factor model. However, extensions to other models are possible and it would be interesting, for instance, to study the performance of IRGMM in more complex diffusion models.

Appendix

Suppose that our sample of observations $\{y_t\}_{t=1,\dots,n}$ is generated from a contaminated model $(1 - \varepsilon)F(\theta) + \varepsilon G$, $0 \leq \varepsilon \leq 1$. In this appendix we provide the technical details to investigate the consistency and the asymptotic distribution of the indirect estimator.

A.1 Asymptotic bias of indirect estimation

Let us define the auxiliary estimator $\tilde{\mu}_n$ as the solution of a maximization criterion

$$\tilde{\mu}_n = \arg \max_{\mu} \rho_n(\{y_t\}, \mu), \quad (17)$$

where ρ_n is an objective function, for instance the log-likelihood function of the auxiliary model $\tilde{F}(\mu)$. Then the auxiliary estimators computed with the simulated pseudo-data are defined by:

$$\tilde{\mu}_n^s(\theta) = \arg \max_{\mu} \rho_n(\{y_t^s(\theta)\}, \mu), \quad s = 1, \dots, S. \quad (18)$$

Let us denote by $\hat{\theta}_{S_n}$ the indirect estimator associated with the auxiliary $\tilde{\mu}_n$ estimator:

$$\hat{\theta}_{S_n}(\Omega) = \arg \min_{\theta} (\tilde{\mu}_n - \tilde{\mu}_{S_n}(\theta))^T \Omega (\tilde{\mu}_n - \tilde{\mu}_{S_n}(\theta)), \quad (19)$$

and assume that the objective function $\rho_n(\{y_t\}, \mu)$ satisfies the following *regularity conditions*:

- (i) $\rho_n(\{y_t\}, \mu)$ and $\rho_n(\{y_t^s(\theta)\}, \mu)$ tend almost surely and uniformly respectively to $\eta_\varepsilon(\theta_0, \mu)$ and to $\eta(\theta, \mu)$ when $n \rightarrow \infty$;
- (ii) these limit functions have unique maxima with respect to μ :

$$b_\varepsilon(\theta_0) = \arg \max_{\mu} \eta_\varepsilon(\theta_0, \mu),$$

$$b(\theta) = \arg \max_{\mu} \eta(\theta, \mu);$$

- (iii) $\rho_n, \eta_\varepsilon, \eta$ are differentiable with respect to μ , and

$$\frac{\partial \eta_\varepsilon}{\partial \mu}(\theta_0, \mu) = \lim_{n \rightarrow \infty} \frac{\partial \rho_n}{\partial \mu}(\{y_t\}, \mu)$$

$$\frac{\partial \eta}{\partial \mu}(\theta, \mu) = \lim_{n \rightarrow \infty} \frac{\partial \rho_n}{\partial \mu}(\{y_n^s(\theta)\}, \mu)$$

- (iv) the only solutions of the asymptotic first order conditions are $b_\varepsilon(\theta_0)$ and $b(\theta)$:

$$\frac{\partial \eta_\varepsilon}{\partial \mu}(\theta_0, \mu) = 0 \Rightarrow \mu = b_\varepsilon(\theta_0)$$

$$\frac{\partial \eta}{\partial \mu}(\theta, \mu) = 0 \Rightarrow \mu = b(\theta)$$

- (v) b is bijective, differentiable with respect to θ and the inverse of $\frac{\partial b}{\partial \theta}(\theta_0^*)$ exists at $\theta_0^* = b^{-1}(b_\varepsilon(\theta_0))$.

Proposition A.1 *Under conditions (i)-(v) above the indirect estimator $\hat{\theta}_{S_n}$ has the asymptotic bias:*

$$\text{asbias}(\hat{\theta}_{S_n}, \varepsilon) = \theta_0^* - \theta. \quad (20)$$

Proof: From assumptions (i)-(v) we have:

$$\tilde{\mu}_n = \arg \max_{\mu} \rho_n(\{y_t\}, \mu) \xrightarrow{n \rightarrow \infty} \arg \max_{\mu} \eta_\varepsilon(\theta_0, \mu) = b_\varepsilon(\theta_0), \quad (21)$$

$$\tilde{\mu}_{S_n}(\theta) = \frac{1}{S} \sum_{s=1}^S \arg \max_{\mu} \rho_n(\{y_t^s(\theta)\}, \mu) \xrightarrow{n \rightarrow \infty} \arg \max_{\mu} \eta(\theta, \mu) = b(\theta). \quad (22)$$

Then:

$$\begin{aligned} \hat{\theta}_{S_n}(\Omega) &= \arg \min_{\theta} (\tilde{\mu}_n - \tilde{\mu}_{S_n}(\theta))^T \Omega (\tilde{\mu}_n - \tilde{\mu}_{S_n}(\theta)) \\ &\xrightarrow{n \rightarrow \infty} \arg \min_{\theta} (b_\varepsilon(\theta_0) - b(\theta))^T \Omega (b_\varepsilon(\theta_0) - b(\theta)) \\ &= \{\theta \mid b_\varepsilon(\theta_0) = b(\theta)\} \\ &= b^{-1}(b_\varepsilon(\theta_0)). \end{aligned} \quad (23)$$

□

A.2 Asymptotic distribution of the indirect estimator

Assume the following additional regularity conditions:

$$\begin{aligned} \text{(vi)} \quad & p \lim_{n \rightarrow \infty} -\frac{\partial^2 \rho_n}{\partial \mu \partial \mu^T}(\{y_t\}, b_\varepsilon(\theta_0)) = -\frac{\partial^2 \eta_\varepsilon}{\partial \mu \partial \mu^T}(\theta_0, b_\varepsilon(\theta_0)) = J_\varepsilon \text{ and} \\ & p \lim_{n \rightarrow \infty} -\frac{\partial^2 \rho_n}{\partial \mu \partial \mu^T}(\{y_t^s(\theta)\}, b(\theta)) = -\frac{\partial^2 \eta}{\partial \mu \partial \mu^T}(\theta, b(\theta)) = J(\theta) \end{aligned}$$

$$\begin{aligned} \text{(vii)} \quad & \sqrt{n} \frac{\partial \rho_n}{\partial \mu}(\{y_t\}, b_\varepsilon(\theta_0)) \xrightarrow[n \rightarrow \infty]{d} N(0, I_\varepsilon) \text{ and} \\ & \sqrt{n} \frac{\partial \rho_n}{\partial \mu}(\{y_t^s(\theta)\}, b(\theta)) \xrightarrow[n \rightarrow \infty]{d} N(0, I(\theta)) \end{aligned}$$

Proposition A.2 *Under conditions (i)-(vii) the indirect estimator $\hat{\theta}_{S_n}$ is asymptotically normally distributed:*

$$\sqrt{n}(\hat{\theta}_{S_n} - \theta_0^*) \xrightarrow[n \rightarrow \infty]{d} N(0, W), \quad (24)$$

where $W = \left(\frac{\partial b(\theta_0^*)}{\partial \theta}\right)^{-1} W_\mu \left(\frac{\partial b(\theta_0^*)}{\partial \theta}\right)^{-1}$ and $W_\mu = J_\varepsilon^{-1} I_\varepsilon J_\varepsilon^{-1} + \frac{1}{S} J(\theta_0^*)^{-1} I(\theta_0^*) J(\theta_0^*)^{-1}$.

Proof: First we derive the asymptotic expansions for $\tilde{\mu}_n$ and $\tilde{\mu}_{S_n}(\theta)$. From the first order conditions we have:

$$\frac{\partial \rho_n}{\partial \mu}(\{y_t\}, \tilde{\mu}_n) = 0 \quad (25)$$

and

$$\frac{\partial \rho_n}{\partial \mu}(\{y_t^s(\theta)\}, \tilde{\mu}_n^s(\theta)) = 0, \quad \forall s = 1, \dots, S. \quad (26)$$

Taylor expansions around $\mu = b_\varepsilon(\theta_0)$ for (25) and around $\mu = b(\theta)$ for (26) give:

$$\frac{\partial \rho_n}{\partial \mu}(\{y_t\}, b_\varepsilon(\theta_0)) + \frac{\partial^2 \rho_n}{\partial \mu \partial \mu^T}(\{y_t\}, b_\varepsilon(\theta_0))(\tilde{\mu}_n - b_\varepsilon(\theta_0)) = o\left(\frac{1}{n}\right) \quad (27)$$

and

$$\frac{\partial \rho_n}{\partial \mu}(\{y_t^s(\theta)\}, b(\theta)) + \frac{\partial^2 \rho_n}{\partial \mu \partial \mu^T}(\{y_t^s(\theta)\}, b(\theta))(\tilde{\mu}_n^s(\theta) - b(\theta)) = o\left(\frac{1}{n}\right) \quad (28)$$

$\forall s = 1, \dots, S.$

From (27) and (28) we obtain

$$\tilde{\mu}_n - b_\varepsilon(\theta_0) = \underbrace{\left(-\frac{\partial^2 \rho_n}{\partial \mu \partial \mu^T}(\{y_t\}, b_\varepsilon(\theta_0))\right)^{-1}}_{\rightarrow J_\varepsilon} \frac{\partial \rho_n}{\partial \mu}(\{y_t\}, b_\varepsilon(\theta_0)) + o\left(\frac{1}{n}\right)$$

and

$$\tilde{\mu}_n^s(\theta) - b(\theta) = \underbrace{\left(-\frac{\partial^2 \rho_n}{\partial \mu \partial \mu^T}(\{y_t^s(\theta)\}, b(\theta))\right)^{-1}}_{\rightarrow J(\theta)} \frac{\partial \rho_n}{\partial \mu}(\{y_t^s(\theta)\}, b(\theta)) + o\left(\frac{1}{n}\right)$$

$\forall s = 1, \dots, S,$

We conclude that

$$\sqrt{n}(\tilde{\mu}_n - b_\varepsilon(\theta_0)) = J_\varepsilon^{-1} \sqrt{n} \frac{\partial \rho_n}{\partial \mu}(\{y_t\}, b_\varepsilon(\theta_0)) + o\left(\frac{1}{n}\right) \quad (29)$$

and

$$\sqrt{n}(\tilde{\mu}_{S_n}(\theta) - b(\theta)) = J(\theta)^{-1} \frac{\sqrt{n}}{S} \sum_{s=1}^S \frac{\partial \rho_n}{\partial \mu}(\{y_t^s(\theta)\}, b(\theta)) + o\left(\frac{1}{n}\right). \quad (30)$$

Next we develop an asymptotic expansion for $\hat{\theta}_{S_n}$. From the definition of the indirect estimator (19) we have

$$\frac{\partial \tilde{\mu}_{S_n}(\hat{\theta}_{S_n})^T}{\partial \theta} \Omega(\tilde{\mu}_n - \tilde{\mu}_{S_n}(\hat{\theta}_{S_n})) = 0. \quad (31)$$

An expansion of $\frac{\partial \tilde{\mu}_{S_n}(\hat{\theta}_{S_n})^T}{\partial \theta}$ around the limit value $\theta_0^* = b^{-1}(b_\varepsilon(\theta_0))$ gives

$$\frac{\partial \tilde{\mu}_{S_n}(\hat{\theta}_{S_n})^T}{\partial \theta} = \frac{\partial \tilde{\mu}_{S_n}(\theta_0^*)^T}{\partial \theta} + \frac{\partial^2 \tilde{\mu}_{S_n}(\theta_0^*)}{\partial \theta \partial \theta^T}(\hat{\theta}_{S_n} - \theta_0^*) + o\left(\frac{1}{n}\right). \quad (32)$$

By replacing (32) in (31), we get

$$\frac{\partial \tilde{\mu}_{S_n}(\theta_0^*)^T}{\partial \theta} \Omega(\tilde{\mu}_n - \tilde{\mu}_{S_n}(\theta_0^*)) - \frac{\partial \tilde{\mu}_{S_n}(\theta_0^*)^T}{\partial \theta} \Omega \frac{\partial \tilde{\mu}_{S_n}(\theta_0^*)}{\partial \theta} (\hat{\theta}_{S_n} - \theta_0^*) = o\left(\frac{1}{n}\right).$$

Finally, with $\frac{\partial \tilde{\mu}_{S_n}(\theta_0^*)}{\partial \theta} \rightarrow \frac{\partial b(\theta_0^*)}{\partial \theta}$, we obtain

$$\begin{aligned}\sqrt{n}(\hat{\theta}_{S_n} - \theta_0^*) &= \left(\frac{\partial b(\theta_0^*)^T}{\partial \theta} \Omega \frac{\partial b(\theta_0^*)}{\partial \theta} \right)^{-1} \frac{\partial b(\theta_0^*)^T}{\partial \theta} \Omega \sqrt{n}(\tilde{\mu}_n - \tilde{\mu}_{S_n}(\theta_0^*)) + o\left(\frac{1}{\sqrt{n}}\right) \\ &= \left(\frac{\partial b(\theta_0^*)}{\partial \theta} \right)^{-1} \sqrt{n}(\tilde{\mu}_n - \tilde{\mu}_{S_n}(\theta_0^*)) + o\left(\frac{1}{\sqrt{n}}\right).\end{aligned}\tag{33}$$

By subtracting (30) from (29), we get

$$\begin{aligned}\sqrt{n}(\tilde{\mu}_n - \tilde{\mu}_{S_n}(\theta_0^*)) &= J_\varepsilon^{-1} \underbrace{\sqrt{n} \frac{\partial \rho_n}{\partial \mu}(\{y_t\}, b_\varepsilon(\theta_0))}_{\xrightarrow[n \rightarrow \infty]{d} N(0, I_\varepsilon)} - \frac{1}{S} (J(\theta_0^*))^{-1} \sum_{s=1}^S \underbrace{\sqrt{n} \frac{\partial \rho_n}{\partial \mu}(\{y_t^s(\theta)\}, b(\theta))}_{\xrightarrow[n \rightarrow \infty]{d} N(0, I(\theta))} + o\left(\frac{1}{\sqrt{n}}\right) \\ &\xrightarrow[n \rightarrow \infty]{d} N\left(0, J_\varepsilon^{-1} I_\varepsilon J_\varepsilon^{-1} + \frac{1}{S} J(\theta_0^*)^{-1} I(\theta_0^*) J(\theta_0^*)^{-1}\right).\end{aligned}$$

By denoting $W_\mu = J_\varepsilon^{-1} I_\varepsilon J_\varepsilon^{-1} + \frac{1}{S} J(\theta_0^*)^{-1} I(\theta_0^*) J(\theta_0^*)^{-1}$, we conclude that

$$\sqrt{n}(\hat{\theta}_{S_n} - \theta_0^*) \xrightarrow[n \rightarrow \infty]{d} N(0, W)\tag{34}$$

with $W = \left(\frac{\partial b(\theta_0^*)}{\partial \theta}\right)^{-1} W_\mu \left(\frac{\partial b(\theta_0^*)}{\partial \theta}\right)^{-1}$. □

A.3 Illustration on the model of Dothan

Let us consider the CKLS model (2) with $\alpha = \beta = 0$ and $\gamma = 1$. We obtain the model of Dothan:

$$dy_t = \sigma y_t dW_t.\tag{35}$$

This stochastic differential equation can be solved explicitly. By dividing (35) by y_t , we obtain

$$\frac{dy_t}{y_t} = \sigma dW_t,$$

and under the hypothesis that $y_t > 0$ and applying Itô's lemma to $f(y_t) = \log y_t$ we have:

$$d \log y_t = -\frac{\sigma^2}{2} dt + \sigma dW_t.$$

By integration between t and $t - 1$, we get

$$\log r_t = -\frac{\sigma^2}{2} + \sigma \epsilon_t,\tag{36}$$

where $r_t = \frac{y_t}{y_{t-1}}$ and ϵ_t is a standard normal variable. Hereafter we will refer to (36) as the *exact discretization* of the model of Dothan.

The model of Dothan is one of the few cases where we can determine the distribution of y_t and, therefore calculate the bias of the estimators. This model enables us

to compare results obtained by way of simulations and those obtained by theoretical calculation.

Let $(n + 1)$ observations of y_t , $t = 0, 1, \dots, n$ to be given. To estimate σ^2 , we proceed to the crude discretization (3) with restrictions: $\alpha = \beta = 0$ and $\gamma = 1$. We obtain

$$y_t = y_{t-1} + \sigma y_{t-1} \epsilon_t, \quad (37)$$

where the ϵ_t are i.i.d. standard normal variables. Then the maximum likelihood estimator of σ^2 is

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (r_t - 1)^2. \quad (38)$$

Let us now consider the model of Dothan when the underlying diffusion model is contaminated *i.e.* when the increments $\int_{t-1}^t dW_t = \epsilon_t$ in (1) are not standard Wiener processes but in a ε -neighborhood $(1 - \varepsilon)\mathcal{N}(0, 1) + \varepsilon\mathcal{N}(0, \tau^2)$.

In this case, there is a bias due to the contamination. Proposition A.3 gives the biases of the auxiliary and indirect estimators of the model of Dothan when the data is contaminated. The auxiliary estimator is based on the maximum likelihood estimator for the crude discretization.

Proposition A.3 *Let us suppose the model (36) with ϵ_t distributed by $\varepsilon\mathcal{N}(0, 1) + (1 - \varepsilon)\mathcal{N}(0, \tau^2)$. Let $\tilde{\sigma}^2$ be the estimator defined by (38). Then, we have:*

$$\text{bias}(\tilde{\sigma}^2, \varepsilon) = e^{\sigma^2} - (1 + \sigma^2) + \varepsilon \left(2 - e^{\sigma^2} + e^{\sigma^2(2\tau^2-1)} - 2e^{\sigma^2(\frac{\tau^2-1}{2})} \right), \quad (39)$$

$$\text{bias}(\hat{\sigma}_I^2, \varepsilon) = \ln \left(e^{\sigma^2} + \varepsilon \left(2 - e^{\sigma^2} + e^{\sigma^2(2\tau^2-1)} - 2e^{\sigma^2(\frac{\tau^2-1}{2})} \right) \right) - \sigma^2 + \mathcal{O}\left(\frac{1}{n}\right). \quad (40)$$

Proof: For the contaminated data, we have:

$$b_\varepsilon(\sigma^2) = E[(r_t - 1)^2] = e^{\sigma^2} - 1 + \varepsilon \left(2 - e^{\sigma^2} + e^{\sigma^2(2\tau^2-1)} - 2e^{\sigma^2(\frac{\tau^2-1}{2})} \right) \quad (41)$$

and (39) follows from (41). For the simulated data we have:

$$b(\sigma^{*2}) = E[(r_t^s(\sigma^{*2}) - 1)^2] = e^{\sigma^{*2}} - 1, \quad (42)$$

$$b^{-1}(b_\varepsilon(\sigma^2)) = \ln(b_\varepsilon(\sigma^2) + 1) = \ln \left(e^{\sigma^2} + \varepsilon \left(2 - e^{\sigma^2} + e^{\sigma^2(2\tau^2-1)} - 2e^{\sigma^2(\frac{\tau^2-1}{2})} \right) \right). \quad (43)$$

Finally, (23) with (43) gives (40). □

A.4 The Choice of the Parameters S and δ

To compute the indirect estimator, we have to chose a constant δ corresponding to the subdivision of the time interval and a number S of simulated pseudo-data. The 'real'

data is generated from the exact discretization with $\sigma^2 = 0.25$. The auxiliary maximum likelihood estimator $\tilde{\sigma}^2 = 0.3097$ and the robust MAD estimator is $\tilde{\sigma}_R^2 = 0.1056$.

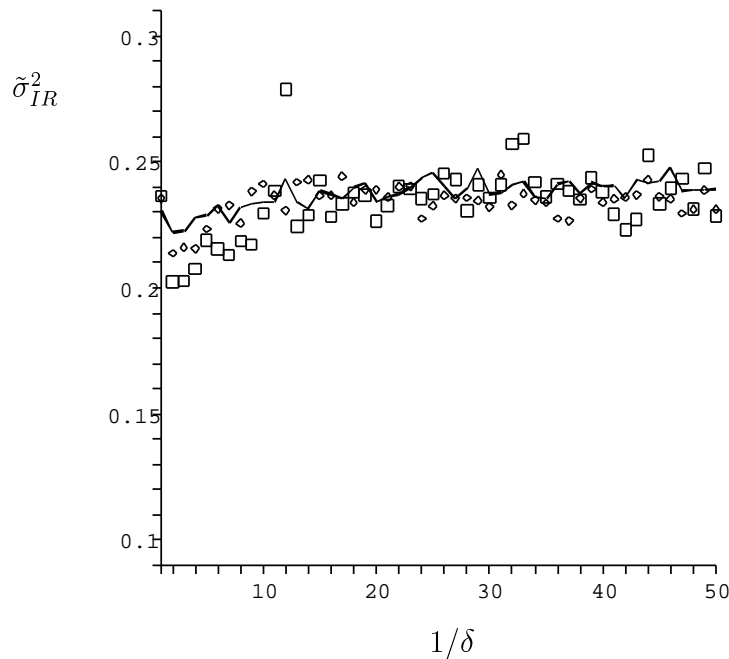


Figure A.4 Indirect robust estimators of σ^2 for $S = 10$ (boxes), for $S = 25$ (diamonds) and for $S = 100$ (curve).

Figure A.4 shows that the indirect robust estimator settles down for $\delta \leq 1/22$. Our choice for the parameter will be $S = 25$ because the gain of stability with $S = 100$ is not significant if we take into account the fact that computation time of an indirect estimator with $S = 100$ is four times longer than with $S = 25$. For the classical indirect estimators the results are more volatile.

References

- Ahn, D.H. and Gao, B. (1999), “A parametric nonlinear model of term structure dynamics”, *The Review of Financial Studies*, **12**, 721–762.
- Aït-Sahalia, Y. (1996), “Testing Continuous-Time Models of the Spot Interest Rate”, *Review of Financial Studies*, **9**, 385–426.
- Aït-Sahalia, Y. (1999), “Transition Densities for Interest Rate and Other Nonlinear Diffusions”, *Journal of Finance*, **54**, 1361–1395.
- Aït-Sahalia, Y. (2002), “Maximum Likelihood Estimation of Discretely Sampled Diffusions: a Closed-form Approximation Approach”, *Econometrica*, **70**, 223–262.
- Andersen, T.G and Lund, J. (1997), “Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate”, *Journal of Econometrics*, **77**, 343–377.
- Brennan, M.J. and Schwartz, E.S. (1977), “Savings Bonds, Retractable Bonds, and Callable Bonds”, *Journal of Financial Economics*, **3**, 133–155.
- Brenner, R.J., Harjes, R.H. and Kroner, K.F. (1996), “Another look at models of the short-term interest rate”, *Journal of Financial and Quantitative Analysis*, **31**, 85–107.
- Broze, L., Scaillet, O. and Zakoian, J.-M. (1995), “Testing for Continuous-time Models of the Short-term Interest Rate”, *Journal of Empirical Finance*, **2**, 199–223.
- Chan, K. C., Karolyi, G. A., Longstaff, F. A. et Sanders, A. B. (1992), “An Empirical Comparison of Alternative Models of the Short-Term Interest Rate”, *The Journal of Finance*, **47**, 1209–1227.
- Chapman, D., Long, J. and Pearson, N. (1999), “Using Proxies for the Short-Rate: When are Three Months Like an Instant?”, *Review of Financial Studies*, **12**, 763–806.
- Chapman, D.A. and Pearson, N.D. (2000), “Is the Short Rate Drift Actually Nonlinear?”, *Journal of Finance*, **55** 4, 355–388.
- Conley, T.G., Hansen, L.P., Luttmer, E.G.J. and Scheinkman, J.A. (1997), “Short-Term Interest Rates as Subordinated Diffusions”, *Review of Financial Studies*, **10**, 525–578.
- Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1981), “A re-examination of traditional hypotheses about the term structure of interest rates,” *Journal of Finance*, **36**, 769–799.
- Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985), “A Theory of the Term Structure of Interest Rates”, *Econometrica*, **53**, 385–407.
- Dell’Aquila, R., Ronchetti, E. and Trojani, F. (2003), “Robust GMM Analysis of Models for the Short Rate Process”, *Journal of Empirical Finance*, **10**, 373–397.
- Dothan, U. L. (1978), “On the Term Structure of Interest Rates”, *Journal of Financial Economics*, **6**, 59–69.

- Durham, G.B. (2003), “Likelihood-based specification analysis of continuous-time models of the short-term interest rate”, *Journal of Financial Economics*, **70**, 463–487.
- Durham, G.B., and Gallant, A.R. (2002), “Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes”, *Journal of Business and Economic Statistics*, **20**, 279–316.
- Gallant, R. and Tauchen, G. (1996), “Which Moments to Match?”, *Econometric Theory*, **12**, 657–681.
- Genton, M. G. and Ronchetti, E. (2003), “Robust Indirect Inference”, *Journal of the American Statistical Association*, **98**, 67–76.
- Gouriéroux, C. and Monfort, A. (1996), *Simulation-Based Econometric Methods*, Oxford University Press.
- Gouriéroux, C., Monfort, A. and Renault, E. (1993), “Indirect Inference”, *Journal of Applied Econometrics*, **8**, S85–S118.
- Hong, Y. and Li, H. (2005), “Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates”, *The Review of Financial Studies*, **18**, 37–84.
- Johannes, M. (2004), “The Statistical and Economic Role of Jumps in Interest Rates”, *Journal of Finance*, **59**, 227–260.
- Kalimipalli, M. and Susmel, R. (2004), “Regime-switching Stochastic Volatility and Short-term Interest Rates”, *Journal of Empirical Finance*, **11**, 309–329.
- Pritsker, M. (1998), “Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models”, *Review of Financial Studies*, **11**, 449–87.
- Ronchetti, E. and Trojani, F. (2001), “Robust Inference with GMM Estimators”, *Journal of Econometrics*, **101**, 37–69.
- Stanton, R. (1997), “A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk”, *Journal of Finance*, **52**, 1973–2002.
- Vasicek, O. (1977), “An Equilibrium Characterization of the term structure.”, *Journal of Financial Economics*, **5**, 177–188.