Generalized empirical likelihood for a continuum of moment conditions

Pierre Chaussé *
Université du Québec à Montréal
Département des sciences économiques

May 2007
(preliminary and incomplete)

Abstract

This paper extends the generalized empirical likelihood method to the case in which the moment conditions are defined on a continuum (C-GEL). Using a continuum of moment conditions is a good alternative for estimating continuous time processes, Euler equations in the frequency domain or to deal with the many instruments problem. We show that the method presented in this paper shares the same asymptotic properties as the GMM for a continuum (C-GMM) of Carrasco and Florens (2000). We show also how we can use the C-GEL framework to obtain the continuously updated estimator for a continuum. In oder to compute C-GEL, which requires us to solve a nonlinear ill-posed problem, we show how to apply the regularized Gauss-Newton method to our problem. We conclude with an example in which we estimate the parameters of a diffusion process and compare the results with C-GMM.

Classification JEL: C13, C30

1 Introduction

The generalized method of moment (GMM) has become a popular alternative estimation method over the maximum likelihood based methods (MLE), mainly because it requires much less prior knowledge of the data generating processes. The asymptotic and final sample properties of GMM has been studied intensively since Hansen (1982) brought it to our attention and we can now find countless applications in almost every area of economics and finance. Recently, Carrasco and Florens (2000) extended the area in which GMM can be used by allowing moment conditions to be defined on a continuous support (C-GMM).

*I would like to thank my advisor Alain Guay, Douglas Hodgson and Marine Carrasco for valuable discussions.

So far, Carrasco, Chernov, Florens and Ghysels (2006) are the only ones to have provided us with applications. They estimate the parameters of few affine diffusion processes by using moment conditions based on their conditional characteristic functions (CCF) as suggested by Singleton (2001). Since CCF's are defined on $\mathbb{R}$, we need a method like C-GMM in order to use all the information included in the moment conditions. In particular, Carrasco and Florens (2002) show that C-GMM with CCF as moment conditions can reach the MLE efficiency bound because in that case the score is spanned by the moment conditions.

The business cycle literature could also benefit from methods that deal with continuum of moment conditions. Berkowitz (2001) proposes to estimate Euler equations based on conditions in the frequency domain. According to Euler equations, the error term that appears once we drop the conditional expectation operator is a martingale difference sequence which implies that its spectral density function is constant at all frequencies. We can therefore define the estimator as being the parameter values which minimize the distance between the empirical spectral density and its theoretical counterpart for all frequencies included in a certain interval. Berkowitz (2001) applies this approach to estimate a consumption based CAPM model while Bindelli (2005) estimates the New Keynesian Phillips Curve of several countries using both the interval $[0, \pi]$. As an alternative interval, we could instead consider only the frequencies which are associated with the business cycle. This possibility is worth exploring if it allows us to estimate Euler equations without having to go through a smoothing process to extract the cycle from the series. But, this will be the subject of a future paper. For now, we need to work on developping the necessary tools and, since this method does not provide us with a specification test and does not weight the moment conditions properly, C-GMM should be a better one. At least for now.

So far, the known properties of C-GMM are limited to its asymptotic behaviors. The only study of its finite sample properties is the one by Carrasco and Florens (2002) in which they derive the second order approximation of the solution in order to choose the optimal Tikhonov regularization parameter. However, their expansion is limited to the second order bias. My first guess would be to say that the properties of GMM does not depend on the support on which is defined the moment function. If that is the case, the generalized empirical likelihood estimator based on the same continuum of moment conditions (C-GEL) should perform better than GMM in finite sample. The superiority of some GEL methods over GMM has been demonstrated theoretically for finite number of moment conditions by Newey and Smith (2004) for the iid case and by Anatolyev (2005) for the case of weakly dependent processes, using second order expansions. They show that part of the extra bias of the GMM estimators comes from both the estimation of the optimal weight matrix and the estimation of the initial value of the parameters vector. The same prior estimations are required for the C-GMM which makes us think that we should obtain similar

results if we compare it with C-GEL based methods. Another important result is that the bias of empirical likelihood (EL), which is a special case of GEL, does not seem to grow with the number of instruments which is an attractive properties when we consider methods that deal with an infinity of moment conditions.

It is far from being straightforward to compare the finite sample properties of C-GMM and C-GEL and we have much to learn before we can make such analysis. The intuition for the superiority of EL over GMM that has been presented in the literature cannot easily be transfered to the case of a continuum of moment conditions. In fact, the ill-posed problem that needs to be be dealt with makes C-GMM quite different from GMM. The first step is to define the C-GEL and to demonstrate that it shares the same asymptotic properties as the C-GMM. To follow the contribution of Carrasco et al. (2006), we will consider the case of weakly dependent processes. We all know that the main cost of GEL with respect to GMM is the numerical difficulties of solving a saddle point problem. It is therefore expected that C-GEL will face the same computational challenge. In fact, while we have to solve a linear ill-posed problem in order to define the C-GMM objective function, the first order conditions of C-GEL require us to solve a nonlinear ill-posed problem. The second objective of this paper is therefore to present some numerical methods that are required when facing such problem. Finally, the C-GEL will be applied to the CIR square root process that has been estimated with C-GMM by Carrasco et al. (2006). Although this specific example cannot be used to compare the properties of C-GMM and C-GEL in general, it will give us a first impression and help us understand how C-GEL can be implemented in practice.

2 The C-GEL method

The moment conditions that serve as estimating equations for the vector of parameters $\theta_0 \in \Theta$ takes the following general form:

$$E^{P_0}\left[g(x,\tau;\theta_0)\right] = 0 \qquad \forall\ \tau \in \Lambda \tag{1}$$

where Λ defines the structure of the moment conditions. If the space is finitely countable, then we have finite number of conditions. If it is $\mathbb{R}^p$, as it would be the case if we use the CCF, or $[0,\pi]^p$, in the case of moment conditions in the frequency domain, then we have a continuum of moment conditions. Although we could make it more general by allowing all kinds of structure, which would be possible if we modify the measure used in inner products, we will consider only cases in which Λ is continuous. The assumptions on the moment function and on the random variable x, which allows us to apply the law of large number and the functional central limit theorem for weakly dependent processes, follow Carrasco et al. (2006). The reader is invited to consult their paper for the justifications. The following assumption summarizes them:

Assumption 1:
(i) the process x_t, which takes values in the measurable space $(S, \mathcal{S})$, is $\alpha-$mixing with $\sum_{j=1}^{\infty} \alpha_j j^2 < \infty$.
(ii) Θ is compact,
(iii) $g(x_t, \tau; \theta) \in L^2(\pi)$ for all $\theta \in \Theta$ and $x_t \in S$ where $L^2(\pi)$ is the Hilbert space of square integrable functions equipped with the inner product $< f, g >= \int f(\tau) g(\tau) \pi(\tau) d\tau$ and $\pi(\tau)$ is absolutely continuous with respect to the Lebesgue measure.
(iv)$g(x_t, \tau; \theta)$ is continuous with respect to θ and absolutely bounded for all $x_t \in S$ and $\tau \in \Lambda$.
(v) θ_0 is the unique solution of the system defined by equation (1) almost everywhere-π.
(vi) $\sup_\theta \|D_\theta g_t(\theta)\| < \infty$

The integrating density $\pi(\tau)$ allows us to expand the space of admissible functions. In other words, some functions that are not square integrable in the usual L^2 space are square integrable in $L^2(\pi)$. The choice of $\pi(\tau)$ seems to have an impact on the finite sample properties of C-GMM and C-GEL. In fact, it determines the subspace of Λ on which we want to focus to estimate θ. For example, if $\Lambda \equiv \mathbb{R}$ and $\pi(\tau)$ is the standard normal density, only information on the moment conditions $g(x_t, \tau, \theta)$ for $\tau \in [-5, 5]$ becomes relevant for the estimation. Furthermore, it gives more importance to moment conditions near $\tau = 0$. We still have infinitely many moment conditions but the choice of the subspace of Λ is arbitrary and the consequent on the properties of the estimator needs to be analyzed further in future research. For now, we will justify the choice of the integrating density by the fact that it will allow use to integrate analytically some of the integrals that appear in the objective function. It is a significant advantage regarding numerical accuracy and the reduction of computational time especially when multiple integrations are needed. We will come back to this aspect in the section on numerical applications bellow.

As it has been shown by Smith (1997) and Kitamura (1997), we need to modify the estimating equations when the stochastic process is weakly dependent in order for GEL to reach asymptotic efficiency. There are more than one way to make that adjustment. According to Anatolyev (2005), the best approach consists in smoothing the moment conditions using an appropriate kernel. He shows that this adjustment produces the smallest asymptotic bias. Moreover, smoothing the moment conditions seems to reduce the bias even if the stochastic process is iid. We will therefore replace the moment function $g(x_t, \tau; \theta)$ by the following:

$$g^\kappa(x_t, \tau; \theta) = \sum_{j=-m}^{m} \kappa(j) g(x_{t-j}, \tau; \theta) \tag{2}$$

where m must converge to infinity at a rate slower than n. See Smith (1997), Kitamura (1997), Andrews (1991) and Newey and West (1987a) for more details on how to choose the smooth transformation. Anatolyev (2005) show also that the choice of kernel can help reduce the bias further. If there is no serial cor-

relation, then the weight function $\kappa()$ is the Dirac delta function. The moment conditions given by equation (1) becomes:

$$E^{P_0}\Big[g^{\kappa}(x,\tau;\theta)\Big] = 0 \tag{3}$$

Carrasco et al. (2006) consider such serial correlation for the C-GMM. One of the implication is that the covariance operator K which is introduced by Carrasco and Florens (2000) with the kernel

$$k(\tau_1,\tau_2) = E^{P_0}[g(x,\tau_1;\theta)\overline{g(x,\tau_2;\theta)}]$$

is now defined with the following kernel:

$$k(\tau_1,\tau_2) = \sum_{j=-\infty}^{\infty} E^{P_0}[g(x_t,\tau_1;\theta)\overline{g(x_{t-j},\tau_2;\theta)}] \tag{4}$$

where $\overline{g()}$ means complex conjugate. The estimation of this kernel is in the same spirit as the estimation of the HAC covariance matrix. So, the choice of $\kappa()$ in equation(3) most be such that:

$$\frac{1}{n}\sum_{t}[g^{\kappa}(x_t,\tau_1;\theta)\overline{g^{\kappa}(x_t,\tau_2;\theta)}] \Longrightarrow \sum_{j=-\infty}^{\infty} E^{P_0}[g(x_t,\tau_1;\theta)\overline{g(x_{t-j},\tau_2;\theta)}]$$

as m and n go to infinity. See Carrasco et al. (2006) for more details. The following assumption is required to derive the properties of C-GEL and to assure the existence of a sequence of operators which converges to K:

Assumption 2:*(i) The operator K with kernel defined in equation(4) is an Hilbert-Schmidt operator. (ii) The null space of K is $\{0\}$. which implies, since it is a covariance operator, that all its eigenvalues are strictly positive.*

We define the C-GEL estimator $\hat{\theta}$ as follows:

$$\hat{\theta} = \arg\min_{\Theta}\sup_{\Omega_n(\theta)} P_n(\theta,\lambda) = \sum_{t=1}^{n}\rho\left(<\lambda, g_t^k(\theta)>\right) \tag{5}$$

where $g_t^k(\theta) \equiv g^k(x_t,.,\theta)$, $\rho(.)$ is a concave function defined on the set $\mathcal{V}$ containing 0, $\Omega_n(\theta) = \{\lambda :< \lambda, g_t^k(\theta) >\in \mathcal{V}, t = 1,\cdots,n\} \subset L^2(\pi)$,

$$< \lambda, g_t^k(\theta) >= \int_{\Lambda} g^k(x_t,\tau;\theta)\lambda(\tau)\pi(\tau)d\tau \tag{6}$$

and λ represents a Lagrange function. It measures the level of restriction imposed by the moment conditions. If they are easily satisfied by the sample, the estimation of λ should be close to zero. It could therefore be a useful tool for analyzing the specifications of the model we want to estimate. We could see,

for example, which subspace of Λ is responsable for the rejection of a model. In particular, if we estimate an Euler equation in the frequency domain and reject the model, we could see which frequencies are responsable.
The difference between GEL and C-GEL is the structure of λ. In our case, we have, for a given θ, to solve an optimization problem in a function space. One way to deal will such problem is to work directly with first order conditions that we apply to operators such as Fréchet derivatives. This approach results in a slightly more compact representation of the solution and allows us to have a notation that is compatible with the one used by Carrasco and Florens (2000). In appendix, we give a definition of a Fréchet derivative and proof some results relative to it that will be used throughout the paper. For a more complete coverage of optimization in function space, see Luenberger (1997). The first order conditions can be written as follows:

$$D_\lambda P_n(\theta, \lambda) = 0 \tag{7}$$

$$D_\theta P_n(\theta, \lambda(\theta)) = 0 \tag{8}$$

where $\lambda(\theta)$ is the solution of the system (7) and $D_\lambda P_n()$ is the Fréchet derivative of $P_n(\theta, \lambda)$. It is an operator from $L^2(\pi)$ to $\mathbb{R}$. To see that, let h be an element of $L^2(\pi)$, then:

$$\begin{aligned} D_\lambda P_n(\theta, \lambda)h &\equiv \left(\sum_{t=1}^{n} \rho'(< g_t^k(\theta), \lambda >) g_t^k(\theta) \right) h \\ &= \int_\Lambda \sum_{t=1}^{n} \rho'(< g_t^k(\theta), \lambda >) g_t^k(\theta, \tau) h(\tau) \pi(\tau) d\tau \end{aligned}$$

The biggest challenge is to solve the system (7) which can be seen as a nonlinear operator from $L^2(\pi)$ to $L^2(\pi)$. In order to clarify what it means, it is better to consider specific C-GEL cases.
According to Anatolyev (2005) and Newey and Smith (2004), the GEL performs better in finite sample. Among all GEL class of estimators, they focussed on the continuous updating GMM estimator (CUE) of Hansen, Heaton and Yaron (1996), which can be obtain if we define $\rho(v)$ as a quadratic function, and the empirical likelihood estimator (EL) of Owen (2001) for $\rho(v) = \ln(1 - v)$. The advantage of these two methods over GMM is their relatively smaller second order asymptotic bias. In the following subsection, we therefore define the C-GEL for those two specific cases.

2.1 The continuous updating GMM for a continuum of moment conditions (C-CUE)

Before introducing the C-CUE, we will review the C-GMM which is a restricted version of it. C-GMM is a two step GMM for a continuum of moment conditions. Its estimator is defined as:

$$\hat{\theta}_{cgmm} = \arg\min_{\Theta} \| K_n^{\alpha_n}(\tilde{\theta})^{-1/2} \hat{g}(\theta) \| \tag{9}$$

where $\hat{g}(\theta) = (1/n)\sum_t g(x_t, ., \theta)$, $\|.\|$ is the norm in $L^2(\pi)$,

$$K_n^{\alpha_n}(\tilde{\theta})^{-1/2} = \left[\left(K(\tilde{\theta})^2 + \alpha_n I\right)^{-1} \hat{K}(\tilde{\theta})\right]^{1/2},$$

$K(\tilde{\theta})$ is the covariance operator with kernel equals to a consistent estimator of (4), which could be any HAC estimator, and $\tilde{\theta}$ is a first step consistent estimator of θ. Usually it is obtain by replacing $K_n^{\alpha_n}(\tilde{\theta})^{-1/2}$ by the identity operator. α_n is a regularization parameters which smooths the inversion of the covariance operator. More will be said about this parameter bellow.
Another way to define the estimator is to modify the Hilbert space and its inner product. It simplifies the notation and defines more precisely the space of functions that satisfy the regularity conditions. The new space is called the reproducing kernel Hilbert space (RKHS) $\mathcal{H}(K)$ and is defined by the covariance operator K. All we need to know about this space is that the inner product is defined as [1]:

$$< f, g >_{\mathcal{H}(k)} = \left\langle K^{-1/2} f, K^{-1/2} g \right\rangle_{L^2} \tag{10}$$

and that it restricts the $L^2(\pi)$ space to functions for which the $L^2(\pi)$ norm of $K^{-1/2}f$ is bounded. We can rewrite the objective function as follows:

$$\hat{\theta}_{cgmm} = \arg\min_{\Theta} \|\hat{g}(\theta)\|_{\mathcal{H}(K_n^{\alpha_n}(\tilde{\theta}))}$$

We know that the CUE does not require a prior estimation of the matrix of weights. It is one reason why it has a smaller bias than GMM. We let the algorithm compute the matrix of weights and θ simultaneously. Similarly, we can define the C-CUE estimator $\hat{\theta}_{ccue}$ as follows:

$$\hat{\theta}_{ccue} = \arg\min_{\Theta} \| \hat{g}(\theta) \|^2_{\mathcal{H}(k_n^{\alpha_n}(\theta))} \tag{11}$$

The following theorem and the proof follow Newey and Smith (2004):

Theorem 1: *If (i) $K_n^{\alpha_n}(\tilde{\theta})^{-1}$ converges to K^{-1} as n goes to infinity and that (ii) $\rho(v)$ is a quadratic function of v, then,*

$$\hat{\theta}_{ccue} = \arg\min_{\Theta} \left(a_1 + a_2 \| \hat{g}(\theta) \|^2_{\mathcal{H}(k_n^{\alpha_n}(\theta))} + o_p(1)\right) \tag{12}$$

where $a_2 > 0$ and a_1 is some bounded constant.

See the appendix for the proof. Notice that the conditions (i) has been limited to a convergence requirement. All the conditions that are needed are enumerated in Carrasco et al. (2006). It depends mainly on the speed of convergence of α_n. We dont want the inverse to become unstable before n reaches infinity. Therefore, α_n needs to go to zero slower than n goes to infinity. The choice of kernel for the HAC estimation affects the rate of convergence but does not prevent it. It follows from the above theorem that

[1] For more details, see Parzen (1970)

2.2 The empirical likelihood method for a continuum of moment conditions (C-EL)

We begin by presenting the EL method using the approach presented by Qin and Lawless (1994). As it is pointed out by Newey and Smith (2004), it represents one of the dual problems of GEL and it gives us a good interpretation of the superiority of GEL over GMM. Let us rewrite the condition on the continuum of estimating equations $g(x, \tau, \theta)$ as follows:

$$E^{P_0} g(X, \tau, \theta_0) \equiv \int_S g(X, \tau, \theta_0) dP_0(X) = 0 \quad \forall\ \tau \in \Lambda \tag{13}$$

where from now on, we omit the superscript k over $g_t()$ for simplicity. As opposed to the C-GMM, the C-EL method seeks an estimator not only for the vector $\theta_0 \in \Theta$ but also for the probability distribution P_0. The difference between this condition and the one presented by Qin and Lawless (1994) is the dimension of the moment condition vector which is no longer finite as it is implied by the structure of Λ. We suppose as above that $g_t(\theta) \in L^2(\pi)\ \forall \theta \in \Theta$. All inner products in this section are defined in $L^2(\pi)$ as

$$< f, g >= \int_\Lambda f(\tau) g(\tau) \pi(\tau) d\tau$$

which defines a norm $\| f \| =< f, f >^{1/2}$. We want to maximize the likelihood-ratio, which compares the likelihood $\prod_{t=1}^n dF(x_t) = \prod_{t=1}^n p_t$ with the likelihood computed from the empirical density $d\hat{F} = 1/n$, subject to the constraint given by the above estimating equations and that $\sum_t p_t = 1$. We therefore face the following problem:

$$\max_{\theta, \lambda, \mu} \mathcal{L}(\theta, \mu, \lambda) = \sum_{t=1}^n \log(p_t) + \mu \left(1 - \sum_{t=1}^n p_t\right) + \left\langle \lambda, \left(\sum_{t=1}^n p_t g(x_t, \theta)\right)\right\rangle \tag{14}$$

where:

$$\left\langle \lambda, \left(\sum_{t=1}^n p_t g(\theta, x_t)\right)\right\rangle = \int_\Lambda \lambda(\tau) \left(\sum_{t=1}^n p_t g(x_t, \tau, \theta)\right) d\tau$$

It is well known that this problem is equivalent to the following C-GEL representation, which is easy to verify since they imply the same first order conditions:

$$\min_\theta \sup_\lambda \frac{1}{n} \sum_{t=1}^n \log\left(1 - \langle g_t(\theta), \lambda \rangle\right)$$

The first order conditions are:

$$\frac{1}{n} \sum_{t=1}^n \frac{1}{1 - < \lambda, g(x_t, \theta) >} g(x_t, \theta) = 0 \tag{15}$$

$$\frac{1}{n} \sum_{t=1}^n \frac{1}{1 - < \lambda, g(x_t, \theta) >} \left\langle \lambda, \left(\frac{\partial g(x_t, \theta)}{\partial \theta}\right)\right\rangle = 0 \tag{16}$$

To simplify we will adopt the linear operator notation in the rest of the article. As long as we know in which space we are, the effect of the operator will be clear. For example, if $A \in (\Lambda \times \mathbb{R}^{n+m})$ with typical element $a_{ij}(\tau)$ and $d \in \mathbb{R}^m$, then in Ad, the operator A is from $\mathbb{R}^m \to \Lambda \times \mathbb{R}^n$) and in Af, for $f \in \Lambda$, the operator is from $\Lambda \to \mathbb{R}^{n \times m}$.
We can rewrite the first order conditions using this notation:

$$\frac{1}{n}\sum_{t=1}^{n} \frac{1}{1 - g_t(\theta)\lambda} g_t(\theta) = 0 \tag{17}$$

$$\frac{1}{n}\sum_{t=1}^{n} \frac{1}{1 - g_t(\theta)\lambda} \left(\frac{\partial g_t(\theta)}{\partial \theta}\right) \lambda = 0 \tag{18}$$

To analyze the asymptotic properties of the solution of this problem, we expand the first order conditions around the true function $\lambda_0 = 0$ (see the appendix for that proof) and $\theta = \theta_0$. To simplify the notation let $Q_1(\lambda, \theta)$ and $Q_2(\lambda, \theta)$ be the first order conditions given by the equations (17) et (18) respectively. Therefore,

$$0 = Q_1(0, \theta_0) + \left(D^1_\lambda \hat{\lambda}\right) + \left(D_\theta(\hat{\theta} - \theta_0)\right) + 0_p(\| \hat{\lambda} \|^2 + \| \hat{\theta} - \theta_0 \|^2) \tag{19}$$

$$0 = \left(D^2_\lambda \hat{\lambda}\right)(k) + 0_p(\| \hat{\lambda} \|^2 + \| \hat{\theta} - \theta_0 \|^2) \tag{20}$$

where the order of the Taylor expansion remainders are derived in the appendix. The above equations have been written using the operator notation . D^1_λ is an operator from $L^2(\pi)$ to $L^2(\pi)$ with the kernel $D^1_\lambda(\tau_1, \tau_2)$ defined as:

$$D^1_\lambda(\tau_1, \tau_2) = \frac{1}{n}\sum_{t=1}^{n} g(x_t, \tau_1, \theta_0)\overline{g(x_t, \tau_2, \theta)}$$

It is therefore the estimator, $\hat{K}_0$, of the covariance operator K evaluated at $\theta = \theta_0$. It implies that:

$$\left(D^1_\lambda \hat{\lambda}\right)(\tau_1) \equiv (\hat{K}\hat{\lambda})(\tau_1) = \int_\Lambda \frac{1}{n}\sum_{t=1}^{n} g(x_t, \tau_1, \theta_0)\overline{g(x_t, \tau_2, \theta_0)}\hat{\lambda}(\tau_2)\pi(\tau_2)d\tau_2$$

The second operator, D_θ is a linear operator from the Hilbert q-dimensional vector space H to $L^2(\pi)$ with the kernel:

$$D_\theta(\tau, k) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial g(x_t, \tau; \theta_0)}{\partial \theta_k}$$

where we made the simplifying assumption that $\Theta = \mathbb{R}^q$. For convenience, we will denote this operator as $\hat{G}$ which follows the usual convention for the first

[1] For a good introduction to operator theory, see Groetsch (1993)

derivative of the moment function. Therefore,

$$\left(\hat{G}(\hat{\theta}-\theta_0)\right)(\tau)=\sum_{k=1}^{q}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(x_t,\tau,\theta_0)}{\partial\theta_k}\right][\hat{\theta}-\theta_0]_k$$

Le third operator D^2_λ is a linear operator from $L^2(\pi)$ to H with the following kernel:

$$D^2_\lambda(k,t)=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g_t(\theta_0,x_i)}{\partial\theta_k}$$

Which is very similar to $\hat{G}$. The only difference is that the domain of $\hat{G}$ is the image of D^2_λ and vice versa. In fact, we can see that the operator $\hat{G}$ is the adjoint of D^2_λ. That is, for all $f\in L^2(\pi)$ and $\delta\in H$,

$$\begin{aligned}
<D_\theta\delta,f>_{L^2} &= \int_\Lambda\left[\sum_{k=1}^{q}\left(\sum_{t=1}^{n}\frac{\partial g(x_t,\tau,\theta_0)}{\partial\theta_k}\right)\delta_k\right]f(\tau)\pi(\tau)d\tau\\
&= \sum_{k=1}^{q}\left[\int_\Lambda\left(\sum_{t=1}^{n}\frac{\partial g(x_t,\tau,\theta_0)}{\partial\theta_k}\right)f(\tau)\pi(\tau)d\tau\right]\delta_k\\
&= <\delta,D^2_\lambda f>_H
\end{aligned}$$

where $<,>_{L^2}$ and $<,>_H$ are the inner products in $L^2(\pi)$ and H respectively. We will therefore define the operator $D^2_\lambda\equiv\hat{G}^*$ to follow the usual notation for adjoint operators. It follows that:

$$\left(\hat{G}^*\hat{\lambda}\right)(k)=\int_\Lambda\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(x_t,\tau,\theta_0)}{\partial\theta_k}\right]\hat{\lambda}(\tau)\pi(\tau)d\tau$$

We can rewrite equations (19) and (20) in matrix form and replace replace $Q_1(0,\theta_0)$ by its value $(1/n)\sum_t g(x_t,\theta_0)\equiv\hat{g}(\theta_0)$.

$$\begin{pmatrix}-\hat{g}(\theta_0)+O_p(n^{-1})\\ O_p(n^{-1})\end{pmatrix}=\begin{pmatrix}\hat{K}_0 & \hat{G}\\ \hat{G}^* & 0\end{pmatrix}\begin{pmatrix}\hat{\lambda}\\ \hat{\theta}-\theta_0\end{pmatrix}\tag{21}$$

We could solve the system if the matrix was invertible and we would get the following solution:

$$\begin{pmatrix}\hat{\lambda}\\ \hat{\theta}-\theta_0\end{pmatrix}=\begin{pmatrix}\hat{K}_0 & \hat{G}\\ \hat{G}^* & 0\end{pmatrix}^{-1}\begin{pmatrix}-\hat{g}(\theta_0)+O_p(n^{-1})\\ O_p(n^{-1})\end{pmatrix}\tag{22}$$

If we were working in a vector space, the only invertible operator would be $\hat{K}_0$ because it would be the only square matrix. However, we know that it represents the estimator of the covariance operator of $g(x_t,\theta_0)$ and, from Carrasco and Florens (2000), that it is not invertible. If we solve the system using the regularization approach of Tikhonov (see the appendix), we obtain:

$$\hat{\lambda}=-\left(K^{\alpha_n}_{0n}\right)^{-1}\left\{I-\hat{G}\left[\hat{G}^*\left(K^{\alpha_n}_{0n}\right)^{-1}\hat{G}\right]^{-1}\hat{G}^*\left(K^{\alpha_n}_{0n}\right)^{-1}\right\}\hat{g}(\theta_0)+O_p(n^{-1})\tag{23}$$

and

$$(\hat{\theta} - \theta_0) = -\left[\hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1} \hat{G}\right]^{-1} \hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1} \hat{g}(\theta_0) + O_p(n^{-1}) \tag{24}$$

where $\left(K_{0n}^{\alpha_n}\right)^{-1}$ is the regularized inverse of $\hat{K}_0$.

2.3 Properties of $\hat{\lambda}$ and $\hat{\theta}$

In the appendix, it is shown that $\|\hat{\lambda}\| = O_p(n^{-1/2})$. We now need to derive its asymptotic distribution. Because λ measures how restrictive is the constraint $\sum_t p_t g_t(\theta_0) = 0$, we can derive a statistics from it to test whether the model is correctly specified. If the assumption 1 is satisfied, we can use the same functional central limit theorem used by Carrasco et al. (2006) which allow us to say:

$$\sqrt{n}\hat{g}(\theta_0) \Rightarrow N(0, K)$$

where $\Rightarrow$ means convergence in law and K is the covariance operator with kernel which is given by equation (4). Remember that even if there is no superscript k to $g_t()$, we suppose as always that the process is weakly dependent and that $g_t()$ is a smoothed transformation of the moment conditions. If we omit the last term that goes to zero in equation (23), we can easily show that:

$$\sqrt{n}\lambda \Longrightarrow N(0, V) \tag{25}$$

where

$$V = \left\{I - K^{-1}G\left(G^*K^{-1}G\right)^{-1}G^*\right\}K^{-1} \tag{26}$$

where

$$G = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \frac{\partial g_t(\theta_0)}{\partial \theta}$$

and G^* is the adjoint of G. Similarly, we have the following result for $\hat{\theta}$

$$\sqrt{n}(\hat{\theta}_\alpha - \theta_0) \Longrightarrow N(0, W)$$

with

$$W = \left\{G^*K^{-1}G\right\}^{-1}$$

So the covariance matrix converge to the GMM covariance matrix $(G^*K^{-1}G)^{-1}$ which can be rewritten as $\|G\|_K^{-2}$ to match the results of Carrasco and Florens (2000). Therefore, as for the discrete case, C-GMM and C-EL estimators share the same asymptotic properties.

3 Estimation procedures

In this section, we analyze different estimation procedures. We have seen in the previous section that the asymptotic properties of the C-EL estimators are similar to the discrete case as long as we choose correctly the regularization

parameter α that is used to invert the covariance operator. It suggests that the estimation procedure has an impact on both the asymptotic and the finite sample properties of the estimator.
The first method uses the Taylor approximation of the estimator of λ while the second is based on the estimation of lambda using a dynamical system method. The latter is an iterated procedure that converges to the solution of nonlinear operator equations when certain regularity conditions are satisfied.

3.1 Taylor approximation

There are many ways to solve this optimization problem. Carrasco and Florens (2000) propose to use the singular value decomposition approach while Carrasco et al. (2006) propose a way to rewrite the objective function as products of vectors and matrices. It is worth considering both methods and compare them. The first can hardly be applied to the iterative procedure, so it will be presented in this section. But first, we need to review some concepts related to operator theory. For more details, see Groetsch (1993).

No iid restriction needs to be imposed on the data generating process. So it may still be weakly dependent. In that case, however, the moment function needs to be smoothed before we start. In order to describe this approach, we will present the C-EL estimator as the solution to the following saddle point problem.

$$\hat{\theta} = arg \min_{\Theta} \left[\sup_{\lambda \in L^2(\pi)} \sum_{t=1}^{n} \rho(< \lambda, g(x_t, \theta) >) \right]$$

where $\rho(v) = \log(1 - v)$. The solution $\lambda(\theta)$ of the sup () problem can be approximated by a first order Taylor expansion about $\lambda = 0$ as long as $\theta \in b(\delta_n, \theta_0)$, which is a ball centered at θ_0 with radius equals to δ_n which converges to zero as n goes to infinity. In other words, for that approximation to be good enough, we need to be not too far from the true value θ_0. It can be shown that it implies:

$$K_n \lambda = -\hat{g}(\theta)$$

where K_n is a covariance operator with kernel

$$k_n(\tau_1, \tau_2) = \frac{1}{n} \sum_{t=1}^{n} g(x_t, \tau_1, \theta) g(x_t, \tau_2, \theta)$$

which is a consistent estimator of K, a self adjoint operator with infinite dimensional range $R(K)$. If we want to solve $Kx = y$ for $x, y \in L^2(\pi)$ we can use the singular system (ν_i, μ_i) of the operator K, where ν_i is an orthonormal eigenfunction and μ_i the associated singular value. Because the dimension of $R(K)$ is infinite, there are infinitely many singular values. Furthermore, these eigenfunctions are complete in $R(K^2) = N(K)^{\perp}$ where $N(K)$ is the null space

of K. It implies that for any $f \in R(K)$:

$$f = \sum_{i=1}^{\infty} < f, \nu_i > \nu_i$$

We can easily see that any solution $\tilde{x}$ of $Kx = y$ has the following form:

$$\tilde{x} = \sum_{i=1}^{\infty} \frac{1}{\mu_i} < y, \nu_i > \nu_i + \varphi$$

where $\varphi \in N(K)$, because:

$$\begin{aligned} K\tilde{x} &= \sum_{i=1}^{\infty} \frac{1}{\mu_i} < y, \nu_i > K\nu_i + K\varphi \\ &= \sum_{i=1}^{\infty} \frac{1}{\mu_i} < y, \nu_i > \mu_i \nu_i + 0 \\ &= \sum_{i=1}^{\infty} < y, \nu_i > \nu_i = y \end{aligned}$$

if $y \in R(K)$. The solution which has minimum norm is:

$$\tilde{x} = \sum_{i=1}^{\infty} \frac{1}{\mu_i} < y, \nu_i > \nu_i$$

If we want to regularize the system to make to solution continuous, we want to solve:

$$(K^2 + \alpha I)x_\alpha = Ky$$

By noting that x_α and Ky are in $R(K)$ we can solve for x_α and obtain:

$$x_\alpha = \sum_{i=1}^{\infty} \left(\frac{\mu_i}{\mu_i^2 + \alpha} \right) < y, \nu_i > \nu_i$$

Therefore, the solution requires us to estimate an infinity of eigenfunctions. But when K is unknown and is replace by K_n, the solution is much simpler. As Carrasco and Florens (2000) show, the dimension of $R(K_n)$ is finite:

$$\begin{aligned} (K_n f)(\tau_2) &= \int_\Lambda k_n(\tau_1, \tau_2) f(\tau_1) \pi(\tau_1) d\tau_1 \\ &= \int_\Lambda \frac{1}{n} \sum_{t=1}^{n} g(x_t, \tau_1, \theta) g(x_t, \tau_2, \theta) f(\tau_1) \pi(\tau_1) d\tau_1 \\ &= \sum_{t=1}^{n} g(x_t, \tau_2, \theta) \left(\int_\Lambda \frac{1}{n} g(x_t, \tau_1, \theta) f(\tau_1) \pi(\tau_1) d\tau_1 \right) \\ &\equiv \sum_{t=1}^{n} \delta_t g(x_t, \tau_2, \theta) \end{aligned}$$

Therefore, $R(K_n)$ is spanned by $\{g(x_1,\theta),\cdots,g(x_n,\theta)\}$. It follows that the singular system of K_n is composed of n eigenfunctions $\nu_i^{(n)}$ and n singular values $\mu_i^{(n)}$. We can extend the previous result to this case and show that the regularized solution to $K_n\lambda = -\hat{g}(\theta)$ is

$$\lambda_\alpha = -\sum_{i=1}^{n}\left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2}+\alpha}\right) < \hat{g}(\theta), \nu_i^{(n)} > \nu_i^{(n)}$$

Because $\nu_i^{(n)} \in R(K_n)$, we can write $\nu_i^{(n)} = 1/n\sum_j \beta_{ij} g(x_j,\theta)$. Carrasco and Florens (2000) show that the vectors β_i for $i = 1,...n$ are the eigenvectors of an nxn matrix C with typical element

$$c_{ij} = \frac{1}{n}\int_\Lambda g(x_i,\tau,\theta)g(x_j,\tau,\theta)\pi(\tau)d\tau$$

and that its eigenvalues are in fact the $\mu_i^{(n)}$ we need. We can therefore obtain the estimator using the following procedure:

1. We construct the nxn matrix C
2. We compute the eigenvectors β_i and eigenvalues $\mu_i^{(n)}$ for $i = 1,...,n$
3. We compute the eigenfunctions of K_n as follows:

$$\nu_i^{(n)} = \frac{1}{n}\sum_{j=1}^{n}\beta_{ji} g(\theta, x_j) \qquad i = 1,...,n$$

4. We compute $\hat{\lambda}_\alpha$:

$$\hat{\lambda}_\alpha = -\sum_{i=1}^{n}\left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2}+\alpha_n}\right) < \hat{g}(\theta), \nu_i^{(n)} > \nu_i^{(n)}$$

5. Me estimate θ by solving the following problem:

$$\hat{\theta}_\alpha = arg\min_\theta \sum_{t=1}^{n}\log\left(1 - \left\langle\hat{\lambda}_\alpha, g(x_t,\theta)\right\rangle\right)$$

Because the solution to the C-EL includes also an estimate of the probability distribution $p_t \quad t = 1,\cdots,n$ with $\sum_t p_t = 1$, which depends on λ, and that we did not obtain the exact solution to equation (17), we may, if we intend for example to use the likelihood ratio $\mathcal{R}(\hat{\theta})$ to construct confidence intervals, have to normalize $\hat{p}_t(\lambda_\alpha)$ as follows

$$\tilde{p_t} = \frac{\hat{p_t}(\lambda_\alpha)}{\sum_{t=1}^{n}\hat{p_t}(\lambda_\alpha)} \equiv \frac{1}{(1+g(x_t,\hat{\theta}_\alpha)\lambda_\alpha)\sum_{i=1}^{n}(1-g(x_t,\hat{\theta}_\alpha)\lambda_\alpha)} \tag{27}$$

We want to analyze the properties of three different approaches which are inspired by the GMM method. Indeed, because we need θ to estimate λ and then λ to estimate θ, the estimation of λ is like the estimation of the weight matrix in the GMM method. The three approaches are as follows: (i) We start with a first estimate θ^* that comes from the C-GMM method, and use it to compute $\hat{\lambda}_\alpha(\theta^*)$. We then use this estimate to compute $\hat{\theta}_\alpha$. (ii) We iterate the previous approach until $\hat{\theta}_\alpha$ converges. We will call this method, iterated EL estimator (C-IEL). (iii) As for the continuously updated estimator, we could find a function $\lambda_\alpha(\theta)$ and estimate θ by solving the following problem:

$$\hat{\theta}_\alpha = arg\min_\theta \sum_{t=1}^{n} \log\left(1 - \left\langle \hat{\lambda}_\alpha(\theta), g(\theta, x_t)\right\rangle\right) \tag{28}$$

We will call this third method, continuously updated EL estimator (C-CUEL). It should be straightforward to show that the three methods share the same asymptotic properties. However, in order to obtain their finite sample properties, we will need to use Monte-Carlo and stochastic expansion. This will be the subjet of future researches. We know from Guggenberger and Hahn (2005) that the two-step EL shares the same second order asymptotic properties as the m-step EL. But we cannot say whether this result can be extended to the case of a continuum of moment conditions. We could however make an educated guess based on Newey and Smith (2004). According to them, iterative GMM does not help to reduce the bias. It only makes the initial estimate of θ more efficient. As we know, the element of the bias associated with the initial estimate does not depend on its efficiency but on the fact that it is random. Therefore, we should avoid this element of the bias with the C-CUEL since it does not require an initial estimate of the parameter vector.
In order to apply this method, it is convenient to rewrite the objective function in matrix notation as in Carrasco et al. (2006). Lets define the $n \times m$ matrix β which contains the eigenvectors of C associated with the m eigenvalues of C which are different from 0 and the $m \times m$ diagonal matrix D with typical element D_{jj}:

$$D_{jj} = \frac{\mu_i^{(n)}}{\mu_i^{(n)^2} + \alpha}$$

The following optimization problem is equivalent to the one given by equation (28):

$$\hat{\theta}_\alpha = arg\min_\theta \sum_{t=1}^{n} \log\left(1 + \frac{1}{n}\iota' C[\beta D \beta']C_{\bullet t}\right) \tag{29}$$

where ι is a $n \times 1$ vector of ones and $C_{\bullet t}$ is the t^{th} column of C (see the appendix for the proof).

The next step is to see how we could estimate θ without having to approximate the solution to equation (17). To do so, we need techniques to solve

nonlinear operator equations which is the subject of the next section. But before that, the singular value decomposition method allows us to construct a test of specification that is inspired by Carrasco and Florens (2000)

3.1.1 Specification test

In the case of finite number of moment conditions, we can construct a test of specification that is based on the likelihood ratio $2\sum_t \log(1- < g_t, \lambda >)$. Qin and Lawless (1994) show that this test is asymptotically χ^2_{r-p}, where r is the number of moment conditions and p the dimension of θ. As for C-GMM, we cannot use that test because the number of conditions is infinite. Carrasco and Florens (2000) propose to construct a test based on the singular value representation of the C-GMM objective function. We can apply the same arguments as the ones used in the lemma 9 and theorem 10 of their paper to build our test. I will only present how to construct the test. A sketch of the proof is given in appendix.
The test is based on the singular value solution of λ given above:

$$\lambda_\alpha = -\sum_{i=1}^{m} \left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2} + \alpha} \right) < \hat{g}(\theta), \phi_i^{(n)} > \phi_i^{(n)}$$

where $\phi_i^{(n)}$ is the orthonormalized eigenfunction $\nu_i^{(n)}/\|\nu_i^{(n)}\|$ with $\|\nu_i^{(n)}\|^2 = \mu_i^{(n)}/n$. Lets define the following variables:

$$p_n = \sum_{i=1^m} \frac{\mu_i^{(n)^3}}{(\mu_i^{(n)^2} + \alpha_n)^2}$$

$$q_n = 2\sum_{i=1}^{m} \frac{\mu_i^{(n)^6}}{(\mu_i^{(n)^2} + \alpha_n)^4}$$

Then the specification test is defined as:

$$t_n = \frac{\|\sqrt{n}\lambda_\alpha\|^2 - p_n}{\sqrt{q_n}} \Longrightarrow N(0,1)$$

under the null that the model is correctly specified.

3.2 Solving a nonlinear operator equation

When we want to solve a nonlinear problem such as $f(x) = 0$, we usually construct an iterative procedure of the form

$$x_i = g(x_{i-1})$$

which converges to the fix point $g(x) = x$, where x is the solution to the initial problem. The simplest method would be to set $g(x) = x + \omega f(x)$. If the

algorithm converges, then we have $f(x) = 0$ as required. The problem with this method is its slow speed of convergence. Furthermore, convergence occurs only if we choose correctly the parameter ω. The Newton method is probably one of the most popular algorithm because it converges always if the solution exists and that we start correctly. This method sets $\omega = -[f'(x)]^{-1}$ so that the algorithm becomes:

$$x_i = x_{i-1} - f'(x_{i-1})^{-1} f(x_{i-1})$$

In order for this method to work, we need the inverse of the first derivative to be bounded. If it is not, we need to regularize the inverse as we did for the linear case. This is similar to the problem we are facing in this section but with the exception that the solution x that we seek is a function from $L^2(\pi)$.
To see that, lets come back to our problem. For a given θ, we need to solve the following problem:

$$\hat{\lambda}(\theta) = arg \sup_{\lambda \in L^2(\pi)} \frac{1}{n} \sum_{t=1}^{n} \log\left(1 - < \lambda, g(x_t, \theta) >\right)$$

which implies that we need to find the solution to the following first order condition:

$$F(\lambda) \equiv \frac{1}{n} \sum_{t=1}^{n} \frac{1}{1 - < \lambda, g(x_t, \theta) >} g(x_t, \theta) = 0 \tag{30}$$

So we need to solve the general nonlinear operator equation $F(\lambda) = 0$, where F is a mapping from $L^2(\pi)$ to $L^2(\pi)$. This problem is ill-posed if the inverse of the operator DF, which is the Fréchet derivative of F, is an unbounded operator. But we show in the appendix that DF is an operator from $L^2(\pi)$ to $L^2(\pi)$ with kernel:

$$DF(\lambda)(\tau_1, \tau_2) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{1 - < g_t(\theta), \lambda >} \right)^2 g_t(\tau_1, \theta) g_t(\tau_2, \theta) \tag{31}$$

which converges to the covariance operator K. Therefore, its inverse is unbounded. Most methods used to solve such problems are iterative. The one that will be presented here uses the Tikhonov regularization approach. The goal of these methods is to minimize

$$\|F(\lambda)\|^2 + \alpha \|\lambda - \lambda_0\|^2$$

so that $F(\lambda)$ is close to 0 and the solution λ sufficiently smooth. α gives more or less importance to the smoothness of the solution. There are many algorithms that reflect this tradeoff. Ramm (2004a) and Ramm (2004b) present the continuous version of such method and give the conditions under which it converges to the solution. The discrete algorithm presented by Airayetpan and Ramm (2000) is a regularized Newton method. It is a Newton method applied to a transformed equation. Indeed, the Newton method solves $F(\lambda) = 0$ with

the algorithm $\lambda_i = \lambda_{i-1} - DF(\lambda_{i-1})^{-1}F(\lambda_{i-1})$ while the regularized Newton method solves $F(\lambda) + \alpha\lambda = 0$ which implies the following algorithm:

$$\lambda_i = \lambda_{i-1} - \omega_i \left[DF(\lambda_{i-1}) + \alpha I\right]^{-1} \left(F(\lambda_{i-1}) + \alpha\lambda_{i-1}\right)$$

where ω_i is a sequence that we need to choose to control the speed of convergence. Another method which uses a regularized inverse which is closer to the one we used above for the linear case has been analyzed by Qi-Nian (2000). This is the method that will be applied to the numerical study of the next section. It is a regularized Gauss-Newton method defined as follows:

$$\lambda_i = \lambda_{i-1} - \left[\alpha I + DF(\lambda_{i-1})^2\right]^{-1} \left\{DF(\lambda_{i-1})F(\lambda_{i-1}) + \alpha\lambda_{i-1}\right\}$$

where the initial value λ_0 has been set equal to its asymptotic value, 0.
In order to apply this algorithm, we will present it in matrix form. As for the previous section, we only need the inner product $< g_t, \lambda >$. Therefore we want to derive the result for the following iterative procedure:

$$< g_t, \lambda_i > = < g_t, \lambda_{i-1} > - \left\langle g_t, \left[\alpha I + DF(\lambda_{i-1})^2\right]^{-1} \left\{DF(\lambda_{i-1})F(\lambda_{i-1}) + \alpha\lambda_{i-1}\right\} \right\rangle \tag{32}$$

Lets define the $n \times n$ diagonal matrix V as:

$$V_{ii} = \left(\frac{1}{1 - < g_i(\theta), \lambda >}\right)^2,$$

the $n \times 1$ vector P as:

$$P_i = \frac{1}{1 - < g_i(\theta), \lambda >},$$

and the $n \times n$ matrix C as usual. Then, we can show that the iterative procedure (32) can be written as follows:

$$< g_t, \lambda_i > = \left\{[CV]^2 + \alpha I\right\}^{-1} \left\{[CV]^2 < g_t, \lambda_{i-1} > - [CV][CP]\right\} \tag{33}$$

with the initial value:

$$< g, \lambda_0 > = - \left\{C^2 + \alpha I\right\}^{-1} C^2 \iota$$

where ι is an $n \times 1$ vector of ones (see the appendix for the proof). This iterative method in included in the objective function:

$$\sum_{t=1}^{n} \log\left(1 - < g_t(\theta), \lambda(\theta)\right)$$

so that the numerical optimizer applies it each time θ is modified. Therefore, it could be a very time consumming method. But if we start with reasonable values for θ and pick an α which is not too small, the iterative method converges very quickly within 4 or 5 iterations. We will come back to it in the next section

3.2.1 computation of C-CUE

Before applying C-EL to a specific example, we will derive a similar algorithm for the C-CUE. In order to do that, we will use the fact that C-CUE is approximately equivalent to the C-GEL with $\rho(v) = -v - 0.5v^2$. Therefore, we define the C-CUE as follows:

$$\hat{\theta}_{ccue} = arg \min_{\Theta} \sup_{L^2(\pi)} \sum_{t=1}^{n} \left[- < g_t(\theta, \lambda > - \frac{1}{2} < g_t(\theta), \lambda > \right]$$

The first order condition for λ implies that we need to solve the following system:

$$F(\lambda) \equiv \hat{g}(\theta) + \frac{1}{n} \sum_{t=1}^{n} < g_t(\theta), \lambda > g_t(\theta) = \hat{g}(\theta) + K_n \lambda = 0$$

Which is the same condition as the one derived from the first order Taylor expansion of C-EL. The difference is that we will apply the regularized inverse to solve for $< g_t, \lambda >$.

4 Numerical study

In this section, we want to compare the properties of the three methods presented above. In fact there are only two methods, C-EL and C-CUE, but we want also to compare the two numerical methods presented in the previous section to compute C-EL. We will refer to C-ELsv when the singular value decomposition applied to the first order Taylor expansion is used and to C-ELgn for the method based on the Gauss-Newton algorithm. The example is taken from Carrasco et al. (2006). We want to estimate the parameters of the Cox-Ingersoll-Ross (CIR) diffusion process. The moment conditions are based on the characteristic function derived by Singleton (2001). What makes characteristic function appealing is the existence of its analytic form for many continuous time processes for which the analytic form of the likelihood does not exist. But to take advantage of the whole set of moments conditions, we need methods like C-GMM or C-GEL.

We will present a small numerical experiment to compare the different estimation procedures. Although, this is not a rigorous way to compare the properties of estimators, it will give us a first impression. The CIR model is defined as follows:

$$dr_t = (\gamma - \kappa r_t)dt + \sigma \sqrt{r_t} dW_t \tag{34}$$

The conditional characteristic function for this process is:

$$E\left(e^{i\tau r_{t+1}} \middle| r_t\right) \equiv \psi(\tau | r_t) = \left(1 - \frac{i\tau}{c}\right)^{-2\gamma/\sigma^2} \exp\left[\frac{i\tau e^{-\kappa}}{1 - \frac{i\tau}{c}} r_t\right] \tag{35}$$

$$c = \frac{2\kappa}{\sigma^2(1 - e^{-\kappa})}$$

We can obtain unconditional moment conditions from the above by using a set of instruments $m(\tau, Y_t)$. The moment function becomes:

$$g(x_t, \tau, \theta) = m(\tau_1, Y_t) \left[e^{i\tau_2 r_{t+1}} - \psi(\tau_2 | r_t) \right] \tag{36}$$

This is the double index version of the moment conditions. As they explain, the estimators reach the Cramér-Rao lower bound if the instruments are $m(\tau_1, Y_t) = e^{i\tau_1 r_t}$. Moreover, the moment function $g_t(\theta)$ defined in such a way is a martingale difference sequence which implies that we don't need to smooth the function as it would be necessary if it was weakly dependent. It follows that the moment conditions are:

$$E^{P_0} g(x_t, \tau, \theta) = E^{P_0} \left(e^{i\tau_1 r_t} \left[e^{i\tau_2 r_{t+1}} - \psi(\tau_2 | r_t) \right] \right) \quad \forall \ \tau \in \mathbb{R}^2 \tag{37}$$

where $\tau = (\tau_1, \tau_2)'$, $x_t = (r_t, r_{t+1})'$ and $\theta = (\gamma, \kappa, \sigma^2)'$. Therefore, the inner product $< f, g >$ in this example is defined as:

$$< f, g >= \int_{\mathbb{R}} \int_{\mathbb{R}} f(\tau_1, \tau_2) g(\tau_1, \tau_2) \pi(\tau_1, \tau_1) d\tau_1 d\tau_2$$

where the integrating density $\pi(\tau_1, \tau_2) = \pi(\tau_1)\pi(\tau_2)$ with $\pi(\tau)$ being the density of a standard normal distribution. The effect of the choice of integrating density on the properties of the estimators is unknown for now. The choice has be driven by the fact that it allows us to get an analytic representation of one of the double integrals that we need to compute for each C_{ij}, $j, i = 1, ..., n$. The simplification is possible only because of this specific moment function. For more details on how C_{ij} can be computed, see the appendix of Carrasco et al. (2006).

The specific model that is simulated in the Monte-Carlo study is:

$$dr_t = (0.02491 - 0.00285 r_t) dt + 0.0275 \sqrt{r_t} dW_t$$

In order to be able to compare our results with Carrasco et al. (2006), the first study is the same as the one that produced the results they reported in their paper. The sample size is 500, the number of iterations is 100 and $\alpha = 0.02$. The results are reported in table 1 of appendix E. For γ and κ, C-GEL seems the have the smallest bias while C-CUE stands between C-GMM and C-GEL. This result is consistent with what we find in the literature. We know from Newey and Smith (2004) and Anatolyev (2005) that the second order asymptotic bias of GMM processes two more terms if we compare it with CUE and the latter has one more term if we compare it with EL. The fact that we use a continuum of moment conditions does not seem to affect this result. What is most suprising is how the mean square error (MSE) of C-GEL and C-CUE performs relative to C-GMM. They are in fact much smaller which contradicts the results found in the literature. However, as for C-GMM, C-GEL and C-CUE have a hard time estimating the volatility parameter. Both the bias and the MSE are by far greater than the ones obtain using MLE which remains for now a mystery.

Finally, there is no significant difference between C-GELsv and C-GELgn. Since the latter has a higher computational cost, the former may be preferred. But for now it is pure speculation.

The second Monte-Carlo study tries to see how the choice of α affects the bias and the MSE. To do so, the previous experiment has been repeated for $\alpha = \{0.001, 0.005, 0.01, 0.02, 0.05, 0.1\}$ and for n=200 and 500. The results are shown in appendix E, figures 1 to 4. There does seem to have a clear tendency. However, it seems that if α is too small the MSE is larger. This is caused by the fact that if we do not regularized enough the system, it becomes unstable. Otherwise, the MSE does not seem to be too sensitive. Similar comments can be made with respect to the bias. It seems however that the choice of α has more impact in larger sample sizes. In such cases, a judicious choice could reduce the bias substantially. But further research needs to be done in order to know what is a judicious choice.

Appendix

A Fréchet derivative

Generally, if we have two normed spaces, X and Y, the Fréchet derivative of a differentiable mapping F: $X \to Y$ at $x \in X$ is the bounded operator D_F: $X \to Y$ which satisfies the following condition:

$$\lim_{h \to 0} \frac{\|F(x+h)) - F(x) - D_F h\|}{\|h\|} = 0$$

We can easily verify that if $F(\lambda) =< g_t(\theta), \lambda >$, the Fréchet derivative DF_λ is $g_t(\theta)$. It can been seen as being an operator if we write $DF_\lambda h = g_t(\theta)h =< g_t(\theta), h > \quad \forall h \in L^2(\pi)$. The proof is straightforward since $F(\lambda)$ is linear:

$$\begin{aligned} \frac{\|F(x+h)) - F(x) - D_F h\|}{\|h\|} &= \frac{\| \langle g_t(\theta), \lambda + h \rangle - \langle g_t(\theta), \lambda \rangle - \langle g_t(\theta), h \rangle \|}{\|h\|} \\ &= 0 \quad \forall h \in L^2(\pi) \end{aligned}$$

A.1 Fréchet derivative of the first order condition of λ

For a given θ, we need to solve the following system:

$$F(\lambda) \equiv \frac{1}{n} \sum_{t=1}^{n} \frac{1}{1 - < g_t(\theta), \lambda >} g_t(\theta) = 0$$

$F(\lambda)$ is therefore a mapping from $L^2(\pi)$ to $L^2(\pi)$. We can expand it about $\lambda = 0$:

$$0 = F(\lambda) = F(0) + DF(0)\lambda + O_p(\|\lambda\|^2)$$

where the last term goes to zero only if θ goes to θ_0. The approximation is therefore valid only if θ is element of the ball $b(\delta_n, \theta_0)$ which shrinks as n goes up. $F(0)$ is the sample average of the moment function $\hat{g}(\theta)$ and $DF(0)$ is the Fréchet derivative of $F(\lambda)$ evaluated at $\lambda = 0$. We will prove that the following is indeed the Fréchet derivative of $F(\lambda)$. Since it is an operator, it is defined by its kernel:

$$DF(\lambda)(\tau_1, \tau_2) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{1 - < g_t(\theta), \lambda) >} \right)^2 g_t(\tau_1, \theta) g_t(\tau_2, \theta)$$

We want to show that:

$$\frac{\|L(\lambda + h) - L(\lambda) - DF(\lambda)h\|}{\|h\|} \Longrightarrow 0$$

as h goes to zero. First we have:

$$\begin{aligned}(DF(\lambda)h)(\tau) &= \frac{1}{n}\sum_{t=1}^{n}\frac{1}{y_t^2}\int_{\Lambda} g_t(\tau)g_t(\tau_1)h(\tau_1)\pi(\tau_1)d\tau_1 \\ &= \frac{1}{n}\sum_{t=1}^{n}\frac{1}{y_t^2} < g_t, h > g_t(\tau)\end{aligned}$$

where $y_t = 1 - < g_t(\theta), \lambda >$ and θ has been remove to simplify the notation,

$$L(\lambda + h) = \frac{1}{n}\sum \frac{1}{y_t - < g_t, h >} g_t$$

Then we have (we set $< g_t, h >= x_t$):

$$\begin{aligned}\frac{\| \frac{1}{n}\sum_{t=1}^{n} g_t \left(\frac{1}{y_t - x_t} - \frac{1}{y_t} - \frac{x_t}{y_t^2} \right) \|}{\|h\|} &\leq \frac{1}{n}\sum_{t=1}^{n} \frac{\| g_t \left(\frac{1}{y_t - x_t} - \frac{1}{y_t} - \frac{x_t}{y_t^2} \right) \|}{\|h\|} \\ &= \frac{1}{n}\sum_{t=1}^{n} \left| \frac{1}{y_t - x_t} - \frac{1}{y_t} - \frac{x_t}{y_t^2} \right| \frac{\|g_t\|}{\|h\|} \\ &= \frac{1}{n}\sum_{t=1}^{n} \left| \frac{x_t^2}{y_t + O(h)} \right| \frac{\|g_t\|}{\|h\|} \\ &\leq \frac{1}{n}\sum_{t=1}^{n} \frac{\|g_t\|^2\|h\|^2}{|y_t + O(h)|} \frac{\|g_t\|}{\|h\|} \\ &= \frac{1}{n}\sum_{t=1}^{n} \frac{\|g_t\|^3\|h\|}{|y_t + O(h)|} \\ &\longrightarrow 0\end{aligned}$$

$DF(\lambda)$ is therefore the Fréchet derivative of (λ). The Taylor approximation is therefore:

$$0 = F(0) + DF(0)\lambda = \hat{g}(\theta) + \left(\frac{1}{n}\sum_{t=1}^{n} g_t(\theta)g_t(\theta) \right) \lambda = \hat{g}(\theta) + \hat{K}\lambda$$

B Proofs

B.1 Proof of Theorem 1

We will follow the proof given by Newey and Smith (2004) to show that the C-GEL is approximately equivalent to C-CUE if $\rho(v)$ is quadratic[2].

[2]In fact, they show that it is exactly equivalent when the number of moment conditions is finite.

Because $\rho(v)$ is a quadratic function, we can write $P(\theta, \lambda)$ from equation (5) as a second order Taylor expansion about 0 as follows:

$$P(\theta, \lambda) = \rho_0 - < \lambda, \hat{g}(\theta) > -\frac{1}{2} < \lambda, \hat{K}(\theta)\lambda > \tag{38}$$

where $\rho_0 \equiv \rho(0)$ and $\rho(v)$ has been normalized so that $\rho'(0) = \rho''(0) = -1$. In general the above equation is a bad approximation of $P(\theta, \lambda)$ because the reminder term, if the third derivative is bounded, is of order $o_p(\| \lambda \|^2)$, where $\| \lambda \|^2$ is $o_p(1)$ only at $\theta = \theta_0$. This will become clear bellow when we analyze the empirical likelihood method.
The first order condition that allows us to solve the supremum problem (5) is:

$$-\hat{g}(\theta) - \hat{K}(\theta)\lambda(\theta) = 0$$

which implies that we need to solve the following linear operator problem:

$$\hat{K}(\theta)\lambda(\theta) = -\hat{g}(\theta) \tag{39}$$

Using the regularized inverse, we obtain

$$\lambda(\theta) = -(\hat{K}^2 + \alpha_n I)^{-1}\hat{K}\hat{g}(\theta) \equiv -K_n^{\alpha_n}(\theta)^{-1}\hat{g}(\theta) \tag{40}$$

This is not the exact solution of the optimization problem, but it is asymptotically equivalent as n goes to infinity and α_n goes to zero. If we substitute the solution in the objective function we obtain:

$$\begin{aligned} P(\theta, \lambda(\theta)) &= \rho_0 + \left\langle K_n^{\alpha_n}(\theta)^{-1}\hat{g}(\theta), \hat{g}(\theta) \right\rangle - \frac{1}{2}\left\langle K_n^{\alpha_n}(\theta)^{-1}, \hat{K}(\theta)K_n^{\alpha_n}(\theta)^{-1}\hat{g}(\theta) \right\rangle \\ &= \rho_0 + \left\langle K_n^{\alpha_n}(\theta)^{-1/2}\hat{g}(\theta), K_n^{\alpha_n}(\theta)^{-1/2}\hat{g}(\theta) \right\rangle \\ &\quad -\frac{1}{2}\left\langle K_n^{\alpha_n}(\theta)^{-1}\hat{g}(\theta), (I + o_p(1))\hat{g}(\theta) \right\rangle \\ &= \rho_0 + \frac{1}{2} \| \hat{g}(\theta) \|^2_{K_n^{\alpha_n}(\theta)} + o_p(1) \end{aligned}$$

where I used the convergence $K_n^{\alpha_n}(\theta)^{-1}$ to K^{-1} so that $K_n^{\alpha_n}(\theta)^{-1}K = I + O_p(\|K_n^{\alpha_n}(\theta) - K\|)$

B.2 Convergence of $\hat{\lambda}$

In this section, we adapt the the proof made by Owen (2001) to the case of a continuum of moments conditions.[3] We want to show that λ is $O_p(n^{-1/2})$ when the model is correctly specified. To begin, we suppose that we know θ_0 and want to find the likelihood function $\mathcal{R}(\theta_0)$ which is defined as follows:

$$\mathcal{R}(\theta_0) = \max_{p_t} \left\{ \prod_{t=1}^{n} np_t \left| \sum_{t=1}^{n} p_t g(x_t, \theta_0) = 0 \ , \ \sum_{t=1}^{n} p_t = 1 \right. \right\}$$

[3]See the appendix for proof applied to the C-GEL methods

The first order condition is given by equation (17) but evaluated at $\theta = \theta_0$ [4]:

$$Q_1(\hat{\lambda}) = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{1 + g_t\hat{\lambda}} g(x_t, \theta_0) = 0 \tag{41}$$

where:

$$g_t\lambda = \int_\Lambda g(x_t, \tau; \theta_0)\lambda(\tau)\pi(\tau)d\tau$$

Let us define $\gamma \in L^2(\pi)$ with $\|\gamma\| = 1$ by the following:

$$\hat{\lambda} = \|\hat{\lambda}\|\gamma$$

and $Y_t = g_t\hat{\lambda}$. In order to prove the desire result, we only need to show that

$$\|\hat{\lambda}\| = \left[\int_\Lambda \hat{\lambda}^2(\tau)\pi(\tau)d\tau\right]^{1/2} = O_p(n^{-1/2})$$

We can rewrite the condition (41) as follows:

$$\begin{aligned}
0 &= \langle Q_1, \gamma\rangle \\
&= \frac{1}{n}\sum_{t=1}^{n} \frac{g_t\gamma}{1 + Y_t} \\
&= \frac{1}{n}\sum_{t=1}^{n} g_t\gamma\left(1 - \frac{Y_t}{1 + Y_t}\right) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left\{g_t\gamma - \frac{g_t\gamma Y_t}{1 + Y_t}\right\} \\
&= \frac{1}{n}\sum_{t=1}^{n}\left\{g_t\gamma - \|\lambda\|\frac{g_t\gamma g_t\gamma}{1 + Y_t}\right\} \\
&= \hat{g}(\theta_0)\gamma - \|\hat{\lambda}\|\ \frac{1}{n}\sum_{ti=1}^{n} \frac{g_t\gamma g_t\gamma}{1 + Y_t} \\
&= \hat{g}(\theta_0)\gamma - \|\hat{\lambda}\|\ \sum_{t=1}^{n} \hat{p}_t(\theta_0) g_t\gamma g_t\gamma
\end{aligned}$$

[4] Notice that λ as been redefined as $-\lambda$ which follows the Owen (2001) notation. It has no effect on the result.

where $\hat{p}$ is the implied probability density. Given our operator notations we have:

$$\begin{aligned}
0 &= \hat{g}(\theta_0)\gamma - \|\hat{\lambda}\| \sum_{t=1}^{n} \hat{p}_t(\theta_0) \int_\Lambda \int_\Lambda g(x_t,\tau_1)\gamma(\tau_1)g(x_t,\tau_2)\gamma(\tau_2)\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2 \\
&= \hat{g}(\theta_0)\gamma - \|\hat{\lambda}\| \int_\Lambda \int_\Lambda \gamma(\tau_1) \left[\sum_{t=1}^{n} \hat{p}_t(\theta_0) g(x_t,\tau_1)g(x_t,\tau_2)\right] \gamma(\tau_2)\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2 \\
&= \hat{g}(\theta_0)\gamma - \|\hat{\lambda}\| \int_\Lambda \int_\Lambda \gamma(\tau_1) K_{n0}(\tau_1,\tau_2)\gamma(\tau_2)\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2
\end{aligned}$$

where $K_{n0}(\tau_1,\tau_2)$ is the estimator of $E^{P_0}(g(x_t,\tau_1,\theta_0)g(x_t,\tau_2,\theta_0)$ which is the kernel associated with the covariance operator K. The index 0 means that it is estimated using our probability density $\hat{p}(\theta_0)$ and the moment function evaluated at the true parameter θ_0. Our relation becomes:

$$0 = \hat{g}(\theta_0)\gamma - \|\hat{\lambda}\| < \gamma, K_{n0}\gamma >$$

$$\hat{g}(\theta_0)\gamma = \|\hat{\lambda}\| < \gamma, K_{n0}\gamma > \tag{42}$$

Since $\gamma \in L^2(\pi)$ and $\|\gamma\| = 1$, we have:

$$\begin{aligned}
\hat{g}(\theta_0)\gamma &= \frac{1}{n}\sum_{t=1}^{n} \int_\Lambda g(x_t,\tau,\theta_0)\gamma(\tau)\pi(\tau)d\tau \\
&= \int_\Lambda \left(\frac{1}{n}\sum_{t=1}^{n} g(x_t,\tau,\theta_0)\right)\gamma(\tau)\pi(\tau)d\tau \\
&= \int_\Lambda O_p(n^{-1/2})\gamma(\tau)\pi(\tau)d\tau \\
&\leq \| O_p(n^{-1/2}) \| \| \gamma \| = O_p(n^{-1/2})
\end{aligned}$$

by Cauchy-Schwartz and the law of large numbers. Therefore,

$$\|\hat{\lambda}\| < \gamma, K_{n0}\gamma > = O_p(n^{-1/2})$$

In order to show that $\|\hat{\lambda}\| = O_p(n^{-1/2})$, we only need $< \gamma, K_{n0}\gamma >$ to be $O_p(1)$. It would be an easy task if $\hat{p}(\theta_0)$ was $1/n$ because we could use the law of large number for iid or weakly dependent processes depending on our assumptions on the data generating process. Let $K_{n0}^{[1/n]}$ be K_{n0} evaluated at $\lambda = 0$ which

implies that $\hat{p}(\theta_0) = 1/n$. $K_{n0}^{[1/n]}$ will be used to find a bound for $< \gamma, K_{n0}\gamma >$.

$$\begin{aligned}
< \gamma, K_{n0}^{[1/n]}\gamma > &= < \gamma, K_{n0}^{[1/n]}\gamma > \left(\frac{\max_t(Y_t+1)}{\max_i(Y_t+1)}\right) \\
&= \int_\Lambda \int_\Lambda \gamma(\tau_1) \left[\sum_{t=1}^{n} \frac{1}{n} g(x_t, \tau_1, \theta_0) g(x_t, \tau_2, \theta_0)\right] \pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2 \left(\frac{\max_t(Y_t+1)}{\max_t(Y_t+1)}\right) \\
&\leq \int_\Lambda \int_\Lambda \gamma(\tau_1) \left[\sum_{t=1}^{n} \frac{1}{n} \frac{g(x_t, \tau_1, \theta_0) g(x_t, \tau_2, \theta_0)}{1+Y_t}\right] \pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2 \left(\max_t(Y_t+1)\right) \\
&= < \gamma, K_{n0}\gamma > \left(\max_t(Y_t+1)\right) \\
&= < \gamma, K_{n0}\gamma > \left(1 + \max_t(g_t\hat{\lambda})\right) \\
&\leq < \gamma, K_{n0}\gamma > (1 + \|\lambda\| Z_n)
\end{aligned}$$

où $Z_n = \max_t \|g_t\|$. It follows that:

$$\begin{aligned}
\|\lambda\| < \gamma, K_{n0}^{[1/n]}\gamma > &\leq \|\hat{\lambda}\| \langle \gamma, K_{n0}\gamma \rangle (1 + \|\lambda\| Z_n) \\
&= \hat{g}(\theta_0)\gamma(1 + \|\hat{\lambda}\| Z_n) = O_p(n^{-1/2}) + \|\hat{\lambda}\| \hat{g}(\theta_0)\gamma Z_n
\end{aligned}$$

which implies

$$\|\hat{\lambda}\| \Big(\Big\langle \gamma, K_{n0}^{[1/n]}\gamma \Big\rangle - \hat{g}(\theta_0)\gamma Z_n \Big) \leq O_p(n^{-1/2})$$

The final result requires that term inside the parentheses does not vanish. Let us consider the first term:

$$\begin{aligned}
< \gamma, K_{n0}^{[1/n]}\gamma > &= \int_\Lambda \int_\Lambda \gamma(\tau_1) \left(\frac{1}{n} \sum_{t=1}^{n} g(x_t, \tau_1, \theta_0) g_s(x_t, \tau_2, \theta_0)\right) \gamma(\tau_2)\pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2 \\
&= o_p(1) + \int_\Lambda \int_\Lambda \gamma(\tau_1) E^{p0}\Big(g(x, \tau_1, \theta_0) g(x, \tau_2, \theta_0)\Big) \gamma(\tau_2)\pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2 \\
&= < \gamma, K\gamma > + o_p(1)
\end{aligned}$$

The second equation follows from the consistency of $K_{n0}^{[1/n]}$. The last term depends on the properties of the covariance operator K. Because $\|\gamma\| = 1$, $< \gamma, K\gamma >$ is bounded from above and below by the highest and the smallest eigenvalues of K respectively. By assumption 2, they are finite and strictly positive which implies that $< \gamma, K\gamma >$ is $O_p(1)$. For the second term, no further assumptions are required to show that it is $o_p(1)$[5]. We have shown that $\|\hat{\lambda}\| = O_p(n^{-1/2})$ which implies that $\hat{\lambda}$ converge to 0 when $\theta = \theta_0$. It can easily be shown, using the weak uniform law of large number, that if we replace θ_0 by a consistent estimator $\hat{\theta} = \theta_0 + o_p(1)$ we obtain the same result for $\hat{\lambda}$.

[5] see Owen (2001)

B.3 Linear approximation of the C-EL first order conditions

We need to solve

$$\begin{pmatrix} -\hat{g}(\theta_0) \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{K}_0 & \hat{G} \\ \hat{G}^* & 0 \end{pmatrix} \begin{pmatrix} \hat{\lambda} \\ \hat{\theta} - \theta_0 \end{pmatrix} \tag{43}$$

Therefore we need to invert the matrix which requires us to solve the following system:

$$\begin{pmatrix} \hat{K}_0 & \hat{G} \\ \hat{G}^* & 0 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I_q \end{pmatrix} \tag{44}$$

where I is the identity operator from $L^2(\pi)$ to $L^2(\pi)$ and I_q is the $q \times q$ identity matrix. The system can be written as:

$$\begin{aligned} \hat{K}_0 A_{11} + \hat{G} A_{12} &= I & (45) \\ \hat{K}_0 A_{12} + \hat{G} A_{22} &= 0 & (46) \\ \hat{G}^* A_{11} &= 0 & (47) \\ \hat{G}^* A_{12} &= I_q & (48) \end{aligned}$$

From equation (45) we have:

$$\hat{K}_0 A_{11} = (I - \hat{G} A_{21})$$

We can solve for A_{11} by using a regularized inverse:

$$A_{11} = (K_{0n}^{\alpha_n})^{-1} (I - \hat{G} A_{21})$$

where

$$(K_{0n}^{\alpha_n})^{-1} = \left(\hat{K}_0^2 + \alpha_n I\right)^{-1} \hat{K}_0$$

Following the same method we obtain from equation (46):

$$A_{12} = -(K_{0n}^{\alpha_n})^{-1} \hat{G} A_{22}$$

and from equation (48):

$$I_q = -\hat{G}^* (K_{0n}^{\alpha_n})^{-1} \hat{G} A_{22}$$

It follows that:

$$A_{22} = -\left[\hat{G}^* (K_{0n}^{\alpha_n})^{-1} \hat{G}\right]^{-1}$$

which exists as long as the score of the moment function has full rank. The three other terms can easily be obtain:

$$A_{12} = (K_{0n}^{\alpha_n})^{-1} \hat{G} \left[\hat{G}^* (K_{0n}^{\alpha_n})^{-1} \hat{G}\right]^{-1}$$

$$A_{21} = \left[\hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1} \hat{G}\right]^{-1} \hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1}$$

$$A_{11} = \left(K_{0n}^{\alpha_n}\right)^{-1} \left\{ I - \hat{G} \left[\hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1} \hat{G}\right]^{-1} \hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1} \right\}$$

It follows that:

$$\hat{\lambda} = -\left(K_{0n}^{\alpha_n}\right)^{-1} \left\{ I - \hat{G} \left[\hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1} \hat{G}\right]^{-1} \hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1} \right\} \hat{g}(\theta_0) + O_p(n^-1)$$

and

$$(\hat{\theta} - \theta_0) = -\left[\hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1} \hat{G}\right]^{-1} \hat{G}^* \left(K_{0n}^{\alpha_n}\right)^{-1} \hat{g}(\theta_0) + O_p(n^{-1})$$

B.4 Consistency of C-GEL

We only adapt the proof made by Newey and Smith (2004) in order to have a more general proof of consistency. The additional assumptions for consistency are:

- $E^{P_0}\left[sup_\Theta \parallel g(x,\theta) \parallel^\alpha \right] < \infty$ for some $\alpha > 2$.
- ρ is twice continuously differentiable in a neighborhood of zero

The only part of their proof that needs to be considered is the lemma A2. The crucial result that needs to be adjusted is the Taylor expansion of $\hat{P}(\theta, \lambda)$ around $\lambda = 0$ for $\theta = \hat{\theta}$:

$$\begin{aligned}
\rho(0) &= \hat{P}(\hat{\theta}, 0) \leq \rho(0) - < \hat{g}(\hat{\theta}), \hat{\lambda} > -\frac{1}{4} < \hat{\lambda}, \hat{K}\hat{\lambda} > \\
&= \rho(0) - < \hat{g}(\hat{\theta}), \hat{\lambda} > -\frac{1}{4} \parallel \hat{\lambda} \parallel^2 < \hat{\phi}, \hat{K}\hat{\phi} > \\
&\leq \rho(0) + \parallel \hat{g}(\hat{\theta}) \parallel \parallel \hat{\lambda} \parallel - \frac{1}{4} \parallel \hat{\lambda} \parallel^2 r(K) \qquad \text{w.p.a.1} \\
&\equiv \rho(0) + \parallel \hat{g}(\hat{\theta}) \parallel \parallel \hat{\lambda} \parallel - C \parallel \hat{\lambda} \parallel^2 \qquad (C > 0)
\end{aligned}$$

where $r(K)$ is the spectral radius of the operator K and $\hat{\phi} \equiv \hat{\lambda} / \parallel \hat{\lambda} \parallel$. Because $\hat{g}(\hat{\theta}) = O_p(n^{-1/2})$, it follows that $\parallel \hat{\lambda} \parallel = O_p(n^{-1/2})$. The consistency theorem 3.1 of Newey and Smith (2004) follows.

C Taylor expansion of the first order condition

We want to expand the first order condition given in equation (17) for a fix θ. We want to show that the remainder is of order $O_p(n^{-1/2})$ if $\theta = \theta_0$ and $\|E^{p0}[g(\theta, x_i)]\|_4 < \infty$. We will need the following theorem given in a paper by Li and He (2005):

Theorem C-1: *Let $L : \Im(L) \to L^2(\pi)$ be a continuous nonlinear operator which is twice Fréchet differentiable. Then for each $f \in L^2(\pi)$, $g \in \Im(F)$, there is a Taylor representation with integral remainder defined has follows:*

$$L(f+g) = L(f) + D_L(f)g + \int_0^1 (1-\delta) D_L^2(f+\delta g) g^2 d\delta$$

where D_L and D_L^2 are the first and second Fréchet derivatives of $L()$.

If we apply this theorem to the equation (17) with $f = 0$ and $g = \hat{\lambda}$ we obtain:

$$\begin{aligned} 0 = \frac{L(\hat{\lambda})}{n} &= \frac{L(0)}{n} + \int_\Lambda \left(-\sum_{t=1}^n \frac{g(x_t, \tau_1, \theta) g(x_t, \tau_2, \theta)}{n} \right) \hat{\lambda}(\tau_2) \pi(\tau_2) d\tau_2 \\ &\quad + \int_0^1 (1-\delta) \frac{D_L^2(\delta\hat{\lambda})}{n} \hat{\lambda}^2 d\delta \end{aligned}$$

where $L(0) = n\hat{g}(\theta)$. Since $D_L^2/n = O_p(\|E^{p0}[g(x_t,\theta)]\|_4)$ which is bounded by assumption, the remainder is therefore of order $\|\hat{\lambda}\|^2$ which is of order $O_p(n^{-1})$ if $\theta = \theta_0$ (see the proof above). It follows that :

$$\hat{g}(\tau_1, \theta_0) = \int_\Lambda \left(\frac{1}{n} \sum_{t=1}^n g_\omega(x_t, \tau_1, \theta_0) g(x_t, \tau_2, \theta_0) \right) \hat{\lambda}(\tau_2) \pi(\tau_2) d\tau_2 + O_p(n^{-1})$$

Therefore, it can easily be shown (see Qin and Lawless (1994)) that

$$0 = \hat{g}(\theta_0) + \left(\hat{K}\hat{\lambda} \right) + \left(\hat{G}(\hat{\theta} - \theta_0) \right) + \underbrace{0_p(\| \hat{\lambda} \|^2 + \| \hat{\theta} - \theta_0 \|^2)}_{O_p(n^{-1})} \tag{49}$$

$$0 = \left(\hat{G}^* \hat{\lambda} \right) + \underbrace{0_p(\| \hat{\lambda} \|^2 + \| \hat{\theta} - \theta_0 \|^2)}_{O_p(n^{-1})} \tag{50}$$

C.1 Distribution of $\hat{\theta}$ and $\hat{\theta}$

This is straightforward once we have derived the results from the previous section. By the strong law of large number for weakly dependent processes we have that $\hat{G}$ converges to G and $K_{0n}^{\alpha_n}$ converges to K almost surely. Therefore:

$$\sqrt{n}\hat{\lambda} \stackrel{a}{=} -K^{-1} \left\{ I - G \left[G^* K^{-1} G \right]^{-1} G^* K^{-1} \right\} \sqrt{n} \hat{g}(\theta_0)$$

where $\stackrel{a}{=}$ means asymptotically equivalent. It is asymptotically normal because of the properties of $\sqrt{n}\hat{g}(\theta_0)$ and its asymptotic variance is:

$$\begin{aligned} \lim_{n\to\infty} Var(\sqrt{n}\hat{\lambda}) &= K^{-1} \left\{ I - G \left[G^* K^{-1} G \right]^{-1} G^* K^{-1} \right\} K \\ &\quad \left\{ I - K^{-1} G \left[G^* K^{-1} G \right]^{-1} G^* \right\} K^{-1} \\ &= \left\{ I - K^{-1} G (G^* K^{-1} G)^{-1} G^* \right\} K^{-1} \end{aligned}$$

From equation (24), we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \stackrel{a}{=} -\left\{G^* K^{-1} G\right\}^{-1} G^* K^{-1} (\sqrt{n}\hat{g}(\theta_0))$$

Therefore,

$$\sqrt{n}(\hat{\theta}_\alpha - \theta_0) \Longrightarrow N(0, W_\alpha)$$

with

$$W_\alpha = \left\{G^* K^{-1} G\right\}^{-1} G^* K^{-1} K K^{-1} G \left\{G^* K - 1 G\right\}^{-1}$$

D Computation of C-GEL

D.1 Computation using the singular value decomposition.

We suppose that C has m eigenvalues different from zero. We define β as the $n \times m$ matrix containing the m eigenvectors associated with the eigenvalues. We can therefore write the solution for λ_α:

$$\lambda_\alpha = -\sum_{i=1}^{m} \left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2} + \alpha} \right) < \hat{g}, \nu_i^{(n)} > \nu_i^{(n)}$$

Numerically, the truncation parameter m can be set equals to the rank of C. This will allow m to increase with the sample size since, as n goes to infinity and K_n converges to K, the rank goes to infinity. Because it is not λ but $< g_t, \lambda >$ which enters the objective function, we only need to compute the latter (θ has been omitted for simplicity):

$$< g_t, \lambda_\alpha > = -\sum_{i=1}^{m} \left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2} + \alpha} \right) < \hat{g}, \nu_i^{(n)} >< g_t, \nu_i^{(n)} >$$

where:

$$\begin{aligned} < g_t, \nu_j^{(n)} > &= \int_\Lambda g_t(\tau) \left(\frac{1}{n} \sum_{i=1}^{m} \beta_{ij} g_i(\tau) \pi(\tau) d\tau \right) \\ &= \frac{1}{n} \sum_{i=1}^{n} \beta_{ij} \int_\Lambda g_t(\tau) g_i(\tau) \pi(\tau) d\tau \\ &= \sum_{i=1}^{n} \beta_{ij} C_{ti} \\ &= C_{t.} \beta_j \end{aligned}$$

where $C_{t.}$ is the t^{th} line of C. We can do the same for the other inner product:

$$\begin{aligned}
< \hat{g}, \nu_j^{(n)} > &= \int_\Lambda \hat{g}(\tau) \left(\frac{1}{n} \sum_{i=1}^{m} \beta_{ij} g_i(\tau) \pi(\tau) d\tau \right) \\
&= \int_\Lambda \left(\frac{1}{n} \sum_{t=1}^{n} g_t(\tau) \right) \left(\frac{1}{n} \sum_{i=1}^{m} \beta_{ij} g_i(\tau) \pi(\tau) d\tau \right) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \beta_{ij} \int_\Lambda g_t(\tau) g_i(\tau) \pi(\tau) d\tau \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{n} \beta_{ij} C_{ti} \\
&= \frac{1}{n} \sum_{t=1}^{n} C_{t.} \beta_j \\
&= \frac{1}{n} \iota' C \beta_j
\end{aligned}$$

where ι is a $n \times 1$ vector of ones. Therefore we can write:

$$\begin{aligned}
< g_t, \lambda_\alpha > &= -\sum_{i=1}^{m} \left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2} + \alpha} \right) \left[\frac{1}{n} \iota' C \beta_i \right] [C_{t.} \beta_i] \\
&= -\frac{1}{n} [\iota' C] \sum_{i=1}^{m} \left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2} + \alpha} \right) [\beta_i \beta_i'] C_{.t} \\
&= -\frac{1}{n} \iota' C (\beta D \beta') C_{.t}
\end{aligned}$$

where D is the diagonal matrix defined in the text. The objective function is therefore:

$$\sum_{t=1}^{n} \log (1 - < \lambda_\alpha, g_t >) = \sum_{t=1}^{n} \log \left(1 + \frac{1}{n} \iota' C (\beta D \beta') C_{.t} \right)$$

D.1.1 Specification test

Starting with the solution of λ:

$$\lambda_\alpha = -\sum_{i=1}^{m} \left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2} + \alpha} \right) < \hat{g}(\theta), \phi_i^{(n)} > \phi_i^{(n)}$$

we can derive the following result:

$$\begin{aligned}
\| \sqrt{n} \lambda_\alpha \|^2 &= \sum_{i=1}^{m} \left(\frac{\mu_i^{(n)^2}}{(\mu_i^{(n)^2} + \alpha)^2} \right) < \sqrt{n} \hat{g}(\theta), \phi_i^{(n)} >^2 \\
&= \sum_{i=1}^{m} \left(\frac{\mu_i^{(n)^3}}{(\mu_i^{(n)^2} + \alpha)^2} \right) \frac{< \sqrt{n} \hat{g}(\theta), \phi_i^{(n)} >^2}{\mu_i^{(n)}}
\end{aligned}$$

where we have used the orthonormalized eigenfunctions. The sketch of the proof is simply to realize that $< \sqrt{n}\hat{g}(\theta), \phi_i^{(n)} > / \sqrt{\mu_i^{(n)}}$ is asymptotically iid N(0,1) which implies that the square is distributed as a χ_1^2. The mean is 1 and variance 2 so we can rearrange to construct a centered distribution that will be asymptotically normal (see Carrasco and Florens (2000). for a more rigorous proof). Is we substract p_n, which is defined in the text, we have:

$$\|\sqrt{n}\lambda_\alpha\|^2 - p_n = \sum_{i=1}^{m} \left(\frac{{\mu_i^{(n)}}^3}{({\mu_i^{(n)}}^2 + \alpha)^2} \right) \left(\frac{< \sqrt{n}\hat{g}(\theta), \phi_i^{(n)} >^2}{\mu_i^{(n)}} - 1 \right)$$

It is clear that the asymptotic variance is q_n which implies that

$$\frac{\|\sqrt{n}\lambda_\alpha\|^2 - p_n}{\sqrt{qn}}$$

is asymptotically N(0,1).

D.2 Computation using the regularized Gauss-Newton method

To simplify the notation, we will set $g_t \equiv g_t(\theta)$, $\lambda \equiv \lambda_{i-1}$, $\lambda' \equiv \lambda_i$, $p_t = 1/(1- < g_t, \lambda >)$ and $p_t' = 1/(1- < g_t, \lambda' >)$. We want to rewrite the following algorithm:

$$\lambda' = \lambda - \left\{ DF(\lambda)^2 + \alpha I \right\}^{-1} \left\{ DF(\lambda)F(\lambda) + \alpha\lambda \right\}$$

which can be written as:

$$\begin{aligned} \left\{ DF(\lambda)^2 + \alpha I \right\} \lambda' &= \left\{ DF(\lambda)^2 + \alpha I \right\} \lambda - DF(\lambda)F(\lambda) - \alpha\lambda \\ &= DF(\lambda)^2\lambda - DF(\lambda)F(\lambda) \end{aligned}$$

What we want to do is to rewrite each term, multiply them by $g_s(\tau_1)\pi(\tau_1)$ and integrate. The first term of the left hand side is:

$$\begin{aligned} DF(\lambda)^2\lambda' &= \int_\Lambda \left\{ \int_\Lambda \left(\frac{1}{n} \sum_{t=1}^{n} p_t^2 g_t(\tau_1) g_t(\tau_2) \right) \left(\frac{1}{n} \sum_{l=1}^{n} p_l^2 g_l(\tau_2) g_l(\tau_3) \right) \pi(\tau_2) d\tau_2 \right\} \lambda'(\tau_3)\pi(\tau_3) d\tau_3 \\ &= \frac{1}{n^2} \sum_t \sum_l p_t^2 p_l^2 g_t(\tau_1) \left(\int_\Lambda g_t g_l \pi d\tau_2 \right) \left(\int_\Lambda g_l \lambda' \pi d\tau_3 \right) \\ &= \frac{1}{n} \sum_t \sum_l p_t^2 p_l^2 g_t(\tau_1) C_{tl} < g_l, \lambda' > \end{aligned}$$

Once we apply the transformation, the term becomes:

$$\begin{aligned}
\left[DF(\lambda)^2\lambda'\right] g_s &= \int_\Lambda \frac{1}{n}\sum_t \sum_l p_t^2 p_l^2 g_t(\tau_1) C_{tl} < g_l, \lambda' > g_s(\tau_1)\pi(\tau_1)d\tau_1 \\
&= \frac{1}{n}\sum_t \sum_l p_t^2 p_l^2 C_{tl} < g_l, \lambda' > \int_\Lambda g_t(\tau_1) g_s(\tau_1)\pi(\tau_1)d\tau_1 \\
&= \sum_t \sum_l p_t^2 p_l^2 C_{tl} < g_l, \lambda' > C_{ts} \\
&= \sum_t C_{ts} p_t^2 \left[C_{t\cdot} V < g, \lambda' >\right] \\
&= C_{s\cdot} VCV < g, \lambda' >
\end{aligned}$$

where V is defined in the text and $< g, \lambda >$ is the $n \times 1$ vector with typical element $< g_t, \lambda >$. Since it has to be valid for all $s = 1, \cdots n$, The first term on the left hand side can be written as follows:

$$\left[DF(\lambda)^2\lambda'\right] g = (CV)^2 < g, \lambda' >$$

It follows that the first term of the right hand side is:

$$\left[DF(\lambda)^2\lambda\right] g = (CV)^2 < g, \lambda >$$

Clearly, the second term of the left hand side is simply $\alpha < g_t, \lambda' >$. The left hand side can therefore be written as:

$$\left\{(CV)^2 + \alpha I\right\} < g, \lambda' >$$

The second term on the right hand side is:

$$\begin{aligned}
DF(\lambda)F(\lambda) &= \int_\Lambda \left[\frac{1}{n}\sum_t p_t^2 g_t(\tau_1) g_t(\tau_2)\right] \left(\frac{1}{n}\sum_l p_l g_l(\tau_2)\right) \pi(\tau_2) \\
&= \frac{1}{n^2}\sum_t \sum_l p_t^2 p_l^2 g_t(\tau_1) \int_\Lambda g_t(\tau_2) g_l(\tau_2)\pi(\tau_2)d\tau_2 \\
&= \frac{1}{n}\sum_t \sum_l p_t^2 p_l^2 g_t(\tau_1) C_{tl}
\end{aligned}$$

If we apply the transformation it becomes:

$$\begin{aligned}
\left[DF(\lambda)F(\lambda)\right] g_s &= \int_\Lambda \frac{1}{n}\sum_t \sum_l p_t^2 p_l^2 g_t(\tau_1) C_{tl} g_s(\tau_1)\pi(\tau_1)d\tau_1 \\
&= \sum_t \sum_l p_t^2 p_l^2 C_{tl} C_{ts}
\end{aligned}$$

For all $s = 1, \cdots, n$ the term can be written as:

$$CVCP$$

where P is defined in the text. We can therefore rewrite the iterative procedure as follows:

$$\left\{(CV)^2 + \alpha I\right\} < g, \lambda' >= (CV)^2 < g, \lambda > -CVCP$$

which implies

$$< g, \lambda' >= \left\{(CV)^2 + \alpha I\right\}^{-1} \left\{(CV)^2 < g, \lambda > -CVCP\right\}$$

If we start with $\lambda_0 = 0$, then $V = I$ and $P = \iota$ which gives us the starting value:

$$< g, \lambda' >= \left\{C^2 + \alpha I\right\}^{-1} \left\{-C^2\iota\right\} = -\left\{C^2 + \alpha I\right\}^{-1} C^2\iota$$

E Numerical results

Table 1: Estimation of the CIR with sample size of 500

True Value	Mean Bias	RMSE
C-GELgn		
$\gamma = 0.02491$	0.001364	0.015882
$\kappa = 0.00285$	0.000149	0.001771
$\sigma = 0.02750$	0.014210	0.017292
C-GELsv		
$\gamma = 0.02491$	0.001605	0.018902
$\kappa = 0.00285$	0.000103	0.002125
$\sigma = 0.02750$	0.016137	0.019095
C-CUE		
$\gamma = 0.02491$	0.001974	0.016623
$\kappa = 0.00285$	0.000144	0.001940
$\sigma = 0.02750$	0.014423	0.017334
C-GMM		
$\gamma = 0.02491$	0.0082	0.0216
$\kappa = 0.00285$	0.0009	0.0025
$\sigma = 0.02750$	0.0134	0.0147
MLE		
$\gamma = 0.02491$	0.0123	0.0125
$\kappa = 0.00285$	0.0014	0.0014
$\sigma = 0.02750$	4e-5	0.0009

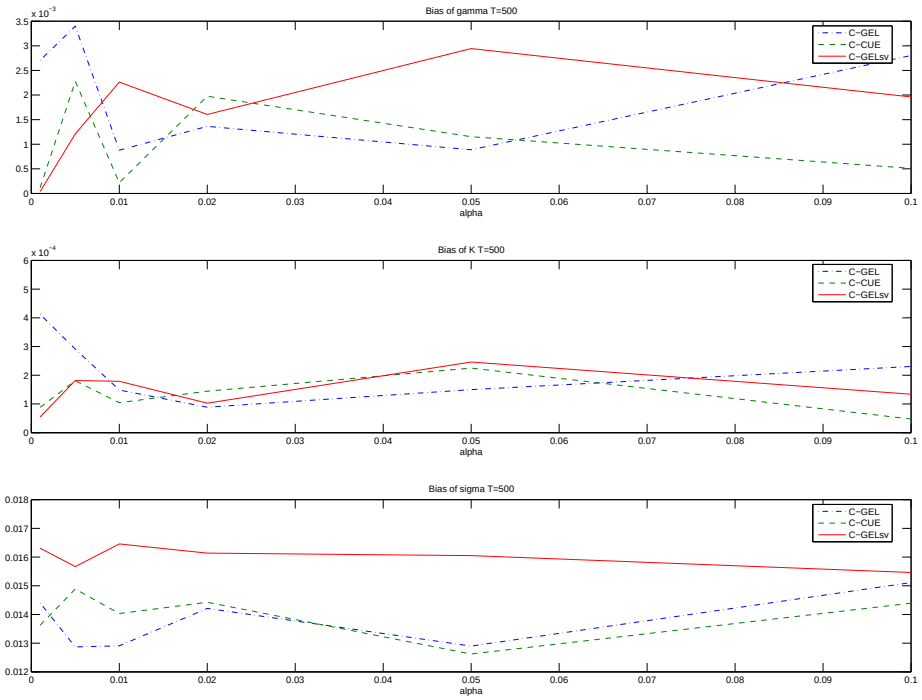

Figure 1: Bias of the different estimators for T=500

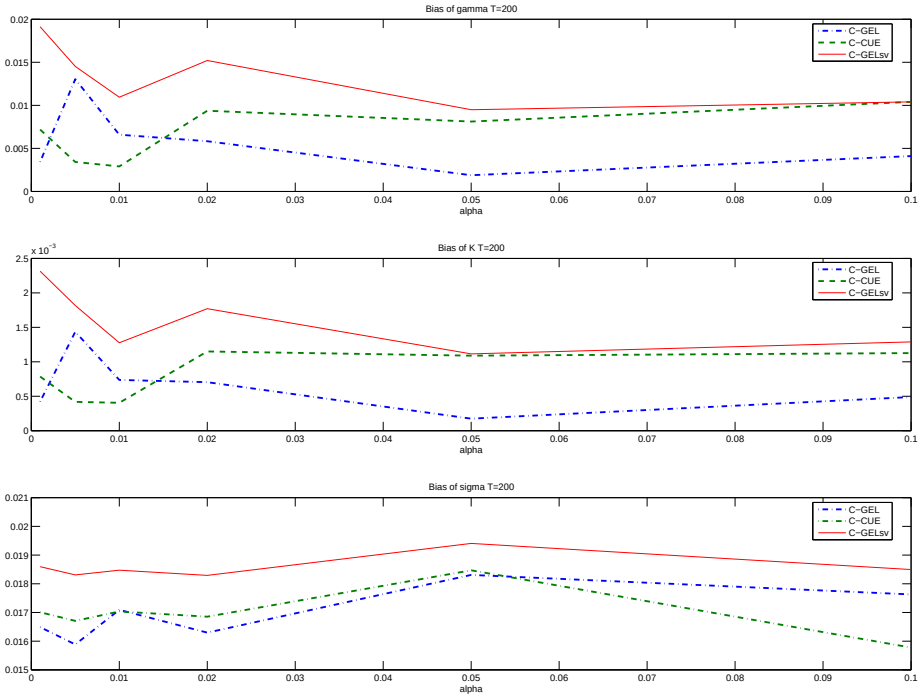

Figure 2: Bias of the different estimators for T=200

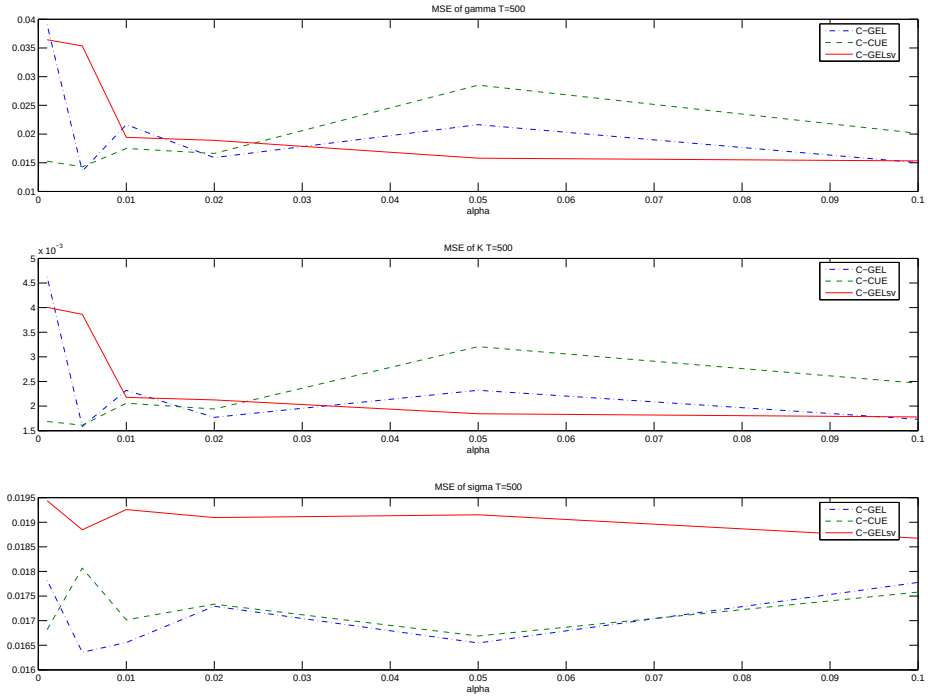

Figure 3: MSE of the different estimators for T=500

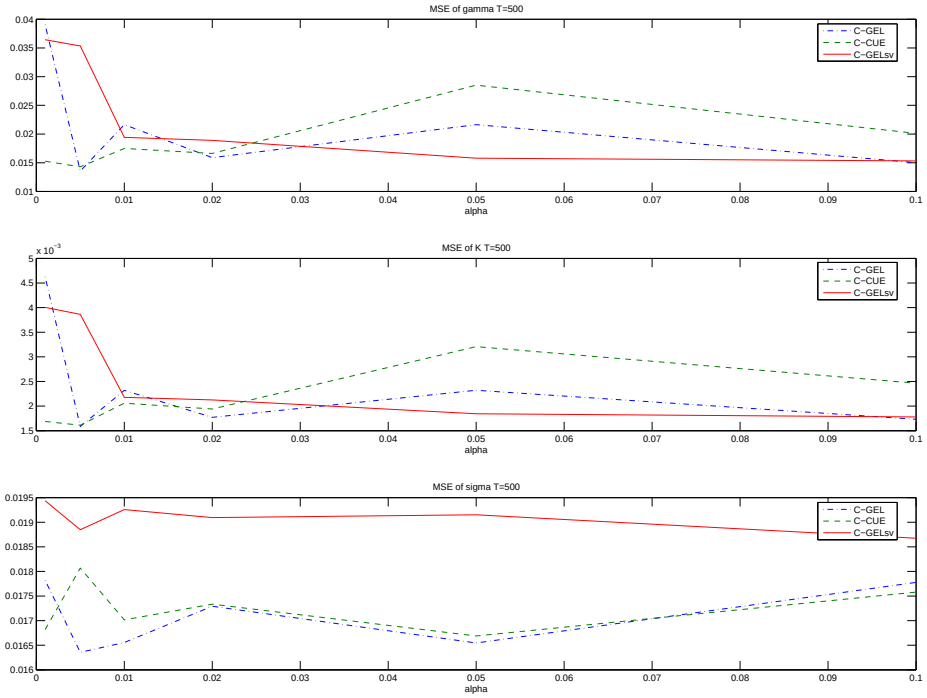

Figure 4: MSE of the different estimators for T=200

References

Airayetpan, R.G. and A.G. Ramm, "Dynamical systems and Discrete Methods for Solving Nonlinear Ill-posed Problem," *Applied Mathematics Review*, 2000, *1*, 491–536.

Anatolyev, S., "GMM, GEL, Serial Correlation, and Asymptotic Bias," *Econometrica*, 2005, *73*, 983–1002.

Andrews, D.W.K., "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 1991, *59*, 817–858.

Berkowitz, J., "Generalized Spectral Estimation of the Consumption based Pricing Model," *Journal of Econometrics*, 2001, *104*, 269–288.

Bindelli, Luca, "Testing the New Keynesian Phillips Curve: a Frequency Domain Approach," 2005. Université de Lausanne. Dernière version: 2005.

Carrasco, M. and J.P. Florens, "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 2000, *16*, 655–673.

___ and ___ , "Efficient GMM Estimation Using the Empirical Characteristic Function," 2002.

___ , M. Chernov, J.P. Florens, and E. Ghysels, "Efficient of Général Dynamic Models with Continuun of Moment Conditions," *forthcoming Journal of Econometrics*, 2006.

Groetsch, C., "Inverse Problems in Mathematical Sciences," *Wiesbaden: Vieweg*, 1993.

Guggenberger, P. and J. Hahn, "Finite Sample Properties of the Two-Step Empirical Likelihood Estimator," *Econometric Reviews*, 2005, *24* (3), 247–263.

Hansen, L.P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50*, 1029–1054.

___ , John Heaton, and Amir Yaron, "Finit-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 1996, *14*, 262–280.

Kitamura, Yuichi, "Empirical Likelihood Methods With Weakly Dependent Processes," *The Annals of Statistics*, 1997, *25* (5), 2084–2102.

Li, Kaitai and Yinnian He, "Taylor Expansion Algorithme for the Branching Solution of the Navier-Stokes Equations," *International Journal Numerical Analysis and Modeling*, 2005, *2* (4), 459–478.

Luenberger, David G., *Optimization by Vector Space Methods*, Wiley and Sons, 1997.

Newey, W. K. and K. D. West, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 1987a, *55*, 703–708.

Newey, W.K. and R.J. Smith, "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 2004, *72*, 219–255.

Owen, Art B., *Empirical Likelihood*, Chapman and Hall, 2001.

Parzen, Emanuel, "Statistical inference on time series by RKHS methods," *Canadian Mathematical Congress*, 1970, *Edited by R. Pyke*, 1–37.

Qi-Nian, Jin, "On the Iterative Regularized Gauss-Newton Method for Solving Nonlinear Ill-posed Problem," *Mathematics of Computation*, 2000, *69* (232), 1603–1623.

Qin, Jing and Jerry Lawless, "Empirical Likelihood and General Estimating Equation," *The Annals of Statistics*, 1994, *22* (1), 300–325.

Ramm, A.G., "Dynamical systems method for solving operator equations," *Comm. Nonlinear Sci. and Numer. Simul.*, 2004a, *9* (N2), 383–402.

——— , "Dynamical systems method for solving nonlinear operator equations," *Int. J. Appl. Math. Sci.*, 2004b, *1*, 97–110.

Singleton, K., "Estimation of Affine Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 2001, (102), 111–141.

Smith, Richard J., "Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation," *The Economic Journal*, 1997, *107*, 503–519.