

ESTIMATION OF THE HISTORICAL MEAN-REVERTING PARAMETER IN THE CIR MODEL

GOURIEROUX, ⁽¹⁾ C., and A., MONFORT⁽²⁾

Fevrier, 2007

¹CREST, CEPREMAP and University of Toronto.

²CNAM and CREST.

Estimation of the Historical Mean-Reverting Parameter in the CIR Model

Abstract

We present a comprehensive analysis of the different estimation methods used to estimate the historical mean-reverting parameter in the CIR model. The nonlinear dynamics and the parameter dependent domain of the observed rates can explain the lack of accuracy and even the nonconsistency of several standard estimation approaches.

Keywords : Cox-Ingersoll-Ross Model, Interest Rate, Nonregular Maximum Likelihood, Domain Restrictions.

JEL number :

1 Introduction

The Cox, Ingersoll, Ross (CIR) model [Cox, Ingersoll, Ross (1985)] is a standard one-factor model for term structure of interest rates, which ensures the positivity of rates at any time-to-maturity and is the basis when building more complicated multifactor affine models [Duffie, Kan (1996), Dai, Le, Singleton (2006), Singleton (2006)], or multivariate extensions to stochastic volatility matrices as the Wishart process [see Gouriéroux, Jasiak, Sufana (2005), Gouriéroux (2006)].

However, its practical implementation reveals a lack of accuracy and stability on the estimated mean-reverting parameter with important consequences on derivative prices, even when applied to the same dataset of US T-bonds for similar periods [see Appendix 2]. The following explanations can be given for the high variability of the estimated mean-reverting parameter :

i) The CIR model can be misspecified. Then, different standard estimation methods, which are consistent if the model were well-specified, will provide different results under misspecification.

ii) Different restrictions can be introduced between the historical and risk-neutral parameters of the models, and different mean-reverting parameters, i.e. historical or risk-neutral, or time units can be considered.

iii) Since the instantaneous interest rate is an artificial unobservable notion, the model can be estimated with different proxies. These proxies can be short term interest rates at 1 or 3 month, for instance, and the estimated mean-reverting parameter can be sensitive to the selected time-to-maturity.

iv) Finally, even if the CIR model is well-specified the standard estimation methods are not necessarily very accurate and some of them can even be inconsistent.

We will focus on the two last possible explanations. The aim of this paper is to present a comprehensive analysis of the properties of the mean-reverting parameter estimators according to the selected estimation method and interest rate data, when the CIR model is well-specified. The estimation methods considered in the literature can be ordinary or generalized least squares, moment method based on the characteristic function, misspecified or exact maximum likelihood approaches, whereas the interest rate data can differ by the number of observations, the observation frequency, or the time-to-maturity.

In Section 2, we briefly review the historical and risk-neutral dynamics of the CIR model and discuss the links between historical and risk-neutral

parameters. In Section 3, we describe the main estimation methods of GMM type proposed in the literature, and derive their asymptotic properties for the mean-reverting parameter in Section 4. The analysis is completed by a Monte-Carlo comparison of the finite sample properties of the estimators.

The maximum likelihood (ML) type methods are presented in Section 5. Their analysis is distinguished from the GMM type methods due to a special feature of interest rate models. Indeed, in the CIR model, the domain of admissible value of a rate with given time-to-maturity depends on the unknown parameter. In such a situation, the standard results on ML estimation are no longer valid. For instance, i) An ML approach is not a special case of GMM; ii) ML method can be inconsistent; iii) ML method can provide estimators, which converge at a faster rate; iv) The efficiency bound does not necessarily exist... Then, the different ML type approaches considered in the literature are introduced and their asymptotic properties are discussed. They are compared by Monte-Carlo in Section 6. Section 7 concludes. The proofs of asymptotic results are given in Appendix 1, and a review of the empirical results derived in the applied literature is provided in Appendix 2.

2 Cox-Ingersoll-Ross model

2.1 Instantaneous short term interest rate

In the CIR model the instantaneous short term interest rate (r_t) satisfies the diffusion equation

$$dr_t = a(b - r_t)dt + \sigma r_t^{1/2}dW_t, \quad (2.1)$$

where (W_t) is a standard Brownian motion, a is the positive mean-reverting parameter, b the positive long run parameter and σ the positive volatility parameter. Moreover, if $\delta = 2ab/\sigma^2 > 1$, a zero interest rate cannot be reached. This condition ensures that zero is not an absorbing state and that the historical distribution of the rate is continuous on $(0, \infty)$. The conditional distribution of r_{t+k} given r_t is explicitly known by means of its conditional Laplace transform, which is given by :

$$\psi_k(u|r_t; \theta) = E[\exp(-ur_{t+k})|r_t]$$

$$= \frac{1}{(1 + c_k u)^\delta} \exp \left[-r_t \frac{\rho^k u}{1 + c_k u} \right], \quad (2.2)$$

where : $\rho = \exp(-a)$, $\delta = (2ab)/\sigma^2$, $c_k = (\sigma^2/2a)[1 - \exp(-ak)]$.

This conditional distribution is a noncentered gamma distribution, up to a change of scale.

The marginal distribution is derived as the limiting case of infinite horizon $k \rightarrow \infty$. The marginal Laplace transform is :

$$\psi(u) = (1 + c_\infty u)^{-\delta}, \text{ where } c_\infty = \sigma^2/2a. \quad (2.3)$$

The marginal distribution is a centered gamma distribution with degree of freedom δ , up to the change of scale.

The conditional Laplace transform can be expanded to get :

$$\begin{aligned} & \psi_k(u|r_t; \theta) \\ &= \exp \left(-\frac{r_t \rho^k}{c_k} \right) \frac{1}{(1 + c_k u)^\delta} \exp \left[\frac{r_t \rho^k}{c_k} \frac{1}{1 + c_k u} \right] \\ &= \sum_{j=0}^{\infty} \left[\exp \left(-\frac{r_t \rho^k}{c_k} \right) \frac{1}{j!} \left(\frac{r_t \rho^k}{c_k} \right)^j \frac{1}{(1 + c_k u)^{\delta+j}} \right]. \end{aligned}$$

This expansion shows that the noncentered gamma distribution is a Poisson mixture of centered gamma distributions. We get :

$$c_k r_{t+k} \sim \gamma(\delta + Z_{t+k}), \quad (2.4)$$

where the stochastic degree of freedom is drawn in a Poisson distribution :

$$Z_{t+k}|r_t \sim \mathcal{P}\left[\frac{r_t \rho^k}{c_k}\right].$$

Thus, the transition density of the instantaneous rate process admits an explicit expression, which involves a series expansion. It is given by

$$\begin{aligned} & g_k(r_{t+k}|r_t; \theta) \\ &= \sum_{j=0}^{\infty} \left[\exp\left(-\frac{r_t \rho^k}{c_k}\right) \frac{1}{j!} c_k^\delta (r_t \rho^k)^j \frac{1}{\Gamma(\delta + j)} \exp(-c_k r_{t+k}) r_{t+k}^{\delta+j-1} \right], \quad (2.5) \end{aligned}$$

where Γ denotes the gamma function.

2.2 Risk-neutral dynamics and affine term structure

In general the literature considers pricing models in which the short term interest rate has a similar risk-neutral evolution, with possibly different parameters. In the sequel we assume a CIR risk-neutral dynamics

$$dr_t = a^*(b^* - r_t)dt + \sigma^* r_t^{1/2} dW_t^*, \quad (2.6)$$

where $\theta^* = (a^*, b^*, \sigma^*)'$ indicates risk-neutral parameters. If the relevant information consists in the current and lagged values of (r_t) , the rates at other maturities $r(t, h)$ are derived by :

$$\exp[-hr(t, h)] = E^* \left\{ \exp\left[-\int_t^{t+h} r_u du\right] | r_t \right\}, \quad (2.7)$$

where E^* denotes the expectation computed under the risk-neutral probability (2.6). The interest rates at other maturities are affine functions of the short term rate :

$$r(t, h) = \alpha_h^*(\theta^*)r_t + \beta_h^*(\theta^*), \quad (2.8)$$

where :

$$\begin{aligned} \alpha_h^*(\theta^*) &= \frac{1}{h} \left[\frac{2}{\gamma^* + a^*} - \frac{4\gamma^*}{\gamma^* + a^*} \frac{1}{(\gamma^* + a^*) \exp(\gamma^* h) + \gamma^* - a^*} \right], \\ \beta_h^*(\theta^*) &= \frac{1}{h} \left[-\frac{a^* b^* (\gamma^* + a^*)}{\sigma^{*2}} h + \frac{2a^* b^*}{\sigma^{*2}} \log[(\gamma^* + a^*) \exp(\gamma^* h) + \gamma^* - a^*] \right. \\ &\quad \left. - \frac{2a^* b^*}{\sigma^{*2}} \log(2\gamma^*) \right], \\ \gamma^* &= \sqrt{a^{*2} + 2\sigma^{*2}}. \end{aligned}$$

As mentioned in the introduction, the instantaneous interest rate is a virtual notion, which is not observed in practice. However, some rates with given time-to-maturity can be observed. The conditional historical distribution of $r(t, h)$ is directly derived by applying the affine transformation (2.8).

Since $\alpha_h^*(\theta^*) > 0$ for any finite time-to-maturity, $[r(t, h)]$ is still a Markov process.

Its conditional Laplace transform is :

$$\begin{aligned} & \psi_k(u, h|r(t, h); \theta, \theta^*) \\ &= E_\theta \{ \exp(-ur(t+k, h)) | r(t, h) \} \\ &= E_\theta \left\{ \exp(-u[\alpha_h^*(\theta^*)r_{t+k} + \beta_h^*(\theta^*)]) | r_t = \frac{r(t, h) - \beta_h^*(\theta^*)}{\alpha_h^*(\theta^*)} \right\}. \end{aligned}$$

We get :

$$\begin{aligned} & \psi_k(u, h|r(t, h); \theta, \theta^*) \\ &= \exp[-u\beta_h^*(\theta^*)] \psi_k(u\alpha_h^*(\theta^*) | \frac{r(t, h) - \beta_h^*(\theta^*)}{\alpha_h^*(\theta^*)}; \theta). \end{aligned} \quad (2.9)$$

We can also derive the transition density of process $r(t, h)$ at horizon k from the transition density of the instantaneous rate by applying the Jacobian formula. We get :

$$\begin{aligned} & g_{k,h}[r(t+k, h)|r(t, h); \theta, \theta^*] \\ &= \frac{1}{\alpha_h^*(\theta^*)} g_k \left[\frac{r(t+k, h) - \beta_h^*(\theta^*)}{\alpha_h^*(\theta^*)} | \frac{r(t, h) - \beta_h^*(\theta^*)}{\alpha_h^*(\theta^*)}; \theta \right] \mathbf{1}_{r(t+k) \geq \beta_h^*(\theta^*)} \end{aligned} \quad (2.10)$$

Whereas the historical distribution of the instantaneous rate (r_t) depends on the historical parameter θ only, the historical distribution of a rate with time-to-maturity $h, h > 0$, depends on both historical and risk-neutral parameters. In the limiting case of infinite time-to-maturity the long run interest rate $r(t, \infty) = \beta_\infty^*(\theta^*)$ depends on the risk-neutral parameters only.

2.3 Link between the historical and risk-neutral parameters

The historical parameters $\theta = (a, b, \sigma)'$ and the risk-neutral parameters $\theta^* = (a^*, b^*, \sigma^*)'$ can be chosen independently [Rogers (1977)]. Thus, in general the

risk-neutral mean-reverting [resp. long run, volatility] parameter differs from the historical mean reverting [resp. long run, volatility] parameter. However, it is usual to constrain these parameters to get a simple interpretation of the change of parameters, that is, of the change of measure, in terms of a path independent "risk premium" parameter λ . These constraints are [see e.g. Cox, Ingersoll, Ross (1985)].

$$a^* = a + \lambda, b^* = \frac{ab}{a + \lambda}, \sigma^* = \sigma, \quad (2.11)$$

and correspond to the change of density :

$$\left(\frac{dP^*}{dP}\right)_t = \exp \left[-\lambda \int_0^t \frac{\sqrt{r_s}}{\sigma} dW_s - \frac{\lambda^2}{2\sigma^2} \int_0^t r_s ds \right], \quad (2.12)$$

and the modified Brownian motion under P^* :

$$W_t^* = W_t + \frac{\lambda}{\sigma} \int_0^t \sqrt{r_s} ds. \quad (2.13)$$

Under these additional restrictions, the model involves a total of 4 parameters, that are a, b, σ, λ , and the historical and risk-neutral mean-reverting parameters still differ. In the sequel, we focus on the **estimation of the historical mean-reverting parameter a** .

Finally, the (geometric) rate $r(t, h) = -(1/h) \log B(t, h)$, where $B(t, h)$ is the price of the zero-coupon bond, and the parameters can be sensitive to the choice of the time unit, e.g. day, month or year. Let us consider a change of time unit $t \rightarrow \alpha t$, say. We get :

$$h \rightarrow \alpha h, r(t, h) \rightarrow r(t, h)/\alpha, r_t \rightarrow r_t/\alpha, dt \rightarrow \alpha dt,$$

$dW_t \rightarrow \alpha^{1/2} dW_t$, and from (2.1), (2.4), $a \rightarrow \alpha a, b \rightarrow b\alpha, \sigma \rightarrow \sigma\alpha, a^* \rightarrow a^*\alpha \dots$. Thus, the mean-reverting parameter a depends on the time unit, whereas the degree of freedom δ is time-unit independent.

3 Estimation of the mean-reverting parameter by GMM type methods.

Let us describe the main moment methods, proposed in the CIR framework, that are weighted or unweighted autoregression, quasi-maximum likelihood

approach, or methods based on the conditional Laplace transform [Singleton (2001), Bates (2006)]. The OLS approach is the only estimation method, which estimates directly the mean-reverting parameter. All other methods are based on the joint estimation of all parameters including the risk premium parameter. They can be expected more efficient, but also more difficult to implement.

3.1 A regression model

The two first conditional moments are deduced from formula (2.1) providing the conditional distribution at horizon k . We get :

$$E(r_{t+k}|r_t) = c_k\delta + \rho^k r_t, \quad (3.1)$$

$$V(r_{t+k}|r_t) = c_k^2\delta + 2\rho^k c_k r_t. \quad (3.2)$$

Equivalently, we can write the conditionally heteroscedastic autoregression model :

$$r_{t+k} = c_k\delta + \rho^k r_t + u_{t,k}, \quad (3.3)$$

where $E(u_{t,k}|r_t) = 0$, $V(u_{t,k}|r_t) = c_k^2\delta + 2\rho^k c_k r_t$.

Since the term structure is affine (see 2.8), similar regression models can also be written for the interest rates with larger time-to-maturity.

$$r(t+k, h) = c_{k,h} + \rho^k r(t, h) + u_{t,k,h}, \quad (3.4)$$

say. Note that the autoregressive parameter depends on the observation frequency, but does not depend on time-to-maturity. However, the conditional variance depends on both features.

3.2 OLS approach

Let us now consider observations of a rate of time-to-maturity h sampled at regular frequency k . By denoting $t = 0$ the date of the first observation, the available data are $r(kn, h)$, $n = 0, 1, \dots, N$, where k denotes the sampling frequency and $N + 1$ the number of observations. The regression model :

$$r(kn, h) = c_{k,h} + \rho^k r[k(n-1), h] + u_{kn,k,h}, \quad (3.5)$$

can be used to estimate the (historical) mean-reverting parameter $a = -\log \rho$ by OLS. The nonlinear OLS estimator is defined as follows :

i) Regress $r(kn, h)$ on $r[k(n-1), h]$, and 1 by OLS to get an estimator of ρ^k . This estimator $\hat{\rho}(k, N)$ does not depend on the time-to-maturity due to the affine relationship between rates.

ii) Then, the nonlinear OLS estimator of a is :

$$\hat{a}(k, N) = -\frac{1}{k} \log \hat{\rho}(k, N). \quad (3.6)$$

3.3 GLS approach

The OLS approach does not account for conditional heteroscedasticity present in regression model (2.10). Thus, the estimation can be improved by following a two-step approach.

i) In a first step, we estimate the conditional variance in the autoregression (3.4). This can be done in the usual way by performing the OLS regression of $r(kn, h)$ on $r[k(n-1), h]$ and 1, and computing the associated residuals $\hat{u}_{kn,k,h}$. Since the volatility is an affine functions of the rate, the parameters of the volatility function are consistently estimated by an OLS regression of $\hat{u}_{kn,k,h}^2$ on $r(k(n-1), h)$ and 1. Then the estimated volatilities are deduced.

ii) In the second step, we regress $r(kn, h)$ on $r(k(n-1), h)$, and 1 by GLS, with the weights associated with the inverse of the estimated conditional variances. This provides another estimator $\hat{\rho}_W(k, h, N)$ of ρ^k , and by transformation another estimator of the (historical) mean-reverting parameter :

$$\hat{a}_w(k, h, N) = -\frac{1}{k} \log \hat{\rho}_W(k, h, N).$$

3.4 QML approach

Another standard estimation method for time series is the quasi (or pseudo) maximum likelihood approach. This method optimizes the log-likelihood function computed as if the error terms in the regression were conditionally Gaussian. Let us denote by $\eta^2(a, b, \sigma, \lambda; r[k(n-1), h])$ the conditional variance of $u_{kn,k,h}$ given $r[k(n-1), h]$ and by $c(a, b, \sigma, \lambda; k, h)$ the intercept in

regression (3.5). It is easily seen that :

$$c(a, b, \sigma, \lambda; k, h) = \alpha_h^*(a, b, \sigma, \lambda)c_k\delta + \beta_h^*(a, b, \sigma, \lambda)(1 - \rho^k) \quad (3.7)$$

$$\eta^2(a, b, \sigma, \lambda; r[k(n-1), h]) = \alpha_h^{*2}c_k^2\delta + 2\rho^k c_k \alpha_h^* [r[k(n-1), h] - \beta_h^*] \quad (3.8)$$

The QML estimator of \mathbf{a} , denoted by $a^*(k, h, N)$, is derived by solving the optimization problem :

$$\text{Max}_{a,b,\sigma,\lambda} QL[r(kn, h)|r(k(n-1), h); h; a, b, \sigma, \lambda],$$

where the quasi log-likelihood for time-to-maturity h is given by :

$$\begin{aligned} & QL[r(kn, h)|r(k(n-1), h); h; a, b, \sigma, \lambda] \\ &= \sum_{n=1}^N \left\{ -\frac{1}{2} \log \eta^2(a, b, \sigma; r[k(n-1), h]) \right. \\ & \left. - \frac{1}{2} \frac{[r(kn, h) - c[a, b, \sigma, \lambda; k, h] - \exp(-ak)r[k(n-1), h]]^2}{\eta^2(a, b, \sigma, \lambda; r(k(n-1), h))} \right\}. \quad (3.9) \end{aligned}$$

3.5 Approximated QML approaches

The exact quasi log-likelihood function (3.9) is often replaced by approximations. Three approximations are used in practice.

The first approximation consists in using the quasi likelihood function corresponding to the instantaneous interest rate and replacing the observations of this rate by a proxy that is a 1-month or 3-month rate. The approximated quasi log-likelihood to be maximized is :

$$\begin{aligned} & AQL^1(a, b, \sigma) \\ &= QL[r(kn, h)|r(k(n-1), h); 0; a, b, \sigma] \\ &= \sum_{n=1}^N \left\{ -\frac{1}{2} \log [c_k^2\delta + 2\rho^k c_k r(k(n-1), h)] \right. \\ & \left. - \frac{1}{2} \frac{[r(kn, h) - c_k\delta - \rho^k c_k r(k(n-1), h)]^2}{c_k^2\delta + 2\rho^k c_k r(k(n-1), h)} \right\}, \quad (3.10) \end{aligned}$$

by using the expressions (3.1), (3.2) of the conditional first-and second-order moments of the instantaneous interest rate. This approximated quasi-likelihood function depends on the historical parameters only and does not involve the risk premium parameter. The substitution of the rate $r(kn, h)$ (with h small, for instance $1/4$), instead of the instantaneous rate, implies a misspecification and the estimator of α obtained by maximizing AQML¹ is not consistent.

The second approximated QML approach uses the conditional moments of $r(kn, h)$ given the instantaneous rate $r_{k(n-1)}$, in which $r_{k(n-1)}$ is replaced by $r[k(n-1), h]$ (with h small for instance $h = 1/4$). More precisely, using $r(t, h) = \alpha_h^* r_t + \beta_h^*$, equation (3.5) can be written :

$$r(kn, h) = c_{k,h}^* + \alpha_h^* \rho^h r_{k(n-1)} + u_{kn,k,h}, \quad (3.11)$$

$$\text{where : } c_{k,h}^* = \alpha^* c_k \delta + \beta_h^*,$$

$$E[u_{kn,k,h} | r_{k(n-1)}] = 0,$$

$$V[u_{kn,k,h} | r_{k(n-1)}] = \alpha_h^{*2} [c_k^2 \delta + 2\rho^k c_k r_{k(n-1)}].$$

The second approximated QML estimator is obtained by maximizing :

$$\begin{aligned} AQL^2(a, b, \sigma, \lambda) = \\ \sum_{n=1}^N \left\{ -\frac{1}{2} \log[\alpha_h^{*2} (c_k^2 \delta + 2\rho^k c_k r(k(n-1), h))] \right. \\ \left. - \frac{1}{2} \frac{[r(kn, h) - c_{k,h}^* - \alpha_h^* \rho^k r[k(n-1), h]]^2}{\alpha_h^{*2} [c_k^2 \delta + 2\rho^k c_k r[k(n-1), h]]} \right\} \end{aligned} \quad (3.12)$$

Note that λ can also be estimated.³

The third QML approach is based on a crude Euler discretization (CED) of the CIR diffusion equation [see e.g. Chan, Karolyi, Longstaff, Sanders (1992), Smith (2002)]. More precisely, the diffusion equation (2.1) is time discretized at frequency k to get :

³In the Monte-Carlo experiment, we also use a QML method based on the conditional moments of $r(kn, h_1)$ for h_1 "large" ($h_1 = 1.5$) given $r_{k(n-1)}$, where $r_{k(n-1)}$ is replaced by $r[k(n-1), h_2]$ with h_2 "small" ($h_2 = 1/4$).

$$r_{t+k} = r_t + a(b - r_t)k + \sigma\sqrt{kr_t^{1/2}}\varepsilon_{t+k}, \quad (3.13)$$

where ε_{t+k} is standard normal. Then, the conditional first-and second-order moments of the instantaneous rate are approximated by the moments corresponding to the Euler discretization (3.13) and r_t is approximated by $r(t, h)$ (h small, for instance $h = 1/4$). We get a third approximated quasi log-likelihood function :

$$\begin{aligned} & AQL^3(a, b, \sigma) \\ &= \sum_{n=1}^N \left\{ -\frac{1}{2} \log[k\sigma^2 r(k(n-1), h)] \right. \\ & \left. - \frac{1}{2} \frac{[r(kn, h) - r(k(n-1), h) - a(b - r(k(n-1), h))k]^2}{k\sigma^2 r(k(n-1), h)} \right\}. \quad (3.14) \end{aligned}$$

The substitution of time discretized expressions of the conditional mean and variance implies an additional misspecification and nonconsistent results. The asymptotic bias of both approximated quasi maximum likelihood approaches can be corrected by simulation methods, such as indirect inference [Broze, Scaillet, Zakoian (1995)].

3.6 Methods based on the Laplace transform

A GMM approach can also be based on conditional moment restrictions associated with the Laplace transform. The conditional moment restrictions are :

$$E_t[\exp[-u_j r(t+k, h)] - \psi_k(u_j, h|r(t, h); \theta, \theta^*)|r(t, h)] = 0, j = 1, \dots, J,$$

where $u_j, j = 1, \dots, J$ are J selected real or complex arguments. When the argument is real (resp.complex) the parameters are calibrated from the expectations of exponential transforms of the rates (resp. sine and cosine transforms). The estimator and its accuracy depend : i) on the selected values, ii) on the instruments used in the approach. As the GLS and QML approaches, the method based on Laplace transform requires the joint estimation of a, b, σ and λ .

When the CIR model is well-specified and the infinitesimal interest rate is stationary, the GMM type estimators⁴ are all consistent, with a rate of convergence \sqrt{N} , and are asymptotic normal. The GMM approaches are not efficient, even if an infinite number of moments is introduced in the Laplace transform based approach. This can seem contradictory with results recently established by Singleton (2001) or Bates (2006) for GMM based on the characteristic function. However, the possibility to reach the asymptotic efficiency assumes a parameter independent domain for the observed rates. As seen, from the expression of the transition density, this assumption is not fulfilled in the CIR model (and more generally for any interest rate model with positive infinitesimal rate [Gourieroux, Monfort (2006)]). Moreover, in some situations consistent estimators with a convergence rate strictly better than \sqrt{N} can be derived (see Section 5).

4 Properties of the GMM type estimators

We first derive the asymptotic variances of the OLS and GLS estimators [see Appendix 1]. The explicit expression of the asymptotic variances are useful to discuss the effects of the selected time-to-maturity and observation frequency. Then, we perform a Monte-Carlo experiment to compare the finite sample properties of the mean-reverting parameters estimated by the different GMM type methods.

4.1 OLS estimator of the historical mean-reverting parameter

Proposition 1 : Under standard regularity conditions, we have

$$\sqrt{N}[\hat{a}(k, N) - a] \simeq N\left[0, \frac{1}{k^2} \frac{1}{\rho^{2k}} [(1 - \rho^k)^2 + 2\rho^k(1 - \rho^k)(1 + 2/\delta)]\right].$$

The asymptotic variance of the OLS estimator depends on k , δ , and ρ .

Figure 1 shows how this asymptotic variance depends on δ , k , ρ (and a); in each case the non varying parameters are fixed at the values used in the Monte-Carlo study (i.e. $\delta = 6$; $k = 1/12$, $\rho = 0.61$ (or $a = 0.5$)).

⁴including the QML estimator, but not the approximated QML estimators.

[Insert Figure 1 : Asymptotic Variance of the OLS Estimator]

i) This is a decreasing function of the degree of freedom δ , which tends to infinity when δ tends to zero. This result is expected. Indeed, the condition $\delta > 0$ ensure that the zero interest rate cannot be reached. If δ tends to zero, we tend to a limiting case in which 0 can be attained and becomes an absorbing state of the process. This renders impossible the consistent estimation of the mean-reverting parameter.

ii) The variance is a skewed U-shape function of k . This function tends to infinity when k tends to either zero, or infinity. Therefore, there exists and optimal frequency function of ρ, δ , which minimizes the variance of the estimator.

iii) The asymptotic variance is a quadratic function of $\rho^{-k} = \exp(ak)$:

$$V_{as}[\sqrt{N}(\hat{a}(k, N) - a)] = \frac{1}{k^2} \left(\frac{1}{\rho^k} - 1 \right) \left(\frac{1}{\rho^k} + 1 + \frac{4}{\delta} \right).$$

This is an increasing function of a , which takes value 0 for $a = 0$, and tends to infinity when a tends to infinity.

iv) Finally, this variance decreases with ρ .

4.2 GLS estimator of the historical mean-reverting parameter

Proposition 2 : Under standard regularity conditions, we have :

$$\sqrt{N}(\hat{a}_W(k, h, N) - a)$$

$$\approx N(0, V_{as}(\hat{a})),$$

$$\text{where } V_{as}(\hat{a}) = (1 - \rho^k)^2 \delta \frac{E\left(\frac{1}{1 + \gamma_k r}\right)}{E\left(\frac{1}{1 + \gamma_k r}\right) E\left(\frac{r^2}{1 + \gamma_k r}\right) - \left[E\left(\frac{r}{1 + \gamma_k r}\right)\right]^2},$$

$\gamma_k = \frac{2\rho^k}{\delta(1 - \rho^k)}$ and r is a random variable with a gamma distribution $\gamma(\delta)$.

Figure 2 displays the asymptotic variance of the GLS estimator as a function of the parameters [again the non varying parameters are fixed at $\delta = 6, k = 1/2, \rho = 0.61$ (or $a = 0.5$)].

[Insert Figure 2 : Asymptotic Variance of the GLS Estimator]

The patterns are similar to the patterns observed for the OLS estimator, except for the effect of time frequency. Indeed, when k decreases it is generally easier to estimate the volatility parameters, but more difficult to estimate the medium run parameters. The balance between both effects explain the specific observed pattern.

4.3 The Monte-Carlo experiment

In the Monte-Carlo experiment, we choose the time unit equal to 1 year and the following values for parameters a, b and σ : $a = 0.5, b = 0.06, \sigma = 0.1$, and consequently $\delta = 6, \rho = 0.61$. These values are very close to the values taken by the ML estimates obtained using the exact discrete time process, with 1-month frequency, derived from the continuous time r_t process [that is, the ARG process with parameters $\delta = \frac{2ab}{\sigma^2}, \rho = \exp(-a/12), c = \frac{\sigma^2}{2a}[1 - \exp(-a/12)]$], and using as a proxy of r_t the rate of maturity 1-month in the CRSP database (1964-06 to 1995-12).

Moreover, we retain the value $\lambda = -0.3$ for risk premium parameter; this value is close to the value obtained from a nonlinear regression of the observed rates at various maturities, on the theoretical ones given by the CIR model with the historical parameters fixed at the values given above. We simulated $S = 500$ replications of a path of length $T = 500$, in the exact discrete time version (frequency 1-month, i.e. $k = 1/12$) of the CIR process $dr_t = -0.5(r_t - 0.06)dt + 0.1r_t^{1/2}dW_t$.

The mean-reverting parameter has been estimated by the different methods described in Section 3, that are the OLS, GLS, QML and the approximated QML approaches. We have not applied the approach based on the characteristic function, which is too sensitive to the selection of the number of moments and the arguments.

In the GLS approach, the estimated conditional variances obtained in the first step are not necessarily positive; we have imposed positivity by

constraining the regression coefficients to be nonnegative.

The QML approach is numerically unstable because of the expression of the conditional variance η^2 given in (3.8). Indeed, this expression can be negative for a large set of parameter values. This is the reason why we have disregarded this method.

The finite sample bias (B), standard error (SE), least absolute deviation (LAD) and root mean squared error (RMSE) of the estimator of a have been first computed for the eight methods given below using only the 3-month rate (see Table 1). These methods are :

- a) the OLS method (OLS), for $h = 1/4$;
- b) the GLS method (GLS), for $h = 1/4$;
- c) the unfeasible first approximated QML method (UAQML1), maximizing (3.10), in which $r(kn, h)$ is replaced by r_{kn} and $r[k(n-1), h]$ by $r_{k(n-1)}$.
- d) the first approximated QML method (AQML1), maximizing (3.10) with $h = 1/4$. (3-month rate);
- e) the unfeasible second approximated QML method (UAQML2) maximizing (3.12) with $h = 1/4$, where $r[k(n-1)]$ has been replaced by $r_{k(n-1)}$;
- f) the second approximated QML method (AQML2) maximizing (3.12) with $h = 1/4$;
- g) the unfeasible third approximated QML method (UAQML3) maximizing (3.14), in which $r(kn, h)$ is replaced by r_{kn} , and $r[k(n-1), h]$ by $r_{k(n-1)}$;
- h) the third approximated QML (AQML3) maximizing (3.14), with $h = 1/4$.

Table 1 : Summary statistics of GMM type methods

	B	SE	LAD	RMSE
OLS	0.109	0.198	0.169	0.225
GLS	0.609	0.351	0.610	0.703
UAQML 1	0.045	0.212	0.168	0.216
AQML 1	0.061	0.211	0.169	0.219
UAQML 2	0.027	0.055	0.046	0.061
AQML 2	-0.136	0.070	0.139	0.153
UAQML 3	0.138	0.209	0.192	0.250
AQML 3	0.135	0.206	0.190	0.247

In finite sample the OLS approach is preferable to the GLS approach for both the bias and standard error. The QML approach in which the instantaneous interest rate is replaced by the 3-month rate shows a rather small bias, even if it is theoretically inconsistent. Finally, the QML approach based on a crude Euler discretization feature a rather large bias. This bias is due more to the discretization than to the replacement of the instantaneous rate by the 3-month rate.

The finite sample p.d.f. of the estimators are given in Figures 3,4,5.

[Insert Figure 3 : PDF of UAQML1 and AQML1 estimators of a]

[Insert Figure 4 : PDF of UAQML2 and AQML2 estimators of a]

[Insert Figure 5 : PDF of UAQML3 and AQML3 estimators of a]

The comparison of the pdf confirms the discussions of Table 1. Note that the normality is almost achieved for all the estimators.

In a second step, the UAQML2 and AQML2 have been used with rates with different time-to-maturity. First we have used the conditional moments of the one year rate ($h_1 = 1$) given the instantaneous rate, then replaced by the three month rate ($h_2 = 1/4$); the methods are called UAQML2.1 and AQML2.1, respectively. Then, we have used $h_1 = 5$ and $h_2 = 1/4$; the methods are called UAQML2.5 and AQML2.5. The summary statistics are given in Table 2, and the p.d.f. of AQML2, AQML2.1, AQML2.5 are compared in Figure 6.

[Insert Figure 6 : PDF of AQML estimator for a]

Table 2 : Summary statistic of GMM type methods based on two rates

	B	SE	LAD	RMSE
UAQML2.1	0.014	0.029	0.019	0.032
AQML2.1	-0.041	0.022	0.042	0.046
UAQML2.5	0.007	0.017	0.008	0.018
AQML2.5	-0.006	0.010	0.010	0.015

The statistical properties of these estimators are improving dramatically when h_1 is getting larger; in particular the RMSE is very small. This result was expected. Indeed in a CIR model, the rates are in a deterministic affine relationships, which allows the estimation of the corresponding coefficients with zero error. The AQML approaches are not the most efficient, but capture a part of this effect.

5 Maximum likelihood estimation methods

We will now consider the estimation methods based on the likelihood function, or on some of its approximations. We first emphasize the non regularity of the likelihood function and its implications. Then, we compare the finite sample properties of different likelihood based estimation methods.

5.1 Nonregular likelihood function

The likelihood function corresponding to observations $r(kn, h), n = 1, \dots, N$ given $r(0, h)$ is :

$$\begin{aligned} l_N(\theta, \theta^*) &= \prod_{n=1}^N g_{k,h}[r(kn, h)|r(k(n-1), h); \theta, \theta^*] \\ &= \prod_{n=1}^N \left\{ \frac{1}{\alpha_h^*(\theta^*)} g_k \left[\frac{r(kn, h) - \beta_h^*(\theta^*)}{\alpha_h^*(\theta^*)} \middle| \frac{r(k(n-1), h) - \beta_h^*(\theta^*)}{\alpha_h^*(\theta^*)}; \theta \right] \right\} \\ &\quad \mathbb{1}_{\inf_n r(kn, h) \geq \beta_h^*(\theta^*)}. \end{aligned} \tag{3.15}$$

The likelihood function depends on the four parameters a, b, σ, λ . This can be a discontinuous or continuous non differentiable function of these parameters when $\beta_h^*(\theta^*) = \inf_n r(kn, h)$.

The standard asymptotic theory for maximum likelihood assumes a continuous differentiable likelihood function. This explains why the standard asymptotic results are not necessarily valid in the CIR framework, where the domain is parameter dependent. There exists no general theory in the statistical literature for parameter dependent domain, but partial results are available and listed below [see e.g. Smith (1985), Hirano, Porter (2003)].

Some results are due directly to the lack of differentiability. For instance, the standard information matrix cannot be defined, since it involves first (or second) order derivatives of the log-likelihood function.

Other results are deduced from the special expression of the likelihood function, which can be written as :

$$l_n(\theta, \theta^*) = \exp[L_N^*(\theta, \theta^*)] \mathbb{1}_{\inf_n r(kn, h) \geq \beta_n^*(\theta^*)} \tag{3.16}$$

where :

$$L_N^*(\theta, \theta^*) = \sum_{n=1}^N \log \left\{ \frac{1}{\alpha_h^*(\theta^*)} g_k \left[\frac{r(kn, h) - \beta_h^*(\theta^*)}{\alpha_h^*(\theta^*)} \middle| \frac{r(k(n-1), h) - \beta_h^*(\theta^*)}{\alpha_h^*(\theta^*)}; \theta \right] \right\}.$$

As usual the normalized quantity $\frac{1}{N} L_N^*(\theta, \theta^*)$ is asymptotically normal, but the likelihood function involves also the indicator function with an asymptotic behaviour corresponding to an extreme distribution (infimum of serially dependent variables). By analogy with the existing literature [see e.g. Smith (1985)], the following results can be expected :

i) The ML estimator which maximizes $l_N(\theta, \theta^*)$ is consistent (but the log-likelihood function does not exist when $\beta_h^*(\theta^*) > \inf_n r(kn, h)$).

ii) The ML estimators can be normalized, but their rates of convergence can differ, that are $1/N$ for the transformed parameter $\beta_h^*(\theta^*)$, $1/\sqrt{N}$ for the transformed parameters asymptotically independent of the estimator of $\beta_h^*(\theta^*)$.

iii) The asymptotic distribution of the normalized estimator can be a complicated transformation of Gaussian and extreme distributions.

iv) The existence of an optimal rate of convergence and of an efficiency bound has not been proved in the literature. Thus, it is not known if the ML estimator is asymptotically efficient. Nevertheless, it is strictly more efficient than the GMM type estimators, which are not using the function $\inf_n r(kn, h)$, which is a component of the asymptotic sufficient statistics for the estimation problem.

5.2 The estimation methods

The exact ML approach presented in Section 5.1 does not seem to have been used previously. Generally, the practitioners or academic researchers propose approximated ML approaches, which can imply nonconsistent estimators. The most frequent ones are given below

i) **Approximate ML with a proxy for the instantaneous rate.**

The idea is to consider the log-likelihood function corresponding to the instantaneous interest rate, and to maximize this approximated log-likelihood after substituting a short term interest rate, e.g. a 3-month rate, to the instantaneous interest rate. The log-likelihood function is :

$$AML^1(\theta) = \sum_{n=1}^N \log g_k[r_k(kn, h)|r(k(n-1), h); \theta]. \quad (5.1)$$

The domain for the instantaneous interest rate is $(0, \infty)$ and is parameter independent. However, the maximization of this misspecified log-likelihood will imply asymptotic biases. Typically, this approximated log-likelihood gives the spurious impression that the only identifiable parameters are the component of θ , while it is known from the correct expression of the likelihood function, that the risk premium parameter λ can also be identified. This approach can be applied with different approximations of the g_k function. A first approximation is obtained by truncating the series expansion defining the density [see e.g. Pearson, Sun (1994)]. A second approximation can be obtained by applying ⁵ the "efficient method of moment" [see e.g. Gallant, Tauchen (1998), Gu, Zivot (2006)].

ii) **Exact ML method ignoring the domain restrictions.**

The estimator is obtained by maximizing the quantity $L_N^*(\theta, \theta^*)$, generally as the consequence of a too crude application of the Jacobian formula. Following Pearson, Sun (1994), this approach has been used or extended by (Duan (1994), (2000), Bruche (2004), Ericsson, Reneby (2005)). According to the tail behaviour of the observations, these estimators can be consistent, or inconsistent [see Smith (1985) for a discussion].

The reason for possible nonconsistency is easily understood. As usual, the estimator is consistent if the associated score is zero-mean, that is, if :

$$E_0 \left[\frac{\partial \log g_{k,h}}{\partial(\theta, \theta^*)} [r(kn, h)|(k(n-1), h); \theta_0, \theta_0^*] \right] = 0.$$

In the standard framework, this condition is deduced from the unit mass restriction. In our framework, this restriction is :

$$\int_{\beta_h^*(\theta^*)}^{\infty} g_{k,h}(r|r[k(n-1), h]; \theta, \theta^*) dr = 1.$$

By computing the partial derivative with respect to a parameter of interest a , say, we get:

⁵A condition for the consistency and asymptotic efficiency of EMM is the differentiability of the likelihood function. This condition is not satisfied in our framework.

$$\begin{aligned}
& - \frac{\partial \beta_h^*(\theta^*)}{\partial a} g_{k,h}(\beta_h^*(\theta^*) | r[k(n-1), h]; \theta, \theta^*) \\
& + E_0 \left(\frac{\partial \log g_{k,h}}{\partial a} [r(kn, h) | r(k(n-1), h); \theta; \theta^* | r(k(n-1), h)] \right) = 0.
\end{aligned}$$

The expected score is equal to zero, if

$$g_{k,h}(\beta_h^*(\theta^*) | r[k(n-1), h]; \theta, \theta^*) = 0,$$

that is, if the likelihood is continuous at $\beta_h^*(\theta^*)$, even if it is not differentiable. In our framework the conditional distribution is deduced from a noncentered gamma distribution. By considering the series expansion of the transition density in a neighbourhood of zero, we see that the dominant term involves the power $r^{\delta-1}$; the continuity is satisfied if the degree of freedom $\delta = 2ab/\sigma^2$ is strictly larger than 1. Thus, the ML method ignoring the domain restriction is consistent in our framework.

However, the transition density is differentiable at zero, if δ is strictly larger than 2. Thus, there exist situations, where the standard Gaussian asymptotic theory cannot be applied.

6 Monte-Carlo comparison of ML type estimation methods.

We consider the same Monte-Carlo experiment as in Section 4. For each path, we estimate the parameters a, b, σ (and possibly λ) by applying the following approaches :

- a) the unfeasible ML (UML) method, which estimates parameters a, b , and σ by the exact ML method applied to the instantaneous rate;
- b) the approximate ML (AML) using the 3-month rate as a proxy;
- c) the "exact" ML (EML) method ignoring the domain restrictions based on the 3-month rate;
- d) the modified maximum likelihood (MML) method based on the 3-month rate. This method is defined as follows :

The MML estimator of β_h^* is $\min_n r(kn, h)$ (with $h = 1/4$) The MML estimators of $a, b\beta_h^*$ and $\eta\beta_h^{*1/2}$ are obtained by using the exact discrete time dynamics of $r(t, h)$, in which the observed values are $r(kn, h) - \min_n r(kn, h)$ (the one corresponding to date τ , where $r(\tau, h) = \min_t r(t, h)$, being suppressed). From the ML estimators of $a, \beta_h^*, b\beta_h^*, \eta\beta_h^{*1/2}$ and $\eta a_h^{1/2}$, we can derive the estimators of a, b, σ, λ .

The finite sample bias (B), standard error (SE), least absolute deviation (LAD) and root mean squared error (RMSE) of the estimated mean-reverting parameter are given in Table 3.

Table 3 : Summary statistics of ML type methods

	B	SE	LAD	RMSE
UML	0.099	0.175	0.149	0.201
AML	0.106	0.175	0.152	0.205
EML	0.096	0.176	0.149	0.200
MML	0.025	0.160	0.121	0.161

The MML method based on the 3-month rate is the best, whereas there is no clear ranking among the other methods. The other methods imply a relative bias of about 20 % (=0.1/0.5), and a significant underestimation of serial dependence. The MML method is also better than the different feasible GMM type methods (see Table 1).

Comparing the correct ML methods based on the instantaneous interest rate (UML) and the 3-month rate (MML), we see that the short term interest rates are less informative than the 3-month interest rates. This is easily understood. Indeed, the lower bound of the domain of the short term interest rate is equal to zero, and, thus, bring no information on the parameters. On the contrary, the lower bound on the 3-month rate allows for estimating accurately one of the parameters, and, thus, there is no important effect of the uncertainty of this parameter on the estimated mean-reversion.

These results are confirmed by the shapes of the estimated finite sample p.d.f. of these estimators given in Figure 7. The pdf of the UML, AML¹, EML estimators are similar.

[Insert Figure 7 : Finite Sample Distributions of the Estimators of a].

7 Conclusion

The aim of this paper was to explore in a systematic way the properties of the estimation methods usually employed to estimate parameters of a term structure model. The discussion has been done for the CIR model for expository purpose, but the results are valid for more complicated term structure models with even more pronounced effects [see e.g. Brandt, Chapman (2002), for a comparison of estimation methods for multifactor models]. In particular,

i) A maximum likelihood approach without taking into account parameter dependent domain restrictions can imply inconsistent parameter estimators;

ii) The substitution of the instantaneous interest rate by a 1-month rate is not innocuous and can imply significant asymptotic and finite sample biases on important parameters.

iii) Several consistent methods are not very accurate, and it is difficult to rely on the associated derivative prices or hedging procedures.

iv) Methods using several rates with different time-to-maturity may have much better properties.

REFERENCES

Anderssen, T., and J., Lund (1977) : "Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate", *Journal of Econometrics*, 77, 343-377.

Antoniou, A., Bernales, A., and D., Beuermann (2005) : "The Dynamics of the Short Term Interest Rate in UK", DP Durham Business School.

Barone, E., Cuoco, D., and E., Zautzik (1991) : "Term Structure Estimation of the Cox, Ingersoll, Ross Model : The Case of Italian Treasury Bonds", *Journal of Fixed Income*, 1, 87-95.

Bates, D. (2006) : "Maximum Likelihood Estimation of Latent Affine Processes", *Review of Financial Studies*, 19, 909-965.

Brandt, M., and D., Chapman (2002) "Comparing Multifactor Models of the Term Structure", DP Wharton School.

Broze, L., Scaillet, O., and J.M., Zakoian (1995) : Testing for Continuous Time Models of the Short Term Interest Rate", *Journal of Empirical Finance*, 2, 199-223.

Brown, S., and P., Dybvig (1986) : "The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates", *The Journal of Finance*, 41, 617-630.

Bruche, M. (2004) : "Estimating Structural Bond Pricing Models via Simulated Maximum Likelihood", London School of Economic DP.

Chan, K., Karolyi, G., Longstaff, F., and A., Sanders (1992) : "The Volatility of Short Term Interest Rates : An Empirical Comparison of Alternative Models of the Term Structure of Interest Rates", *Journal of Finance*, 47, 1209-1227.

Chen, S., and C., Tang (2006) : "Parameters Estimation and Bias Correction for Diffusion Processes", Dept of Statistics, Iowa State Univ.

Cox, J., Ingersoll, J., and S., Ross (1985) : "A Theory of the Term

Structure of Interest Rates”, *Econometrica*, 53, 385-408.

Dai, Q., Le, A., and K., Singleton (2006) ”Discrete Time Dynamic Term Structure Models with Generalized Market Prices of Risks”, Working Paper, Stanford Univ.

Daves, P., and M., Ehrhardt (1993) : ”Joint Cross-Section / Time Series Maximum Likelihood Estimation for the Parameters of the Cox-Ingersoll-Ross Bond Pricing Model”, *The Financial Review*, 28, 203-237.

Davidson, J. (1994) : ”Stochastic Limit Theory”, *Advanced Texts in Econometrics*, Oxford Univ. Press.

Duan, J.C. (1994) : ”Maximum Likelihood Estimation Using Price Data of the Derivative Contract”, *Mathematical Finance*, 4, 155-167.

Duan, J.C. (2000) : ”A Correction”, *Mathematical Finance*, 10, 461-462.

Duffie, D., and P., Kan (1996) : ”A Yield Factor Model of Interest Rates”, *Mathematical Finance*, 6, 379-406.

Ericsson, J., and J., Reneby (2005) : ”Estimating Structural Bond Pricing Models”, *Journal of Business*, 78, 707-736.

Fisher, M., and C., Gilles (1996) : ”Estimating Exponential Affine Models of the Term Structure”, DP.

Gallant, R., and G., Tauchen (1998) : ”Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions”, *Journal of the American Statistical Association*, 93, 10-24.

Gibbons, M., and K., Ramaswamy (1993) : ”A Test of the Cox-Ingersoll and Ross Model of the Term Structure”, 6, 619.

Gourieroux, C. (2006) : ”Wishart Process for Stochastic Risk”, *Econometric Reviews, Special Issue on Stochastic Volatility*, 25, 1-41.

Gourieroux, C., and J., Jasiak (2006) : ”Autoregressive Gamma Process”, *Journal of Forecasting*, 25, 129-152.

Gourieroux, C., Jasiak, J., and R., Sufana (2005) : "The Wishart Autoregressive Process of Multivariate Stochastic Volatility", forthcoming Journal of Econometrics.

Gourieroux, C., and A., Monfort (2006) : "Domain Restriction in Interest Rate Models", CREST DP.

Gu, Y., and E., Zivot (2005) : "A Comparison of Univariate Stochastic Volatility Models for US Short Rates Using EMM Estimation", DP Washington Univ.

Hirano, K., and J., Porter (2003) : "Asymptotic Efficiency in Parameter Structural Models with Parameter Dependent Support", *Econometrica*, 71, 1307-1338.

Nowman, K. (1997) : "Gaussian Estimation of Single Factor Continuous Time Models of the Term Structure of Interest Rates", *The Journal of Finance*, 4, 1695-1706.

Nowman, B. (2002) : "The Volatility of Japanese Interest Rates : Evidence for Certificate of Deposit and Gensaki Rate", *International Review of Financial Analysis*, 11, 29-38.

Pearson, N., and T., Sun (1994) : "Exploiting the Conditional Density in Estimating the Term Structure : An Application to the CIR Model", *Journal of Finance*, 46, 1279-1304.

Rogers, L. (1977) : "The Potential Approach to the Term Structure of Interest Rates and Foreign Exchange Rates", *Mathematical Finance*, 7, 157-176.

Singleton, K. (2001) : "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Functions", *Journal of Econometrics*, 102, 111-141.

Singleton, K. (2006) : "Empirical Asset Pricing Models : Specification and Empirical Assessment", Princeton Univ. Press.

Smith, R. (1985) : "Maximum Likelihood Estimation in a Class of Non-

regular Cases”, *Biometrika*, 72, 67-90.

Smith, D. (2002) : ”Markov Switching and Stochastic Volatility Diffusion Models of Short Term Interest Rates”, *Journal of Business and Economic Statistics*, 20, 183-197.

Yu, J., and P., Phillips (2001) : ”Gaussian Estimation of Continuous Time Models of the Short Term Interest Rate”, Cowles Foundation, DP 1309.

Zhou, H (2001) : ”A Study of the Finite Sample Properties of EMM, GMM, QMLE and MLE for a Square-Root Interest Rate Diffusion Model”, *Journal of Computational Finance*, Vol 5.

Appendix 1
Asymptotic properties of weighted least squares

1. General results

Let us consider a regression model :

$$y_n = x_n\theta + u_n, n = 1, \dots, N,$$

where $(y_n, x_n)'$ is a strongly stationary process, (u_n) is a martingale difference sequence such that :

$$E_{n-1}u_n = 0, V_{n-1}u_n = \eta_{n-1}^2,$$

and E_{n-1} (resp. V_{n-1}) denotes the conditional expectation (resp. conditional variance) given $y_{n-1}, y_{n-2}, \dots, x_n, x_{n-1}, x_{n-2}, \dots$. The parameter θ can be estimated by a weighted least squares approach. The estimator is given by :

$$\hat{\theta}_N = \left[\sum_{n=1}^N x_n' x_n \frac{1}{\sigma_{n-1}^2} \right]^{-1} \sum_{n=1}^N x_n' y_n \frac{1}{\sigma_{n-1}^2},$$

where σ_{n-1}^2 is a (stationary) function of $y_{n-1}, y_{n-2}, \dots, x_n, x_{n-1}, x_n$.

We know that :

$$\hat{\theta}_N = \theta + \left(\sum_{n=1}^N x_n' x_n \frac{1}{\sigma_{n-1}^2} \right)^{-1} \left(\sum_{n=1}^N x_n' u_n \frac{1}{\sigma_{n-1}^2} \right).$$

The OLS (resp. GLS) estimator is obtained for $\sigma_{n-1}^2 = 1$ (resp. $\sigma_{n-1}^2 = \eta_{n-1}^2$).

Under standard regularity conditions [see e.g. Davidson (1994)], we get :

$$\sqrt{N}(\hat{\theta}_N - \theta) \sim [E(x_n' x_n)]^{-1} \left(\frac{1}{\sqrt{N}} \sum_{n=1}^N x_n' u_n \frac{1}{\sigma_{n-1}^2} \right).$$

Since $x_n' u_n \frac{1}{\sigma_{n-1}^2}$ is a martingale difference sequence, an appropriate Central Limit Theorem can be applied to get :

$$\sqrt{N}(\hat{\theta}_{OLS,N} - \theta_k) \approx N[0, [E(x_n' x_n)]^{-1} E(x_n' x_n \eta_{n-1}^2) [E(x_n' x_n)]^{-1}].$$

$$\sqrt{N} (\hat{\theta}_{GLS,N} - \theta) \approx N \left[0, \left[E \left(x'_n x_n \frac{1}{\eta_{n-1}^2} \right) \right] \right]^{-1}.$$

The asymptotic variance of the GLS estimator is smaller than the asymptotic variance of any other weighted least squares estimator. For instance, we have

$$\begin{aligned} & V_{as} \left[\sqrt{N} (\hat{\theta}_{OLS,N} - \theta) \right] - V_{as} \left[\sqrt{N} (\hat{\theta}_{GLS,N} - \theta) \right] \\ &= E \left\{ \left[\left[E(x'_n x_n) \right]^{-1} x'_n - \left[E \left(x'_n x_n \frac{1}{\eta_{n-1}^2} \right) \right]^{-1} \frac{x'_n}{\eta_{n-1}^2} \right] \eta_{n-1}^2 \right. \\ & \quad \left. \left[\left[E(x'_n x_n) \right]^{-1} x'_n - \left[E \left(x'_n x_n \frac{1}{\eta_{n-1}^2} \right) \right]^{-1} \frac{x'_n}{\eta_{n-1}^2} \right]' \right\} \gg 0. \end{aligned}$$

To summarize, the Gauss-Markov theorem can also be applied to regression models with lagged endogenous regressions, if we restrict the question to weighted least squares and asymptotic results.

2. OLS estimator of the mean-reverting parameter.

The result above can be applied to the model :

$$y_n = r_{kn} = \theta_{0k} + \theta_{1k} r_{k(n-1)} + u_{kn,k},$$

where $\theta_{1k} = \rho^k$, $\eta_{n-1}^2 = c_k^2 \delta + 2\rho^k c_k r_{k(n-1)}$.

We get :

$$\begin{aligned} & V_{as} \left[\sqrt{N} (\hat{\theta}_{k,OLS,N} - \theta_k) \right] \\ &= \begin{pmatrix} 1 & E(r) \\ E(r) & E(r^2) \end{pmatrix}^{-1} \begin{pmatrix} E(\eta^2) & E(\eta^2 r) \\ E(\eta^2 r) & E(\eta^2 r^2) \end{pmatrix} \begin{pmatrix} 1 & E(r) \\ E(r) & E(r^2) \end{pmatrix}^{-1}. \end{aligned}$$

We deduce that :

$$\begin{aligned}
& V_{as} \left[\sqrt{N} \left(\hat{\theta}_{1k,OLS,N} - \theta_{1k} \right) \right] \\
&= \frac{1}{[V(r)]^2} \{ E(\eta^2)(Er)^2 - 2E(r)E(\eta^2 r) + E(\eta^2 r^2) \}.
\end{aligned}$$

This asymptotic variance involves only marginal moments of interest rate, that is, it is computed with $c_\infty r \sim \gamma(\delta)$.

Since $c_k = c_\infty(1 - \rho^k)$, $\eta^2 = c_k^2 \delta + 2\rho^k c_k r$, we see that the asymptotic variance is homogenous with degree 0 with respect to c_∞ . Thus, we can choose $c_\infty = 1$. We get :

$$\begin{aligned}
& V_{as} \left[\sqrt{N} \left(\hat{\theta}_{1k,OLS,N} - \theta_{1k} \right) \right] \\
&= \frac{1}{(Vr)^2} \left\{ E[(1 - \rho^k)^2 \delta + 2\rho^k(1 - \rho^k)r](Er)^2 \right. \\
&\quad \left. - 2E(r)E[(1 - \rho^k)^2 \delta r + 2\rho^k(1 - \rho^k)r^2] \right. \\
&\quad \left. + E[(1 - \rho^k)^2 \delta r^2 + 2\rho^k(1 - \rho^k)r^3] \right\}.
\end{aligned}$$

The three first marginal moments of r are :

$$Er = \delta, E(r^2) = \delta^2 + \delta, E(r^3) = \delta^3 + 3\delta^2 + 2\delta.$$

By substitution, we get :

$$\begin{aligned}
& V_{as}[\sqrt{N}(\hat{\theta}_{1k,OLS,N} - \theta_{1k})] \\
&= (1 - \rho^k)^2 + 2\rho^k(1 - \rho^k)(1 + 2/\delta).
\end{aligned}$$

By applying the δ -method, the asymptotic variance of the mean-reverting parameter a is :

$$\begin{aligned}
& V_{as} \left[\sqrt{N}(\hat{a}(k, N) - a) \right] \\
&= \frac{1}{k^2} \frac{1}{\rho^{2k}} V_{as} \left[\sqrt{N}(\hat{\theta}_{1k, OLS, N} - \theta_{1k}) \right] \\
&= \frac{1}{k^2} \frac{1}{\rho^{2k}} \left[(1 - \rho^k)^2 + 2\rho^k(1 - \rho^k)(1 + 2/\delta) \right].
\end{aligned}$$

3. GLS estimator of the mean-reverting parameter.

We get :

$$\begin{aligned}
& V_{as} \left[\sqrt{N}(\hat{\theta}_{k, GLS, N} - \theta_k) \right] \\
&= \begin{bmatrix} E(1/\eta^2) & E(r/\eta^2) \\ E(r/\eta^2) & E(r^2/\eta^2) \end{bmatrix}^{-1}.
\end{aligned}$$

We deduce :

$$V_{as} \left[\sqrt{N}(\hat{\theta}_{1k, GLS, N} - \theta_{1k}) \right] = \frac{E(1/\eta^2)}{E(1/\eta^2)E(r^2/\eta^2) - [E(r/\eta^2)]^2}.$$

This function is still homogenous of degree 0 with respect to c_∞ . We get :

$$\begin{aligned}
& V_{as} \left[\sqrt{N}(\hat{\theta}_{1k, GLS, N} - \theta_{1k}) \right] \\
&= (1 - \rho^k)^2 \delta \frac{E\left(\frac{1}{1 + \gamma_k r}\right)}{E\left(\frac{1}{1 + \gamma_k r}\right) E\left(\frac{r^2}{1 + \gamma_k r}\right) - \left[E\left(\frac{r}{1 + \gamma_k r}\right)\right]^2},
\end{aligned}$$

where : $\gamma_k = \frac{2\rho^k}{\delta(1 - \rho^k)}$ and $r \sim \gamma(\delta)$.

Appendix 2

Estimated mean-reverting parameter in the literature

The literature is summarized in Table a.1. The first column refers to the research papers listed among the references. The second column gives the time unit : week (W), month(M), year(Y) or not available (n.a.) The third column gives the type of mean-reverting parameter : historical (H), risk neutral (RN), or non available (n.a.) Column 4 describes the link assumed between the historical and risk-neutral parameters : standard (S), that is, as in the main part of the text, 0, for an assumed zero risk premium,.. Column 5 mentions the database, and the period of observations. Column 6 and column 7 give the observation frequency and time-to-maturity, respectively. The estimation method is given in column 8 : OLS, GLS, characteristic function (CF), Efficient Method of Moments (EMM), Maximum Likelihood (ML)..., and the estimated values of δ and a in columns 9 and 10, respectively. "With proxy" means that an estimation method is used after direct substitution of the instantaneous rate by a 3-month rate for instance. The methods with proxy are not consistent. For comparison we have also provided estimation results based on either cross sectional approaches, or the Chan, Karolyi, Longstaff, Sanders (CKLS) model, which extend the CIR model by allowing any power volatility. In both cases, the δ parameter cannot be identified or defined (n.d).

Table a.1 : Estimated Mean-Reverting Parameter

Research	t.u.	H / RN	link	data base	freq.	h	estim	δ	\hat{a}
GT (98)	?	H	n.a.	STATLIB 92-96	W	3-month.	EMM with proxy	8.8	0.006
CKLS(92)	?	H	n.a.	CRSP 64 – 89	M	1-month	QML	4.97	0.23
S(02)	?	H	n.a.	? 64 – 89	M	?	OLS	n.a.	0.05
PS(94)		H	n.a.	CRSP 71-86	M	portf. nominal	ML, proxy	1.8	0.87
ABB(2005)	?	H	n.a.	UK. Datastream 83 – 03	W	1-month	QML, CED proxy	325	0.0015
N(2002)	?	H	n.a.	Japan CD	?	1-month	QML, CED proxy	1.5	0.088
N(2002)	?	H	n.a.	Japan CD	W	3-month	QML, CED proxy	0.099	0.049
N(2002)	?	H	n.a.	Gensaki	W	1-week	QML, CED proxy	0.67	0.055
N(2002)	?	H	n.a.	Gensaki	?	1-month	QML, CED proxy	0.67	0.049
N(97)	Y	H	n.a.	CRSP 64 – 89	M	1-month	QML, CED proxy	6.2	0.037
N(97)	?	H	n.a.	UK 75 – 95	M	1-month	QML, CED proxy	8.3	0.028
BSZ (95)	?	H	n.a.	CRSP 72 – 91	M	1-month	QML, CED proxy	9.9	0.04
BCZ(91)	?	RN	S	Milan	D	several	NLLS Cross.Sect.	n.d.	0.24
CKLS(92)	?	H	n.a.	CRSP 64 – 89	M	1-month	QML CED, proxy CKLS	n.d.	0.59
S(02)	?	H	n.a.	? 64 – 89	M	?	QMLE CED; Proxy CKLS	n.d.	0.03
GZ(06)	?	H	n.a.	Fed.Res 54 – 04	W	3-month	EMM with proxy CKLS	n.d.	0.19
YP(01)	?	H	n.a.	UK 75 – 95	M	1-month	QMLE CED; Proxy CKLS	n.d.	0.35
YP(01)	?	H	n.a.	UK 75 – 95	M	1-month	QMLE Proxy CKLS	n.d.	0.34
YP(01)	?	H	n.a.	CRSP 64 – 89	M	1-month	QMLE CED; Proxy CKLS	n.d.	0.33
YP(01)	?	H	n.a.	CRSP 64 – 89	M	1-month	QMLE Proxy CKLS	n.d.	0.33

