Empirical likelihood estimation of interest rate diffusion model

MASTER THESIS

Lukáš Lafférs

COMENIUS UNIVERSITY, BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS
DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS

9.1.9 Economic and Financial Mathematics

doc. Mgr. Marian Grendár PhD.

BRATISLAVA 2009

Odhad parametrov difúzneho modelu úrokovej miery metódou empirickej vierhodnosti

DIPLOMOVÁ PRÁCA

Lukáš Lafférs

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
KATEDRA APLIKOVANEJ MATEMATIKY A ŠTATISTIKY

9.1.9 Ekonomická a Finančná Matematika

doc. Mgr. Marian Grendár PhD.

BRATISLAVA 2009

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics
Department of Applied Mathematics and Statistics
Economic and Financial Mathematics

Empirical Likelihood Estimation of Interest Rate Diffusion model

(Master Thesis)

Lukáš Lafférs

Supervisor: Marian Grendár

Bratislava, 2009

I declare this thesis was written on my own, with the only help provided by my supervisor and the referred-to literature and sources.

. . . .

Acknowledgement

I am grateful to Marian for his excellent leading and insightful comments.

I hereby thank and acknowledge Qingfeng Liu for the source code and discussions, this was a great help for me.

My thanks and appreciation also goes to Zuzka, my family and friends for love and support.

Abstract

Empirical Likelihood (EL) combined with Estimating Equations (EE) provides a modern semi-parametric alternative to classical estimation techniques like Maximum Likelihood Estimation (ML).

In the diploma work we use closed form of conditional expectation and conditional variance of Interest Rate Diffusion model (Vasicek model and Cox-Ingersoll-Ross model) to perform Maximum Empirical Likelihood (MEL) estimation and Maximum Euclidean Empirical Likelihood (EEL) estimation of parameters of the models. Problem of singularity in EEL is resolved my means of Moore-Penrose pseudoinverse. Monte Carlo simulations show that MEL and EEL provide competitive performance to parametric alternatives. Moreover, it turns out that a set of estimating equations employed here provides increased stability compared to recent approach, which utilizes closed form of conditional characteristic function. Calibration of CIR model on European Over-Night Interest Rate Average data by means of MEL and EEL appears to be sufficiently plausible. Steady state mean and standard deviation implied by obtained estimates are consistent with sample mean and standard deviation from the data.

Keywords: Empirical likelihood, Euclidean empirical likelihood, Estimating equations, Vasicek model, CIR model

Abstrakt

Metóda empirickej vierhodnosti spolu s odhadovými rovnicami je modernou alternatívou ku klasickým odhadovacím prístupom ako napríklad metóda maximálnej vierhodnosti.

V tejto diplomovej práci používame explicitné vyjadrenie podmienenej strednej hodnoty a podmienenej disperzie difúzneho modelu úrokových mier (Vašíčkov model a Cox-Ingersoll-Ross model) na odhad parametrov modelov metódou maximálnej empirickej vierodnosti (MEL) a maximálnej euklidovskej empirickej vierohodnosti (EEL). Použitie Moore-Penrose-ovej pseudoinverznej matice vyriešilo problémy so singularitou v EEL. Monte Carlo simulácie ukazujú, že MEL a EEL sú porovnateľné s parametrickými alternatívami. Naviac, ukazuje sa, že použitie daných odhadovacích rovníc zvýšilo stabilitu v porovnaní s moderným prístupom, ktorý využíva explicitné vyjadrenie podmienenej charakteristickej funkcie. Kalibrácia CIR modelu na dátach európskeho priemeru jednodňových úrokových mier použitím MEL a EEL sa javí ako dostatočne hodnoverná. Stredná hodnota a štandardná odchýlka stacionárneho stavu vypočítaná na základe odhadnutých parametrov je konzistentná s výberovou strednou hodnotou a štandardnou odchýlkou z dát.

Kľúčové slová: Empirická vierohodnosť, Euklidovská empirická vierohodnosť, Odhadovacie rovnice, Vašíček model, CIR model

Contents

Chapter 1

Introduction

Estimation of a parameter of interest which affects distribution of measured data usually begins with a specification of a model, i.e., a family of probability distributions parametrized by the parameter. In this setting, the estimation is most commonly performed by the Method of Maximum Likelihood (ML). Resulting ML estimators enjoy excellent asymptotic properties: they are asymptotically unbiased, asymptotically normally distributed and asymptotically efficient, so that the given information is fully exploited. Unfortunately, in many cases the family of distributions that generates data is unknown. In this respect it should be noted that the theory of Quasi Maximum Likelihood studies conditions under which ML estimators retain at least consistency property, when they are erroneously based on gaussian model.

If a researcher refuses to make distributional assumptions then semi-parametric methods can be used. Suppose that the information about the parameters of interest is in form of unbiased moment functions. All what is known is that the expectation of the moment functions, which are functions of data and vector of parameters, is zero. Resulting equations are called (Unbiased) Estimating Equations (EE). EE define a set of probability distributions which form the model. In order to relate the model and data, an empirical analogue of EE is formed by replacing the expectations by the average. If the number of equations is equal to dimension of vector of parameters, which is known as the exactly identified case, then the set of equations can be solved. An estimator which is obtained this way is known as the Method of Moments estimator.

Hansen (1982) extended MM estimation and inference to the over-identified case, where the number of the moment conditions (encoding the information we have in our disposal) is greater than number of parameters. In this case it is not possible to satisfy all the EE at once, but it is meaningful to find a pseudo-solution which is as close to zero in all the EE, as possible. The closeness is measured by a weighted Euclidean distance. The resulting estimator is known as the Generalised Method of Moments (GMM) estimator. If the statistical model is not misspecified (i.e., it contains the true data-generating distribution), then GMM estimator is asymptotically normally distributed, with a known covariance matrix.

In GMM framework every observation is given the equal weight $1/n$, where n is the number of observations. However, it is meaningful to assign unequal weights to the data. If the drawn data are IID, then the likelihood of a random sample is simply the product of all the assigned weights (probabilities). Then, an objective may be to jointly set these probabilistic weights and vector of parameters to maximize the likelihood of the sample subject to empirical Estimating Equations. This way a parametrized probability mass function (pmf) from the model (i.e., the set of parametrized pmf's which are supported by the sample and satisfy empirical EE) with highest value of likelihood is selected. The parametric component of the pmf serves as an estimator of the parameter of interest. The estimator is known as the Maximum Empirical Likelihood estimator. Note that the data itself chooses which observation should be given a higher or lower probability (weight). The Empirical Likelihood approach combines reliability of semi-parametric models with efficiency of likelihood based methods.

Chapter 2

Empirical Likelihood

2.1 Empirical Likelihood for the mean

In this section we will explain empirical likelihood approach to estimation and inference in more technical terms. Presentation is based on Owens' book [Owe01] and an article [QL94] by Qin and Lawless. In his first explanation of EL Owen ([Owe01], Chapter 2) focuses on confidence intervals for mean of a random variable, which is sufficient also for our purposes here. The presentation will be restricted to the discrete case, because the idea of empirical likelihood is there very clear and easily understandable.

First of all, we will show, that given no information, Empirical cumulative distribution function (ECDF) is the most likely distribution or nonparametric maximum likelihood estimate (NPMLE).

Let X be random variable, the cumulative distribution function is $F(x) = P(X \leq x)$ for $-\infty < x < \infty$. We also denote $F(x-) = P(X < x)$ and $P(X = x) = F(x) - F(x-)$. Let $1_{A(x)}$ stands for the function which is equal to 1 if proposition $A(x)$ holds, otherwise it is 0.

Definition 1. *Let $X_1, X_2, ..., X_n \in R$ are random variables. Then the empirical cumulative distribution function of X is*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i \leq x},$$

for $-\infty < x < \infty$.

Definition 2. *Given $X_1, X_2, ..., X_n \in \mathbb{R}$, which are assumed independent with common CDF F_0, the nonparametric likelihood of the CDF F is*

$$L(F) = \prod_{i=1}^{n} \left(F(X_i) - F(X_i-)\right).$$

Next theorem shows that nonparametric likelihood is maximized by the ECDF.

Theroem 3. *Let $X_1, X_2, ..., X_n \in \mathbb{R}$ be independent random variables with common CDF F_0. Let F_n be their ECDF and F be any CDF. If $F \neq F_n$, then $L(F) < L(F_n)$.*

Proof. Let $z_1, z_2, ..., z_m$ be distinct values in $\{X_1, X_2, ..., X_n\}$. Let n_j be the number of X_i that are equal to z_j. Denote $p_j = F(z_j) - F(z_j-)$ and $\hat{p}_j = n_j/n$. If $p_j = 0$ for any $j = 1, ..., m$ then $L(F) = 0$ and inequality holds. We also use that $\log(x) \leq x - 1$. Suppose that $p_j > 0$ for all $j = 1, ..., m$. So

$$\begin{aligned}
\log\left(\frac{L(F)}{L(F_n)}\right) &= \log\left(\frac{\prod_{j=1}^{m} p_j^{n_j}}{\prod_{j=1}^{m} \hat{p}_j^{n_j}}\right) \\
&= \sum_{j=1}^{m} n_j \log\left(\frac{p_j}{\hat{p}_j}\right) \\
&= n \sum_{j=1}^{m} \hat{p}_j \log\left(\frac{p_j}{\hat{p}_j}\right) \\
&< n \sum_{j=1}^{m} \hat{p}_j \left(\frac{p_j}{\hat{p}_j} - 1\right) = 0,
\end{aligned}$$

therefore $L(F) < L(F_n)$.

There is a strong intuition behind this result. If there is no information about the underlying distribution, other than the observed random sample, then ECDF is the most likely distribution generating the data.

Now we define the empirical likelihood ratio (ELR), which is a basis for parameter estimates, tests and confidence intervals in EL framework.

Definition 4. *For distribution F, we define empirical likelihood ratio $R(F)$ as*

$$R(F) = \frac{L(F)}{L(F_n)}.$$

Let there be a random sample $x_1, x_2, ..., x_n$ of size n from distribution F. Let w_i denote weight that is assigned to i-th observation; I.E., $w_i = F(x_j) - F(x_j-)$. It can be easily shown, that

$$R(F) = \prod_{i=1}^{n} nw_i,$$

even if x-s are not distinct.

Assume that there is an interest in a parameter $\theta = T(F)$, where T is some function of distributions F.

Definition 5. *For distribution $F \in \mathcal{F}$, we define profile empirical likelihood ratio $R(F)$ as*

$$\mathcal{R}(\theta) = \sup\{R(F)|T(F) = \theta, F \in \mathcal{F}\}.$$

Once we know the distribution of ELR, we are ready to make statistical inference. Therefore, we proceed with Empiricial likelihood theorem (ELT), proof can be found in [Owe01]. It is a non-parametric analogue of Wilks's theorem. ELT concerns the population mean of some distribution. We will extend this to more general cases using estimating functions later on.

Theroem 6. *Let $X_1, X_2, ..., X_n$ be independent random variables with common distribution F_0. Let $\mu_0 = E(X_i)$, and suppose that $0 < Var(X_i) < \infty$. Then $-2\log(\mathcal{R}(\mu_0))$ converges in distribution to $\chi^2_{(1)}$ as $n \to \infty$.*

Based on this information we can construct confidence intervals for mean.

$$R(\mu) = \max_{w_i} \left\{ \sum_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i X_i = \mu, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right\}$$

$$\{\mu | \mathcal{R} \geq r_0\} = \left\{ \sum_{i=1}^{n} w_i X_i | \sum_{i=1}^{n} nw_i \geq r_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right\}.$$

Therefore confidence interval is the set of all μ that are plausible enough, more than the threshold r_0, which depends on our significance level.

2.2 Estimating equations and EL

This part presents estimating equations (EE) linked with empirical likelihood as extremely flexible tool for parameters estimation; cf. [QL94], [Owe01], [MJM00].

Note that with this framework we are also able to incorporate the prior information about the underlying distribution.

Suppose that information about distribution F is in form of unbiased estimating functions $m(X, \theta)$. Let $X \in \mathbb{R}^d$ be a random variable, $\theta \in \mathbb{R}^p$ vector of parameters of interest and vector-valued function $m(X, \theta) \in \mathbb{R}^s$ such that

$$E(m(X, \theta)) = 0.$$

There are three different cases that can occur, we will focus on the last one.

- Under-identified case $p > s$ - in this case do not sufficient information to identify θ, but we can reduce the size of the space of parameters.
- Just-identified case $p = s$ - Method of Moments can be used since the number of the restrictions is the same as the number of parameters of interest
- Over-identified case $p < s$ - this is of crucial importance in Econometrics, where several methods have been developed to deal with this case; e.g. Generalized Method of Moments or Empirical Likelihood.

Few examples of estimating equations

- $m(X, \theta) = X - \theta$ for estimation of the mean
- for estimation of the variance $\theta = (\mu, \sigma)$, $m_1(X, \theta) = X - \mu$, $m_2(X, \theta) = (X - \mu)^2 - \sigma^2$
- we require that α-quantile is 4:
  $m(X, \theta) = 1_{X \leq 4} - \alpha$
- for estimation of the mean of a symmetric distribution $\theta = \mu$, $m_1(X, \theta) = X - \mu$, $m_2(X, \theta) = 1_{X \leq \mu} - 0.5$

EE can cover broad type of information. Note that by selecting $m(X, \theta) = \frac{\partial}{\partial \theta} \log f(x, \theta)$, maximum likelihood estimator can be obtained, whenever it is defined by the score equations.

In this case profile empirical likelihood is in form

$$\mathcal{R}(\theta) = \max_{w_i} \left\{ \prod_{i=1}^{n} n w_i | \sum_{i=1}^{n} w_i m(X_i, \theta) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right\}.$$

Following theorem is natural extension of 6.

Theroem 7. *Let $X_1, X_2, ..., X_n$ be independent random variables with common distribution F_0. For $\theta \in \Theta \subseteq \mathbb{R}^p$, and $X \in \mathbb{R}^d$, let $m(X, \theta) \in \mathbb{R}^s$. Let $\theta_0 \in \Theta$ be such that $Var(m(X_i, \theta))$ is finite and has rank $q > 0$. If θ_0 satisfies $E(m(X_i, \theta)) = 0$, then $-2\log(\mathcal{R}(\mu_0))$ converges in distribution to $\chi^2_{(1)}$ as $n \to \infty$.*

Proof is straightforward from 6.

2.3 Maximum Empirical Likelihood Estimator

Empirical likelihood is method of statistical inference, however in this work we are interested more in estimation than in tests and confidence regions by EL. This section draws on [Owe01] and [QL94].

Maximum Empirical Likelihood Estimator (MEL) is defined as the value of θ that maximizes the profile empirical likelihood

$$\hat{\theta}_{EL} = \arg\max_{\theta \in \Theta} \mathcal{R}(\theta).$$

We can see that MEL is result of two interdependent optimization problems.

For the inner loop, which is maximization over weights (probabilities) w_i that we assign to particular observations, we can solve dual problem.

Since log-transformation is monotonous, let us rewrite the inner loop using Lagrangian in the following form

$$G = \sum_{i=1}^{n} \log(nw_i) - n\lambda' \left(\sum_{i=1}^{n} w_i m(X_i, \theta) \right) - \gamma \left(\sum_{i=1}^{n} w_i - 1 \right).$$

Note that space of vectors of weights is convex set $S_{n-1} = \{(w_1, ..., w_n) | \sum_{i=1}^{n} w_i = 1, w_i \geq 0\}$ and log-transformed objective function is strictly concave. We solve FOC for this optimization problem

$$\begin{aligned} \frac{\partial G}{\partial w_i} &= \frac{1}{w_i} - n\lambda' m(X_i, \theta) - \gamma = 0 \\ \sum_{i=1}^{n} w_i \frac{\partial G}{\partial w_i} &= n - \gamma = 0 \Rightarrow \gamma = n, \end{aligned}$$

so

$$w_i = \left(\frac{1}{n} \right) \frac{1}{1 + \lambda' m(X_i, \theta)},$$

with restriction

$$0 = \sum_{i=1}^{n} w_i m(X_i, \theta) \quad = \quad \left(\frac{1}{n}\right) \sum_{i=1}^{n} \frac{1}{1 + \lambda' m(X_i, \theta)} m(X_i, \theta),$$

therefore we can think about the lagrange multipliers λ as a function of θ, $\lambda = \lambda(\theta)$. In order to ensure that $0 \leq w_i$, for fixed θ, vector λ has to satisfy

$$\forall i: \quad 1 + \lambda' m(X_i, \theta) > 0. \tag{2.1}$$

We omitted case in which $w_i = 0$, because this cannot be result of our minimization, since the objective function approaches $-\infty$, so $\forall i: \ w_i \neq 1$ and we can use strict inequality sign in (2.1).

If we substitute this w_i into $\log R(F)$ we get

$$\log R(F) = -\sum_{i=1}^{n} \log(1 + \lambda' m(X_i, \theta)) \equiv L(\lambda).$$

In this dual problem we seek minimum of $L(\lambda)$ over λ. So we have changed maximization over n-weights subject to $d+1$ constraints to minimization over d variables λ subject to n constraints (2.1), note that we eliminated $\gamma = n$.

Now we face the following constrained optimization problem

$$\min_{\lambda \in \mathbb{R}^d} L(\lambda) \quad s.t. \ \ \forall i: \ 1 + \lambda' m(X_i, \theta) > 1/n.$$

Owen in his book [Owe01] provides a trick which change this problem into the unconstrained optimization. Let us define a pseudo-logarithm function

$$\log_*(z) \quad = \quad \begin{cases} \log(z), & \text{if } z \geq 1/n, \\ \log(1/n) - 1.5 + 2nz - (nz)^2/2, & \text{if } z \leq 1/n, \end{cases}$$

$$L_*(\lambda) \quad \equiv \quad -\sum_{i=1}^{n} \log_*(1 + \lambda' m(X_i, \theta)). \tag{2.2}$$

The function is unchanged for arguments greater than $1/n$ and it is quadratic if argument is less than $1/n$, which corresponds to $w_i > 1$ and therefore will not affect optimiziation. This transformation may significantly reduce the computational burden.

Using convex duality theorem, we have thus obtained another optimization problem that leads to MEL

$$\hat{\theta}_{EL} = \arg\max_{\theta\in\Theta} \max_{w_i\in S_{n-1}} \left\{ \sum_{i=1}^{n} \log(nw_i) | \sum_{i=1}^{n} w_i m(X_i,\theta) = 0 \right\}, \quad (2.3)$$

$$\Leftrightarrow$$

$$\hat{\theta}_{EL} = \arg\max_{\theta\in\Theta} \min_{\lambda\in\mathbb{R}^d} \sum_{i=1}^{n} L_*(\lambda),$$

$$\Leftrightarrow$$

$$\hat{\theta}_{EL} = \arg\min_{\theta\in\Theta} \max_{\lambda\in\mathbb{R}^d} \sum_{i=1}^{n} \log_*(1+\lambda' m(X_i,\theta)). \quad (2.4)$$

Equation (2.4) will be the basis for computational part of this thesis.

2.4 Empirical Likelihood as GMC

Previous sections describe Empirical Likelihood in discrete case, where it is a very intuitive concept. To extend EL into continuous case, we have to use a more theoretical framework. This seems to be necessary to avoid a discrete-continuous conflict, which results from the fact that we are optimizing over discrete distributions subject to constraints involving continuous pdf's. This section provides a justification for use of EL also in the continuous case. It is based on convex duality and subsequent replacement of the original measure μ by its sample counterpart μ_n (ECDF). Bickel, Klassen, Ritov and Wellner [BKRW93] pointed out, that Maximum Likelihood estimate may be subsummed under Generalized Minimum Contrast (GMC) estimation procedure. Kitamura in his comprehensive study [Kit06] showed how EL can be included into GMC scheme. We will follow his argument and notation. Suppose we have a convex function ϕ which measures a divergence between two probability measures P and Q

$$D(P,Q) = \int \phi\left(\frac{dP}{dQ}\right) dQ. \quad (2.5)$$

We denote x as IID p-observations from the true probability measure μ, M is the set of all possible probability measures on $\mathbb{R}^p$ and

$$\mathcal{P}(\theta) = \left\{ P \in M : \int m(x,\theta) dP = 0 \right\}.$$

Let $\mathcal{P} = \cup_{\theta\in\Theta}\mathcal{P}(\theta)$ is the set of all measures that are consistent with the moment restriction. Statistical model $\mathcal{P}$ is correctly specified if $\mu \in \mathcal{P}$.

Our goal is to find the value of parameter θ which solves the GMC optimization

$$\inf_{\theta\in\Theta} \rho(\theta,\mu), \quad \rho(\theta,\mu) = \inf_{P\in\mathcal{P}(\theta)} D(P,\mu).$$

Note that inner loop of this optimization problem contains a variational problem and therefore it is difficult to compute. We will show in detail how Lagrange duality can be used to transform this problem to finite dimensional unconstrained convex optimization problem. If we set $p = \frac{dP}{d\mu}$, so $D(P,\mu) = \int \phi(p)d\mu$, then the primal problem is infinite dimensional optimization. Note that since we are in the inner loop, parameter θ remains fixed

$$v(\theta) = \inf_{p\in\mathcal{P}} \int \phi(p)d\mu \quad s.t. \quad \int m(x,\theta)pd\mu = 0, \int pd\mu = 1. \tag{2.6}$$

We write down Lagrangian

$$\begin{aligned}
L(p,\lambda,\gamma) &= \int \phi(p)d\mu - \lambda' \int m(x,\theta)pd\mu - \gamma\left(\int pd\mu - 1\right), \\
L(p,\lambda,\gamma) &= \gamma + \int (\phi(p) - \lambda' m(x,\theta)p - \gamma p)d\mu, \\
\inf_{p\in\mathcal{P}} L(p,\lambda,\gamma) &= \gamma + \int \inf_{p\in\mathcal{P}}[\phi(p) - (\lambda' m(x,\theta) + \gamma)p]d\mu
\end{aligned}$$

and since [1]

$$\begin{aligned}
f^*(y) &= \sup_x [xy - f(x)], \\
-f^*(y) &= \inf_x [f(x) - xy],
\end{aligned}$$

the objective function in dual problem may be rewritten as

$$\inf_{p\in\mathcal{P}} L(p,\lambda,\gamma) = \gamma - \int \phi^*(\gamma + \lambda' m(x,\theta))d\mu,$$

so we obtained computationally more convenient dual problem

$$v^*(\theta) = \max_{\gamma\in\mathbb{R},\lambda\in\mathbb{R}^q} \left[\gamma - \int \phi^*(\gamma + \lambda' m(x,\theta)d\mu\right]. \tag{2.7}$$

[1]Function $f^*(y)$ is convex conjugate of f

Note that by the Fenchel duality theorem (Borwein and Lewis (1991), [BL91]) $v(\theta) = v^*(\theta)$. The probability measure that solve the optimization problem (2.6) is in form

$$\tilde{p} = (\phi')^{-1}(\tilde{\gamma} + \tilde{\lambda}' m(x, \theta)), \tag{2.8}$$

where $\tilde{\gamma}, \tilde{\lambda}$ is the solution of dual problem (2.7).

Therefore we face the following problem

$$\inf_{\theta \in \Theta} v^*(\theta) = \inf_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}, \lambda \in \mathbb{R}^q} \left[\gamma - \int \phi^*(\gamma + \lambda' m(x, \theta) d\mu \right].$$

Here we take empirical measure μ_n as valid approximation of true measure μ. This will lead to sample version of GMC problem

$$\min \frac{1}{n} \sum_{i=1}^{n} \phi(np_i), \quad s.t. \quad \sum_{i=1}^{n} p_i m(x, \theta) = 0, \sum_{i=1}^{n} p_i = 1, \theta \in \Theta.$$

So the GMC estimator for θ is defined as

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \inf_{p_i, p \in S_{n-1}, \sum_{i=1}^{n} p_i m(x_i, \theta) = 0} \frac{1}{n} \sum_{i=1}^{n} \phi(np_i).$$

We can use Lagrange duality to form computationally convenient but equivalent dual representation of GMC estimator

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}, \lambda \in \mathbb{R}^q} \left[\gamma - \sum_{i=1}^{n} \phi^*(\gamma + \lambda' m(x, \theta) \right].$$

Now different choices of function $\phi(x)$ yield different estimators. If we set $\phi(x) = -\log(x)$ the GMC estimator is empirical likelihood estimator

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \inf_{p_i, p \in S_{n-1}, \sum_{i=1}^{n} p_i m(x_i, \theta) = 0} \frac{1}{n} \sum_{i=1}^{n} -\log(np_i)$$

or

$$\begin{aligned} \hat{\theta} &= \arg\min_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}, \lambda \in \mathbb{R}^q} \left[\gamma + 1 + \frac{1}{n} \sum_{i=1}^{n} \log(-\gamma - \lambda' m(x_i, \theta)) \right] \\ &= \arg\min_{\theta \in \Theta} \max_{\lambda \in \mathbb{R}^q} \left[\frac{1}{n} \sum_{i=1}^{n} \log(1 + \lambda' m(x_i, \theta)) \right] \end{aligned} \tag{2.9}$$

in convenient dual representation.

2.5 Euclidean Empirical Likelihood

Quadratic choice of $\phi(x) = \frac{1}{2}(x^2 - 1)$ in (2.5) implies so called Euclidean Likelihood [Owe01], where the nonparametric likelihood function is in its position of objective function replaced by the euclidean distance $\frac{1}{2n}\sum_{i=1}^{n}(np_i - 1)^2$; cf. [BC98]. For euclidean distance an explicit solution of the inner loop exists that significantly reduces the computational burden. Euclidean likelihood is in fact quadratic approximation of empirical likelihood, as is shown below.

The EL problem is to maximize

$$l(\theta) = \min_{p_i, p \in S_{n-1}, \sum_{i=1}^{n} p_i m(x_i,\theta)=0} \left[-\sum_{i=1}^{n} \frac{1}{n} \log(np_i) \right] = \left[\frac{1}{n} \sum_{i=1}^{n} \log(1 + \lambda' m(x_i, \theta)) \right] \tag{2.10}$$

over θ, where $\lambda = \arg\max_{\lambda \in D} \sum_{i=1}^{n} \log(1+\lambda' m(x_i, \theta))$ and $D = \{\lambda | 1+\lambda' m(x_i, \theta) > 0\}$. For short $m(x_i, \theta)$ the first argument of m is ommited.

Expand $l(\theta)$ using Taylor series near $p_i = \frac{1}{n}$

$$l(\theta) = \sum_{i=1}^{n} \left(p_i - \frac{1}{n} \right) + \frac{1}{2n} \sum_{i=1}^{n} n^2 \left(p_i - \frac{1}{n} \right)^2 + \dots \tag{2.11}$$

The dominant term of $l(\theta)$ is $\frac{1}{2n}\sum_{i=1}^{n}(np_i - 1)^2$.

Euclidean Empirical Likelihood is defined as
$eel(\theta) = \min_{p_i, p \in S_{n-1}, \sum_{i=1}^{n} p_i m(x_i,\theta)=0} \sum_{i=1}^{n}(np_i - 1)^2$.

Thanks to its quadratic form an explicit solution for the the inner loop exists. Indeed, since $\sum_{i=1}^{n}(np_i - 1)^2 = n^2 \sum_{i=1}^{n} p_i^2 - n$, our optimization problem is

$$\max_{p_i} \sum_{i=1}^{n} p_i^2, \quad s.t. \quad \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i m_i(\theta) = 0 \quad \forall i : p_i > 0.$$

Writing down Lagrangian, where α_j are Lagrange multipliers yields

$$L(p, \alpha) = \sum_{i=1}^{n} p_i^2 + \alpha_0 \left(\sum_{i=1}^{n} p_i - 1 \right) + \alpha' \sum_{i=1}^{n} p_i m_i(\theta).$$

From the first order conditions it can be seen that

$$p_i = -\left(\alpha_0 + \sum_{j=1}^{q} \alpha_j m_{ij}(\theta) \right). \tag{2.12}$$

Denote $\bar{\alpha}' = (\alpha_0, \alpha_1, ..., \alpha_q)$,$V_j = \sum_{i=1}^n m_{ij}(\theta)$,$V' = (V_1, ..., V_q)$,$R = (R_{jj'})_{q\times q}$ and $R_{jj'} = \sum_{i=1}^n m_{ij}(\theta) m_{ij'}(\theta)$.

Conditions

$$\sum_{i=1}^n p_i = 1 \quad \sum_{i=1}^n p_i m_i(\theta) = 0$$

can be rewritten in matrix form as (there we introduce e_1 and B)

$$e_1 = (1\ 0\ \dots\ 0)' = -\frac{1}{2}\begin{pmatrix} n & V' \\ V & R \end{pmatrix}\alpha = -\frac{1}{2}B\alpha.$$

Rewriting (2.12) yields

$$p_i = (1, m_i'(\theta))B^{-1}e_1 = \frac{1}{n} + \frac{1}{n}\left(\frac{1}{n}V - m_i(\theta)\right)' H^{-1}V,$$

where $H = R - n^{-1}VV^T$.

Using simple algebra it can be shown [BC98], that

$$eel(\theta) = V'H^{-1}V,$$

so Maximum Euclidean Empirical Likelihood estimator is

$$\hat{\theta}_{EEL} = \arg\min_{\theta\in\Theta} eel(\theta) = \arg\min_{\theta\in\Theta} V'H^{-1}V.$$

2.6 Asymptotic properties of EL with estimating equations

Qin and Lawless [QL94] obtained the main asymptotic results for EL with EE. We state Theorem 1 and result of Theorem 3 from [QL94] (two main asymptotic results), saying about asymptotic normality and efficiency in the sense of Bickel, Klaassen, Ritov and Wellner [BKRW93].

Let us denote $\hat{\theta}_{EL}$

$$\hat{\theta}_{EL} = \arg\min_{\theta\in\Theta}\max_{\lambda\in\mathbb{R}^d}\sum_{i=1}^n \log_*(1 + \lambda' m(X_i, \theta))$$

and $\hat{\lambda}$ as the value of $\lambda \in R^d$ that optimizes the inner loop.

Theroem 8. *Assume that $E[m(x,\theta_0)m'(x,\theta_0)]$ is positive definite, $\partial m(x,\theta)/\partial\theta$ and $\partial^2 m(x,\theta)/\partial\theta\partial\theta'$ are continuous in θ in a neighborhood of the true value θ_0, $\| \partial m(x,\theta)/\partial\theta\|^2$, $\| \partial^2 m(x,\theta)/\partial\theta\partial\theta'\|$ and $\|m(x,\theta)\|^3$ are bounded by some integrable function $G(x)$ in this neighborhood, and the rank of $\| \partial m(x,\theta)/\partial\theta\|$ is p. Then*

$$\sqrt{n}(\hat{\theta}_{EL}-\theta_0) \to N(0,V), \quad \sqrt{n}(\hat{\lambda}-0) \to N(0,U),$$

$$\sqrt{n}(F_n - F(x)) \to N(0,W(x)),$$

where

$$F_n = \sum_{i=1}^{n} \hat{p}_i 1_{(x_i<x)},$$

$$\hat{p}_i = \left(\frac{1}{n}\right)\frac{1}{1+\hat{\lambda}' m(x_i,\hat{\theta}_{EL})},$$

$$V = \left[E\left(\frac{\partial m}{\partial\theta}\right)'(Emm')^{-1}E\left(\frac{\partial m}{\partial\theta}\right)\right]^{-1},$$

$$W(x) = F(x)(1-F(x)) - B(x)UB'(x),$$

$$B(x) = E\left[m(x_i,\theta_0 1_{(x_i<x)})\right],$$

$$U(x) = [E(mm')]^{-1}\left\{I - E\left(\frac{\partial m}{\partial\theta}\right)VE\left(\frac{\partial m}{\partial\theta}\right)'[E(mm')]^{-1}\right\}$$

and $\hat{\theta}_{EL}$ and $\hat{\lambda}$ are asymptotically uncorelated.

Another asymptotic result from [QL94] is the fact that under conditions of Theorem 8 and some other mild regularity conditions MEL's for both the parameters and the distribution function are asymptotically efficient in the sense of Bickel, Klaassen, Ritov and Wellner [BKRW93].

[2] $\| A\|$ denotes the euclidean norm of A.

Chapter 3

Interest Rate Models

This chapter introduces three interest rate models, for which we are interested in parameters estimation. Evolution of interest rates is driven by the time and random components.

3.1 Short term interest rate

Riskless bond is the base of the pricing of all financial derivatives. The prices of the riskless bonds on the market determines the term structure of the interest rates $R(t,T)$ [eSM09]

$$P(t,T) = e^{-R(t,T)(T-t)}, \tag{3.1}$$

where $P(t,T)$ is the price of the bond at time t which pays one unit at the maturity time T and $R(t,T)$ is the corresponding interest rate

$$R(t,T) = -\frac{\log P(t,T)}{T-t}.$$

Then short term interest rate or short rate is defined as the starting point of the term structure of the interest rate

$$r_t = \lim_{t \to T} R(t,T).$$

Term structure of interest rates is usually named by the capital of the country e.g. LIBOR (London Interbank Offered Rate), PRIBOR (Praha Interbank Offered Rate) or BRIBOR (Bratislava Interbank Offered Rate). Depending on the maturity, term structure of interest rates can have different shapes [eSM09].

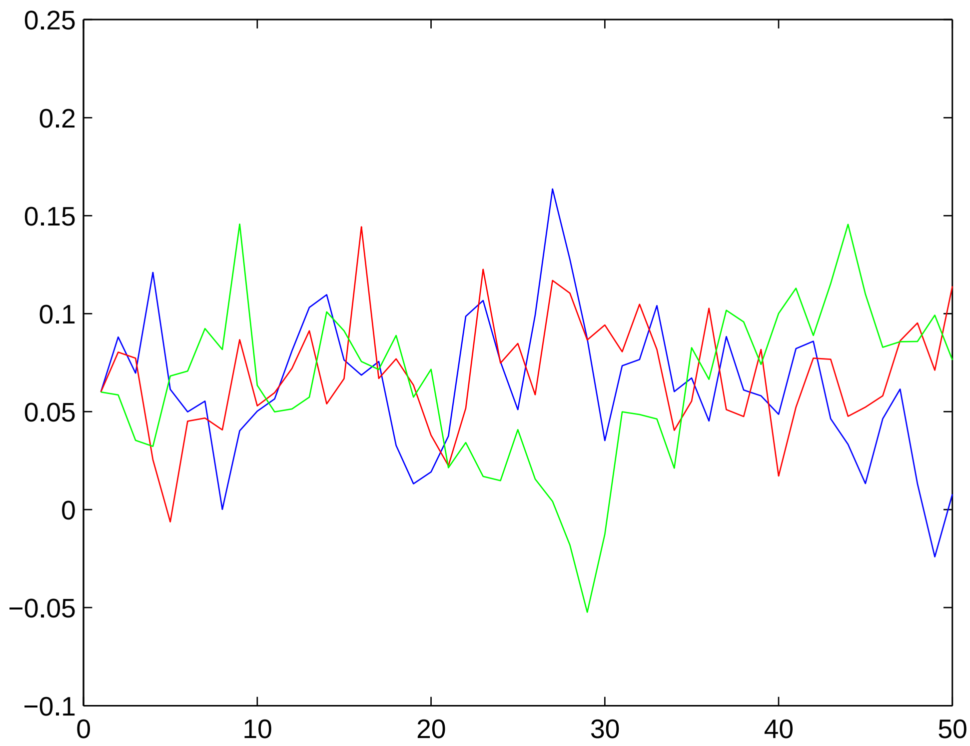
3.2 Vasicek model

Vasicek model for instantenous interest rates was introduced by Oldrich Vasicek in 1977 [Vas77]. This process has interesting mean reversion property, which causes that interest rate cannot rise to infinity

$$dr_t = (\delta - \kappa r_t)dt + \sigma dW_t.$$

Given the information about the interest rate in time t-r_t, using Itô lemma one can derive the conditional density, which is normal with the following mean and variance (we set time increment $\tau = 1$):

$$\begin{aligned} E(r_{t+1}|r_t, t>0) &= r_t e^{-\kappa} + \frac{\delta(1-e^{-\kappa})}{\kappa}, \\ Var(r_{t+1}|r_t, t>0) &= \frac{\sigma^2(1-e^{-2\kappa})^2}{2\kappa}. \end{aligned} \tag{3.2}$$



Three simulations of Vasicek model $\delta = 0.03$, $\kappa = 0.5$ and $\sigma = 0.0367$. Vasicek model allows r_t to drop below zero.

3.3 Cox-Ingersol-Ross model

Cox-Ingersol-Ross model (CIR) is short-term interest rate model which was developed in 1985 [CIR85]. The economic theory behind this model involves anticipations, risk aversion, investment alternatives and preferences about the timing of consumption. All these factors determine bond prices [CIR85].

Random component in CIR is represented by the Wiener process increment

$$dr_t = (\delta - \kappa r_t)dt + \sigma\sqrt{r_t}dW_t. \tag{3.3}$$

Parameters in this equation are positive, $\delta > 0, \kappa > 0, \sigma > 0$.

CIR captures two important properties of real short-term interest rate dynamics:

- Mean reversion - interest rate tends to fluctuate over long-run trend δ/κ,
- Volatility is not constant, but increases with interest rate r_t.

In order to estimate the parameters of this model by maximum likelihood method, it is necessary to find conditional density function. This was done in [CIR85]:

$$f(r_{t+\tau}|r_t, t \geq 0, \tau > 0) = ce^{-c(u+r_{t+\tau})}\left(\frac{r_{t+\tau}}{u}\right)^{q/2} I_q(2c\sqrt{ur_{t+\tau}}), \tag{3.4}$$

$$c = \frac{2\kappa}{\sigma^2(1-e^{-\kappa\tau})},\ u = r_t e^{\kappa\tau},\ q = \frac{2\delta}{\sigma^2 - 1},$$

where $I_q(\cdot)$ denotes the Bessel function of the first kind of order q. If we set $\tau = 1$, then the conditional mean and conditional variance are

$$\begin{aligned} E(r_{t+1}|r_t, t > 0) &= r_t e^{-\kappa} + \frac{\delta(1-e^{-\kappa})}{\kappa}, \\ Var(r_{t+1}|r_t, t > 0) &= \frac{r_t\sigma^2 e^{-\kappa}(1-e^{-\kappa})}{\kappa} + \frac{\sigma^2\delta(1-e^{-\kappa})^2}{2\kappa^2}. \end{aligned} \tag{3.5}$$

If we want to simulate data that follows CIR dynamics, we use the fact that $2cr_{t+1}|r_t \sim \chi^2(2q+2, 2cu)$. Zhou [Zho01] recommends to discard some data from the beginning so that the interest rate series can "forget" its initial value r_0.

The steady state density of the CIR model is

$$f(r_0) = \frac{w^v}{\Gamma(v)} r_0^{v-1}\exp(-wr_0), \tag{3.6}$$

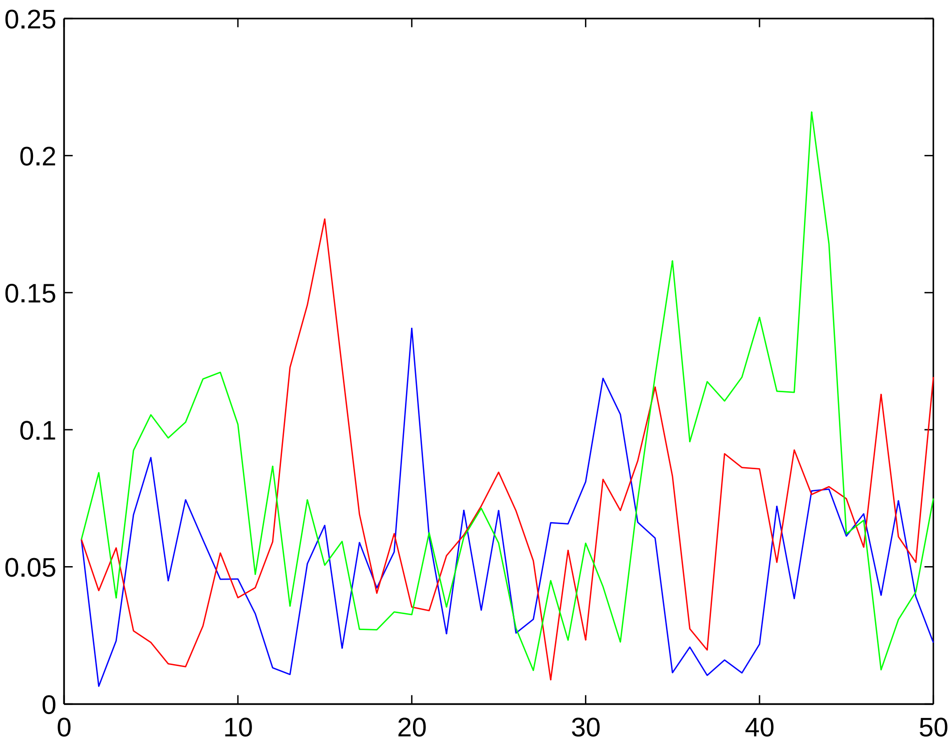
where

$$v = \frac{2\delta}{\sigma^2} \quad , \quad w = \frac{2\kappa}{\sigma^2}.$$

Mean and variance of this marginal density are

$$E(r_0) = \frac{\delta}{\kappa}, \tag{3.7}$$

$$Var(r_0) = \frac{\sigma^2\delta}{2\kappa^2}. \tag{3.8}$$



Three simulations of CIR model $\delta = 0.03$, $\kappa = 0.5$ and $\sigma = 0.15$. Decreasing r_t diminishes the volatility.

3.4 Vasicek Exponential Jump model

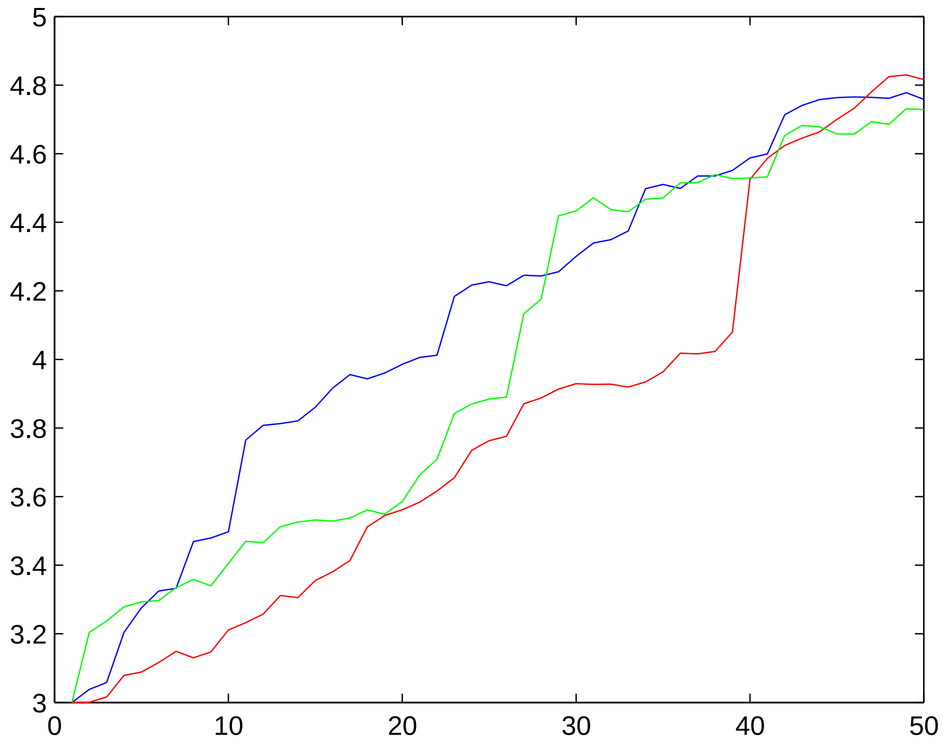
Vasicek Exponential Jumps model (VEJ) was proposed by Das and Foresi [DF96] (1996). In the model, size of random jumps is exponentially distributed, distribution of direction of the jumps is binomial and the frequency of these jumps is represented by the Poisson increment. These jumps can model higher order moments of the conditional distribution. Note that for Vasicek model (3.2) conditional density is normally distributed, therefore skewness and kurtosis are equal to 0 and 3, respectively. In practice, however, normality of data is usually not the case; [DF96].

Short term interest rate in VEJ satisfies the following stochastic differential equation

$$dr_t = (\delta - \kappa r_t)dt + \sigma dW_t + J_t dN_t,$$

$|J_t| \sim \text{Exp}(\alpha)$, $sign(J_t) \sim \text{Bin}(\beta)$, $N_t \sim \text{Poi}(\lambda)$.

Despite the fact that the conditional density cannot be obtained in explicit form, the Conditional Characteristic Function [DF96] can be obtained explicitly, and consequently utilized for estimation of the model parameters; cf. [LN08]. More details can be found in chapter 6.



Three simulations o VEJ CIR model $\theta = 0.02949$, $\kappa = 0.00283$, $\sigma = 0.022$, $\alpha = 0.1$, $\beta = 1$ and $\lambda = 0.28846$. Note that $\beta = 1$ causes that there are only upward jumps.

Chapter 4

Small sample properties of different estimators of parameters of Vasicek and CIR models

Small sample properties of possible estimators of of parameters of Vasicek and CIR models can be obtained by means of a Monte Carlo study. Zhou [Zho01] performed an extensive MC comparison of Efficient Method of Moments (EMM) with other estimation methods. Zhou has not included into studied semi-parametric methods the Empirical likelihood method, which we do here.

4.1 Estimators

4.1.1 Maximum Likelihood Estimator

Since the underlying conditional distribution are known, (3.2 and 3.4) the information can be fully utilized in parameter estimation by ML method, which maximizes the log-likelihood function.

Suppose we observe $x_1, ..., x_n$ from the distribution with probability density function $f(x|\theta)$

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \log f(x_i|\theta).$$

We will be able to compare our estimators with the asympotical efficient ML, which will be the base of our comparison.

4.1.2 Quasi-Maximum Likelihood Estimator

QML assumes that the distribution is normal with the conditional mean and conditional variance given at (3.2 and 3.5)

$$\hat{\theta}_{QML} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log f(x_i|\theta), \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)}{2\sigma^2}}.$$

In the case of Vasicek model, the conditional density is normal, hence QML and ML coincide.

4.1.3 Maximum Empirical Likelihood Estimator

To form a model, estimating equations from [Zho01]:

$$m_t(\theta) = \begin{bmatrix} r_{t+1} - E(r_{t+1}|r_t) \\ r_t \left[r_{t+1} - E(r_{t+1}|r_t) \right] \\ V(r_{t+1}|r_t) - \left[r_{t+1} - E(r_{t+1}|r_t) \right]^2 \\ r_t \left\{ V(r_{t+1}|r_t) - \left[r_{t+1} - E(r_{t+1}|r_t) \right]^2 \right\} \end{bmatrix}. \quad (4.1)$$

were used. The estimating equations utilize the explicit form of conditional mean and conditional variance shown in section 3.3. Unconditional moments are constructed from the conditional ones. We know that

$$E(r_{t+1}|r_t) = 0 \Rightarrow E(r_{t+1}|g(r_t)) = 0.$$

So that choices $g(x) = 0$ and $g(x) = x$ lead to estimating equations (4.1).

MEL estimator (2.10) is given as ($\theta = [\delta, \kappa, \sigma]$)

$$\hat{\theta}_{EL} = \arg\min_{\theta \in \Theta} \max_{\lambda \in \mathbb{R}^q} \left[\frac{1}{n} \sum_{i=1}^{n} \log_*(1 + \lambda' m_i(\theta)) \right].$$

4.1.4 Maximum Euclidean Empirical Likelihood Estimator

The detailed derivation of EEL is in section 2.5

$$\hat{\theta}_{EEL} = \arg\min_{\theta \in \Theta} eel(\theta) = \arg\min_{\theta \in \Theta} V' H^{-1} V,$$

where $V_j = \sum_{i=1}^{n} m_i j(\theta)$,$V' = (V_1, ..., V_q)$,$R = (R_{jj'})_{q\times q}$ and $R_{jj'} = \sum_{i=1}^{n} m_{ij}(\theta)m_{ij'}(\theta)$. We encountered singularity problems. Similar problems were noted in [BC98] where GMM was used; and it is known that EEL and GMM are closely related methods. To avoid it, we used the Moore-Penrose pseudoinverse [Pen55] [1]. The same moments 4.1 (the same information) were used as in the MEL case.

A connection between Continuous GMM and EEL is discussed in [Kit01].

4.2 Monte Carlo simulation results

A 2000 samples of interest-rate series of length 1000 were generated by CIR and Vasicek model, with parameter values $\delta = 0.03, \kappa = 0.5, \sigma = 0.15$ for CIR. These values, according to Ait-Sahalia [AS02], fit in with the US interest rates. For Vasicek model the same values of $\delta = 0.03, \kappa = 0.5$ were used, but the diffusion parameter was changed into $\sigma = 0.15\sqrt{\frac{\delta}{\kappa}} = 0.0367$ so that it takes into account the long run trend. Then both CIR and Vasicek model yield similar interest rate series. Simulations were performed in Matlab 7.7.0 on CPU T2130 with 2GB RAM. We used BFGS method with cubic line search (Matlab built-in function `fminunc`) in the inner loop of the optimization, utilizing the gradient information and Nelder-Mead simplex method (`fminsearch`) in the outer loop. For all the estimators the true parameters [0.03 0.5 0.0367] (Vasicek model) and [0.03 0.5 0.15] (CIR model) were used as starting points in the numerical optimization.

Results are in the following tables

[1] Moore-Penrose pseudoinverse of real valued matrix A is defined as the matrix A^+ satisfying following criteria:
i. $AA^+A = A$
ii. $A^+AA^+ = A$
iii. $(AA^+)' = AA^+$
iv. $(A^+A)' = A^+A$.

Vasicek Model			
True values $\delta = 0.03$ $\kappa = 0.5$ and $\sigma = 0.0367$			
	ML	MEL	EEL
Mean Bias			
δ	4.1495E-004	4.2457E-004	4.2252E-004
κ	6.4294E-003	6.5863E-003	6.5411E-003
σ	5.6962E-005	5.5853E-005	-1.1096E-005
Standard deviation			
δ	2.8175E-003	2.8287E-003	2.8322E-003
κ	4.2991E-002	4.3189E-002	4.3282E-002
σ	1.0568E-003	1.0612E-003	1.0617E-003
RMSE			
δ	2.8478E-003	2.8604E-003	2.8635E-003
κ	4.3469E-002	4.3689E-002	4.3773E-002
σ	1.0583E-003	1.0627E-003	1.0618E-003

Cox-Ingersoll-Ross Model				
True values $\delta = 0.03$ $\kappa = 0.5$ and $\sigma = 0.15$				
	ML	QML	MEL	EEL
Mean Bias				
δ	2.8216E-004	3.1435E-004	3.3184E-004	3.3701E-004
κ	6.0902E-003	6.6657E-003	7.2612E-003	8.3903E-003
σ	1.7335E-004	1.4081E-004	2.4684E-004	-6.7843E-005
Standard deviation				
δ	2.3778E-003	2.8270E-003	2.8866E-003	2.8721E-003
κ	4.3983E-002	5.1322E-002	5.1985E-002	5.1731E-002
σ	4.3338E-003	4.9144E-003	4.8903E-003	4.8733E-003
RMSE				
δ	2.3944E-003	2.8445E-003	2.9056E-003	2.8919E-003
κ	4.4402E-002	5.1752E-002	5.2489E-002	5.2407E-002
σ	4.3372E-003	4.9164E-003	4.8965E-003	4.8738E-003

The results are in accord with asymptotic efficiency of ML, as expected. Interestingly, for the diffusion parameter δ, EEL provided the best absolute mean bias, even better than ML. MEL and EEL provide competitive performance to QML. In our opinion the advantage of these two methods is clearer intuition behind the semi-parametric EL approach.

A table below shows the approximated mean CPU time for every parameter estimation.

CPU time in seconds			
Vasicek Model			
ML	MEL	EEL	
0.1594	3.8919	5.9029	
Cox-Ingersoll-Ross Model			
ML	QML	MEL	EEL
1.0066	0.1654	3.9534	5.8267

Replacing inversion by the Moore-Penrose pseudoinversion in computation of EEL increased CPU time and consequently diminished the computational advantage of EEL. Another interesting fact is that QML estimation was extremely fast.

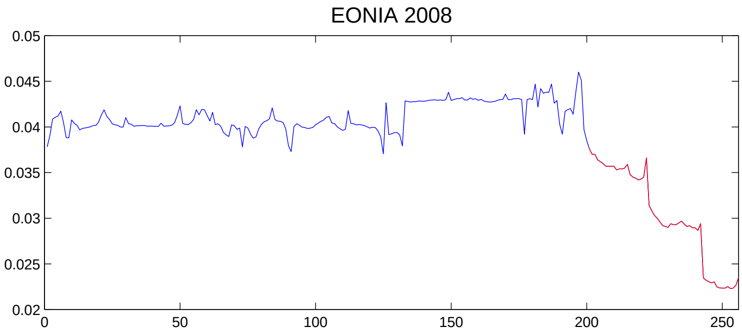
4.3 Sensitivity to starting point

To test how robust are the estimation techniques to choice of starting point, we used different starting points, keeping an interest rate series fixed. This is important if the next step is a calibration of the model on a real data. All estimation methods ML, QML, EL and EEL were very resistant to change of starting point. An instability started to occur when the diffusion parameter was very low, close to zero (6% of the original value). In these cases ML and QML failed to converge and diffusion parameter in EEL was estimated with opposite sign. This might indicate an advantage of MEL, since insensitivity to choice of starting point might be crucial when calibrating the model on real data.

Chapter 5

Real-Data calibration

Euro Overnight Index Average (EONIA) was chosen for calibration of CIR model. The period is from 1.1.2008 to 6.10.2008 (200 working days). Note that the steep fall in interest rate caused by financial crisis after the September is omitted, since non-standard techniques has to be used to model non-standard behavior of the real world.



EONIA series for year 2008, red part is omitted from the estimation.

5.1 Starting point

When calibrating real data, choice of the starting point is crucial. We use Ordinary Least Squares Estimate (OLS) on discretized version of (3.3), Matlab implementation is done in [Kla07], time increment is set to 1.

$$r_{t+1} - r_t = (\delta - \kappa r_t) + \sigma \sqrt{r_t} \epsilon_t, \tag{5.1}$$

where ϵ_t is a white noise.
Equation (5.1) can be transformed into

$$\frac{r_{t+1} - r_t}{\sqrt{r_t}} = \frac{\delta}{\sqrt{r_t}} - \kappa\sqrt{r_t} + \sigma\epsilon_t. \tag{5.2}$$

The initial estimates of δ and κ can be found by minimization of the residual sum of squares

$$(\hat{\delta}, \hat{\kappa}) = \arg\min_{\delta,\kappa} \sum_{t=1}^{N-1} \left(\frac{r_{t+1} - r_t}{\sqrt{r_t}} - \frac{\delta}{\sqrt{r_t}} + \kappa\sqrt{r_t} \right)^2, \tag{5.3}$$

and the diffusion parameter σ is estimated as the standard deviation of residuals. Using basic algebra we obtain the initial estimates $(\hat{\delta}, \hat{\kappa}, \hat{\sigma})$

$$\begin{aligned}
\hat{\delta} &= \frac{(N-1)\sum_{t=1}^{N-1} r_{t+1} - \sum_{t=1}^{N-1} \frac{r_{t+1}}{r_t} \sum_{t=1}^{N-1} r_t}{N^2 - 2N + 1 - \sum_{t=1}^{N-1} r_t \sum_{t=1}^{N-1} \frac{1}{r_t}}, \\
\hat{\kappa} &= \frac{N^2 - 2N + 1 + \sum_{t=1}^{N-1} r_{t+1} \sum_{t=1}^{N-1} \frac{1}{r_t} - \sum_{t=1}^{N-1} r_t \sum_{t=1}^{N-1} \frac{1}{r_t} - (N-1)\sum_{t=1}^{N-1} \frac{r_{t+1}}{r_t}}{N^2 - 2N + 1 - \sum_{t=1}^{N-1} r_t \sum_{t=1}^{N-1} \frac{1}{r_t}}, \\
\hat{\sigma} &= \sqrt{\frac{1}{N-2} \sum_{t=1}^{N-1} \left(\frac{r_{t+1} - r_t}{\sqrt{r_t}} - \frac{\hat{\delta}}{\sqrt{r_t}} + \hat{\kappa}\sqrt{r_t} \right)^2}.
\end{aligned}$$

5.2 Estimation Results

As it was pointed out by Kladivko [Kla07], numerical issues might arise when using function `besseli(.,.)` in Maximum Likelihood estimation. Kladivko suggests to use directly density of non-central chi-square distribution, which is implemented in Matlab as a function `ncx2pdf(.,.)`. This increased stability of ML.

Results from the estimations are in the following table:

EONIA 2008 - CIR					
	start(OLS)	ML	QML	MEL	EEL
δ	0.01020	0.01159	0.01175	0.01042	0.01113
κ	0.24819	0.28203	0.28581	0.25352	0.27137
σ	0.00514	0.00586	0.00589	0.00574	0.00579

The next table compares mean and standard deviation of the data with the marginal (steady state) mean and variance standard deviation using (3.7) for different estimators. Note that for "good" estimators these values should approximately match.

Real and implied mean and standard deviation

	DATA	start(OLS)	ML	QML	MEL	EEL
mean	0.04107	0.04110	0.04110	0.04110	0.04110	0.04102
st. dev.	0.001616	0.001480	0.001582	0.001578	0.001635	0.001592

MEL is slightly different from other estimators, but implied standard deviation is closer to the one computed from data. The graphical representation of likelihood funcions for different estimators is included in Appendix (chapter 8). Conclusion of this calibration is that MEL and EEL provided reasonable alternative to ML and QML.

Chapter 6

Estimation of parameters of VEJ model by EL with conditional characteristic function

Not for all models of interest rate the conditional density can be obtained in explicit form. Consequently, it is not possible to rely on Maximum Likelihood method for estimation of parameters of the models. However, if there is an information in form of estimating equations, it can be exploited for estimation by the Empirical Likelihood method. For some interest rate models for which there is no explicit form of conditional density, there might be explicit form of Conditional Characteristic Function (CFF) which can be utilized in formulating a set of estimating equations. For instance Vasicek Exponential Jump 3.4 model is such a model and we will show this approach on VEJ. The idea of combining EL with estimating equations based on CCF comes from [LN08].

Define CCF of process as

$$\psi(\omega, \tau|\theta, r_t) = E^{\theta}(e^{i\omega r_{t+\tau}}|r_t).$$

Note that there is one-to-one correspondence between CCF and underlying conditional density, therefore CCF captures all the information about the dynamics of the interest rate movement. According to Das and Foresi [DF96], CCF of the VEJ takes the following form (time increment τ was set to 1)

$$\psi(\omega|\theta, r_t) = e^{(A(\omega)+B(\omega)r_t)}, \quad A(\omega) = \frac{i\omega\delta}{\kappa}(1-e^{-\kappa}) - \frac{\omega^2\sigma^2}{4\kappa}\left(1-e^{-2\kappa}\right)$$

$$+\frac{i\lambda(1-2\beta)}{\kappa}(\arctan(\omega\alpha e^{-\kappa})-\arctan(\omega\alpha))+\frac{\lambda}{2\kappa}\log\left(\frac{1+\omega^2\alpha^2e^{-2\kappa}}{1+\omega^2\alpha^2}\right),$$

$$B(\omega)=i\omega e^{-\kappa},$$

where β was set to 1, so that there are only upward jumps.

Once we know CCF we can construct following conditional moments

$$\begin{aligned} E\left[\Re(\psi(\omega|\omega,r_t)-\exp(i\omega r_{t+1}))|r_t\right] &= 0,\\ E\left[\Im(\psi(\omega|\omega,r_t)-\exp(i\omega,r_{t+1}))|r_t\right] &= 0,\end{aligned}$$

so both real an imaginary part must equal zero.

The idea of approximating conditional moments by the sequence of unconditional ones comes from Donald et al. [DIN03].

Intuition behind this is expressed by this impication:

$$E(r_{t+1}|r_t)=0\Rightarrow E(r_{t+1}|g(r_t))=0.$$

So the information about CCF can be transformed into unconditional moment restrictions [LN08]

$$\begin{aligned} E\left[Re(\psi(\omega|\omega,r_t)-\exp(i\omega r_{t+1}))\otimes q_K(r_t)\right] &= 0,\\ E\left[Im(\psi(\omega|\omega,r_t)-\exp(i\omega r_{t+1}))\otimes q_K(r_t)\right] &= 0,\end{aligned}$$

where $\otimes$ denotes kronecker product and $q_K(x)$ is vector of approximating functions. In [LN08], $q_K(x)$ is chosen as follows

$$q_K(x)=(1,2,x^2,x^3,1_{x-s_1}(x-s_1)^3,\ldots,1_{x-s_{K-3-1}}(x-s_{K-3-1})^3),$$

where s_i are the points set empirically.

Note that several things are subject to empirical choice in this estimation method:

- we have to choose vector of approximating functions (q_K and s_i)
- we need to decide in which ω will we evaluate the moment functions ([LN08] set $\omega=[0.1\ 0.3\ 0.7\ 1.1\ 1.5]$)

- we need to choose appropriate time increment and sample size, since the computation is numerically very demanding.

Computing MEL using the information about CCF on real data might be challenging, but nowadays computational complexity allows us to provide it only on simulated data. The starting point in the optimization had to be very close, so it was set directly as the parameter from simulation.

We tried to move the starting point and found out that for the one fixed interest rate series even a very slight change (0.1%) of the starting value lead to different function values. In the light of these facts there might be a doubt whether the computed values are true MEL. Several reasons can explain these finding [1]:

- The objective function for MEL is very flat
- We do not know the shape of the objective function with parameters faraway from the true values
- It is dificult to estimate a jump process because the intensity is low, we can not observe so many jumps
- The process do not satisfies the assumptions of asymptotic theory
- Maybe we need better optimization tool.

Therefore when the closed form of conditional mean and conditional variance are known, we propose using moments from 4.1.3 instead of those based on CCF, since the former lead to estimation method which is more stable and reliable.

[1] The author is grateful to Dr. Q. Liu for these (and many other) helpful comments.

Chapter 7

Summary

The thesis explores possible applications of Empirical Likelihood for estimation of parameters of interest rate diffusion processes.

A Monte Carlo study of small sample properties was conducted in order to compare performance of the Maximum Likelihood and Quasi Maximum Likelihood estimators with that of Maximum Empirical Likelihood and Euclidean Empirical Likelihood estimators under conditional mean and variance estimating equations. This was done for both Vasicek's model and Cox-Ingersoll-Ross model.

Singularity problems in computations of Euclidean Empirical Likelihood were resolved by means of the Moore-Penrose pseudoinverse, which, in turn, adversely affected speed of computation of the estimator. The Monte Carlo study revealed that both MEL and EEL provide competitive performance compared to asymptotically efficient parametric ML. Furthermore MEL appeared to be more stable than ML in some scenarios.

Calibration of EONIA interest rates for year 2008 indicates that MEL and EEL can be considered as a competitive alternative to widely used ML and QML.

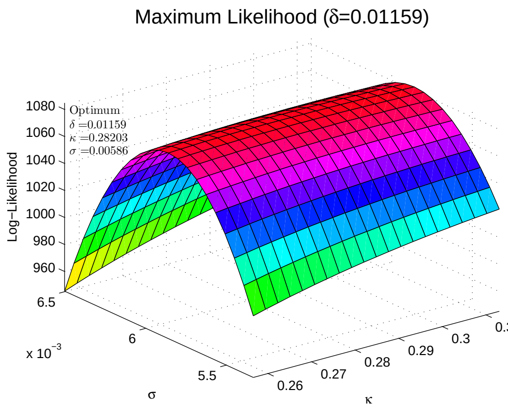
Finally, it was shown that a recent application [LN08] of EL with estimating equations that are based on Conditional Characteristic Function faces a serious numerical instability problem, which renders results of the simulation study presented in [LN08] unreliable.

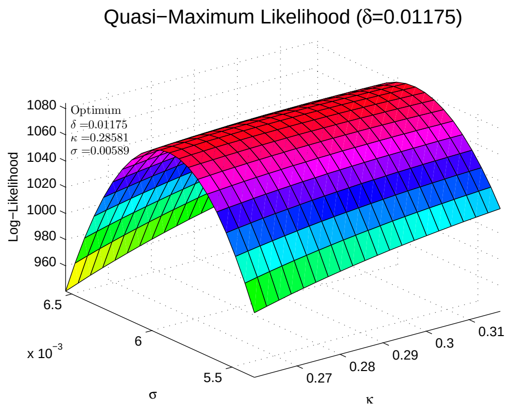
Chapter 8

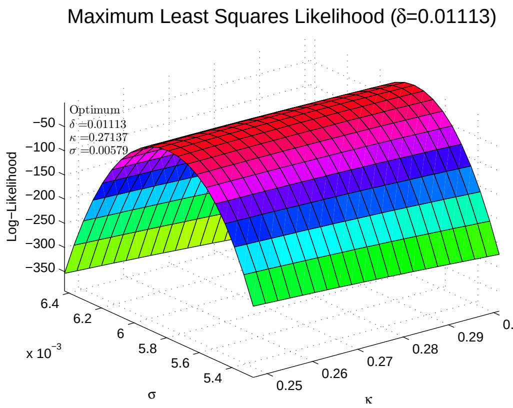
Appendix

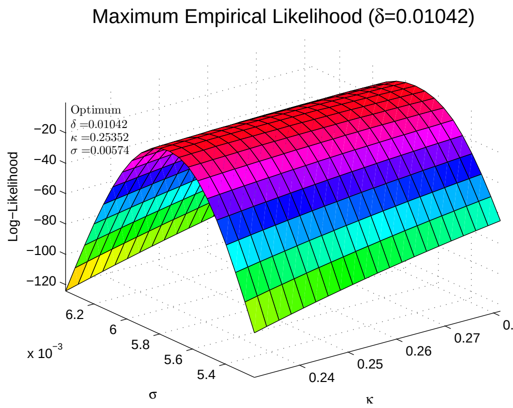
8.1 Likelihood functions from real-data calibration

8.1.1 Fixed delta

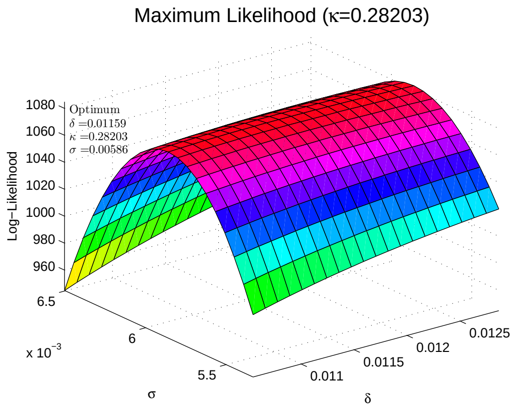


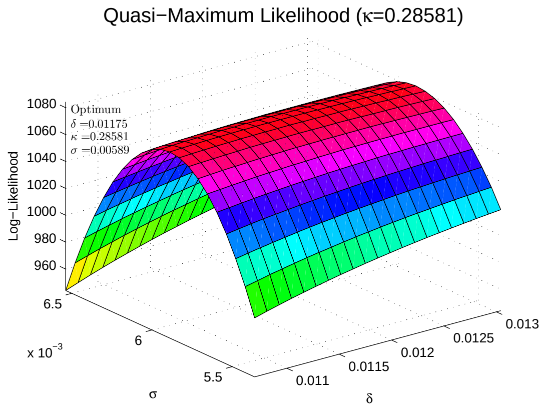


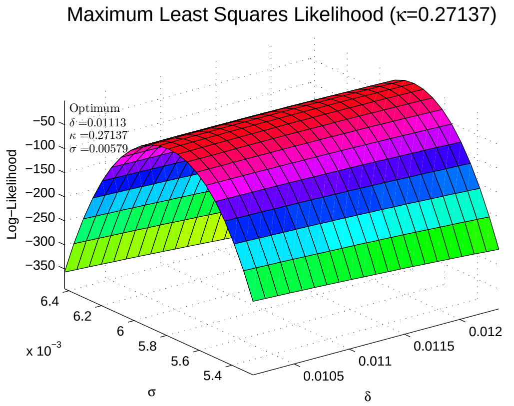


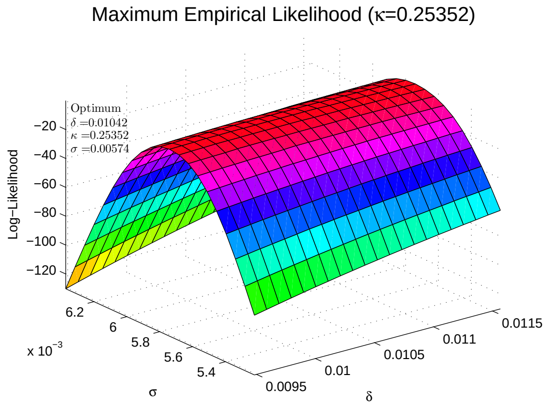


8.1.2 Fixed kappa

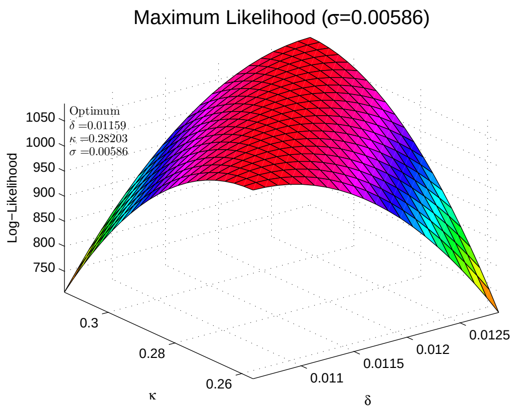


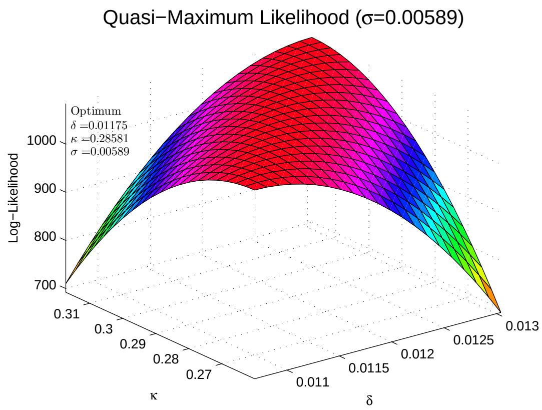


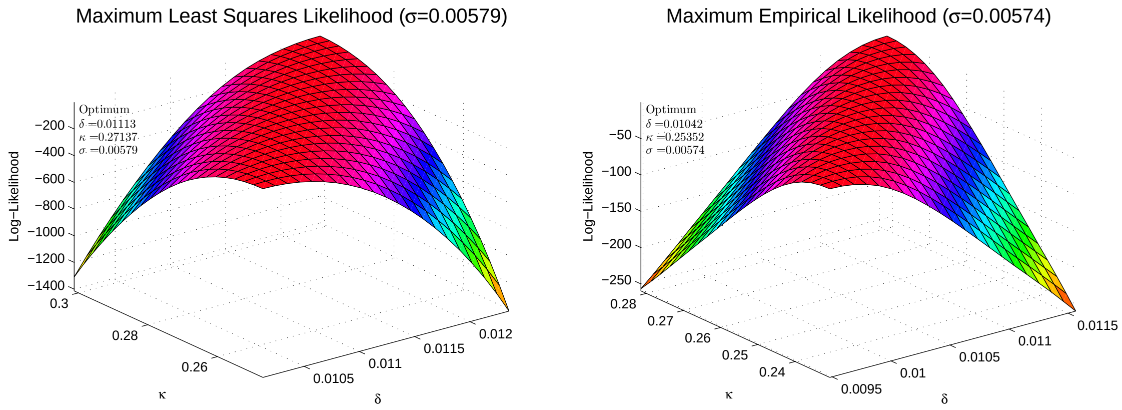




8.1.3 Fixed sigma






Maximum Least Squares Likelihood (σ=0.00579)
Optimum
δ =0.01113
κ =0.27137
σ =0.00579
Log−Likelihood
−200
−400
−600
−800
−1000
−1200
−1400
0.3
0.28
0.26
κ
0.0105
0.011
0.0115
0.012
δ
Maximum Empirical Likelihood (σ=0.00574)
Optimum
δ =0.01042
κ =0.25352
σ =0.00574
Log−Likelihood
−50
−100
−150
−200
−250
0.28
0.27
0.26
0.25
0.24
κ
0.0095
0.01
0.0105
0.011
0.0115
δ

Bibliography

[AS02] Yacine Ait-Sahalia. Maximum-likelihood estimation of discretely sampled diffusions: A closed-form approach. *Econometrica*, 70:223–262, 2002.

[BC98] Bruce Brown and Song Chen. Combined and least squares empirical likelihood. *Annals of the Institute of Statistical Mathematics*, 50:697–714, 1998.

[BKRW93] Peter J. Bickel, Chris A. J. Klaassen, Ya'acov Ritov, and Jon A. Wellner. *Efficient and adaptative estimation for semiparametric model.* Baltimore: Johns Hopkins Press, 1993.

[BL91] Jonathan M. Borwein and Adrian S. Lewis. Duality relationships for entropy-like minimization problems. *SIAM J. Control Optim.*, 29(2):325–338, 1991.

[CIR85] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–408, 1985.

[DF96] Sanjiv R. Das and Silverio Foresi. Exact Solutions for Bond and Option Prices with Systematic Jump Risk. *Review of derivatives research*, 1(1), 1996.

[DIN03] Stephen G. Donald, Guido W. Imbens, and Whitney K. Newey. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117(1):55–93, 2003.

[eSM09] Daniel Ševčovič, Beáta Stehliková, and Karol Mikula. *Analytické a numerické metódy oceňovania finančných derivátov.* STU, 2009.

[IW07] Guido Imbens and Jeffrey Wooldridge. *Generalized Method of Moments and Empirical Likelihood*. 2007. http://www.nber.org/WNE/lect_15_el.pdf.

[Kit01] Yuichi Kitamura. Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, 69:1661–1672, 2001.

[Kit06] Yuichi Kitamura. Empirical likelihood methods in econometrics: Theory and practice. (1569), 2006. http://ideas.repec.org/p/cwl/cwldpp/1569.html.

[Kla07] Kamil Kladivko. Maximum likelihood estimation of the cox-ingersoll-ross process: The matlab implementation. 2007. http://dsp.vscht.cz/konference_matlab/MATLAB07/.

[LN08] Qingfeng Liu and Yoshihiko Nishiyama. Maximum empirical likelihood estimation of continous-time models with conditional characteristic functions. *Mathematics and Computers in Simulation 78*, pages 341–350, 2008.

[MJM00] Ron C. Mittelhammer, George G. Judge, and Douglas J. Miller. *Econometric foundations*. Cambridge Univ. Press, 2000.

[Owe01] Art B. Owen. *Empirical Likelihood*. CRC Press, 2001.

[Pen55] Roger Penrose. A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51:406–13, 1955.

[QL94] Jing Qin and Jerry Lawless. Empirical likelihood and general estimating equations. *Annals of Statistics, Vol.22, No.1*, pages 300–325, 1994.

[Vas77] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

[Zho01] Hao Zhou. A study of the finite sample properties of emm, gmm, qmle, and mle for a square-root interest rate diffusion model. *The Journal of Computational Finance*, 5(2), 2001.