Comparing Maximum Likelihood and Generalized Method of Moments in Short Term Interest Rate Models

Severin Hanevik

University of Bergen

Department of Mathematics

31 January, 2016

Abstract

In this thesis we will look at some different continuous models for predicting the short term interest rate, and focus on the method of parameter estimation in such models. A particular focus will be placed on the method of Maximum Likelihood Estimation (MLE), which has not been the most common method of estimation in this context. Furthermore, we will compare MLE to the Generalized Method of Moments (GMM), and both methods will be used in simulation experiments and on different sets of real data. Our starting point for investigation will be the model used in Chan, Karolyi, Longstaff, and Sanders [1992], a model in which several of the most popular short rate models can be nested. We will use a discrete-valued approximation of this model to facilitate the estimation of parameters.

In conclusion, this thesis argues that the MLE method of parameter estimation in short term interest rate models deserves more attention than it is currently given. In simulation experiments, the MLE method produced more accurate and precise estimates than the GMM method. Specifically, the bias in estimating the mean-reversion parameter is smaller using the MLE method. These results are particularly interesting as the GMM method is currently the common approach to parameter estimation in short term interest rate models. However, as will be shown in this thesis, it may very well be that the MLE method constitutes a better approach than GMM for such estimations. The results in this thesis may therefore contribute to the theoretical perspective as well as the real-world applications of methods of parameter estimation in predicting the short term interest rate.

Acknowledgements

I would like to express my gratitude to my supervisor Hans A. Karlsen, who has provided excellent advice and guidance during my work with this Master's thesis.

Thanks also to Tore S. Kleppe for introducing me to the - no pun intended - interesting topic of short term interest rate models.

Furthermore, I truly appreciate the patience and support I have received from my family and friends.

Finally, I would like to thank my girlfriend Inger Karin Hjellbakk Arntzen for her love and encouragement, and for proofreading earlier drafts.

Severin Hanevik

Bergen, 31 January 2016

Contents

Chapter 1

Introduction

1.1 The short term riskless interest rate

One of the most important quantities of economics and finance is the short term riskless interest rate. The riskless interest rate is in theory the rate of return on a completely safe investment. Money is assumed to be able to buy less commodity at some point in the future compared to the same amount of money in present. Thus investors would want to be compensated for tying up resources in an investment, even if it is completely risk-free. The riskless interest rate is the least rate of return on such an investment that the investors will accept. In reality, of course, no investment is completely risk-free. In practice, government bonds such as the US treasury bonds, are used as a closest estimate of the riskless interest rate. One expects that the government will default its bonds with such a low probability that it is considered to be almost riskless. Being able to predict the behaviour of the riskless interest rate is naturally a big advantage, for example in the valuing of contingent claims. The Black Scholes formula (Black and Scholes [1973])

$$C(S,t) = N(d_1)S - N(d_2)Ke^{-r(T-t)}$$

$$d_j = \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S}{K}\right) + (r + (-1)^{j-1}\frac{\sigma^2}{2})(T-t)\right], \qquad j = 1, 2.$$

which is used to price European call options, depends on the riskless rate r. In this equation S is the spot price of the underlying asset, K is the strike price, $N(\cdot)$ is the cumulative distribution function of a standard normally distributed variable, T is the time of maturity, σ is the volatility of returns of the underlying asset and $C(S,t)$ is the value of an European call option at time t.

Another example is the price of a zero-coupon bond. A zero-coupon bond is simply a bond which promises the holder of the bond a payment of 1 at time of maturity T. The price of such a bond at time t under the risk-neutral measure Q is given by

$$p(t,T) = \mathbb{E}^{Q}\left[\exp\left(-\int_t^T r(s)ds\right)\middle|\mathcal{F}_t\right]$$

where $r(s)$ is the riskless rate process, and $\mathcal{F}$ is its natural filtration.

As the riskless rate is such an important value in finance, the task of modeling it with a stochastic process has been subject to a lot of attention from economists. A number of different models, both continuous and discrete, have been proposed during the past half-century. In this thesis we will look at some different continuous models for predicting the short term interest rate, and focus on the method of parameter estimation in such models. Particular emphasis will be placed on the method of Maximum Likelihood Estimation (MLE).

1.2 Research question

The main research question of this thesis is to what extent possible advantages outweigh the disadvantages of using maximum likelihood estimation in short term interest rate models. Answering this question requires analysing alternative methods for parameter estimation. This thesis will include a comparative analysis of MLE and the Generalized Method of Moments (GMM). We will use both MLE and GMM in simulation experiments and on real data collected from US treasury bill interest rates. As will be shown, the examination of the methods together with the results of practical testing revealed some interesting results, indicating that MLE might be the superior approach of the two to parameter estimation in short term interest rate models.

1.3 A model for the short term interest rate

GMM is the estimation method used in Chan et al. [1992], which will serve as the main reference for this thesis. We will examine MLE and GMM using both simulated data and real data collected from US Treasury bill interest rates, and subsequently compare the results.

Our starting point for investigation will be the model used in Chan et al. [1992], given as the stochastic differential

$$dr = (\alpha + \beta r)dt + \sigma r^{\gamma} dW \tag{1.1}$$

where W is a Wiener process, and α, β, σ and γ are parameters.

This model is particularly useful as a basic framework for investigation due to the fact that several of the most well-known models of the short term interest rate can be nested within this equation. Thus the different models can conveniently be benchmarked and compared, as has been done in Chan et al. [1992]. The nested models can easily be obtained from (1.1) by introducing proper restrictions on the parameter space.

To estimate the parameters of this continuous model it is convenient to introduce a *discrete approximation* of it. This is done because real life data on interest rates are necessarily provided as a finite number of observations, and not as a continuous stream of data. The observations can be recorded daily, weekly, monthly or in other intervals. When collected and sorted these observations represent a time series which we may examine. It is possible to estimate the parameters of the continuous model directly as well, but we choose to introduce the discrete approximation to avoid the problem of calculating transition probabilities in a continuous setting.

The assumptions of the discrete approximation we use are slightly different depending on whether we want to use maximum likelihood estimation or GMM for parameter estimation. In the case of GMM, there is no need to assume any specific distribution for the error terms. We simply use the approximation from Chan et al. [1992] given by

$$r_{t+1} - r_t = \alpha + \beta r_t + \sigma r_t^{\gamma} \epsilon_{t+1}, \qquad \mathbb{E}[\epsilon_{t+1}] = 0, \qquad \mathbb{E}[\epsilon_{t+1}^2] = \sigma^2 r_t^{2\gamma}. \tag{1.2}$$

In the case of maximum likelihood estimation, however, we will explicitly specify a distribution function for the error terms. This is because we have to calculate the *likelihood function*, which is the essential component of the maximum likelihood procedure. Although a complete specification is used for the MLE, the resulting estimates can be consistent in a more general context. Thus for the maximum likelihood estimation, our approximation will assume that the error terms follow a normal distribution, specifically we have

$$r_{t+1} - r_t = \alpha + \beta r_t + \sigma r_t^{\gamma} \cdot z_{t+1}, \qquad z_{t+1} \sim \mathcal{N}(0, 1). \tag{1.3}$$

The fact that the maximum likelihood procedure demands this additional assumption on the noise component is perhaps its main drawback compared to GMM. In this thesis we will examine whether the possible advantages of using maximum likelihood estimation can outweigh its disadvantages.

Estimating the parameters of a continuous model using a discrete approximation does mean that there will be some approximation error. However, the the amount of approximation error in this type of model can be shown to be small and of second-order importance if changes in the interest rate is measured over short periods of time (Campbell [1984]). Some other issues of discretisation will be discussed in Chapter 3.

1.4 Outline

After this introductory chapter, the following chapter will introduce the Maximum Likelihood and Generalized Method of Moments procedures. Other concepts used in this thesis will also be explained. The Chan, Karolyi, Longstaff and Sanders (CKLS) model for short term interest rates will be introduced and investigated in chapter 3, together with the other models that can be nested within its structure. In addition, the role of the different parameters in the model will be elaborated upon. Chapter 4 will set out the specifics of each of the parameter estimation procedures. Subsequently, chapter 5 will present the practical results of using the discussed methods on real life data from US Treasury bills. This section will also include a comparison of the results from MLE to the results from GMM. In conclusion, the final chapter will summarise the findings of this thesis and point out how these results support the hypothesis that MLE is superior to GMM as a method of parameter estimation in short term interest rate models.

Chapter 2

Theory and concepts

2.1 Maximum Likelihood Estimators

Maximum Likelihood Estimation (MLE) is one of the most well-known procedures for producing an estimate of the parameters used in a given model. To use MLE we look at what is known as the *likelihood function* of the model. The likelihood function is in essence the same function as the simultaneous density function of the model, except the roles of the parameter and observations are switched so that the observations are held fixed, while the parameter is variable. Now the idea is to find the parameter values that *are most likely to produce the given observations*. Thus our objective is to maximise the likelihood function with respect to the parameters. Given a model specified by the probability density function $f(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector of q dimensions, and a random sample $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ generated from this pdf, the joint probability density of the sample is

$$f(x_1, x_2, ..., x_n|\boldsymbol{\theta}) = f(x_1|\boldsymbol{\theta}) \cdot f(x_2|\boldsymbol{\theta}) \cdot \ldots \cdot f(x_n|\boldsymbol{\theta}) = \prod_{t=1}^{n} f(x_t|\boldsymbol{\theta}).$$

The likelihood function is the same function as the pdf, but now $\boldsymbol{\theta}$ is the changing variable, while $x_1, x_2, ..., x_n$ is held fixed.

$$L(\boldsymbol{\theta}|x_1, x_2, ..., x_n) = \prod_{t=1}^{n} f(x_t|\boldsymbol{\theta}).$$

It is practical to look at the logarithm of the likelihood function (log-likelihood function) because the product in the joint density then turns into a sum. This makes it easier to calculate the derivatives.

$$l(\boldsymbol{\theta}|x_1, x_2, ..., x_n) = \log(\prod_{t=1}^{n} L(\boldsymbol{\theta}|x_t) = \sum_{t=1}^{n} \log L(\boldsymbol{\theta}|x_t).$$

Now to obtain the MLE, we need to find the $\boldsymbol{\theta}$ that maximizes the value of the log-likelihood function. We do this by calculating derivatives, demanding they be equal to zero, and then solving the resulting system of equations:

$$\frac{\partial l}{\partial \theta_j} = 0, \qquad j = 1, \ldots, q.$$

The solution yields the Maximum Likelihood Estimate of $\boldsymbol{\theta}$, given the data $x_1, x_2, ..., x_n$.

Maximum Likelihood estimators have several desirable properties when the number of observations approaches infinity, assuming some regularity conditions:
1. They are consistent. That is, the maximum likelihood estimates converge in probability to the true value of the parameter.
2. The distributions of the estimators are normal, with mean equal to the true value of the parameter, and covariance matrix equal to the inverse Fisher information matrix divided by n:

$$\hat{\boldsymbol{\theta}}_{MLE} \sim \mathcal{N}(\theta, n^{-1}\mathcal{I}^{-1}).$$

The MLE also has the *invariance property*. This means that if $\boldsymbol{\theta}_{MLE}$ is the MLE of $\boldsymbol{\theta}$, then for any function $\tau(\boldsymbol{\theta})$, the MLE of $\tau(\boldsymbol{\theta})$ is $\tau(\hat{\boldsymbol{\theta}}_{MLE})$. (Casella and Berger [2001] p. 320)

Disadvantages of the maximum likelihood estimators include:
1. It is necessary to know the exact pdf of the random variables. In some cases this can be difficult to obtain. Also, the derivatives may be hard to calculate analytically.
2. If the likelihood function is flat around the maximum, the variance of the estimate can be large. In that case the Fisher information matrix will have small values.

2.1.1 Likelihood function for time-homogeneous Markov processes

In the previous section we assumed that the observations were independent and identically distributed. This need not be the case. Let us consider the case where the observations are assumed to be dependent, but has the Markov property. Then the simultaneous density of the observations can be written as

$$\begin{aligned} f(x_1, ..., x_n, \boldsymbol{\theta}) &= f(x_1|\boldsymbol{\theta}) f(x_2|x_1, \boldsymbol{\theta}) f(x_3|x_2, x_1, \boldsymbol{\theta}) \cdots f(x_n|x_{n-1}, x_{n-2}, \ldots x_1, \boldsymbol{\theta}) \\ &= f(x_1|\boldsymbol{\theta}) \prod_{t=2}^{n} f(x_t|x_{t-1}, \boldsymbol{\theta}). \end{aligned}$$

Following, the log-likelihood can be expressed as

$$l(\boldsymbol{\theta}|x_1,\ldots,x_n) = \log\left(f(x_1|\boldsymbol{\theta})\prod_{t=2}^{n} f(x_t|x_{t-1},\boldsymbol{\theta})\right) = \log f(x_1|\boldsymbol{\theta}) + \sum_{t=2}^{n}\log f(x_t|x_{t-1},\boldsymbol{\theta}).$$

2.1.2 The score function

The gradient of the log-likelihood function with respect to $\boldsymbol{\theta}$ is called the *score function*. The score function describes how sensitively the likelihood function depends on the parameter $\boldsymbol{\theta}$. Assuming some regularity conditions, it can be shown that the expected value of the score is zero:

$$\mathbb{E}\left[\frac{\partial l}{\partial \theta_j}\right] = 0, \;\; j = 1,\ldots,q.$$

In the case of a stationary time-homogeneous Markov process, let $\pi(x|\boldsymbol{\theta})$ denote its stationary distribution. Then we have

$$\begin{aligned}
\mathbb{E}_\pi\left(\frac{\partial}{\partial\theta}\log f(x_t|x_{t-1},\theta)\right) &= \int\int\left[\frac{\partial}{\partial\theta}\log f(x_t|x_{t-1},\theta)\right]\pi(x_{t-1}|\theta)f(x_t|x_{t-1},\theta)dx_{t-1}dx_t \\
&= \int\int\left[\frac{\partial}{\partial\theta}f(x_t|x_{t-1},\theta)\right]\pi(x_{t-1}|\theta)dx_{t-1}dx_t \\
&= \int\left[\int\frac{\partial}{\partial\theta}f(x_t|x_{t-1},\theta)dx_t\right]\pi(x_{t-1}|\theta)dx_{t-1} \\
&= \int\left(\frac{\partial}{\partial\theta}\left[\int f(x_t|x_{t-1},\theta)dx_t\right]\right)\pi(x_{t-1}|\theta)dx_{t-1} \\
&= \int\left(\frac{\partial}{\partial\theta}1\right)\pi(x_{t-1}|\theta)dx_{t-1} \\
&= 0.
\end{aligned}$$

If we assume that the process starts at some value x_0, and x_1 depends on x_0, the expectation of the score function of the full likelihood is just a sum of terms that equals zero, and thus also zero itself.

2.1.3 The theoretical Fisher information matrix

The Fisher Information $\mathcal{I}$ is a number or a matrix that measures how much information about the parameter $\boldsymbol{\theta}$ is contained in the sample $\boldsymbol{x}$. It is defined as the variance of the score, which here is equal to the second moment of the score due to the fact that the expected value of the score is zero. For a stationary time-homogeneous Markov process we define:

$$\mathcal{I}(\boldsymbol{\theta}) = \mathbb{E}_\pi\left[\nabla\log f(x_1|x_0,\boldsymbol{\theta})\nabla^T\log f(x_1|x_0,\boldsymbol{\theta})\right], \tag{2.1}$$

$$\mathcal{I}_n(\boldsymbol{\theta}) = \mathbb{E}_\pi\bigg[\sum_{t=1}^{n} \nabla \log f(x_t|x_{t-1}, \boldsymbol{\theta}) \nabla^T \log f(x_t|x_{t-1}, \boldsymbol{\theta})\bigg].$$

Under some regularity conditions we also have:

$$\mathcal{I}(\boldsymbol{\theta}) = \mathbb{E}_\pi\bigg[- \nabla\nabla^T \log f(x_1|x_0, \boldsymbol{\theta})\bigg], \tag{2.2}$$

$$\mathcal{I}_n(\boldsymbol{\theta}) = \mathbb{E}_\pi\bigg[\sum_{t=1}^{n} -\nabla\nabla^T \log f(x_t|x_{t-1}, \boldsymbol{\theta})\bigg].$$

The Fisher information matrix is useful for calculating the variance of maximum likelihood estimates. The asymptotic distribution of the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{MLE}$ is normally distributed with variance given by the inverse Fisher information matrix:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}_0) \sim \mathcal{N}(0, \mathcal{I}^{-1}),$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$.

2.1.4 The observed Fisher information matrix

The observed Fisher information matrix $\mathcal{J}$ is a sample equivalent to the theoretical Fisher information matrix $\mathcal{I}$. Denote the log-likelihood function by $l_n(\hat{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \log f(x_t|x_{t-1}, \boldsymbol{\theta})$. The observed Fisher information is defined for a stationary time-homogeneous Markov process at $\hat{\boldsymbol{\theta}}$ as

$$\mathcal{J}_n(\hat{\boldsymbol{\theta}}) = -\nabla\nabla^T l_n(\hat{\boldsymbol{\theta}})$$

$$= -\bigg\{\frac{\partial^2 l_n}{\partial\theta_i \partial\theta_j}, \ 1 \leq i, j \leq q\bigg\}.$$

evaluated at $\hat{\boldsymbol{\theta}}$. To avoid calculating second derivatives we use the expression

$$\mathcal{J}_n^{'}(\hat{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \nabla \log f(x_t|x_{t-1}, \hat{\boldsymbol{\theta}}) \nabla^T \log f(x_t|x_{t-1}, \hat{\boldsymbol{\theta}}).$$

From the definition, it is evident that

$$\mathbb{E}_\pi(\mathcal{J}_n(\hat{\boldsymbol{\theta}})) = \mathbb{E}_\pi(\mathcal{J}_n^{'}(\hat{\boldsymbol{\theta}})) = \mathcal{I}_n(\hat{\boldsymbol{\theta}})$$

As n approaches infinity, we assume that $n^{-1}\mathcal{J}_n(\hat{\boldsymbol{\theta}})$ and $n^{-1}\mathcal{J}_n^{'}(\hat{\boldsymbol{\theta}})$ will to converge to the same value, namely $\mathcal{I}(\hat{\boldsymbol{\theta}})$. We can use this to create an estimate of the theoretical Fisher information matrix form the observed Fisher information, and thus obtain information regarding the variances and covariances of the parameter estimates.

2.2 Generalized Method of Moments

2.2.1 Pearson's Method of Moments

The Generalized Method of Moments (GMM) is a method of estimation that extends Pearson's Method of Moments. Given a sample $X_1, X_2, ..., X_n$ from a population with pdf $f(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is q-dimensional, the original Method of Moments equates the q first moments with its sample counterpart.

$$\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} x_k, \quad k = 1, \ldots, q.$$

As the number of equations is equal to the number of parameters, this system can be solved exactly for $\boldsymbol{\theta}$, and the solution is the Method of Moments estimator of $\boldsymbol{\theta}$.

$$\hat{m}_j = m_j(\hat{\boldsymbol{\theta}}), \; j = 1, \ldots, q.$$

2.2.2 The Generalized Method of Moments

Hansen [1982] extends the method of Pearson by considering any population condition instead of only the moments. For GMM we require a vector $\boldsymbol{g}(x_t, \boldsymbol{\theta})$ of s dimensions such that

$$\boldsymbol{m}(\boldsymbol{\theta}_0) = \mathbb{E}[\boldsymbol{g}(x_t, \boldsymbol{\theta}_0)] = 0.$$

holds for the true parameter $\boldsymbol{\theta}_0$. We then, in a similar fashion to the original Method of Moments, replace the expectation with the sample average of $\boldsymbol{g}$:

$$\hat{\boldsymbol{m}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}(x_t, \boldsymbol{\theta}).$$

In the case of GMM, we allow s to be larger than q. This means that the resulting system of equations will not have a single solution. Instead, we try to minimize a *norm*:

$$Q_n(\boldsymbol{\theta}) = \hat{\boldsymbol{m}}_n(\boldsymbol{\theta})^T \mathbf{W} \hat{\boldsymbol{m}}_n(\boldsymbol{\theta}),$$

where $\mathbf{W}$ is a weighting matrix. The GMM estimator is actually an entire class of estimators that depends on this $\mathbf{W}$. It is formally defined as

$$\hat{\boldsymbol{\theta}}_{GMM}(\mathbf{W}) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, Q_n(\boldsymbol{\theta}).$$

Chapter 3

The interest rate model

3.1 The model

The model we use as a framework for investigation in this thesis is the Chan, Karolyi, Longstaff and Sanders (CKLS) model, presented in Chan et al. [1992]. It can be represented by the stochastic differential equation

$$dr = (\alpha + \beta r)dt + \sigma r^{\gamma} dW,$$

where W is a Wiener process, and α, β, σ and γ are parameters. This is an example of a diffusion process (Bjork [2009] p. 40). This model is of particular interest because many of the well known continuous short rate models can be nested within this equation, creating a framework for comparison between them. Models that fit inside this framework include: the model of Merton (Merton [1973]), the Vasicek model (Vasicek [1977]), the Cox, Ingersoll and Ross square root model (CIR SR) (Cox, Ingersoll, and Ross [1985]), Dothan model (Dothan [1978]), geometric Brownian motion (GBM) (Black and Scholes [1973]), the Brennan-Schwartz model (Brennan and Schwartz [1980]), CIR variable rate model (CIR VR) (Cox, Ingersoll, and Ross [1980] and the constant elasticity of variance model (CEV) (Cox [1975] and Cox and Ross [1976]).

3.2 The parameters

The parameter α in this model can be interpreted as a scale parameter for the long term average level of the interest rate process. In the long run the process will, for parameter values ensuring stationarity, repeatedly return towards and fluctuate around a level given by $-\alpha/\beta$. This effect is called *mean reversion* and is featured in many of the most common short rate models. Thus, the α parameter can be increased or decreased to manipulate the typical size and average level of the interest rate in the long run. When β is equal to 0, α is simply the drift parameter of the process.

The parameter β decides the speed of which the process will return to the mean level $-\alpha/\beta$. A larger absolute value of β will force the process to return to its average more quickly after hitting a more extreme value. Conversely, a smaller value of β will enable the process to stay for a longer time at a level that is away from the process mean. β needs to be negative to ensure stability of the process.

The parameter σ increases or decreases the size of the random component of the model, thus in effect deciding the variance between two adjacent steps of the interest rate r_t.

The parameter γ is a measure of the sensitivity of the volatility to the level of r, a greater γ means that the size of the random component will be more sensitive to the level of r. However, for typical values of r, that is $0 < r < 1$, a greater value of γ will also *reduce* the overall size of the random component. This is because increasing the value of the exponent of a number between zero and one will result in a smaller number. Due to this effect, a large value of γ will often be accompanied by a large value of σ, and smaller values of γ with smaller values of σ. Consequently, if you want to increase the sensitivity of the volatility by increasing the value of γ - without also decreasing the overall size of the variance - the σ parameter must also be increased to compensate. This relationship between σ and γ can be useful to keep in mind while examining results of parameter estimation. If γ is equal to zero, then the size of the random component will be completely independent of the level of r. A γ equal to one will result in the random component being proportional to the absolute value of r. The Cox, Ingersoll and Ross (CIR) model uses $\gamma = 0.5$.

The following plots show how one can expect the process to behave when the parameters are changed. The parameters are assumed to be $\alpha = 0.004$, $\beta = -0.05$, $\sigma = 0.05$ and $\gamma = 0.5$ where they are not given specifically. The generated standard normal error terms are the same for all the plots.

Figure 3.1: *Simulations with the CKLS process for selected values of α*

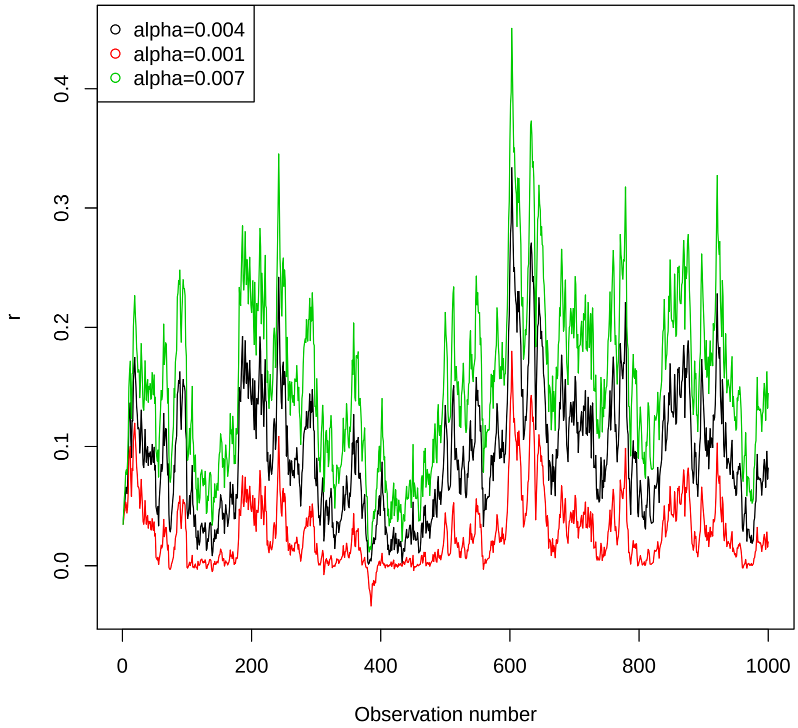


Figure 3.2: *Simulations with the CKLS process for selected values of β and α*

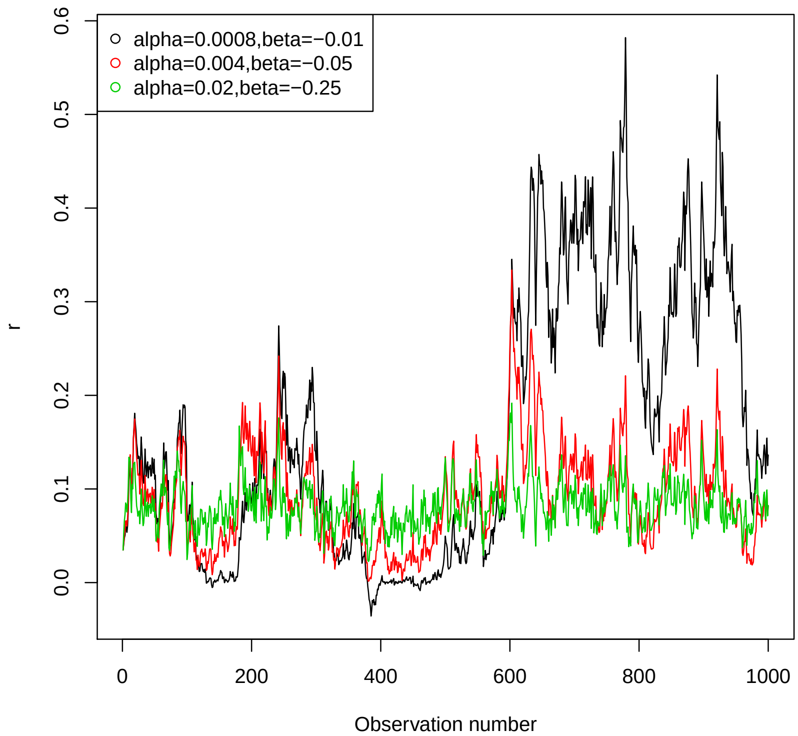


Figure 3.3: *Simulations with the CKLS process for selected values of* σ

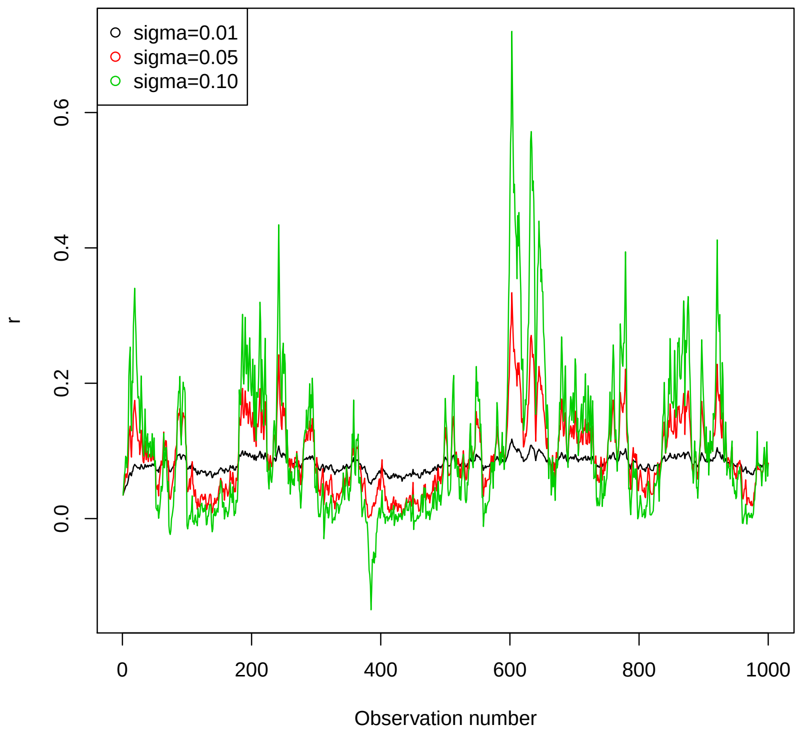
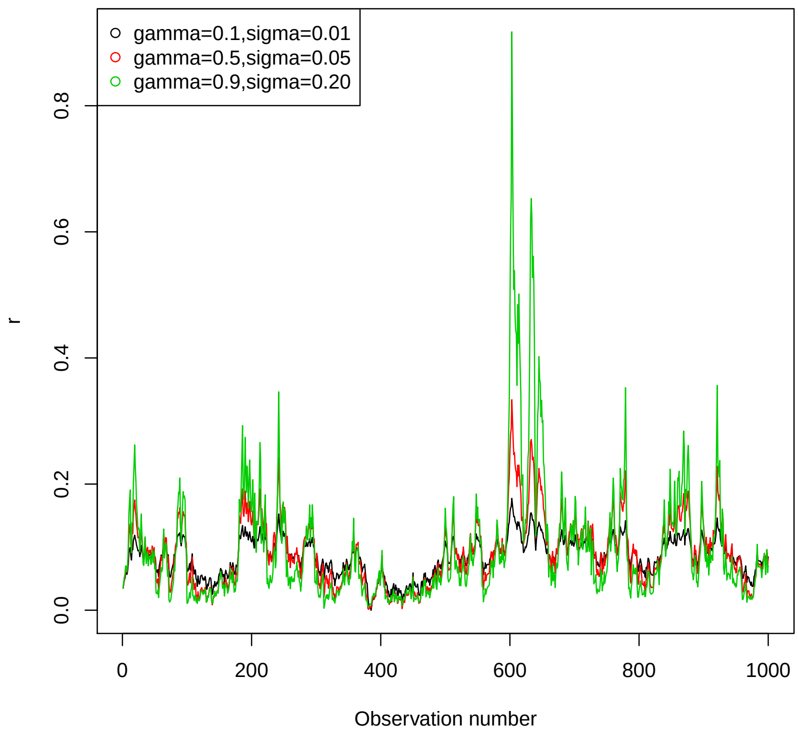


Figure 3.4: *Simulations with the CKLS process for selected values of* γ



3.3 Nested models

The equations defining the various nested models are contained within the equation for the full CKLS model by restricting parameters to certain values, specific to each nested model. The following table describe how the various nested models relate to the full CKLS model by listing their restrictions:

Table 3.1: *Parameter restrictions on various models that can be nested within the CKLS model.*

Model	α	β	σ^2	γ
Merton		0		0
Vasicek				0
CIR SR				0.5
Dothan	0	0		1
GBM	0			1
Brennan-Schwartz				1
CIR VR	0	0		1.5
CEV	0			

Here, a '0' means that the parameter is not used in the specified model. A number different from '0' indicates a design parameter of the model.

The following sections will discuss some of the models in a little more detail.

3.3.1 The Vasicek and Merton models

The Vasicek and Merton models both introduce the restriction on the full model that $\gamma = 0$. This means that the r in the noise component disappears, and thus the conditional variance of the changes in the interest rate is constant and independent of the current level of the interest rate. The consequence is that even if the interest rate grows to bigger values, it will still change with the same distribution as it did at a lower value. This can be seen as a major shortcoming of these models, although it does simplify the models alot. The Merton model additionally requires β to be zero, which means that it can be characterised simply as a Brownian motion with drift. The discrete approximation is a random walk with drift α. Because it lacks the β parameter, it does not possess the mean-reverting property of the Vasicek model. Whether mean-reversion is an important feature of a short rate model has been debated. For example Chan et al. [1992] report that the results from their parameter estimation on the full model indicates that mean reversion might not be crucial to successful modeling. In any case, mean-reversion of the short term interest rate makes sense intuitively, and many models incorporate its use.

3.3.2 The Cox, Ingersoll and Ross square root (CIR SR) process

The only restriction of the Cox, Ingersoll and Ross (CIR) process on the full model is that γ be equal to $\frac{1}{2}$. This means that the size of the random component of the predicted interest rate level will be proportional to the current interest rate level, because we have $\mathbb{E}(\epsilon_{t+1}^2) = \sigma^2 r_t$.

3.4 The discrete approximation

To be able to make inferences about this model, we can approximate the continuous model by a discrete version. The approximation we use is given by the time series

$$r_{t+1} - r_t = \alpha + \beta r_t + \sigma |r_t|^\gamma \cdot \epsilon_{t+1},$$

where ϵ_t is a sequence of independent, normally distributed random variables, each with mean equal to zero and variance equal to one. To highlight the prediction of the next value of r, the model can also be written as

$$r_{t+1} = \alpha + (1 + \beta) r_t + \sigma |r_t|^\gamma \cdot \epsilon_{t+1}.$$

Approximating the continuous model with a discrete version does introduce some issues. In models where the interest rate is usually not allowed to be negative in the continuous case (for example in the CIR SR process with $\alpha > 0$ and $\beta < 0$), the discrete approximation can violate this imposed restriction. This is due to the fact that in the discrete model the interest rate may 'jump' from a positive value to a negative one, whereas in the continuous model the constantly increased pressure to return to a higher level would stop the process from turning negative.

3.4.1 Relation with well-known time series

It can be interesting to see how the discrete approximation of the CKLS process relates to more well-known time series. We introduce the autoregressive (AR) and autoregressive conditional heteroskedasticity (ARCH) processes: *The autoregressive process of order q* (AR(q)) can be written

$$X_t = a + \sum_{i=1}^{q} b_i X_{t-i} + Z_t,$$

where $a, b_1, \ldots, b_n$ are parameters and Z_t is a white noise process.

The terms of the *autoregressive conditional heteroskedasticity process of order q* (ARCH(q)) is given by

$$X_t = c_t Z_t,$$

with

$$c_t^2 = d_0 + \sum_{i=1}^{q} d_i Z_{t-1}^2,$$

with $d_0 > 0$ and $d_j \geq 0$, $j = 1, \ldots, q$.

With these definitions in place, we note that the CKLS model with γ fixed at 1 (the Brennan-Schwartz model) can be described as a combination of AR and ARCH models. More specifically, it is quite similar to an AR(1) time series with noise components similar to an ARCH(1) process. In such a model, a would correspond to α in the CKLS model, b_1 to $1 + \beta$ and d_1 to σ^2, $d_0 = 0$. Note, however, that ARCH models generally require $d_0 > 0$.

3.4.2 Stationarity

The following table shows which parameters values the discrete approximation of the CKLS model is stationary for:

Table 3.2: *Stationarity of the discrete approximation of the CKLS model*

	$\beta < 0$	$\beta = 0,\ \alpha = 0$	$\beta = 0,\ \alpha > 0$	$\beta > 0,\ \alpha > 0$
$\gamma = 0$	Positive recurrent	Null recurrent	Transient	Transient
$0 < \gamma \leq 1$	Positive recurrent	Transient	Transient	Transient
$\gamma > 1$	Transient	Transient	Transient	Transient

Generally, the process will be transient if the β parameter is equal to or larger than zero, except when the α and γ parameters are also zero. In that case the model is just a random walk process, which is null recurrent. Values of γ larger than 1 will make the noise component of the process grow out of control when r takes a value larger than 1. Thus, the process will be transient.

Chapter 4

Estimating the parameters

4.1 Maximum Likelihood Estimation

4.1.1 The likelihood function

In this section we will derive the log-likelihood function for the discrete-valued approximation to the CKLS model. The model is specified for $r_1, r_2, \ldots, r_n$ by

$$r_t = r_{t-1}(1+\beta) + \alpha + \epsilon_t,$$

with

$$\mathbb{E}[\epsilon_t] = 0,\ \mathbb{E}[\epsilon_t^2] = \sigma^2 r_t^{2\gamma}, \qquad \epsilon_t \sim \mathcal{N}(0, (\sigma r_{t-1}^{\gamma})^2).$$

This is the same as

$$r_t = r_{t-1}(1+\beta) + \alpha + \sigma|r_{t-1}|z_t, \qquad z_t \sim \mathcal{N}(0,1)$$

We see that the model has the Markov property and find that the simultaneous density is given by

$$f(r_1, ..., r_n, \boldsymbol{\theta}) = f_{r_t|r_{t-1}}(r_1|r_0, \boldsymbol{\theta}) \prod_{i=2}^{n} f_{r_t|r_{t-1}}(r_t|r_{t-1}, \boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = [\alpha, \beta, \sigma, \gamma]^T$ and r_0 is an initial value for the process.

$$r_t|r_{t-1} \sim \mathcal{N}(r_{t-1}(1+\beta) + \alpha, \sigma^2 |r_{t-1}|^{2\gamma})$$

$$\implies f_{r_t|r_{t-1}}(r_t|r_{t-1}) = \frac{1}{\sigma|r_{t-1}|^{\gamma}\sqrt{2\pi}} \exp\left(- \frac{r_t - (r_{t-1}(1+\beta)+\alpha))^2}{2\sigma^2|r_{t-1}|^{2\gamma}} \right).$$

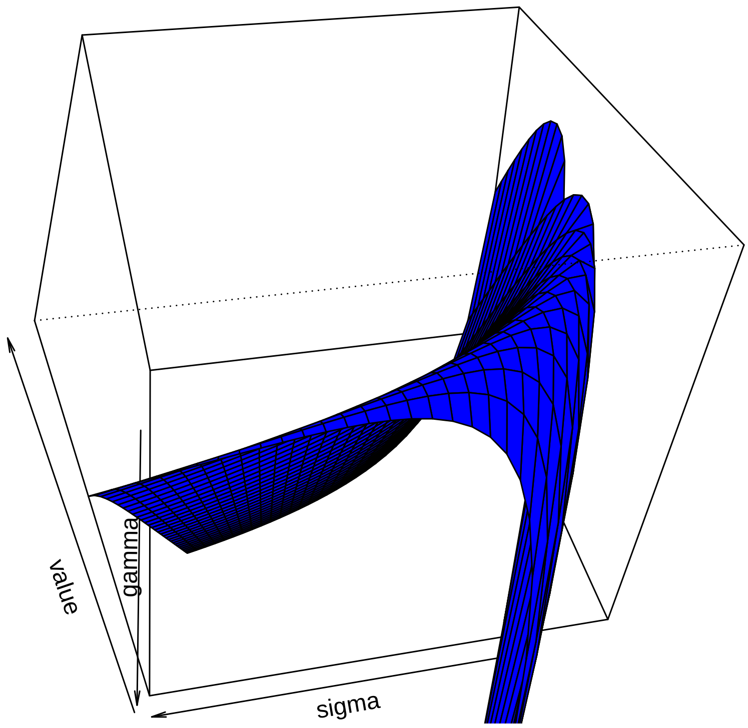
Thus, the likelihood is given as

$$L(\boldsymbol{\theta}|r_0, r_1, ..., r_n) = \prod_{t=1}^{n} \frac{1}{\sigma |r_{t-1}|^{\gamma} \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2 |r_{t-1}|^{2\gamma}} (r_t - (r_{t-1}(1+\beta) + \alpha))^2 \right).$$

When we finally take the logarithm of this expression we obtain the log-likelihood function:

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}; r_0, r_1, ..., r_n) = \sum_{t=1}^{n} \left(\log\left(\frac{1}{\sigma |r_{t-1}|^{\gamma} \sqrt{2\pi}} \right) - \frac{1}{2\sigma^2 |r_{t-1}|^{2\gamma}} (r_t - (r_{t-1}(1+\beta) + \alpha))^2 \right).$$

The figure below shows a plot of the log-likelihood function using real data, with α and β held fixed.



4.1.2 The score function

The score function V is the derivative of the log-likelihood function, that is

$$V = \nabla l(\boldsymbol{\theta}).$$

In our case, the score function looks like this:

$$
\begin{aligned}
\frac{\partial l}{\partial \alpha} &= \sum_{t=1}^{n} -\frac{1}{\sigma^2 |r_{t-1}|^2\gamma}(\alpha + r_t - r_{t-1} + r_{t-1}\beta) \\
\frac{\partial l}{\partial \beta} &= \sum_{t=1}^{n} -\frac{1}{\sigma^2 |r_{t-1}|^2\gamma}(-r_t r_{t-1} + \alpha r_{t-1} + r_{t-1}^2 + \beta r_{t-1}^2) \\
\frac{\partial l}{\partial \sigma} &= \sum_{t=1}^{n} -\frac{1}{\sigma} + \frac{1}{\sigma^3 |r_{t-1}|^2\gamma}(r_t - (r_{t-1}(1+\beta) + \alpha))^2 \\
\frac{\partial l}{\partial \gamma} &= \sum_{t=1}^{n} -\log|r_{t-1}| + \frac{\log|r_{t-1}||r_{t-1}|^{-2\gamma}}{\sigma^2}(r_t - (r_{t-1}(1+\beta) + \alpha))^2
\end{aligned}
\tag{4.1}
$$

The score function can next be used to calculate the observed Fisher information matrix.

4.1.3 Obtaining the covariance matrix of the MLE estimate

Let $\mathcal{I}$ denote the Fisher information matrix for our model. We assume that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}_0) \sim \mathcal{N}(0, \mathcal{I}^{-1}), \tag{4.2}$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. In our case, we can not directly compute the Fisher information matrix by formulas (2.1) or (2.2) because we do not know the stationary distribution of the process. Instead, we calculate an estimate of $\mathcal{I}$ using the observed Fisher information as in section (2.1.4). Specifically, we estimate $\mathcal{I}$ by the matrix:

$$\hat{\mathcal{I}} = \frac{1}{n}\mathcal{J}_n(\hat{\boldsymbol{\theta}}_{MLE}) = \frac{1}{n}\left(\sum_{t=1}^{n} \nabla_\theta \log f(x_t|x_{t-1}, \hat{\boldsymbol{\theta}})\nabla_\theta^T \log f(x_t|x_{t-1}, \hat{\boldsymbol{\theta}})\right).$$

Subsequently it follows from (4.2) that the estimate $\hat{\boldsymbol{\theta}}_{MLE}$ calculated from n observations has the approximate distribution

$$\mathcal{N}(\theta_0, n^{-1}\hat{\mathcal{I}}^{-1}) = \mathcal{N}(\theta_0, n^{-1}(n^{-1}\mathcal{J}_n)^{-1}) = \mathcal{N}(\theta_0, \mathcal{J}_n^{-1}).$$

4.2 Generalized Method of Moments

4.2.1 Model and assumptions

As a starting point, we will use the same discrete approximation of the continuous CKLS model as used under the discussion of the method of maximum likelihood. However, for the generalized method of moments, we do not need to assume that the noise components are distributed normally. Given observations $r_1, r_2, \ldots, r_n$, we thus consider the model

$$r_{t+1} = r_t(1+\beta) + \alpha + \epsilon_{t+1}$$

where the assumptions on the noise are

$$\mathbb{E}[\epsilon_{t+1}] = 0 \tag{4.3}$$

$$\mathbb{E}[\epsilon_{t+1}^2] = \sigma^2 r_t^{2\gamma} \tag{4.4}$$

The fact that we do not need to make any further assumptions about the distribution of the random component of the model makes GMM a attractive choice for estimating parameters. Not being restricted to any specific distribution, GMM can accurately provide estimates even if we do not have detailed information about the random nature of the process we are trying to model.

4.2.2 The moment conditions

The moment conditions we use are the same as used in Chan et al. [1992]. In addition to equations (4.3) and (4.4) we also know that

$$\mathbb{E}[\epsilon_{t+1} r_t] = 0$$

$$\mathbb{E}[(\epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}) r_t] = 0$$

Because $\epsilon_{t+1} = r_{t+1} - r_t - \alpha - \beta r_t$, we can use these four equations as our moment conditions and we write

$$g_t(\boldsymbol{\theta}) = \begin{pmatrix} \epsilon_{t+1} \\ \epsilon_{t+1} r_t \\ \epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma} \\ (\epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}) r_t \end{pmatrix}$$

and

$$\boldsymbol{m}_t(\boldsymbol{\theta}) = \mathbb{E}[g_t(\boldsymbol{\theta})] = 0$$

Next we replace the expectation in the moment conditions with the empirical sample equivalent and use the law of large numbers to obtain the equation

$$\hat{\boldsymbol{m}}_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^{n} g_t(\hat{\boldsymbol{\theta}}) = 0 \tag{4.5}$$

This estimation of the moment conditions by sample equivalents is the basis of the both the standard method of moments and the generalized method of moments.

Now, for the full unrestricted model, we have four parameters and four moment conditions. The number of unknowns and the number of equations match, and this means that the system can be solved exactly. From this we obtain a single solution for the parameters α, β, σ^2 and γ, and this will be our GMM estimate. In this case, apart from using different moment conditions, the GMM procedure is not that different from the standard method of moments.

4.2.3 Minimisation of the norm

When the number of moment conditions are *larger* than the number of parameters (which is the case for all of the nested models), we will not have a unique solution to the system of equations in (4.5). When this is the case, we instead try to come as close as possible to a solution by minimising a norm. Intuitively, we are trying to make the left side of (4.5) as close as possible to zero. It does not necessarily make sense to weigh the closeness of each equation the same however, because the values of some of the moment conditions may vary more than some of the others. We can thus introduce a class of estimators which are defined as the value of $\boldsymbol{\theta}$ that minimises the quadratic form

$$Q_n(\boldsymbol{\theta}) = \hat{\boldsymbol{m}}_n(\boldsymbol{\theta})^T \mathbf{W} \hat{\boldsymbol{m}}_n(\boldsymbol{\theta})$$

where $\hat{\boldsymbol{m}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}_t(\boldsymbol{\theta})$. The next question is: which $\mathbf{W}$ to choose? Any positive-definite symmetric matrix will be a valid choice, but Hansen [1982] shows that using $\mathbf{W} = \left(\mathbb{E}[\boldsymbol{g}_t(\boldsymbol{\theta})^T \boldsymbol{g}_t(\boldsymbol{\theta})]\right)^{-1}$ will yield the GMM estimator with the smallest possible covariance matrix. This will be the choice of $\mathbf{W}$ that we will use in the calculation of estimates for parameters in the nested models.

4.2.4 Calculating the weighting matrix

We can not directly compute $\left(\mathbb{E}[\boldsymbol{g}(\boldsymbol{\theta})^T \boldsymbol{g}(\boldsymbol{\theta})]\right)^{-1}$, because this expression includes $\boldsymbol{\theta}$, which is the unknown parameter. There are several ways of getting around this problem, and the simplest is the 'two-step' procedure proposed by Hansen [1982]. This method suggests first setting $\mathbf{W}$ equal to the identity matrix I, and then use this $\mathbf{W}$ to produce a preliminary estimate $\hat{\boldsymbol{\theta}}_{(1)}$ of $\boldsymbol{\theta}$. Then we may calculate $\hat{\boldsymbol{W}}$ by

$$\hat{\boldsymbol{W}} = \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}_t(\hat{\boldsymbol{\theta}}_{(1)})^T \boldsymbol{g}_t(\hat{\boldsymbol{\theta}}_{(1)}))^{-1}\right).$$

This $\hat{\mathbf{W}}$ can then be used to calculate the GMM estimate of $\boldsymbol{\theta}$.

The estimation of $\hat{W}$ can be repeated, this is called 'Iterated GMM'. Doing this will produce estimates $\hat{\boldsymbol{\theta}}_{(1)}, \hat{\boldsymbol{\theta}}_{(2)}, \ldots, \hat{\boldsymbol{\theta}}_{(i)}$. This procedure is equivalent to solving the system of equations (Imbens, Johnson, and Spady [1995]):

$$\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g_t}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_{(i)})\right)^T \left(\frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\boldsymbol{\theta}}_{(i)})^T g_t(\hat{\boldsymbol{\theta}}_{(i)}))^{-1}\right)\left(\frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\boldsymbol{\theta}}_{(i)})\right) = 0.$$

The 'two-step' procedure of Hansen [1982] is the most common method, so that is the method we will use to calculate the GMM estimates presented in this thesis.

4.2.5 Obtaining the covariance matrix of the GMM estimate

After having calculated an estimate for the weighting matrix $\mathbf{W}$, we can find the covariance matrix for the GMM parameter estimate. If $\mathbf{D}(\boldsymbol{\theta})$ is the Jacobian matrix of $\hat{\boldsymbol{m}}_n(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, we have

$$\hat{\mathcal{I}}_{GMM} = \frac{1}{n}\left(\mathbf{D}^T(\hat{\theta})\hat{\mathbf{W}}\mathbf{D}(\hat{\theta})\right).$$

We assume that this will converge to the theoretical information $\mathcal{I}_{GMM}$ as n approaches infinity. Asymptotically, assuming stationarity and that common regularity conditions are satisfied, we have for a GMM estimate $\hat{\boldsymbol{\theta}}_{GMM}$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0) \sim \mathcal{N}(0, \mathcal{I}_{GMM}^{-1}),$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. Approximating $\mathcal{I}_{GMM}$ by $\hat{\mathcal{I}}_{GMM}$, we have that an approximate distribution for the GMM estimate $\hat{\boldsymbol{\theta}}_{GMM}$ estimated from n observations is $\mathcal{N}(\boldsymbol{\theta}_0, \hat{\mathcal{I}}_{GMM}^{-1})$.

Chapter 5

Results and discussion

5.1 Simulations

In this section we will use R to perform parameter estimation on generated data. The goal is to compare the properties and efficiency of the maximum likelihood and GMM methods in a setting where we already know what the optimal results will be. This makes it easy to highlight how the methods works and how they compare in practical use. In particular, we will look at the empirical variance of the parameter estimates and potential bias, as these features are the most important to incite confidence in the results.

5.1.1 Simulation details

We will perform two main simulations. In the first we generate a series of 10000 observations for each of the models considered. The same series of randomly generated standard normal distributed noise will be used for all models, but the generated dataset will be tailored to each models specifications. For the full CKLS model, data will be generated with parameters set at $\alpha = 0.0004$, $\beta = -0.005$, $\sigma^2 = 0.01$ and $\gamma = 0.5$. For the submodels, the dataset will also be generated from these parameters, but with their own restrictions imposed. For example, the data for the Geometrical Brownian Motion will be generated with parameters: $\alpha = 0$, $\beta = -0.005$, $\sigma^2 = 0.01$ and $\gamma = 1$. For each simulation experiment we estimate the parameters both with MLE and GMM. In total we have used 50 repetitions and then calculated the empirical mean and empirical variance of the parameters. This is done for each of the methods. We will also generate covariance matrices derived from the Fisher information matrices for the MLE estimates. These covariance matrices along with the empirical covariance matrices can be found in the appendix. As the primary goal of this experiment is to test variance and bias of parameter estimates, the initial values for the optimising functions in

R are just set to the true values. It could be interesting to perform experiments testing the robustness of the methods and consequences of using different initial values. While the time constraints in this thesis did not allow for such analysis, it is indeed a topic that deserves further study.

As a second simulation we will repeat the first one, but this time with only 500 observations for each of the 50 repetitions. The intention is that the first and longer simulation with 10000 observations will highlight the performance of the estimation methods given sufficient data, while the shorter one of 500 observations is closer to the typical number of data points given for example for monthly observations of real data.

It should be noted that the Merton, Dothan and CIR VR models are not stationary processes, and lacks a stationary distribution. This means that the theory reviewed earlier in this thesis which depends on the process having a stationary distribution does not necessarily hold in these cases, and we have not yet established any theoretical reason for the methods to work properly. The other models except Merton, Dothan and CIR VR may be stationary or non-stationary depending on the parameters, but in this case the parameters chosen for data generation guarantees that they will be stationary. It follows that the previously reviewed theory will apply in these cases.

5.1.2 The results

The Merton model $[\beta = 0, \gamma = 0]$

Table 5.1: *Merton model estimates from large sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.000478	$9.64e-07$	$1.02e-06$
MLE σ^2	0.01	0.00998	$3.05e-08$	$2.02e-08$
GMM α	0.0004	0.000472	$9.80e-07$	
GMM σ^2	0.01	0.00997	$2.99e-08$	

Results of simulations with the Merton model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 10000 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a Merton model with $\alpha = 0.0004$ *and* $\sigma^2 = 0.01$.

For the Merton model, there seems to be very little difference in the results of the maximum likelihood and GMM methods. We notice that the estimates for σ^2 has significantly less variance than the ones for α, indicating that it is easier to obtain confidence in the estimates for σ^2 than for α in the Merton model. The α estimate means of 0.000478 and 0.000472 may seem to indicate bias, but the error is likely caused by the fact that the standard deviation of the estimates is about twice as big as the actual value of α, and thus the discrepancy is not statistically significant. Note that the Merton model is not stationary, and because of this, that previously reviewed theory does not necessarily apply here.

The Vasicek model $[\gamma = 0]$

Table 5.2: *Vasicek model estimates from large sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.00054	$1.16e-06$	$1.02e-06$
MLE β	-0.005	-0.00548	$1.04e-06$	$9.96e-07$
MLE σ^2	0.01	0.00997	$3.04e-08$	$2.02e-08$
GMM α	0.0004	0.000535	$1.16e-06$	
GMM β	-0.005	-0.00548	$1.05e-06$	
GMM σ^2	0.01	0.00997	$3.04e-08$	

Results of simulations with the Vasicek model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 10000 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a Vasicek model with $\alpha = 0.0004$, $\beta = -0.005$ *and* $\sigma^2 = 0.01$.

Maximum likelihood and GMM yield almost identical results also in the context of the Vasicek model. The deviation in the β parameter is big enough though to indicate a possible bias towards bigger absolute values. Estimates for σ^2 are very good. The variance gathered from the Fisher information matrix is close to the empirical variance, which is in accord with theory.

The CIR SR model $[\gamma = 0.5]$

Table 5.3: *CIR SR model estimates from large sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.000409	$2.05e-09$	$1.66e-09$
MLE β	-0.005	-0.00591	$6.65e-06$	$5.12e-06$
MLE σ^2	0.01	0.00997	$3.03e-08$	$2.02e-08$
GMM α	0.0004	0.000665	$1.22e-07$	
GMM β	-0.005	-0.00804	$1.78e-05$	
GMM σ^2	0.01	0.0102	$3.01e-05$	

Results of simulations with the CIR SR model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 10000 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a CIR SR model with $\alpha = 0.0004$ $\beta = -0.005$ *and* $\sigma^2 = 0.01$.

The CIR SR model seems to be the first where the maximum likelihood and GMM estimates differ significantly. We can clearly see that the maximum likelihood estimates has a much lower variance than the GMM estimates. There is an obvious bias towards larger absolute values in the GMM estimates of the β parameter. Though this bias is also evident in the maximum likelihood β estimate, it is much smaller. There is also bias in the GMM α estimate.

The Dothan model $[\alpha, \beta = 0, \gamma = 1]$

Table 5.4: *Dothan model estimates from large sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE σ^2	0.01	0.00998	$3.03e-08$	$2.02e-08$
GMM σ^2	0.01	0.00976	$7.96e-07$	

Results of simulations with the Dothan model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 10000 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a Dothan model with $\sigma^2 = 0.01$.

Both maximum likelihood and GMM yield good consistent estimates in the Dothan model. Maximum likelihood estimates has somewhat smaller variance, though, and is thus more efficient. Note that the Merton model is not stationary, and because of this we have no guarantee that the previously reviewed theory applies here.

The GBM model $[\alpha = 0, \gamma = 1]$

Table 5.5: *GBM model estimates from large sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE β	-0.005	-0.00490	$9.16e-07$	$1.02e-06$
MLE σ^2	0.010	0.00997	$3.05e-08$	$2.02e-08$
GMM β	-0.005	-0.00732	$4.03e-05$	
GMM σ^2	0.010	0.0101	$4.24e-06$	

Results of simulations with the GBM model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 10000 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a GBM model with $\beta = -0.005$ and $\sigma^2 = 0.01$.

In the GBM model, the maximum likelihood method results in good consistent estimates. GMM estimates has significantly bigger variances and also exhibits a clear bias towards larger absolute values in the β parameter. Overall, maximum likelihood gives clearly superior results in our simulation.

The Brennan-Schwartz model $[\gamma = 1]$

Table 5.6: *Brennan-Schwartz model estimates from large sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.000418	$3.14e-09$	$2.87e-09$
MLE β	-0.005	-0.00536	$2.78e-06$	$2.91e-06$
MLE σ^2	0.01	0.00997	$3.03e-08$	$2.02e-08$
GMM α	0.0004	0.000693	$2.12e-08$	
GMM β	-0.005	-0.00920	$6.00e-06$	
GMM σ^2	0.01	0.0100	$1.29e-07$	

Results of simulations with the Brennan-Schwartz model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 10000 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a Brennan-Schwartz model with $\alpha = 0.0004$, $\beta = -0.005$ *and* $\sigma^2 = 0.01$.

The GMM estimates for α and β display obvious bias towards larger absolute values. Some bias seems to be present in the maximum likelihood estimates of these parameters as well, but in a much smaller degree. Estimates of σ^2 are consistent and stable for both estimation procedures.

The CIR VR model $[\alpha, \beta = 0, \gamma = 1.5]$

Table 5.7: *CIR VR model estimates from large sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE σ^2	0.01	0.00998	$3.03e-08$	$2.02e-08$
GMM σ^2	0.01	0.00997	$6.99e-08$	

Results of simulations with the CIR VR model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 10000 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a CIR VR model with $\sigma^2 = 0.01$.

The CIR VR model has much in common with the Dothan model in the way that it only estimates one parameter, namely σ^2. This is reflected in the maximum likelihood estimates in this model, as the estimates and variance are identical with the ones from the Dothan model. In the estimates generated by GMM though, we notice a smaller variance in the CIR VR setting than in the Dothan model. This indicates that the GMM method finds it easier to confidently estimate the σ^2 parameter when γ is fixed at 1.5 rather than 0.5. It should be noted that the CIR VR model is not stationary as the γ-parameter exceeds 1, and that this means the previously reviewed theory does not necessarily apply for the CIR VR model.

The CEV model $[\alpha = 0]$

Table 5.8: *CEV model estimates from large sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE β	-0.005	-0.00604	$7.54e-06$	$6.25e-06$
MLE σ^2	0.010	0.00998	$9.19e-08$	$6.93e-08$
MLE γ	0.500	0.500	$1.24e-05$	$1.06e-05$
GMM β	-0.005	-0.00706	$1.13e-05$	
GMM σ^2	0.010	0.00954	$5.74e-06$	
GMM γ	0.500	0.480	$1.87e-02$	

Results of simulations with the CEV model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 10000 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a CEV model with $\beta = -0.005$, $\sigma^2 = 0.01$ and $\gamma = 0.5$.

The CEV model is the only model except the full CKLS model that does not impose restrictions on the γ parameter. The results reveal bias in the estimates for the β parameter for both estimation methods but again it is more pronounced in the GMM estimates. We also note that the variance of estimates are consistently lower for the maximum likelihood estimates.

The CKLS (full) model

Table 5.9: *CKLS model estimates from large sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.000409	$2.06e-09$	$1.66e-09$
MLE β	-0.005	-0.00591	$6.68e-06$	$5.13e-06$
MLE σ^2	0.01	0.00998	$9.40e-08$	$6.54e-08$
MLE γ	0.5	0.500	$1.31e-05$	$1.14e-05$
GMM α	0.0004	0.000496	$1.21e-07$	
GMM β	-0.005	-0.00725	$1.10e-05$	
GMM σ^2	0.01	0.0102	$1.92e-06$	
GMM γ	0.5	0.514	$6.80e-03$	

Results of simulations with the full CKLS model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 10000 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a CKLS model with $\alpha = 0.0004$, *beta* $= -0.005$, $\sigma^2 = 0.01$ *and* $\gamma = 0.5$.

Finally, we investigate the results of estimation in the full CKLS model. We notice again bias towards larger absolute values in both maximum likelihood and GMM estimates of the β parameter, and that the bias is more severe in the GMM estimates. This issue of overestimation of the mean-reversion speed in the CKLS model has previously been discussed in Faff and Gray [2006]. Our results confirm these notions. There is also some evidence of bias in the α parameter estimate, although not as obvious as for the β parameter.

Maximum likelihood gives less variability in all four parameter estimates. This might not come as a surprise, as the maximum likelihood uses the actual likelihood functions in its calculations and one should expect that information to be worth something in terms of precision in estimates. In any case our results confirm that there is a substantial increase in parameter estimate confidence to be gained from using maximum likelihood over GMM.

5.1.3 Simulation with smaller samples

In our simulations of 50 times 500 observations we observe that we as expected lose precision in estimates. Another effect though from the shortening of sample length is that the issue of bias in the β estimates evidently seems to be aggravated compared to the longer samples. This is especially clear in the GMM estimates. There is no question that this is an issue that is very important to be aware of when dealing with estimates in these models. This has already been noted by Faff and Gray [2006]. GMM estimates also exhibits bias in the α parameter.

On a more positive note, we observe that the other estimations generally yield reasonable results. This is good news, and shows that even relatively short samples can give us much information about the parameters. In all cases, however, maximum likelihood returns the same or better precision in the estimates, and also generally suffers less from the issue of bias.

Table 5.10: *Merton model estimates from small sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.000236	$1.57e-05$	$1.85e-05$
MLE σ^2	0.01	0.0101	$4.38e-07$	$3.26e-07$
GMM α	0.0004	0.000241	$1.67e-05$	
GMM σ^2	0.01	0.00994	$5.04e-07$	

Results of simulations with the Merton model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 500 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a Merton model with $\alpha = 0.0004$ and $\sigma^2 = 0.01$.

Table 5.11: *Vasicek model estimates from small sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.000436	$1.12e-04$	$3.58e-05$
MLE β	-0.005	-0.0164	$1.34e-04$	$2.43e-05$
MLE σ^2	0.01	0.0101	$4.30e-07$	$3.38e-07$
GMM α	0.0004	0.000394	$1.12e-04$	
GMM β	-0.005	-0.0165	$1.37e-04$	
GMM σ^2	0.01	0.0100	$4.70e-07$	

Results of simulations with the Vasicek model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 500 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a Vasicek model with $\alpha = 0.0004$, $\beta = -0.005$ and $\sigma^2 = 0.01$.

Table 5.12: *CIR SR model estimates from small sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.000447	$1.02e-07$	$1.13e-07$
MLE β	-0.005	-0.0189	$6.40e-04$	$4.26e-05$
MLE σ^2	0.01	0.0101	$4.40e-07$	$3.28e-07$
GMM α	0.0004	0.000946	$4.17e-06$	
GMM β	-0.005	-0.0347	$7.92e-04$	
GMM σ^2	0.01	0.00444	$7.90e-05$	

Results of simulations with the CIR SR model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 500 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a CIR SR model with $\alpha = 0.0004$ $\beta = -0.005$ *and* $\sigma^2 = 0.01$.

Table 5.13: *Dothan model estimates from small sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE σ^2	0.01	0.0102	$4.41e-07$	$3.25e-07$
GMM σ^2	0.01	0.00983	$5.85e-07$	

Results of simulations with the Dothan model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 500 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a Dothan model with $\sigma^2 = 0.01$.

Table 5.14: *GBM model estimates from small sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE β	-0.005	-0.00521	$1.50e-05$	$1.85e-05$
MLE σ^2	0.010	0.0101	$4.38e-07$	$3.26e-07$
GMM β	-0.005	-0.00675	$5.29e-05$	
GMM σ^2	0.010	0.00974	$9.09e-07$	

Results of simulations with the GBM model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 500 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a GBM model with $\beta = -0.005$ and $\sigma^2 = 0.01$.

Table 5.15: *Brennan-Schwartz model estimates from small sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.000785	$2.65e-07$	$1.07e-07$
MLE β	-0.005	-0.0149	$1.88e-04$	$4.81e-05$
MLE σ^2	0.01	0.0101	$4.33e-07$	$3.30e-07$
GMM α	0.0004	0.00122	$6.31e-07$	
GMM β	-0.005	-0.0222	$3.06e-04$	
GMM σ^2	0.01	0.0102	$5.11e-07$	

Results of simulations with the Brennan-Schwartz model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 500 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a Brennan-Schwartz model with $\alpha = 0.0004$, $\beta = -0.005$ and $\sigma^2 = 0.01$.

Table 5.16: *CIR VR model estimates from small sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE σ^2	0.01	0.0102	$4.41e-07$	$3.25e-07$
GMM σ^2	0.01	0.00993	$4.3e-07$	

Results of simulations with the CIR VR model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 500 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a CIR VR model with $\sigma^2 = 0.01$.

Table 5.17: *CEV model estimates from small sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE β	-0.005	-0.0194	$7.21e-04$	$5.42e-05$
MLE σ^2	0.010	0.00981	$1.60e-06$	$6.11e-07$
MLE γ	0.500	0.496	$2.09e-04$	$2.52e-04$
GMM β	-0.005	-0.0287	0.00101	
GMM σ^2	0.010	0.0137	0.00041	
GMM γ	0.500	0.493	0.02090	

Results of simulations with the CEV model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 500 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a CEV model with $\beta = -0.005$, $\sigma^2 = 0.01$ *and* $\gamma = 0.5$.

Table 5.18: *CKLS model estimates from small sample simulation*

Parameter	True value	Empirical mean	Empirical variance	Fisher variance
MLE α	0.0004	0.000449	$1.06e-07$	$1.14e-07$
MLE β	-0.005	-0.0192	$6.52e-04$	$4.33e-05$
MLE σ^2	0.01	0.00971	$1.54e-06$	$5.29e-07$
MLE γ	0.5	0.494	$2.61e-04$	$2.31e-04$
GMM α	0.0004	0.000997	$4.11e-06$	
GMM β	-0.005	-0.0340	$8.07e-04$	
GMM σ^2	0.01	0.0104	$1.33e-05$	
GMM γ	0.5	0.497	$6.87e-03$	

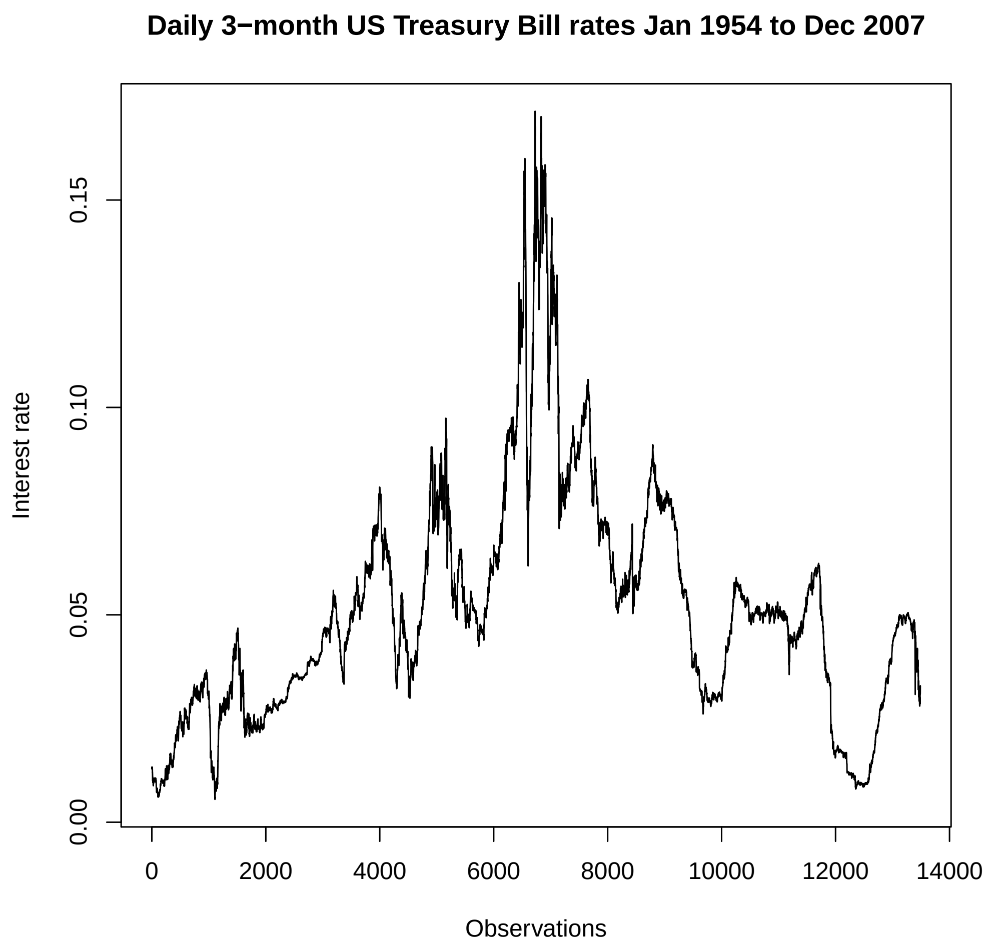
Results of simulations with the full CKLS model. The table shows the empirical mean and variance of both MLE and GMM estimates from 50 performed simulations of 500 observations each, plus the variance of the MLE estimate derived from the Fisher information matrix. Data was generated from a CKLS model with $\alpha = 0.0004$, $\beta = -0.005$, $\sigma^2 = 0.01$ *and* $\gamma = 0.5$.

5.2 Estimating parameters from real data

In this section we will estimate parameters from real life data using both MLE and GMM, and compare the results.

5.2.1 The data

The data is from 3-month US Treasury bills from the period of January 1954 to December 2007, collected from the webpages of the Federal Reserve Bank of St.Louis. (http://research.stlouisfed.org/fred2/categories/116). These are recorded with daily, weekly and monthly frequences. For the daily data, data is from workdays only. In the 1979-82 period, there was a temporary change in the monetary policy of the Federal Reserve (Dell'Aquila, Ronchetti, and Trojani [2003]). To obtain results without addressing the issue of a structural break in this period, we also estimate parameters from data from January 1983 to December 2007.



5.2.2 The results

Parameter estimates are given for the full CKLS model and the specified 8 nested models. Standard deviations of MLE estimates derived from the observed Fisher information matrix and standard deviations of GMM estimates are included in parenthesises.

Table 5.19: *Maximum likelihood estimates, daily data, Jan 1954-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.000122	-0.00197	0.0000976	0.841
	(0.00000503)	(0.000201)	(0.00000198)	(0.00299)
Merton	0.00000151	0	0.000000890	0
	(0.00000813)		(0.00000000277)	
Vasicek	0.0000336	-0.000626	0.000000890	0
	(0.0000209)	(0.000199)	(0.00000000277)	
CIR SR	0.0000205	-0.000370	0.0000122	0.5
	(0.0000102)	(0.000180)	(0.0000000546)	
Dothan	0	0	0.000274	1
			(0.00000125)	
GBM	0	0.000208	0.000274	1
		(0.000143)	(0.00000125)	
Brennan-Schwartz	0.0000187	-0.000318	0.000274	1
	(0.00000384)	(0.000221)	(0.00000125)	
CIR VR	0	0	0.0120	1.5
			(0.0000313)	
CEV	0	0.000112	0.0000602	0.769
		(0.000133)	(0.00000117)	(0.00302)

Maximum likelihood parameter estimates for daily data of US Treasury bills in the period January 1954 to December 2007 (13488 observations). Standard deviations derived from the observed Fisher information matrix are shown in parenthesises.

Table 5.20: *GMM estimates, daily data, Jan 1954-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.0000280 (0.0000310)	-0.000518 (0.000725)	0.0480 (0.0190)	2.11 (0.0878)
Merton	0.480 (0.00000617)	0	0.230 (0.0000000532)	0
Vasicek	0.488 (0.00305)	0.190 (0.0592)	0.248 (0.00000854)	0
CIR SR	0.0741 (0.000509)	1.30 (0.00921)	0.407 (0.00240)	0.5
Dothan	0	0	0.000205 (0.0000105)	1
GBM	0	-0.000861 (0.000144)	0.000235 (0.0000106)	1
Brennan-Schwartz	-0.000107 (0.0000316)	0.00267 (0.000737)	0.000141 (0.0000107)	1
CIR VR	0	0	0.00321 (0.000154)	1.5
CEV	0	0.000764 (0.000144)	0.183 (0.0950)	2.44 (0.121)

GMM parameter estimates for daily data of US Treasury bills in the period January 1954 to December 2007 (13488 observations). Standard deviations are shown in parenthesises.

Table 5.21: *Maximum likelihood estimates, weekly data, Jan 1954-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.000188	-0.00338	0.000261	0.771
	(0.0000236)	(0.000839)	(0.0000108)	(0.00609)
Merton	0.00000689	0	0.00000398	0
	(0.0000380)		(0.0000000306)	
Vasicek	0.000151	-0.00280	0.00000398	0
	(0.000100)	(0.000942)	(0.0000000304)	
CIR SR	0.0000945	-0.00171	0.0000537	0.5
	(0.0000444)	(0.000789)	(0.000000596)	
Dothan	0	0	0.00125	1
			(0.0000129)	
GBM	0	0.000936	0.00125	1
		(0.000666)	(0.0000129)	
Brennan-Schwartz	0.0000961	-0.00175	0.00125	1
	(0.0000171)	(0.00105)	(0.0000130)	
CIR VR	0	0	0.0600	1.5
			(0.000000146)	
CEV	0	0.000463	0.000224	0.741
		(0.000616)	(0.00000947)	(0.00644)

Maximum likelihood parameter estimates for weekly data of US Treasury bills in the period January 1954 to December 2007 (2817 observations). Standard deviations derived from the observed Fisher information matrix are shown in parenthesises.

Table 5.22: *GMM estimates, weekly data, Jan 1954-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.000168	-0.00312	0.216	2.09
	(0.000159)	(0.00382)	(0.150)	(0.149)
Merton	0.0105	0	0.0000186	0
	(0.0000421)		(0.000000329)	
Vasicek	0.491	0.186	0.251	0
	(0.00299)	(0.0579)	(0.00000758)	
CIR SR	-0.0306	0.633	0.00416	0.5
	(0.00299)	(0.0676)	(0.000748)	
Dothan	0	0	0.000824	1
			(0.0000823)	
GBM	0	0.000121	0.000775	1
		(0.000730)	(0.0000849)	
Brennan-Schwartz	0.000138	-0.00276	0.000846	1
	(0.000159)	(0.00382)	(0.0000878)	
CIR VR	0	0	0.0139	1.5
			(0.00128)	
CEV	0	0.000563	0.354	2.21
		(0.000731)	(0.257)	(0.159)

GMM parameter estimates for weekly data of US Treasury bills in the period January 1954 to December 2007 (2817 observations). Standard deviations are shown in parenthesises.

Table 5.23: *Maximum likelihood estimates, monthly data, Jan 1954-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.000418 (0.000143)	-0.00744 (0.00410)	0.000986 (0.0000799)	0.725 (0.0118)
Merton	0.0000250 (0.000187)	0	0.0000201 (0.000000321)	0
Vasicek	0.000743 (0.000506)	-0.0139 (0.00475)	0.0000200 (0.000000327)	0
CIR SR	0.000440 (0.000200)	-0.00803 (0.00364)	0.000261 (0.00000655)	0.5
Dothan	0	0	0.00635 (0.0000997)	1
GBM	0	0.00427 (0.00337)	0.00634 (0.000108)	1
Brennan-Schwartz	0.000435 (0.0000995)	-0.00789 (0.00539)	0.00624 (0.000139)	1
CIR VR	0	0	0.333 (0.00265)	1.5
CEV	0	0.00192 (0.00294)	0.000963 (0.0000754)	0.719 (0.0107)

Maximum likelihood parameter estimates for monthly data of US Treasury bills in the period January 1954 to December 2007 (648 observations). Standard deviations derived from the observed Fisher information matrix are shown in parenthesises.

Table 5.24: *GMM estimates, monthly data, Jan 1954-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.000752 (0.000541)	-0.0140 (0.0124)	0.813 (0.996)	2.02 (0.242)
Merton	0.00294 (0.000180)	0	0.0000688 (0.00000221)	0
Vasicek	0.485 (0.00344)	0.197 (0.0663)	0.245 (0.00000576)	0
CIR SR	0.0753 (0.00871)	-1.09 (0.166)	0.0309 (0.00273)	0.5
Dothan	0	0	0.00324 (0.000455)	1
GBM	0	0.00212 (0.00366)	0.00321 (0.000461)	1
Brennan-Schwartz	0.000335 (0.000484)	-0.00542 (0.0114)	0.00316 (0.000461)	1
CIR VR	0	0	0.0600 (0.00884)	1.5
CEV	0	0.00246 (0.00366)	0.444 (0.515)	1.87 (0.223)

GMM parameter estimates for monthly data of US Treasury bills in the period January 1954 to December 2007 (648 observations). Standard deviations are shown in parenthesises.

Table 5.25: *Maximum likelihood estimates, daily data, Jan 1983-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.0000514 (0.00000452)	-0.00474 (0.000446)	0.0883 (0.00327)	1.81 (0.00410)
Merton	-0.00000733 (0.00000743)	0	0.000000340 (0.00000000218)	0
Vasicek	0.0000147 (0.0000223)	-0.000438 (0.000354)	0.000000340 (0.00000000218)	0
CIR SR	0.0000124 (0.0000131)	-0.000392 (0.000300)	0.00000679 (0.0000000395)	0.5
Dothan	0	0	0.000179 (0.000000933)	1
GBM	0	-0.0000511 (0.000170)	0.000179 (0.000000936)	1
Brennan-Schwartz	0.0000116 (0.00000559)	-0.000369 (0.000284)	0.000179 (0.000000943)	1
CIR VR	0	0	0.00724 (0.0000301)	1.5
CEV	0	-0.000145 (0.000150)	0.00000752 (0.000000401)	0.516 (0.00823)

Maximum likelihood parameter estimates for daily data of US Treasury bills in the period January 1983 to December 2007 (6250 observations). Standard deviations derived from the observed Fisher information matrix are shown in parenthesises.

Table 5.26: *GMM estimates, daily data, Jan 1983-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.0000239 (0.0000158)	-0.000619 (0.000344)	0.0851 (0.0935)	2.18 (0.219)
Merton	0.0994 (0.00000918)	0	0.00953 (0.0000000496)	0
Vasicek	0.485 (0.00302)	0.197 (0.0599)	0.245 (0.00000570)	0
CIR SR	0.211 (0.0371)	0.0718 (0.739)	0.926 (0.0327)	0.5
Dothan	0	0	0.0000854 (0.00000590)	1
GBM	0	0.00141 (0.000153)	0.000105 (0.00000620)	1
Brennan-Schwartz	0.0000327 (0.0000164)	-0.00235 (0.000365)	0.000200 (0.00000794)	1
CIR VR	0	0	0.00106 (0.0000768)	1.5
CEV	0	0.000853 (0.000153)	0.234 (0.289)	2.38 (0.248)

GMM parameter estimates for daily data of US Treasury bills in the period January 1983 to December 2007 (6250 observations). Standard deviations are shown in parenthesises.

Table 5.27: *Maximum likelihood estimates, weekly data, Jan 1983-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.00101 (0.0000447)	-0.0211 (0.00157)	0.000196 (0.0000288)	0.767 (0.0225)
Merton	-0.0000362 (0.0000327)	0	0.00000119 (0.0000000163)	0
Vasicek	0.0000448 (0.0000972)	-0.00161 (0.00150)	0.00000119 (0.0000000163)	0
CIR SR	0.0000388 (0.0000568)	-0.00149 (0.00124)	0.0000225 (0.000000356)	0.5
Dothan	0	0	0.000548 (0.00000859)	1
GBM	0	-0.000422 (0.000700)	0.000547 (0.00000930)	1
Brennan-Schwartz	0.0000348 (0.0000249)	-0.00138 (0.00113)	0.000547 (0.00000930)	1
CIR VR	0	0	0.0196 (0.000253)	1.5
CEV	0	-0.000664 (0.000651)	0.0000526 (0.00000560)	0.638 (0.0175)

Maximum likelihood parameter estimates for weekly data of US Treasury bills in the period January 1983 to December 2007 (1304 observations). Standard deviations derived from the observed Fisher information matrix are shown in parenthesises.

Table 5.28: *GMM estimates, weekly data, Jan 1983-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.0000137	-0.000993	0.0112	1.66
	(0.0000874)	(0.00192)	(0.0191)	(0.344)
Merton	0.406	0	0.164	0
	(0.0000333)		(0.000000182)	
Vasicek	0.485	0.197	0.245	0
	(0.00285)	(0.0565)	(0.00000531)	
CIR SR	0.0559	1.21	0.280	0.5
	(0.000691)	(0.0194)	(0.00393)	
Dothan	0	0	0.000325	1
			(0.0000549)	
GBM	0	0.000452	0.000250	1
		(0.000713)	(0.0000569)	
Brennan-Schwartz	0.000204	-0.00513	0.000365	1
	(0.0000820)	(0.00188)	(0.0000588)	
CIR VR	0	0	0.00365	1.5
			(0.000688)	
CEV	0	0.00161	0.245	2.41
		(0.000733)	(1.42)	(1.20)

GMM parameter estimates for weekly data of US Treasury bills in the period January 1983 to December 2007 (1304 observations). Standard deviations are shown in parenthesises.

Table 5.29: *Maximum likelihood estimates, monthly data, Jan 1983-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.000443 (0.000210)	-0.0136 (0.00543)	0.000176 (0.0000740)	0.578 (0.0624)
Merton	-0.000164 (0.000147)	0	0.00000526 (0.000000337)	0
Vasicek	0.000180 (0.000373)	-0.00680 (0.00575)	0.00000524 (0.000000338)	0
CIR SR	0.000186 (0.000228)	-0.00691 (0.00531)	0.000104 (0.00000722)	0.5
Dothan	0	0	0.00279 (0.000135)	1
GBM	0	-0.00181 (0.00308)	0.00278 (0.000137)	1
Brennan-Schwartz	0.000184 (0.000118)	-0.00688 (0.00575)	0.00277 (0.000146)	1
CIR VR	0	0	0.110 (0.00321)	1.5
CEV	0	-0.00325 (0.00291)	0.000100 (0.0000389)	0.493 (0.0595)

Maximum likelihood parameter estimates for monthly data of US Treasury bills in the period January 1983 to December 2007 (300 observations). Standard deviations derived from the observed Fisher information matrix are shown in parenthesises.

Table 5.30: *GMM estimates, monthly data, Jan 1983-Dec 2007*

Model	α	β	σ^2	γ
Unrestricted	0.000166 (0.000525)	-0.00650 (0.0104)	0.943 (2.76)	2.19 (0.577)
Merton	0.551 (0.000222)	0	0.303 (0.000000804)	0
Vasicek	0.485 (0.00258)	0.197 (0.0507)	0.245 (0.00000441)	0
CIR SR	0.0259 (0.00956)	-0.553 (0.164)	0.00477 (0.000888)	0.5
Dothan	0	0	0.00154 (0.000240)	1
GBM	0	-0.00449 (0.00438)	0.00174 (0.000278)	1
Brennan-Schwartz	0.000761 (0.000453)	-0.0188 (0.00935)	0.00173 (0.000280)	1
CIR VR	0	0	0.0200 (0.00339)	1.5
CEV	0	-0.00240 (0.00441)	0.426 (1.22)	2.07 (0.572)

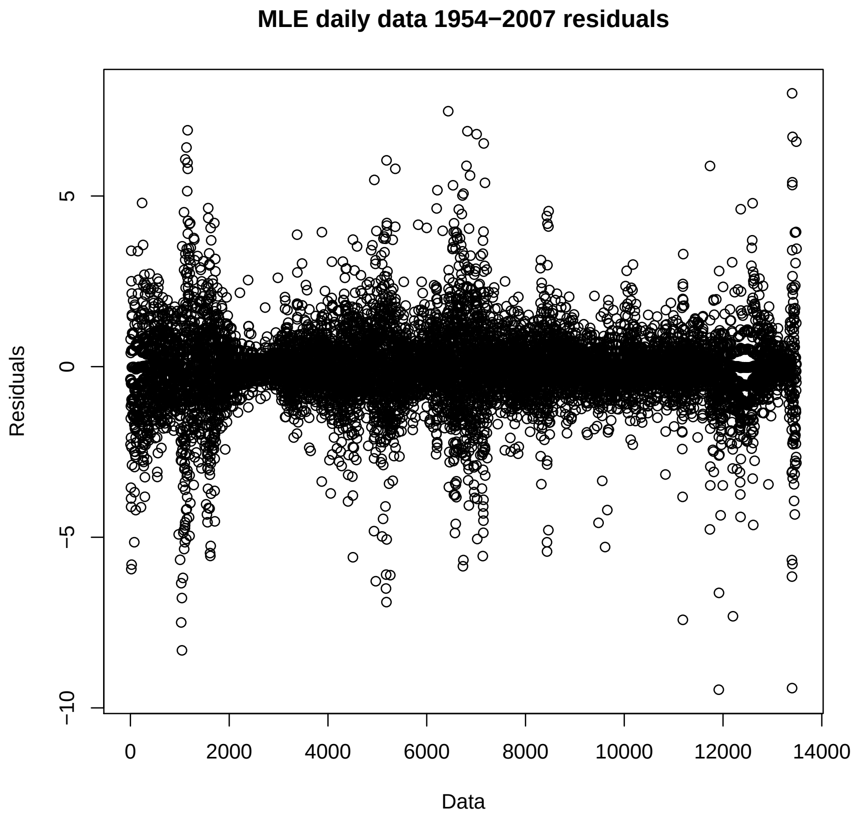
GMM parameter estimates for monthly data of US Treasury bills in the period January 1983 to December 2007 (300 observations). Standard deviations are shown in parenthesises.

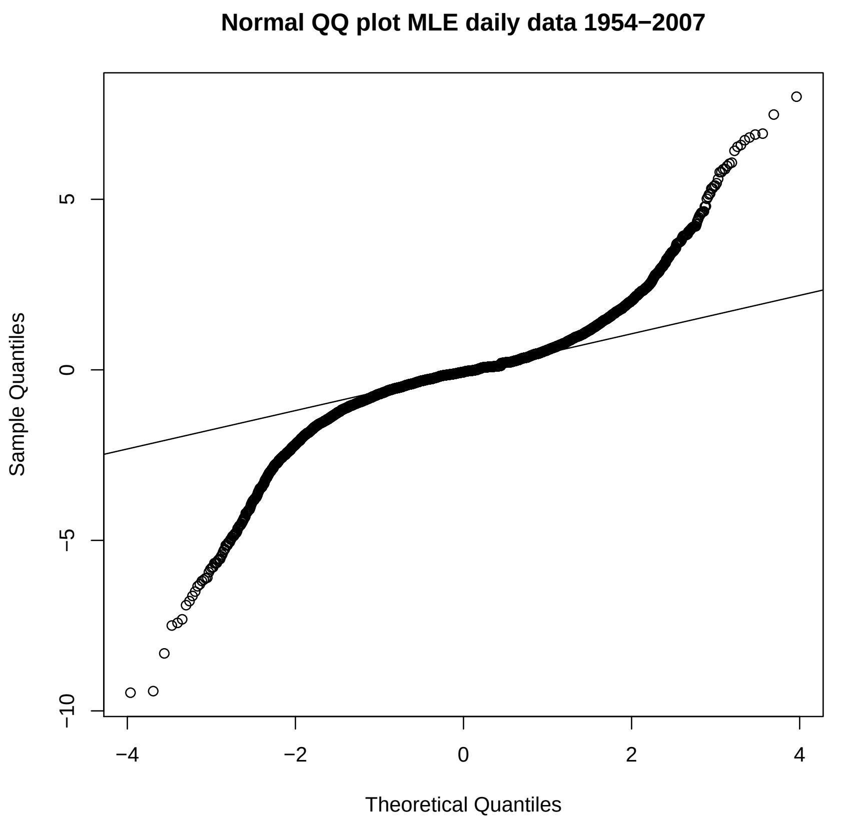
Comparing the estimates obtained using MLE to the estimates obtained using GMM, we find that they differ substantially. This arouses suspicion that something has gone wrong in the estimation procedure of one of the methods. Preliminary error testing reveals that initial values used in numerical optimisation during GMM procedure might be responsible for this. Originally, optimisation methods for both MLE and GMM were given the following initial values of search: $\alpha = 0$, $\beta = 0$, $\sigma^2 = 1$ and $\gamma = 1$. When re-estimating with GMM for some of the models using the MLE estimates as initial values, GMM estimates proved to converge towards values more in line with the MLE estimates. Changing the initial values for MLE optimisation resulted in no significant change in estimates though. In conclusion, this indicates a certain lack of robustness in the numerical optimisation part of the GMM estimation procedure. As the same optimisation code was used for both GMM and MLE estimates, it seems natural to assume that there might be some feature of the GMM equations that possibly makes it less robust in numerical optimisation situations. Time constraints prohibit this thesis from going into further detail on the matter, but an investigation would be interesting as further work.

Analysing the MLE estimates in the full CKLS model we find some evidence of mean-reversion, although the values of the β parameter are relatively small. This means that even though the short term interest rate process might be mean-reverting, it seems to do so at a slow pace. MLE estimates generally sets the γ parameter to be smaller than 1. This contrasts somewhat with the results of Chan et al. [1992]. Estimating from a smaller set of US Treasury bill rates, Chan et al. [1992] find the γ parameter to be about 1.5.

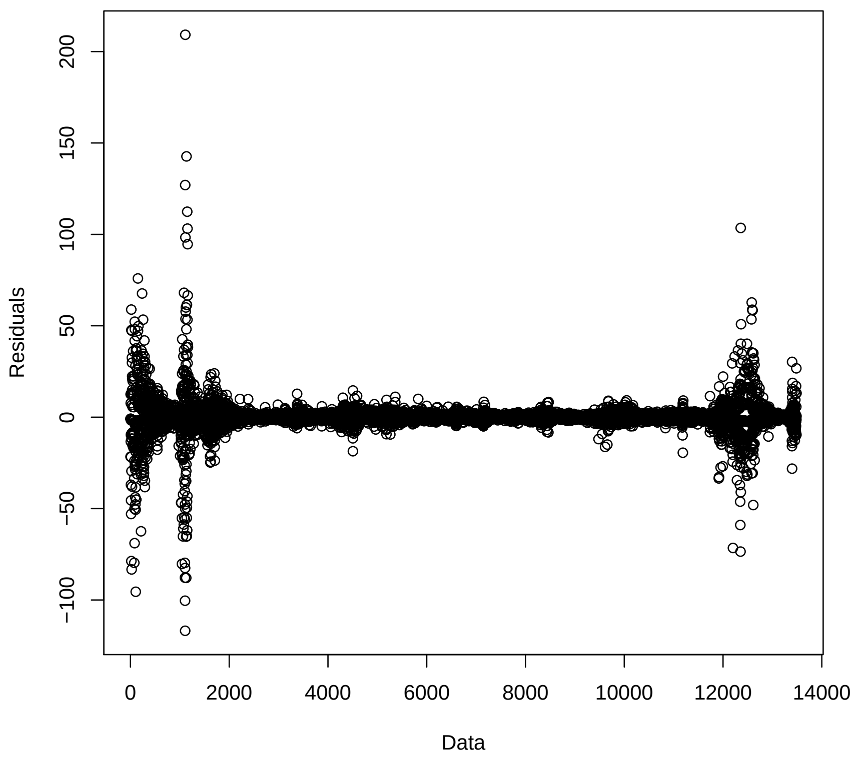
5.2.3 Residuals and QQ-plots

In this section, residual plots and normal QQ-plots are displayed for both the MLE and GMM estimates in the full CKLS model. This done for the daily data from January 1954 to December 2007, and for the monthly data form the same time period. The normal QQ-plots show that the residuals are not normally distributed. The plots for GMM estimates further point toward serious error in the calculation of the GMM estimates in this section.

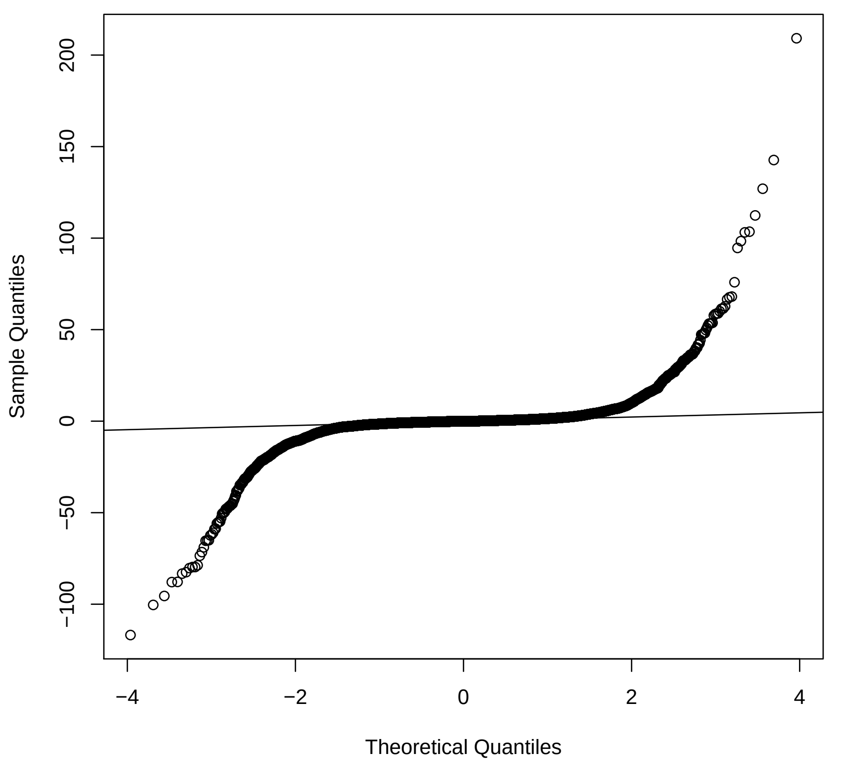

MLE daily data 1954−2007 residuals
Residuals
Data
0
2000
4000
6000
8000
10000
12000
14000
5
0
−5
−10

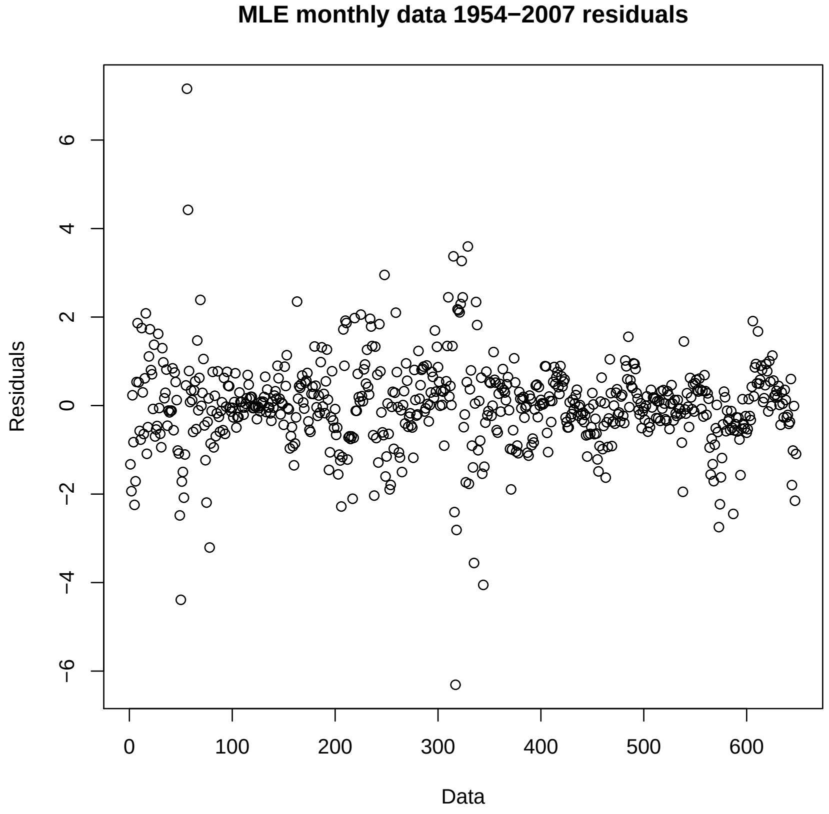

Normal QQ plot MLE daily data 1954−2007
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
5
0
−5
−10

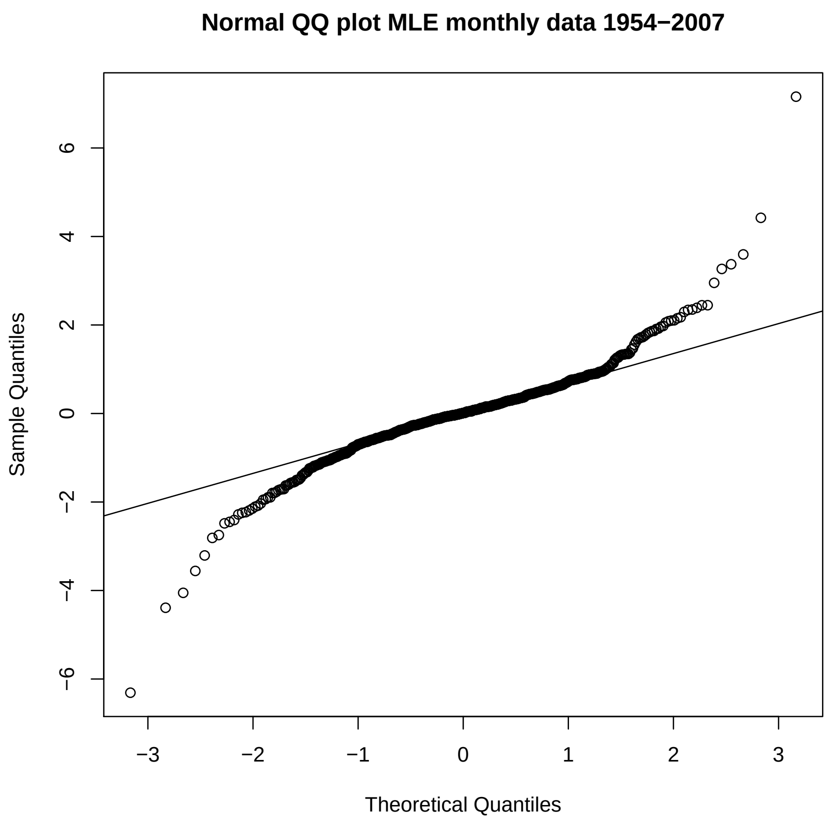
GMM daily data 1954−2007 residuals

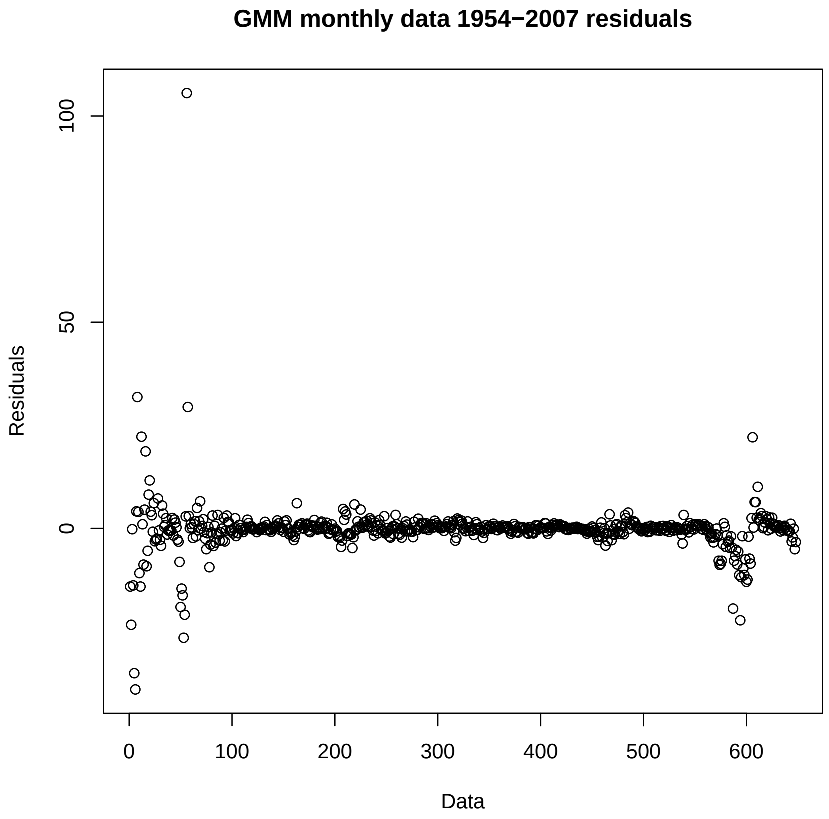


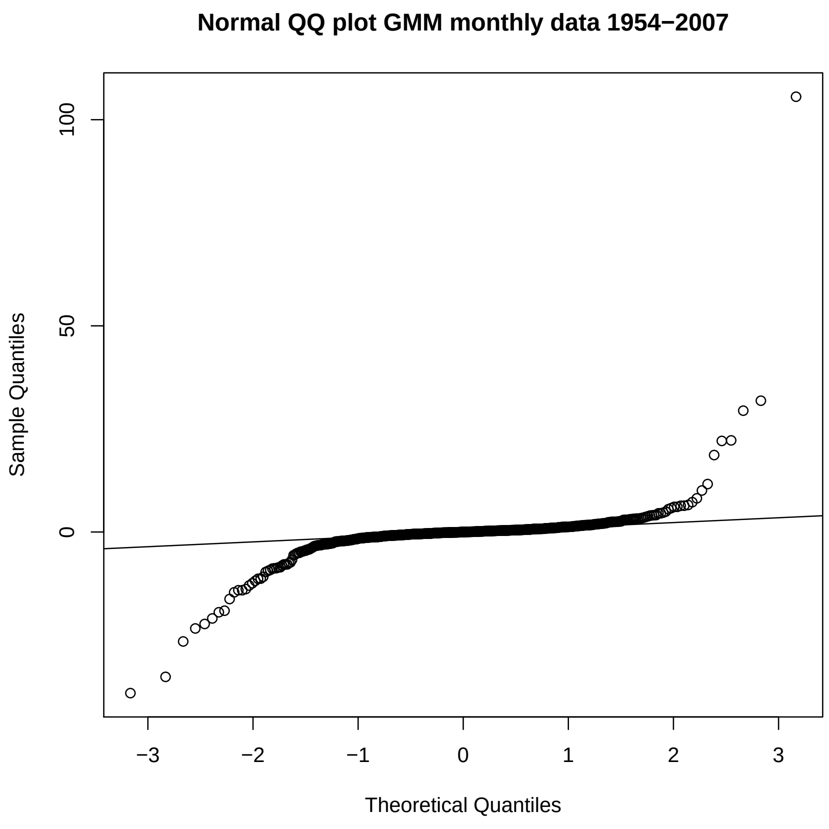
Normal QQ plot GMM daily data 1954−2007




MLE monthly data 1954−2007 residuals
Residuals
Data
−6
−4
−2
0
2
4
6
0
100
200
300
400
500
600


Normal QQ plot MLE monthly data 1954−2007
Sample Quantiles
Theoretical Quantiles
−6
−4
−2
0
2
4
6
−3
−2
−1
0
1
2
3


GMM monthly data 1954−2007 residuals
Residuals
Data
0
50
100
0
100
200
300
400
500
600


Normal QQ plot GMM monthly data 1954−2007
Sample Quantiles
Theoretical Quantiles
0
50
100
−3
−2
−1
0
1
2
3

Chapter 6

Conclusions

Our simulation experiments conclude that MLE estimates have smaller variance than the GMM estimates in the CKLS model, and smaller or equal variance in all the nested models. This is not very surprising, as the MLE method uses more information. More specifically it uses the information of the likelihood function in its calculation of estimates. Theory of asymptotics dictates that the variance of MLE estimates reach the Cramer-Rao bound, assuming stationarity of the model and some regularity conditions. GMM estimates are slower to converge, but are still efficient in most cases where the available data samples are large enough. However, GMM seems to have trouble estimating the β-parameter in the full CKLS model, as well as in some of the nested models. Bias towards larger absolute values for this parameter is evident. This is also noted in Faff and Gray [2006]. The bias effect is also found in MLE estimates, but it is much smaller than in GMM. This should be a good reason to consider MLE as an alternative in place of GMM for parameter estimation in these models. The findings of this thesis is that the extra work in calculating likelihood functions is sufficiently rewarded with a more sensible estimate of particularly the β-parameter.

Looking at the results of estimating parameters from real life data, we find that the differences between the MLE and GMM parameter estimates calculated from the US Treasury bills datasets are large. We assume from preliminary error testing that the GMM estimates seem to be wrong and that this is likely due to a lack of robustness in the part of the estimation procedure involving numerical optimisation. As the same optimisation code was used for both GMM and MLE, we suspect that the lack of robustness might be linked to some feature of the GMM system of equations. An investigation on this matter would be interesting for further work.

Analysing the MLE estimates in the full CKLS model we find some evidence of a slow mean-reversion. Moreover, our MLE estimates generally set the γ parameter to be smaller than 1. This contrasts somewhat with the results of Chan et al. [1992]. Estimating from a smaller set of US Treasury bill rates, Chan et al. [1992] find the γ parameter to be about 1.5.

Appendix: Covariance matrices from simulation experiments

In this section the covariance matrices obtained from the simulation experiments performed in Chapter 5.1 are displayed. From these we may calculate the observed correlation between parameter estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ by

$$\hat{\rho}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Cov}(\hat{\theta}_1, \hat{\theta}_2)}{\sqrt{\text{Var}(\hat{\theta}_1)\text{Var}(\hat{\theta}_2)}} \quad (1)$$

Covariance matrices from both the experiment with longer samples (50 sets of 10000 observations) as well as from the experiment with shorter samples (50 sets of 500 observations) are included.

Covariance matrices from simulations with 50 sets of 10000 observations

Table 1: *Empirical covariance matrix for the MLE estimates in the Merton model simulation with 50 sets of 10000 observations*

MERTONcovarMLE	α	σ^2
α	$9.64e-07$	$-3.17e-10$
σ^2	$-3.17e-10$	$3.05e-08$

Table 2: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the Merton model simulation with 50 sets of 10000 observations*

MERTONfisherVar	α	σ^2
α	$1.02e-06$	$7.72e-10$
σ^2	$7.72e-10$	$2.02e-08$

Table 3: *Empirical covariance matrix for the GMM estimates in the Merton model simulation with 50 sets of 10000 observations*

MERTONcovarGMM	α	σ^2
α	$9.8e-07$	$1.50e-09$
σ^2	$1.5e-09$	$2.99e-08$

Table 4: *Empirical covariance matrix for the MLE estimates in the Vasicek model simulation with 50 sets of 10000 observations*

VASICEKcovarMLE	α	β	σ^2
α	$1.16e-06$	$-1.70e-07$	$-4.57e-09$
β	$-1.70e-07$	$1.04e-06$	$1.44e-08$
σ^2	$-4.57e-09$	$1.44e-08$	$3.04e-08$

Table 5: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the Vasicek model simulation with 50 sets of 10000 observations*

VASICEKfisherVar	α	β	σ^2
α	$1.02e-06$	$-2.55e-08$	$7.13e-10$
β	$-2.55e-08$	$9.96e-07$	$2.28e-09$
σ^2	$7.13e-10$	$2.28e-09$	$2.02e-08$

Table 6: *Empirical covariance matrix for the GMM estimates in the Vasicek model simulation with 50 sets of 10000 observations*

VASICEKcovarGMM	α	β	σ^2
α	$1.16e-06$	$-1.70e-07$	$-3.63e-09$
β	$-1.70e-07$	$1.05e-06$	$1.42e-08$
σ^2	$-3.63e-09$	$1.42e-08$	$3.04e-08$

Table 7: *Empirical covariance matrix for the MLE estimates in the CIR SR model simulation with 50 sets of 10000 observations*

CIRSRcovarMLE	α	β	σ^2
α	$2.05e-09$	$6.23e-09$	$-1.44e-09$
β	$6.23e-09$	$6.65e-06$	$-5.13e-08$
σ^2	$-1.44e-09$	$-5.13e-08$	$3.03e-08$

Table 8: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the CIR SR model simulation with 50 sets of 10000 observations*

CIRSRfisherVar	α	β	σ^2
α	$1.66e-09$	$-1.73e-09$	$1.11e-11$
β	$-1.73e-09$	$5.12e-06$	$5.44e-09$
σ^2	$1.11e-11$	$5.44e-09$	$2.02e-08$

Table 9: *Empirical covariance matrix for the GMM estimates in the CIR SR model simulation with 50 sets of 10000 observations*

CIRSRcovarGMM	α	β	σ^2
α	$1.22e-07$	$-6.61e-07$	$1.19e-06$
β	$-6.61e-07$	$1.78e-05$	$-6.82e-06$
σ^2	$1.19e-06$	$-6.82e-06$	$3.01e-05$

Table 10: *Empirical covariance for the MLE estimates in the Dothan model simulation with 50 sets of 10000 observations*

DOTHANcovarMLE	σ^2
σ^2	$3.03e-08$

Table 11: *Covariance derived from the observed Fisher information matrix for the MLE estimates in the Dothan model simulation with 50 sets of 10000 observations*

DOTHANfisherVar	σ^2
σ^2	$2.02e-08$

Table 12: *Empirical covariance for the GMM estimates in the Dothan model simulation with 50 sets of 10000 observations*

DOTHANcovarGMM	σ^2
σ^2	$7.96e-07$

Table 13: *Empirical covariance matrix for the MLE estimates in the GBM model simulation with 50 sets of 10000 observations*

GBMcovarMLE	β	σ^2
β	$9.16e-07$	$2.87e-09$
σ^2	$2.87e-09$	$3.05e-08$

Table 14: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the GBM model simulation with 50 sets of 10000 observations*

GBMfisherVar	β	σ^2
β	$1.02e-06$	$7.72e-10$
σ^2	$7.72e-10$	$2.02e-08$

Table 15: *Empirical covariance matrix for the GMM estimates in the GBM model simulation with 50 sets of 10000 observations*

GBMcovarGMM	β	σ^2
β	$4.03e-05$	$-5.00e-06$
σ^2	$-5.00e-06$	$4.24e-06$

Table 16: *Empirical covariance matrix for the MLE estimates in the Brennan-Schwartz model simulation with 50 sets of 10000 observations*

BRENNANcovarMLE	α	β	σ^2
α	$3.14e-09$	$-7.30e-08$	$2.59e-11$
β	$-7.30e-08$	$2.78e-06$	$-5.22e-09$
σ^2	$2.59e-11$	$-5.22e-09$	$3.03e-08$

Table 17: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the Brennan-Schwartz model simulation with 50 sets of 10000 observations*

BRENNANfisherVar	α	β	σ^2
α	$2.87e-09$	$-7.36e-08$	$-8.36e-11$
β	$-7.36e-08$	$2.91e-06$	$2.92e-09$
σ^2	$-8.36e-11$	$2.92e-09$	$2.02e-08$

Table 18: *Empirical covariance matrix for the GMM estimates in the Brennan-Schwartz model simulation with 50 sets of 10000 observations*

BRENNANcovarGMM	α	β	σ^2
α	$2.12e-08$	$-2.19e-07$	$1.63e-09$
β	$-2.19e-07$	$6.00e-06$	$-7.79e-09$
σ^2	$1.63e-09$	$-7.79e-09$	$1.29e-07$

Table 19: *Empirical covariance for the MLE estimates in the CIR VR model simulation with 50 sets of 10000 observations*

CIRVRcovarMLE	σ^2
σ^2	$3.03e-08$

Table 20: *Covariance derived from the observed Fisher information matrix for the MLE estimates in the CIRVR model simulation with 50 sets of 10000 observations*

CIRVRfisherVar	σ^2
σ^2	$2.02e-08$

Table 21: *Empirical covariance for the GMM estimates in the CIR VR model simulation with 50 sets of 10000 observations*

CIRVRcovarGMM	σ^2
σ^2	$6.99e-08$

Table 22: *Empirical covariance matrix for the MLE estimates in the CEV model simulation with 50 sets of 10000 observations*

CEVcovarMLE	β	σ^2	γ
β	$7.54e-06$	$-1.08e-07$	$-1.23e-06$
σ^2	$-1.08e-07$	$9.19e-08$	$8.66e-07$
γ	$-1.23e-06$	$8.66e-07$	$1.24e-05$

Table 23: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the CEV model simulation with 50 sets of 10000 observations*

CEVfisherVar	β	σ^2	γ
β	$6.25e-06$	$2.41e-08$	$2.22e-07$
σ^2	$2.41e-08$	$6.93e-08$	$7.23e-07$
γ	$2.22e-07$	$7.23e-07$	$1.06e-05$

Table 24: *Empirical covariance matrix for the GMM estimates in the CEV model simulation with 50 sets of 10000 observations*

CEVcovarGMM	β	σ^2	γ
β	$1.13e-05$	$2.72e-07$	$1.52e-05$
σ^2	$2.72e-07$	$5.74e-06$	$3.19e-04$
γ	$1.52e-05$	$3.19e-04$	$1.87e-02$

Table 25: *Empirical covariance matrix for the MLE estimates in the full CKLS model simulation with 50 sets of 10000 observations*

CKLScovarMLE	α	β	σ^2	γ
α	$2.06e-09$	$5.60e-09$	$-2.48e-09$	$-1.46e-08$
β	$5.60e-09$	$6.68e-06$	$4.16e-08$	$1.26e-06$
σ^2	$-2.48e-09$	$4.16e-08$	$9.40e-08$	$9.10e-07$
γ	$-1.46e-08$	$1.26e-06$	$9.10e-07$	$1.31e-05$

Calculating the observed correlation between the MLE estimates of σ^2 and γ in the CKLS model by

$$\hat{\rho}(\hat{\sigma^2}, \hat{\gamma}) = 9.10e-07/\sqrt{9.40e-08 * 1.31e-05} \approx 0.82,$$

it is evident that the estimates of σ^2 and γ are strongly correlated. This confirms the relationship between σ^2 and γ that was described in Section 3.2.

Table 26: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the full CKLS model simulation with 50 sets of 10000 observations*

CKLSfisherVar	α	β	σ^2	γ
α	$1.66e-09$	$-1.79e-09$	$-1.99e-10$	$-3.34e-09$
β	$-1.79e-09$	$5.13e-06$	$1.82e-08$	$2.02e-07$
σ^2	$-1.99e-10$	$1.82e-08$	$6.54e-08$	$7.18e-07$
γ	$-3.34e-09$	$2.02e-07$	$7.18e-07$	$1.14e-05$

Table 27: *Empirical covariance matrix for the GMM estimates in the full CKLS model simulation with 50 sets of 10000 observations*

CKLScovarGMM	α	β	σ^2	γ
α	$1.21e-07$	$-2.92e-07$	$1.80e-08$	$-2.87e-06$
β	$-2.92e-07$	$1.10e-05$	$-1.65e-06$	$-4.31e-05$
σ^2	$1.80e-08$	$-1.65e-06$	$1.92e-06$	$7.31e-05$
γ	$-2.87e-06$	$-4.31e-05$	$7.31e-05$	$6.80e-03$

Covariance matrices from simulations with 50 sets of 500 observations

Table 28: *Empirical covariance matrix for the MLE estimates in the Merton model simulation with 50 sets of 500 observations*

MERTONcovarMLE	α	σ^2
α	$1.57e-05$	$1.29e-07$
σ^2	$1.29e-07$	$4.38e-07$

Table 29: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the Merton model simulation with 50 sets of 500 observations*

MERTONfisherVar	α	σ^2
α	$1.85e-05$	$-1.25e-07$
σ^2	$-1.25e-07$	$3.26e-07$

Table 30: *Empirical covariance matrix for the GMM estimates in the Merton model simulation with 50 sets of 500 observations*

MERTONcovarGMM	α	σ^2
α	$1.67e-05$	$-1.90e-08$
σ^2	$-1.90e-08$	$5.04e-07$

Table 31: *Empirical covariance matrix for the MLE estimates in the Vasicek model simulation with 50 sets of 500 observations*

VASICEKcovarMLE	α	β	σ^2
α	$1.12e-04$	$-4.98e-06$	$1.98e-06$
β	$-4.98e-06$	$1.34e-04$	$4.14e-07$
σ^2	$1.98e-06$	$4.14e-07$	$4.30e-07$

Table 32: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the Vasicek model simulation with 50 sets of 500 observations*

VASICEKfisherVar	α	β	σ^2
α	$3.58e-05$	$-2.05e-05$	$-5.70e-07$
β	$-2.05e-05$	$2.43e-05$	$5.28e-07$
σ^2	$-5.70e-07$	$5.28e-07$	$3.38e-07$

Table 33: *Empirical covariance matrix for the GMM estimates in the Vasicek model simulation with 50 sets of 500 observations*

VASICEKcovarGMM	α	β	σ^2
α	$1.12e-04$	$-1.82e-06$	$2.08e-06$
β	$-1.82e-06$	$1.37e-04$	$7.75e-07$
σ^2	$2.08e-06$	$7.75e-07$	$4.70e-07$

Table 34: *Empirical covariance matrix for the MLE estimates in the CIR SR model simulation with 50 sets of 500 observations*

CIRSRcovarMLE	α	β	σ^2
α	$1.02e-07$	$-1.08e-07$	$1.94e-08$
β	$-1.08e-07$	$6.40e-04$	$3.49e-06$
σ^2	$1.94e-08$	$3.49e-06$	$4.40e-07$

Table 35: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the CIR SR model simulation with 50 sets of 500 observations*

CIRSRfisherVar	α	β	σ^2
α	$1.13e-07$	$-1.87e-07$	$-9.63e-09$
β	$-1.87e-07$	$4.26e-05$	$2.78e-07$
σ^2	$-9.63e-09$	$2.78e-07$	$3.28e-07$

Table 36: *Empirical covariance matrix for the GMM estimates in the CIR SR model simulation with 50 sets of 500 observations*

CIRSRcovarGMM	α	β	σ^2
α	$4.17e-06$	$-1.94e-08$	$1.46e-05$
β	$-1.94e-08$	$7.92e-04$	$-3.46e-05$
σ^2	$1.46e-05$	$-3.46e-05$	$7.90e-05$

Table 37: *Empirical covariance for the MLE estimates in the Dothan model simulation with 50 sets of 500 observations*

DOTHANcovarMLE	σ^2
σ^2	$4.41e-07$

Table 38: *Covariance derived from the observed Fisher information matrix for the MLE estimates in the Dothan model simulation with 50 sets of 500 observations*

DOTHANfisherVar	σ^2
σ^2	$3.25e-07$

Table 39: *Empirical covariance for the GMM estimates in the Dothan model simulation with 50 sets of 500 observations*

DOTHANcovarGMM	σ^2
σ^2	$5.85e-07$

Table 40: *Empirical covariance matrix for the MLE estimates in the GBM model simulation with 50 sets of 500 observations*

GBMcovarMLE	β	σ^2
β	$1.50e-05$	$1.52e-07$
σ^2	$1.52e-07$	$4.38e-07$

Table 41: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the GBM model simulation with 50 sets of 500 observations*

GBMfisherVar	β	σ^2
β	$1.85e-05$	$-1.25e-07$
σ^2	$-1.25e-07$	$3.26e-07$

Table 42: *Empirical covariance matrix for the GMM estimates in the GBM model simulation with 50 sets of 500 observations*

GBMcovarGMM	β	σ^2
β	$5.29e-05$	$9.77e-07$
σ^2	$9.77e-07$	$9.09e-07$

Table 43: *Empirical covariance matrix for the MLE estimates in the Brennan-Schwartz model simulation with 50 sets of 500 observations*

BRENNANcovarMLE	α	β	σ^2
α	$2.65e-07$	$-5.92e-06$	$2.19e-08$
β	$-5.92e-06$	$1.88e-04$	$8.53e-07$
σ^2	$2.19e-08$	$8.53e-07$	$4.33e-07$

Table 44: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the Brennan-Schwartz model simulation with 50 sets of 500 observations*

BRENNANfisherVar	α	β	σ^2
α	$1.07e-07$	$-1.78e-06$	$-1.93e-08$
β	$-1.78e-06$	$4.81e-05$	$1.95e-07$
σ^2	$-1.93e-08$	$1.95e-07$	$3.30e-07$

Table 45: *Empirical covariance matrix for the GMM estimates in the Brennan-Schwartz model simulation with 50 sets of 500 observations*

BRENNANcovarGMM	α	β	σ^2
α	$6.31e-07$	$-1.04e-05$	$-6.19e-09$
β	$-1.04e-05$	$3.06e-04$	$3.19e-07$
σ^2	$-6.19e-09$	$3.19e-07$	$5.11e-07$

Table 46: *Empirical covariance for the MLE estimates in the CIR VR model simulation with 50 sets of 500 observations*

CIRVRcovarMLE	σ^2
σ^2	$4.41e-07$

Table 47: *Covariance derived from the observed Fisher information matrix for the MLE estimates in the CIR VR model simulation with 50 sets of 500 observations*

CIRVRfisherVar	σ^2
σ^2	$3.25e-07$

Table 48: *Empirical covariance for the GMM estimates in the CIR VR model simulation with 50 sets of 500 observations*

CIRVRcovarGMM	σ^2
σ^2	$4.3e-07$

Table 49: *Empirical covariance matrix for the MLE estimates in the CEV model simulation with 50 sets of 500 observations*

CEVcovarMLE	β	σ^2	γ
β	$7.21e-04$	$1.09e-05$	$9.32e-05$
σ^2	$1.09e-05$	$1.60e-06$	$1.53e-05$
γ	$9.32e-05$	$1.53e-05$	$2.09e-04$

Table 50: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the CEV model simulation with 50 sets of 500 observations*

CEVfisherVar	β	σ^2	γ
β	$5.42e-05$	$8.94e-07$	$1.73e-05$
σ^2	$8.94e-07$	$6.11e-07$	$8.46e-06$
γ	$1.73e-05$	$8.46e-06$	$2.52e-04$

Table 51: *Empirical covariance matrix for the GMM estimates in the CEV model simulation with 50 sets of 500 observations*

CEVcovarGMM	β	σ^2	γ
β	0.001010	−0.000278	−0.00105
σ^2	−0.000278	0.000410	0.00214
γ	−0.001050	0.002140	0.02090

Table 52: *Empirical covariance matrix for the MLE estimates in the full CKLS model simulation with 50 sets of 500 observations*

CKLScovarMLE	α	β	σ^2	γ
α	$1.06e-07$	$-1.97e-07$	$9.27e-10$	$-5.47e-07$
β	$-1.97e-07$	$6.52e-04$	$8.96e-06$	$7.94e-05$
σ^2	$9.27e-10$	$8.96e-06$	$1.54e-06$	$1.66e-05$
γ	$-5.47e-07$	$7.94e-05$	$1.66e-05$	$2.61e-04$

Table 53: *Covariance matrix derived from the observed Fisher information matrix for the MLE estimates in the full CKLS model simulation with 50 sets of 500 observations*

CKLSfisherVar	α	β	σ^2	γ
α	$1.14e-07$	$-1.59e-07$	$5.46e-09$	$5.11e-07$
β	$-1.59e-07$	$4.33e-05$	$6.57e-07$	$1.29e-05$
σ^2	$5.46e-09$	$6.57e-07$	$5.29e-07$	$6.82e-06$
γ	$5.11e-07$	$1.29e-05$	$6.82e-06$	$2.31e-04$

Table 54: *Empirical covariance matrix for the GMM estimates in the full CKLS model simulation with 50 sets of 500 observations*

CKLScovarGMM	α	β	σ^2	γ
α	$4.11e-06$	$6.62e-06$	$1.67e-07$	$1.77e-05$
β	$6.62e-06$	$8.07e-04$	$-3.28e-05$	$-5.58e-04$
σ^2	$1.67e-07$	$-3.28e-05$	$1.33e-05$	$2.68e-04$
γ	$1.77e-05$	$-5.58e-04$	$2.68e-04$	$6.87e-03$

Bibliography

Tomas Bjork. *Arbitrage Theory in Continuous Time*. Oxford University Press, 3 edition, 2009. URL http://EconPapers.repec.org/RePEc:oxp:obooks:9780199574742.

Fischer Black and Myron S Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–54, May-June 1973. URL https://ideas.repec.org/a/ucp/jpolec/v81y1973i3p637-54.html.

Michael Brennan and Eduardo S. Schwartz. Analyzing convertible bonds. *Journal of Financial and Quantitative Analysis*, 15(04):907–929, 1980. URL http://EconPapers.repec.org/RePEc:cup:jfinqa:v:15:y:1980:i:04:p:907-929_01.

John Y. Campbell. A defense of traditional hypotheses about the term structure of interestrates. Working Paper 1508, National Bureau of Economic Research, November 1984. URL http://www.nber.org/papers/w1508.

George Casella and Roger Berger. *Statistical Inference*. Duxbury Resource Center, June 2001. ISBN 0534243126.

K. C. Chan, G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *JOURNAL OF FINANCE*, 47(3):1209–1227, 1992.

John C. Cox. Notes on option pricing i: constant elasticity of variance diffusions. 1975.

John C. Cox and Stephen A. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3(1-2):145–166, 1976. URL http://EconPapers.repec.org/RePEc:eee:jfinec:v:3:y:1976:i:1-2:p:145-166.

John C Cox, Jonathan E Ingersoll, and Stephen A Ross. An analysis of variable rate loan contracts. *Journal of Finance*, 35(2):389–403, 1980. URL http://EconPapers.repec.org/RePEc:bla:jfinan:v:35:y:1980:i:2:p:389-403.

John C Cox, Jonathan E Ingersoll, and Stephen A Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985. URL http://EconPapers.repec.org/RePEc:ecm:emetrp:v:53:y:1985:i:2:p:385-407.

Rosario Dell'Aquila, Elvezio Ronchetti, and Fabio Trojani. Robust gmm analysis of models for the short rate process. *Journal of Empirical Finance*, 10(3):373–397, May 2003. URL http://www.sciencedirect.com/science/article/B6VFG-480CSHT-1/1/bb906fd52874a39033b677a45c14d3d4.

L. Uri Dothan. On the term structure of interest rates. *Journal of Financial Economics*, 6(1): 59–69, 1978. URL http://EconPapers.repec.org/RePEc:eee:jfinec:v:6:y:1978:i:1:p:59-69.

Robert Faff and Philip Gray. On the estimation and comparison of short-rate models using the generalised method of moments. *Journal of Banking & Finance*, 30(11): 3131–3146, November 2006. URL http://www.sciencedirect.com/science/article/B6VCY-4KGG5WY-4/1/916b31533da10e2a89fe74984d6714f1.

Lars Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–54, 1982. URL http://EconPapers.repec.org/RePEc:ecm:emetrp:v:50:y:1982:i:4:p:1029-54.

Guido W. Imbens, Phillip Johnson, and Richard H. Spady. Information theoretic approaches to inference in moment condition models. Working Paper 186, National Bureau of Economic Research, October 1995. URL http://www.nber.org/papers/t0186.

Robert C. Merton. Theory of Rational Option Pricing. *Bell Journal of Economics*, 4(1):141–183, Spring 1973. URL https://ideas.repec.org/a/rje/bellje/v4y1973ispringp141-183.html.

Oldrich Vasicek. An equilibrium characterization of the term structure, 1977.