Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions*

Marine Carrasco
University of Rochester†

Mikhail Chernov
Columbia Business School‡

Jean-Pierre Florens
University of Toulouse§

Eric Ghysels
University of North Carolina¶

First Draft: November 2000
This Revision: December 16, 2002

*We would like to thank Yacine Aït-Sahalia, Ruslan Bikbov, Lars Hansen, Mike Johannes, Nour Meddahi, Benoit Perron, Eric Renault, Tano Santos, Ken Singleton, the conference and seminar participants at the Canadian Econometric Study Group in Quebec city, CIREQ-CIRANO-MITACS Conference on Univariate and Multivariate Models for Asset Pricing, Econometric Society North American winter meeting in New Orleans, Western Finance Association in Tucson, Chicago, Princeton, ITAM, Michigan State, Montréal, and USC for comments. We also thank Jun Pan for providing us Matlab routines to compute characteristic functions of affine term structure models. Carrasco acknowledges financial support from the National Science Foundation, grant # 0211418.

†Department of Economics, Harkness Hall 226, Rochester, NY 14627, USA, Phone: (585) 275-7307, Fax: (585) 256-2309, Email: csco@troi.cc.rochester.edu

‡Division of Finance and Economics, 413 Uris Hall, 3022 Broadway, New York, NY 10027, USA, Phone: (212) 854-9009, Fax: (212) 316-9180, Email: mc1365@columbia.edu

§Manufacture des Tabacs, Bâtiment F, 21 allée de Brienne, 31000 Toulouse - France, Phone: (0)5.61.12.85.96, Fax: (0)5.61.12.86.37, Email: florens@cict.fr

¶Department of Economics, Gardner Hall, CB 3305, Chapel Hill, NC 27599, USA, Phone: (919) 966-5325, Fax: (919) 966-4986, Email: eghysels@unc.edu

Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions

Abstract

A general estimation approach combining the attractive features of method of moments with the efficiency of ML is proposed. The moment conditions are computed via the characteristic function. The two major difficulties with the implementation is that one needs to use an infinite set of moment conditions leading to the singularity of the covariance matrix in the GMM context, and the optimal instrument yielding the ML efficiency was previously shown to depend on the unknown probability density function. We resolve the two problems simultaneously in the framework of C-GMM (GMM with a continuum of moment conditions). First, we prove asymptotic properties of the C-GMM estimator applied to dependent data and then provide a reformulation of the estimator that enhances its computational ease. Second, we propose to span the unknown optimal instrument by an infinite basis consisting of simple exponential functions. Since the estimation framework already relies on a continuum of moment conditions, adding a continuum of spanning functions does not pose any problems. As a result, we achieve ML efficiency when we use the values of conditional CF indexed by its argument as moment functions. We also introduce HAC-type estimators so that the estimation methods are not restricted to settings involving martingale difference sequences. Hence, our methods apply to Markovian and non-Markovian dynamic models. Finally, a simulated method of moments type estimator is proposed to deal with the cases where the characteristic function does not have a closed-form expression. Extensive Monte-Carlo study based on the models typically used in term-structure literature favorably documents the performance of our methodology.

Introduction

Recent advances in estimation of univariate diffusions have highlighted the shortcomings of many standard continuous time models often used in asset pricing.[1] As a consequence additional factors, such as stochastic volatility or jumps, are required to account for these shortcomings. Unfortunately, the extant univariate econometric methods can not be easily extended to the multivariate case.

These developments prompted the introduction of new estimation methods. In principle, the generalized method of moments (GMM) estimation approach is quite general because, despite unknown expressions for the probability density function (p.d.f.), moment conditions are available in analytical form for many multifactor models of practical interest. The choice of the appropriate moments typically is a challenge because the efficiency may vary with the set of moment conditions. Of course, the maximum likelihood (ML) method is efficient and for this reason is more attractive than GMM. Since ML is not feasible in most multivariate settings several simulation-based maximum likelihood methods have been introduced recently.[2]

This paper proposes a general estimation approach which combines the attractive features of method of moments estimation with the efficiency of ML in one framework. The method exploits the moment conditions computed via the characteristic function (CF) of a stochastic process instead of the likelihood function, as in the recent work by Chacko and Viceira (1999), Jiang and Knight (2002), and Singleton (2001). The most obvious advantage of such an approach is that in many cases the CF is available in analytic form, while the likelihood is not, the most celebrated example being the class of affine diffusion models. Moreover, the CF contains the same information as the likelihood function up to the Fourier transform. Therefore, a clever choice of moment conditions should provide the same efficiency as ML.

Another advantage of the CF-based estimation is that it applies to many different settings. There are two cases where the advantages of the CF-based estimation are the most notable.

[1] Applications of the parametric and non-parametric methods in finance include among others, Aït-Sahalia (1996), Aït-Sahalia (2000), Conley et al.(1997), Hansen and Scheinkman (1995), Lo (1988).

[2] See Brandt and Santa-Clara (2000), Durham and Gallant (2000), Elerian, Chib and Shephard (2001), Eraker, Johannes, and Polson (2000) for various implementations of simulation-based ML.

One typical example is an N-factor affine term-structure model with N observed yields corresponding to different maturities. In this case, yields are a linear function of the state variables (see e.g. Duffie and Kan, 1996), and, therefore, these state variables are effectively observed. A second application involves a jump component. Very often jump-diffusion specifications imply that asset prices come from a mixture of distributions. In this case, the likelihood function is not bounded, and ML estimation is not feasible (see Honoré, 1998). On the other hand, it is known that GMM using ad hoc moment conditions does not achieve the efficiency associated with ML. This paper shows that GMM based on CF achieves ML efficiency. The applicability of our method is not limited to the two previous examples. For instance, it can be applied to randomly sampled continuous time processes and stochastic volatility models as well.

The main contribution of this paper is the resolution of two major difficulties with the estimation via the CF. The first one is related to the intuition that the more moments one generates by varying the CF argument, the more information one uses, and, therefore, the estimator becomes more efficient. However, as one refines the range of CF argument values, the associated covariance matrix approaches singularity. The second difficulty is that in addition to a large set of CF-based moment conditions, one requires an optimal instrument to achieve the ML efficiency. Prior work (e.g. Feuerverger and McDunnough, 1981 or Singleton, 2001) derived the optimal instrument, which is a function of the unknown probability density. Such an estimator is clearly hard to implement.

We use the framework of Carrasco and Florens (2000a), known as C-GMM, to rely on a continuum of moment conditions in a GMM procedure. This allows us to address the two problems simultaneously. First, the original work of Carrasco and Florens deals with covariance matrix singularity by replacing it with a covariance operator. Computing the unbounded inverse operator is known in functional analysis as an ill-posed problem and can be resolved by regularizing the operator. This method was initially developed in an iid framework where the moment functions were indexed by an index parameter in an interval of $\mathbb{R}$. We reformulate the GMM objective function to facilitate the implementation of the estimation technique. We also allow the moment functions to be complex valued and be functions of an index parameter taking its values in $\mathbb{R}^d$ for an arbitrary $d \geq 1$ in order to accommodate the specific features of CF. Finally, we extend the method in various directions in order to be able to resolve the second problem of the instrument choice. We distinguish two cases depending on whether the observable variables are Markov or not.

In the Markov case, the moment conditions are based on conditional CF. Therefore, we derive the asymptotic properties of the C-GMM estimator applied to dependent data. We then propose to span the unknown optimal instrument by an infinite basis consisting of simple exponential functions. Since the estimation framework already relies on a continuum of moment conditions, adding a continuum of spanning functions does not pose any problems. As a result, we achieve ML efficiency when we use the values of conditional CF indexed by its argument as moment functions. We propose a simulated method of moments type estimator for the cases when CF is unknown. If one is able to draw from the true conditional distribution, then the conditional CF can be estimated via simulations and ML efficiency obtains. This approach can be thought as a simple alternative to the Indirect Inference proposed by Gouriéroux, Monfort and Renault (1993) and the Efficient Method of Moments (EMM) suggested by Gallant and Tauchen (1996).

If the observations are not Markov, it is not possible to construct the conditional CF. Therefore, we propose to use joint CF of a particular number of data lags, which do not have to be martingale differences. Hence, we extend the C-GMM methodology to autocorrelated moment functions. While we were not able to obtain optimal moment functions yielding ML efficiency in this case, we derived an upper bound on the variance of the resulting estimator. In the worst case scenario, if one uses the joint CF for estimation, the variance of the C-GMM estimator corresponds to that of the ML estimator based on the joint density of the same data lags. As the joint CF is often unknown, a simulated method of moments becomes especially useful. The simulation scheme differs from that used in the Markov case. Instead of simulating conditionally on the observable data, we simulate the full time-series as it is done in Duffie and Singleton (1993).

The paper is organized as follows. The first section provides motivating examples and reviews issues related to the estimation via CF. Section 3 extends the C-GMM proposed by Carrasco and Florens (2000a) to the case where the moment functions are correlated. It shows how to estimate the long-run covariance and how to implement the C-GMM estimator in a simple way. Section 4 specializes to the cases where the moment conditions are based either on the conditional characteristic function or joint characteristic function. In the first case, we establish which choice of the instrument function yields ML efficiency. These results can be applied in a straightforward manner in case of fully observed vector of state variables, i.e. a Markov process. Section 5 considers simulation-based CF estimation which is of greatest importance for partially observed state vector (non-Markov) processes. Finally, a Monte Carlo comparison of the C-GMM estimator with other popular estimators is reported in Section 6. The last section concludes.

1 Motivating Examples

We provide motivating examples that are of interest in many applications and for which there is no feasible maximum likelihood estimation available. The characteristic function based methods, henceforth CF-based, will provide feasible estimators that attain ML efficiency in each of these cases. The first class of processes are multivariate diffusions prominently used in the term structure literature and also other continuous time multiple-asset pricing models. These models have typically been estimated via QMLE, or simulation-based method of moments. Next we consider diffusion processes augmented with a jump component. ML estimation of such processes has several difficulties that can be circumvented via CF-based estimators. The final subsection covers subordinated processes, also traditionally challenging for the implementation of efficient estimation procedures.

1.1 Multivariate Affine Diffusions

Suppose the sequence X_t, $t = 1, \dots, T$ is observed, where $X_t \in \mathbb{R}^p$ with $p \geq 1$. It is assumed the process X_t is Markov and satisfies the following stochastic differential equation:

$$dX_t = \mu\left(X_t, \theta_0\right) dt + \sigma\left(X_t, \theta_0\right) dW_t$$

where the function μ is the drift, σ^2 is the diffusion matrix and $\{W_t\}$ is a standard Brownian motion. Finally, $\theta \in \mathbb{R}^q$ is the parameter of interest and θ_0 is the true value of θ. The diffusion is assumed to be affine. Loosely speaking, this means that the drift and variance functions are linear in X_t (for a more formal characterization see Duffie, Filipovic and Schachermayer, 2002).

Multivariate affine diffusions, which date back to the works of Vasicek (1977) and Cox, Ingersoll and Ross (1985), play a key role in modeling the term structure of interest rates. This general rich class of models yields essentially closed-form expressions for zero-coupon-bond prices (see Duffie and Kan (1996) or Dai and Singleton (2000)) and characteristic functions (see Duffie, Pan and Singleton (2000)).

Let $\psi_\theta(\tau|X_t)$ denote the characteristic function of X_{t+1} conditional on X_t :

$$\psi_\theta(\tau|X_t) \equiv E^\theta\left(e^{i\tau X_{t+1}}|X_t\right) \tag{1.1}$$

By stationarity of X_t, $\psi_\theta(\tau|x)$ does not depend[3] on t. Under suitable regularity conditions, given in Proposition 1 of Duffie, Pan and Singleton (2000), one can show that the conditional characteristic function equals:

$$\psi_\theta(\tau|X_t) = \exp\left(A(\tau) + B(\tau)X_t\right) \tag{1.2}$$

where $A(\tau)$ and $B(\tau)$ satisfy complex-valued ordinary differential equations (ODE) that can either be solved explicitly or numerically (see Equations (2.5) and (2.6) in Duffie, Pan and Singleton (2000) or Equations (7) and (8) in Singleton (2001) for further details).

For multivariate processes a closed-form solution of the probability density $f(X_{t+1}|X_t,\theta)$ is not available and, therefore, MLE is not feasible. Consequently, estimation involves either quasi-MLE or approximations to the likelihood function. An example of the former is Duffee (2002) who uses a Gaussian density involving analytic expressions for the first and second conditional moments. Examples of the latter include Dai and Singleton (2000) who use simulation-based EMM and Aït-Sahalia and Kimmel (2002) who use closed-form polynomial expansions to the likelihood function proposed in Aït-Sahalia (2002a,b). In all these multivariate cases we will provide asymptotic efficient estimators that are equivalent to MLE.

[3]We use the notation $\psi_\theta(\tau|x)$ for $E^\theta\left(e^{i\tau X_{t+1}}|X_t = x\right)$.

1.2 Jump Diffusion Processes

There is a mounting evidence in the empirical literature that a jump component is an important modeling component for financial times series.[4] However, even if the analytical likelihood is available, estimation of this component presents certain challenges, which can be resolved by relying on the CF rather than ML methods. The issue goes back to at least Kiefer (1978) and relates to the mixture of normal distributions.

We will sacrifice some generality to discuss the issues, namely consider the univariate jump-diffusion, also known as the Merton (1976) model.[5] The Merton model discretized over interval Δ takes the form:

$$\Delta X_t = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta + \sigma\Delta W_t + J_t \Delta N_t \tag{1.3}$$

where $\Delta W_t \sim \mathcal{N}(0, \Delta)$, $J_t \sim \mathcal{N}(\mu_J, \delta_J^2)$, and ΔN_t follows a Bernoulli with probability $\lambda\Delta <$ 1. As a result, each of the sample observations ΔX_t, $t = 1, \ldots, T$, comes from a normal distribution with parameters $\mu_1 = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta$, $\sigma_1 = \sigma\sqrt{\Delta}$, with probability $\gamma = \lambda\Delta$, and from a normal distribution with parameters $\mu_2 = \left(\alpha - \frac{1}{2}\sigma^2\right)\Delta + \mu_J$, and $\sigma_2 = \sqrt{\sigma^2\Delta + \delta_J^2}$ with probability $1 - \gamma$.

Honoré (1998) points out the difficulties that arise with ML estimation in this case because the likelihood is unbounded. As a remedy, he proposes to tie up the two unknown volatilities σ_1, and σ_2 via a multiplicative parameter and re-estimate the whole model for each value of this parameter from a selected grid. Alternatively, Quandt and Ramsey (1978), recognizing the same issue, suggest to rely on the method of moments, where moment conditions are based on the CF because, contrary to the likelihood, the CF is always bounded. The CF of ΔX_t is the weighted sum of the CF of two normal distributions, namely:

$$\psi_{(\mu_1,\mu_2,\sigma_1,\sigma_2,\gamma)}(\tau) = \gamma\psi_{(\mu_1,\sigma_1)}(\tau) + (1-\gamma)\psi_{(\mu_2,\sigma_2)}(\tau) \tag{1.4}$$

where $\psi_{(\mu_l,\sigma_l)}(\tau) = \exp\left(i\tau\mu_l - \tau^2\sigma_l^2/2\right)$, $l = 1, 2$.

[4] The most recent examples include, among others, Johannes (2000) and Piazzesi (2000) for interest rates, and Chernov et al. (2002), Johannes, Eraker, and Polson (2001) and Pan (2002) for equities.

[5] See Duffie, Filipovic and Schachermayer (2002) for the most general affine jump-duffusion specifications and regularity conditions.

In fact, often the mixture of distributions can not be computed analytically, while the respective CF is available. One example is a combination of normal and exponential distributions, which is encountered in the models of jumps to volatility (Duffie, Pan, and Singleton, 2000; Eraker, Johannes, and Polson, 2001). The convolution of the two distributions is only available as an approximation, while the CF can be computed via (1.2) where $A(\tau)$ and $B(\tau)$ satisfy complex-valued ODE that are augmented with the jump component parameters. Inference can be performed using the method described in this paper and we show that the resulting estimator is efficient.[6] Hence, the best of the two worlds can be achieved: ML efficiency in the framework of the method of moments which avoids the likelihood function unboundedness. This view on the jump component simplifies many estimation problems recently encountered in finance.[7]

1.3 Subordinated Diffusions

Since asset prices are driven by information arrivals there is a long tradition in finance to consider subordinated processes, an idea originated in the work of Mandelbrot and Taylor (1967) and Clark (1973). They argued that since the number of transactions in any time period is random, one may think of asset price movements as a realization of a process $X_t = Y_{\mathcal{T}_t}$. The nondecreasing stochastic process $\mathcal{T}_t$ is a directing process related to the number of transactions or, more fundamentally, to the arrival of information. Obviously, as noted by Mandelbrot and Taylor, time deformation is also related to the mixture of distributions model (see e.g. Tauchen and Pitts (1983)).[8]

For illustration, assume $\mathcal{T}_t = \int_0^t X_u^* du$ where X_u^* takes on positive values. Moreover X_u^* is

[6] Note that the maximum likelihood approach of Aït-Sahalia (2002b) does not apply to jump diffusion processes.

[7] Schaumburg (2000) is also concerned with the estimation of Lévy processes. He proposes a procedure which approximates the likelihood function based on a representation of a vector in a Hilbert space using Fourier series. However, the closed form expression of the CF is available for all Lévy processes via the Lévy-Khintchine formula. Hence CF-based estimation is much more simple and intuitive.

[8] There is now a substantial literature on time deformation, recent examples include Madan and Seneta (1990), Geman and Yor (1993), Anderson (1996), Ghysels and Jasiak (1996), Ghysels, Gouriéroux and Jasiak (1997), Carr, Geman, Madan and Yor (2001) and Carr and Wu (2002).

supposed to be observed, e.g. it could model the volume of transactions. Assume that both Y_t and X_t^* are affine diffusion processes:

$$\begin{cases} dY_t = \mu\left(Y_t, \theta_0\right) dt + \sigma\left(Y_t, \theta_0\right) dW_t \\ dX_t^* = \mu^*\left(X_t^*, \theta_0\right) dt + \sigma^*\left(X_t^*, \theta_0\right) dW_t^* \end{cases}$$

where W_t and W_t^* are independent Brownian motions. It can be shown that (X_t, X_t^*) is a bivariate diffusion with drift $[X_t^* \mu\left(X_t\right), \mu^*\left(X_t^*\right)]'$ and diagonal diffusion matrix

$$\begin{bmatrix} X_t^* \sigma\left(X_t\right)^2 & 0 \\ 0 & \sigma^*\left(X_t^*\right)^2 \end{bmatrix}.$$

Using (1.2), one can derive the conditional CF of $\left(X_{t+1}, X_{t+1}^*\right)$ given (X_t, X_t^*) and estimate efficiently the parameters of both diffusions. Other examples of subordinations (possibly endogenous) resulting in Markov processes are discussed in Carrasco, Hansen, and Chen (1998).

To conclude, it is also worth noting that the framework of subordinated diffusions easily adapts to that of randomly sampled data.[9] In this case the calendar time spacing of data is the directing process of the diffusion process. Hence, the CF approach easily allows to take into account the random nature of data sampling.

2 Overview of the Methodology

In this section we discuss the major unresolved issues pertaining to estimation via CF and explain how we propose to tackle them via GMM based on the continuum of moment conditions (C-GMM). We then intuitively describe how to use the continuum of moment in practice. Finally, we give an overview of the main results of this paper. Our discussion is based on the most simple case where the conditional CF is known in closed-form. In this case, we will be able to abstract from many technical details and provide the most transparent expression for the C-GMM objective function. More general results and regularity conditions will be discussed in subsequent sections.

[9] Aït-Sahalia and Mykland (2002) emphasize the importance of taking into account the data sampling scheme for the asymptotic properties of estimators. See also Duffie and Glynn (2001), who rely on random sampling to construct their GMM estimators.

2.1 Estimation based on Characteristic Function: The issues

Since the conditional characteristic function is available in all the aforementioned cases as well as many others, one may think of using the CF to generate a set of moment conditions to estimate θ. Assume that X_t is scalar to simplify. Equation (1.1) implies that the following unconditional moment conditions are satisfied:

$$E^{\theta} h_t(\tau;\theta) = 0 \text{ for all } \tau \in \mathbb{R}$$

with

$$h_t(\tau;\theta) \equiv h(\tau, X_t, X_{t+1}, \theta) = (e^{i\tau X_{t+1}} - \psi_\theta(\tau|X_t))m(\tau, X_t) \tag{2.1}$$

where $m(\tau, X_t)$ is an arbitrary instrument. There are two issues of interest here: the choice of τ and the choice of the instrument $m(\tau, X_t)$.

The generalized method of moments is the simplest estimation procedure, which can exploit these moment conditions. Consider a discrete set of τ_i, $i = 1, \ldots, N_\tau$. The estimator of θ is obtained as:

$$\hat{\theta}_T = \underset{\theta}{\text{argmin}}\ \hat{h}_T(\theta)' W_T \hat{h}_T(\theta) \tag{2.2}$$

where the vector $\hat{h}_T(\theta)$ with ith individual element $\hat{h}_T(\tau_i;\theta)$ is known as a set of sample moment functions:

$$\hat{h}_T(\tau_i;\theta) = \frac{1}{T}\sum_{t=1}^{T} h_t(\tau_i;\theta) \tag{2.3}$$

with $h_t(\tau_i;\theta)$ denoting the unconditional moments and W_T is the weighting matrix.[10]

We first discuss the selection of an appropriate set of τ_i. Different sets of τ_i may lead to different values of the θ estimates because they describe different aspects of the data generating process (henceforth DGP). Increasing the number of moments of the type (2.1) should describe the properties of the DGP distribution more and more accurately. Indeed, if τ goes through all real numbers (N_τ approaches infinity), the probability density can be recovered. The problem with such an approach is that, as the number of moment conditions

[10] This approach is taken by Chacko and Viceira (1999), who select the unity instrument, $m(\tau, X_t) = 1$.

increases, they become more and more correlated with each other, and their covariance matrix, K_T, becomes singular, therefore the optimal weighting matrix $W_T = K_T^{-1}$ can not be computed.

The selection of a grid for the real-valued index τ is not the only problematic issue. The optimal choice of instruments $m(\tau, X_t)$ is cumbersome as well. Feuerverger and McDunnough (1981), and Singleton (2001) discuss under which conditions the CF-based estimator achieves the Cramér-Rao lower bound and show that the instrument that attains the bound is:

$$m(\tau, X_t) = \frac{1}{(2\pi)^p} \int e^{-i\tau x} \frac{\partial \ln f_\theta}{\partial \theta} (x|X_t)\, dx \tag{2.4}$$

The drawback of this approach is that the instrument m requires the knowledge of the unknown likelihood function f_θ. [11]

In this paper we will be able to address the two raised issues – (i) potential covariance matrix singularity, and (ii) optimal selection of instrument without relying in the unknown probability density function – using the framework of Carrasco and Florens (2000a), who proposed using a continuum of moment conditions in the context of GMM.

2.2 Extending GMM to a continuum of moment conditions

Extension of the regular GMM to the one utilizing the continuum of moment conditions can be understood as follows. In the framework of regular GMM τ_i, which indexes individual moment conditions in (2.2), (2.3), can be selected to be equal to i/N_τ. Note that for finite N_τ, the expression in (2.2) with the optimal weighting matrix is equivalent to:

$$\hat{\theta}_T = \underset{\theta}{\text{argmin}}\, ||K_T^{-1/2} \hat{h}_T(\theta)|| \tag{2.5}$$

[11] There are certain parallels between the raised issues and the estimation of univariate subordinated diffusions via an infinitesimal generator in Conley, Hansen, Luttmer, and Scheinkman (1997). They show that, assuming a continous sampling, constructing moment conditions by applying the generator to the likelihood score of the marginal distribution is optimal and, in particular, is more efficient than building moments via the score directly. Being unable to implement in practice the corresponding optimal instrument (or test function) for the discrete sampling case, they still use the score for the empirical application.

where $||\cdot||$ is the euclidean norm on $\mathbb{R}^{N_\tau}$:

$$||f||^2 = \sum_{\tau=1}^{N_\tau} f^2(\tau). \tag{2.6}$$

Then letting N_τ increase without bound, the entire continuum set of moment conditions can be recovered. When N_τ is infinite the objective function (2.2) can not be constructed. The extension to infinite N_τ is easier to perform given the representation in (2.5) because it suggests that one has to consider an objective function based on a different norm, which takes into account the infiniteness of the set of possible values of τ :

$$||f||^2 = \int_{\mathbb{R}} f(\tau)\overline{f(\tau)}\pi(\tau)d\tau \tag{2.7}$$

where $\bar{f}$ denotes the complex conjugate of f, and π denotes a probability density function (pdf), which is typically selected to be Gaussian.

This new norm gives a flavor of how one could use the continuum set of moment conditions in a GMM framework. However, it is still not clear how one could compute the infinite-dimensional covariance matrix K_T. It turns out that this is feasible in a new space of moment conditions, or more accurately, moment functions endowed with a norm (2.7) (Assumption A.2). In this space, K_T is understood as the covariance operator, and the problem of finding its inverse is well studied in functional analysis. The next subsection will discuss these issues in detail.

Applying the continuum of moment conditions principle to the CF-based moment conditions (2.1), one is able to exploit all the information contained in CF. The appropriate choice of m, which happens not to depend on the unknown p.d.f. $f(X_{t+1}|X_t;\theta)$, also becomes easy in this setting. As we will show in Section 4, viewing τ as a double index $(r,s)'$, and constructing the C-GMM estimator based on moment functions:

$$h_t(\tau;\theta) = (e^{isX_{t+1}} - \psi_\theta(s|X_t))m(r,X_t) \tag{2.8}$$

with

$$m(r,X_t) = e^{irX_t} \tag{2.9}$$

yields an ML-efficient estimator.

Such a choice of instrument is quite intuitive. Although we can not construct the optimal instrument in (2.4), we can span it via a set of basis functions. The utilization of the continuum of moment conditions is precisely what allows us to perform this spanning. Moreover, as we show in Appendix C, because of the simple functional form of the new optimal instrument (2.9), the introduction of the double index does not increase the computational complexity of the estimation procedure: all elements associated with the index r can be computed in analytic form.

In order to further our understanding of the C-GMM methodology we next provide informal statements of the key results, which show how to construct the estimator in the case where the conditional CF is available in analytical form. Since moment conditions (2.1) based on conditional CF form martingale difference sequence, all the results are particularly simple. A rigorous statement of more general results will be provided in section 3.[12]

2.3 An Example of Computing the C-GMM Estimator

We assume that the stationary process X_t is a $p \times 1$-vector of random variables which represents the data-generating process indexed by a finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^q$.

The C-GMM estimator is based on the arbitrary set of moment conditions:

$$E^{\theta_0} h_t(\tau; \theta_0) = 0 \tag{2.10}$$

where $h_t(\tau; \theta)$ and index $\tau \in \mathbb{R}^d$. We will refer to $h_t(\tau; \theta_0)$ as moment function. Let $\hat{h}_T(\tau; \theta_0) = \sum_{t=1}^{T} h_t(\tau; \theta_0)/T$ denote the sample mean of the moment functions. Having the CF-based moment conditions (2.1) in mind, we assume here that the moment functions form martingale difference sequences.

As discussed above, our goal is to consider all moment functions associated with different values of the index τ. This requirement implies an infinity of possible functions. The most

[12] We will need more general results because CF-based estimation is not limited to analytic conditional CF. In the following sections we will develop CF-based estimators when the observable data is not Markov, and hence conditional CF is not available, and when there is no analytic expression for CF, i.e. simulated C-GMM.

convenient way to work with such infinite set is to impose a Hilbert space structure, and in particular to define the inner product, which leads to the norm discussed in (2.7). Assumption A.2 introduces a space $\mathbb{L}^2(\pi)$ to which $h_t(.;\theta_0)$ belongs as a function of τ. The inner product in this space is defined as

$$\langle f, g\rangle = \int f(\tau)\,\overline{g(\tau)}\pi(\tau)\,d\tau \tag{2.11}$$

where $\overline{g(\tau)}$ denotes the complex conjugate of $g(\tau)$ and π is a pdf usually selected to be Gaussian. The norm corresponding to the inner product is $\| f \|^2 = \langle f, f\rangle$ coincides with the one in (2.7).

Anticipating the GMM asymptotic results, we need to think of an analogue of the optimal weighting matrix, which features in the objective function (2.5). The covariance operator K is the counterpart of the covariance matrix in finite dimension. It is an integral operator that can be written as

$$Kf(\tau_1) = \int k(\tau_1, \tau_2)\, f(\tau_2)\,\pi(\tau_2)\,d\tau_2 \tag{2.12}$$

with

$$k(\tau_1, \tau_2) = E^{\theta_0}\left(h_t(\tau_1;\theta_0)\,\overline{h_t(\tau_2;\theta_0)}\right) \tag{2.13}$$

In order to implement the C-GMM estimator with the optimal weighting operator (2.5), we have to estimate K. Given a simple form of the operator (2.12), (2.13), we can estimate it via the usual two-step procedure. In the remainder we let $\hat{\theta}_T^1$ be a $T^{1/2}-$consistent first step estimate of θ_0 :

$$\hat{\theta}_T^1 = \underset{\theta}{\text{argmin}} \left\|\hat{h}_T(\tau;\theta)\right\| \tag{2.14}$$

We construct the second-step covariance operator estimator by estimating k in (2.12) by

$$\hat{k}_T(\tau_1, \tau_2) = \frac{1}{T}\sum_{t=1}^{T} h_t\left(\tau_1;\hat{\theta}_T^1\right)\overline{h_t\left(\tau_2;\hat{\theta}_T^1\right)} \tag{2.15}$$

Finding the inverse of the covariance operator involves solving equation

$$Kg = f \tag{2.16}$$

with respect to g. Thus, the issue of covariance matrix singularity in regular GMM is replaced by the issue of covariance operator invertibility. As discussed in Carrasco and Florens (2000a), since the inverse of K is not bounded, the solution of this equation is not continuous in f, in other words, it is unstable to small perturbations of f. We, therefore, consider the regularized version of the inverse, involving a penalizing term α_T. Namely, the operator K is replaced by some nearby operator that has a bounded inverse. For $\alpha_T > 0$, the equation:

$$\left(K_T^2 + \alpha_T I\right) g = Kf \tag{2.17}$$

has a unique stable solution for each $f \in \mathbb{L}^2(\pi)$. The Tikhonov approximation of the generalized inverse to K is given by:

$$\left(K_T^{\alpha_T}\right)^{-1} = \left(K_T^2 + \alpha_T I\right)^{-1} K_T$$

In order to implement the square-root of the inverse of the covariance operator we have to represent it in terms of the eigenvalues, $\hat{\lambda}_j$ and corresponding eigenfunctions (principal components), $\hat{\phi}_j$, of K_T :

$$\left(K_T^{\alpha_T}\right)^{-1/2} f = \sum_{j=1}^{T} \frac{\sqrt{\hat{\lambda}_j}}{\sqrt{\hat{\lambda}_j^2 + \alpha_T}} \left\langle f, \hat{\phi}_j \right\rangle \hat{\phi}_j. \tag{2.18}$$

This expression shows that α_T is used to discard the smallest, i.e. the least informative, principal components $\hat{\phi}_j$, and, this way the analogue of the covariance matrix singularity problem is resolved. The choice of α_T is clearly important: if it is too large the generalized inverse will be far away from the actual inverse, and if it is too small the generalized inverse will be unstable. We determine the rate at which α_T should converge to zero. However note that the penalizing term α_T really plays a role only to compute the optimal weighting operator and hence to obtain an optimal C-GMM estimator. An estimator obtained for an arbitrary fixed $\alpha_T > 0$ will be still consistent but will have a larger variance

Having understood how to estimate the covariance operator and how to approximate its inverse, one way to implement the C-GMM estimator is to minimize the objective function:

$$\hat{\theta}_T = \underset{\theta}{\operatorname{argmin}} \left\| \left(K_T^{\alpha_T}\right)^{-1/2} \hat{h}_T(\tau;\theta) \right\| = \underset{\theta}{\operatorname{argmin}} \sum_{j=1}^{T} \frac{\hat{\lambda}_j}{\hat{\lambda}_j^2 + \alpha_T} \left| \left\langle \hat{h}_T(\tau;\theta), \hat{\phi}_j \right\rangle \right|^2. \tag{2.19}$$

Such an estimator will have the usual $\sqrt{T}$ asymptotic normality properties. However, the computation of eigenvalues and eigenfunctions can be difficult in large samples. It turns out that it is possible to rewrite the objective function (2.19) in terms of matrices and vectors.

Proposition 2.1 *Solving (2.19) is equivalent to solving*

$$\min_{\theta} \underline{v}'(\theta) \left[I_T - C \left[\alpha_T I_T + C^2 \right]^{-1} C \right] \underline{v}(\theta) \tag{2.20}$$

where C is a $T \times T-$matrix with (t,l) element $c_{tl}/(T-q)$, $t,l = 1,...,T$, I_T is the $T \times T$ identity matrix, $\underline{v} = [v_1,...,v_T]'$ with

$$\begin{aligned} v_t(\theta) &= \left\langle \hat{h}_T(\tau;\theta), h_t\left(\tau;\hat{\theta}_T^1\right) \right\rangle, \\ c_{tl} &= \left\langle h_l\left(\tau;\hat{\theta}_T^1\right), h_t\left(\tau;\hat{\theta}_T^1\right) \right\rangle. \end{aligned}$$

Further computational simplifications are obtained by explicit substitution of the moment functions (2.8) with optimal instruments (2.9). These computations are provided in the Appendix C. Standard errors are computed in a similar fashion (see section 3.3).

Section 3 develops these ideas formally. We apply the martingale difference results discussed in this section to derive the ML-efficient estimator based on the conditional CF in Section 4.1. Then, in Section 4.2, we use more general results to derive the properties of the estimator based on joint CF, which is relevant for processes with latent variables. Finally, Section 5 establishes properties of the simulation-based estimators when CF is not available in analytic form.

3 C-GMM with dependent data

This section will extend the results of Carrasco and Florens (2000a) in the i.i.d. case to the case where the data are weakly dependent. The first subsection proves asymptotic normality and consistency of the C-GMM estimator, introduces the covariance operator and its regularized version, which is known to yield the C-GMM estimator with the smallest variance. The next subsection derives the convergence rate of the estimator of the covariance operator. The third subsection proposes a simpler way to compute the C-GMM objective function in large samples in terms of matrices and vectors as opposed to the computation of

Carrasco and Florens (2000a) in terms eigenvalues and eigenfunctions. The last subsection discusses the choice of moment conditions to achieve ML efficiency.

All regularity conditions used in this section are collected in Appendix A. All the proofs are provided in Appendix D.

3.1 General asymptotic theory

The data are assumed to be weakly dependent (see Assumption A.1 for a formal definition). The C-GMM estimator is based on the arbitrary set of moment conditions:

$$E^{\theta_0} h_t(\tau;\theta_0) = 0 \tag{3.1}$$

where $h_t(\tau;\theta) \equiv h(\tau, Y_t;\theta)$ with $Y_t = (X_t, X_{t+1}, .., X_{t+L})'$ for some finite integer L, and index $\tau \in \mathbb{R}^d$. [13] As a function of τ, $h_t(.;\theta_0)$ is supposed to belong to the set $\mathbb{L}^2(\pi)$ as described in definition A.2. Moreover all parameters are identified by the moment conditions (3.1), see Assumption A.3. Let $\hat{h}_T(\tau;\theta_0) = \sum_{t=1}^{T} h_t(\tau;\theta_0)/T$. In the sequel, we write the functions $h_t(.;\theta_0)$, $\hat{h}_T(.;\theta_0)$ as $h_t(\theta_0)$ and $\hat{h}_T(\theta_0)$ or to simplify h_t and $\hat{h}_T$. $\{h_t(\theta_0)\}$ is supposed to satisfy the set of Assumptions A.4, in particular h_t should be a measurable function of Y_t. Since L is finite, h_t inherits the mixing properties of X_t. Finally, h_t is assumed to be scalar because the CF itself is scalar and hence we do not need results for a vector h_t.

These assumptions allow us to establish the asymptotic normality of the moment functions.

Lemma 3.1 *Under regularity conditions A.1 to A.3, and A.4(i)(ii) we have*

$$\sqrt{T}\hat{h}_T(\theta_0) \Rightarrow \mathcal{N}(0, K)$$

as $T \to \infty$, in $\mathbb{L}^2(\pi)$ where $\mathcal{N}(0,K)$ is the Gaussian random element of $\mathbb{L}^2(\pi)$ with a zero mean and the covariance operator $K : \mathbb{L}^2(\pi) \to \mathbb{L}^2(\pi)$ satisfying

$$\begin{aligned} &\langle Kf, g\rangle \\ = \;& E^{\theta_0}\left[\langle f, h_t(\theta_0)\rangle \overline{\langle g, h_t(\theta_0)\rangle}\right] + \sum_{j=2}^{\infty} E^{\theta_0}\left[\langle f, h_1(\theta_0)\rangle \overline{\langle g, h_j(\theta_0)\rangle}\right] + \sum_{j=2}^{\infty} E^{\theta_0}\left[\langle f, h_j(\theta_0)\rangle \overline{\langle g, h_1(\theta_0)\rangle}\right] \end{aligned} \tag{3.2}$$

[13] In the previous section we discussed the case corresponding to $L = 1$.

for all f and g in $\mathbb{L}^2(\pi)$.[14] *Moreover the operator K is a Hilbert-Schmidt operator.*[15]

We can now establish the standard properties of GMM estimators: consistency, asymptotic normality and optimality.

Proposition 3.1 *Assume the regularity conditions A.1 to A.4 hold. Moreover, let B be a bounded linear operator defined on $\mathbb{L}^2(\pi)$ or a subspace of $\mathbb{L}^2(\pi)$. The null space of $B : \{f \in \mathbb{L}^2(\pi) : Bf = 0\} = \{0\}$. Let B_T be a sequence of random bounded linear operators converging to B. The C-GMM estimator*

$$\hat{\theta}_T = \underset{\theta}{argmin} \left\| B_T \hat{h}_T(\theta) \right\|$$

has the following properties:

1. *$\hat{\theta}_T$ is consistent and asymptotically normal such that*

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V)$$

   with

$$\begin{aligned} V &= \left\langle BE^{\theta_0}(\nabla_\theta h), BE^{\theta_0}(\nabla_\theta h)\right\rangle^{-1} \\ &\quad \times \left\langle BE^{\theta_0}(\nabla_\theta h), (BKB^*) BE^{\theta_0}(\nabla_\theta h)\right\rangle \\ &\quad \times \left\langle BE^{\theta_0}(\nabla_\theta h), BE^{\theta_0}(\nabla_\theta h)\right\rangle^{-1}. \end{aligned}$$

2. *Among all admissible weighting operators B, there is one yielding an estimator with minimal variance. It is equal to $K^{-1/2}$, where K is the covariance operator defined in (3.2).*

As discussed in Carrasco and Florens (2000a), the operator $K^{-1/2}$ does not exist on the whole space $\mathbb{L}^2(\pi)$ but only on a subset, denoted $\mathcal{H}(K)$, which corresponds to the so-called reproducing kernel Hilbert space (RKHS) associated with K (see Parzen, 1970, for details).

[14] Definition A.1 describes a Hilbert-space valued random element.

[15] For a definition and the properties of Hilbert-Schmidt operators, see Dautray and Lions (1988) or Dunford and Schwartz (1988). As K is a Hilbert-Schmidt operator, it can be approached by a sequence of bounded operators denoted K_T. This property will become important when we discuss how to estimate K.

The inner product defined on $\mathcal{H}(K)$ is denoted $\langle f,g\rangle_K$.[16] Since the inverse of K is not bounded, the regularized version of the inverse, involving a penalizing term α_T, is considered (see the discussion of Equation (2.17)).

In order to implement the C-GMM estimator with the optimal weighting operator, we have to estimate K, which can be done via bounded operator K_T approaching K as the sample size grows because K is a Hilbert-Schmidt operator (see Lemma 3.1). We postpone the explicit construction of K_T, which is very similar to the GMM procedure, until the next subsection and establish the asymptotic properties of the optimal C-GMM operator, given K_T, first.

Proposition 3.2 *Assume the regularity conditions A.1 to A.5 hold. Let K_T denote a consistent estimator of K that satisfies $\|K_T - K\| = O_p(T^{-a})$ for some $a \geq 0$ and $(K_T^{\alpha_T})^{-1} = (K_T^2 + \alpha_T I)^{-1} K_T$ denote the regularized estimator of K^{-1}.*[17] *The optimal GMM estimator of θ is obtained by:*

$$\hat{\theta}_T = \underset{\theta}{argmin} \left\| (K_T^{\alpha_T})^{-1/2} \hat{h}_T(\theta) \right\| \tag{3.3}$$

and satisfies

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(\left\langle E^{\theta_0}(\nabla_\theta h), E^{\theta_0}(\nabla_\theta h)\right\rangle_K\right)^{-1}\right) \tag{3.4}$$

as T and $T^a \alpha_T^{5/4}$ go to infinity and α_T goes to zero. [18]

A simple estimator of the asymptotic variance of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ will be discussed in Subsection 3.3. Proposition 3.2 gives a rate of convergence of α_T but does not indicate how to choose α_T in practice. Recall that the estimator will be consistent for any $\alpha_T > 0$ but its variance will be the smallest for the α_T decreasing to zero at the right rate. In the simulations, we choose an arbitrary α_T relatively small. Of course a data-driven selection method of α_T would be preferable. Carrasco and Florens (2000b) propose a cross-validation method to select α_T by minimizing the mean square error of $\hat{\theta}_T$. This method is developed in an iid

[16] The properties of the RKHS norm associated with this inner product are discussed in Appendix B.

[17] See the discussion of equations (2.16)-(2.18).

[18] Let $\theta = (\theta_1, ..., \theta_q)'$. By a slight abuse of notation, $\left\langle E^{\theta_0}(\nabla_\theta h), E^{\theta_0}(\nabla_\theta h)\right\rangle_K$ in (3.4) denotes the $q \times q-$matrix with (i,j) element $\left\langle E^{\theta_0}(\nabla_{\theta_i} h), E^{\theta_0}\left(\nabla_{\theta_j} h\right)\right\rangle_K$.

context and its adaptation to time-series seems to be beyond the scope of the current paper. Simulations in Carrasco and Florens (2000b) show that the estimator is not very sensitive to the choice of α_T.

3.2 Convergence rate of the estimator of the covariance operator

Note that the covariance operator defined in (3.2) is an integral operator that can be written as

$$Kf(\tau_1) = \int k(\tau_1, \tau_2) f(\tau_2) \pi(\tau_2) d\tau_2 \tag{3.5}$$

with

$$k(\tau_1, \tau_2) = \sum_{j=-\infty}^{\infty} E^{\theta_0} \left(h_t(\tau_1; \theta_0) \overline{h_{t-j}(\tau_2; \theta_0)} \right) \tag{3.6}$$

The function k is called the kernel of the integral operator K. To estimate K, we use a kernel estimator of the type studied by Andrews (1991). Given the first step estimator $\hat{\theta}_T^1$ from (2.14), we estimate the kernel of the covariance operator at the second step via:

$$\hat{k}_T(\tau_1, \tau_2) = \frac{T}{T-q} \sum_{j=-T+1}^{T-1} \omega\left(\frac{j}{S_T}\right) \hat{\Gamma}_T(j) \tag{3.7}$$

with

$$\hat{\Gamma}_T(j) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} h_t\left(\tau_1; \hat{\theta}_T^1\right) \overline{h_{t-j}\left(\tau_2; \hat{\theta}_T^1\right)}, & j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} h_{t+j}\left(\tau_1; \hat{\theta}_T^1\right) \overline{h_t\left(\tau_2; \hat{\theta}_T^1\right)}, & j < 0 \end{cases} \tag{3.8}$$

where ω is a kernel and S_T is a bandwidth that will be allowed to diverge at a certain rate. The kernel ω is required to satisfy the regularity conditions A.6, which are based on Assumptions B and C of Andrews (1991).

Denote $f(\lambda)$ the spectral density of Y_t at frequency λ and $f^{(\nu)}$ its νth derivative at $\lambda = 0$. Denote $\omega_\nu = (1/\nu!) \left(d^\nu \omega(x)/dx^\nu\right)\big|_{x=0}$.

Proposition 3.3 *Assume that the regularity conditions A.1 to A.6 hold and that* $S_T^{2\nu+1}/T \to \gamma \in (0, +\infty)$ *for some* $\nu \in (0, +\infty)$ *for which* $\omega_\nu, \left\| f^{(\nu)} \right\| < \infty$. *Then*

$$\|K_T - K\| = O_p\left(T^{-\nu/(2\nu+1)}\right).$$

For the Bartlett kernel, $\nu = 1$ and for the Parzen, Tuckey-Hanning and QS kernels, $\nu = 2$. To obtain the result of Proposition 3.3, we have selected the value of S_T that delivers the fastest rate for K_T. For this S_T, we then select α_T such that $T^a \alpha_T^{5/4}$ goes to infinity according to Proposition 3.2. Instead, we could have chosen S_T and α_T simultaneously. However, from Proposition 3.2, it seems that the faster the rate for K_T, the faster the rate for α_T. So this approach seems to guarantee the fastest rate for α_T.

Note that if $\{h_t\}$ are uncorrelated (as in Section 2.3), then the kernel of K simplifies to $k(\tau_1, \tau_2) = E^{\theta_0}\left(h_t(\tau_1; \theta_0)\overline{h_t(\tau_2; \theta_0)}\right)$ and can be estimated by the sample average. The resulting estimator will satisfy $\|K_T - K\| = O_p\left(T^{-1/2}\right)$, hence $a = 1/2$ in that case. When $\{h_t\}$ are correlated, the convergence rate of K_T is slower and accordingly the rate of convergence of α_T to zero is slower.

3.3 Simplified Computation of the C-GMM Estimator

Carrasco and Florens (2000a) propose to write the objective function in terms of the eigenvalues and eigenfunctions of the operator $K_T^{\alpha_T}$. The computation of eigenvalues and eigenfunctions can be burdensome, particularly in large samples. We review briefly this method before turning to a more attractive approach that consists in rewriting the objective function in terms of matrices and vectors.

Note that $\hat{k}_T$ is a degenerate kernel that can be rewritten as

$$\hat{k}_T(\tau_1, \tau_2) = \frac{1}{T-q}\sum_{t=1}^{T} h_t\left(\tau_1; \hat{\theta}_T^1\right) U h_t\left(\tau_2; \hat{\theta}_T^1\right)$$

where

$$U h_t\left(\tau; \hat{\theta}_T^1\right) = \omega(0)\overline{h_t\left(\tau; \hat{\theta}_T^1\right)} + \sum_{j=1}^{T}\omega\left(\frac{j}{S_T}\right)\left(\overline{h_{t-j}\left(\tau; \hat{\theta}_T^1\right)} + \overline{h_{t+j}\left(\tau; \hat{\theta}_T^1\right)}\right)$$

using the convention that $h_t\left(\tau; \hat{\theta}_T^1\right) = 0$ if $t \leq 0$ or $t > T$. Hence K_T has at most T eigenvalues (denoted $\hat{\lambda}_j$) different from zero. They can be calculated by solving the equation $K_T \hat{\phi}_j = \hat{\lambda}_j \hat{\phi}_j$. It turns out that $\hat{\lambda}_1, ..., \hat{\lambda}_T$ are the eigenvalues of the $T \times T-$matrix C defined in Proposition 3.4 below. The eigenfunctions of K_T take the form $\hat{\phi}_j(\tau) = \sum_{t=1}^{T} \beta_{jt} h_t\left(\tau; \hat{\theta}_T^1\right)$ where $\underline{\beta}_j = (\beta_{j1}, ..., \beta_{jT})'$, $j = 1, ..., T$ are the T eigenvectors of C. The $\hat{\phi}_j$ are orthogonal but need to be normed, let $\hat{\phi}_j$ denote the orthonormalized eigenfunctions of K_T. The objective function in equation (3.3) becomes

$$\sum_{j=1}^{T} \frac{\hat{\lambda}_j}{\hat{\lambda}_j^2 + \alpha_T} \left| \left\langle \hat{h}_T(\theta), \hat{\phi}_j \right\rangle \right|^2. \tag{3.9}$$

The objective function (3.9) can be rewritten as a quadratic form involving only matrices and vectors.

Proposition 3.4 *Solving (3.3) is equivalent to solving*

$$\min_\theta \underline{w}'(\theta) \left[I_T - C \left[\alpha_T I_T + C^2 \right]^{-1} C \right] \underline{v}(\theta) \tag{3.10}$$

where C is a $T \times T-$matrix with (t,l) element $c_{tl}/(T-q)$, $t,l = 1,...,T$, I_T is the $T \times T$ identity matrix, $\underline{v} = [v_1, ..., v_T]'$ and $\underline{w} = [w_1, ..., w_T]'$ with

$$\begin{aligned} v_t(\theta) &= \int U h_t\left(\tau; \hat{\theta}_T^1\right) \hat{h}_T(\tau;\theta) \pi(\tau) d\tau, \\ w_t(\theta) &= \left\langle h_t\left(\tau; \hat{\theta}_T^1\right), \hat{h}_T(\tau;\theta) \right\rangle, \\ c_{tl} &= \int U h_t\left(\tau; \hat{\theta}_T^1\right) h_l\left(\tau; \hat{\theta}_T^1\right) \pi(\tau) d\tau. \end{aligned}$$

Note that in the case where the $\{h_t\}$ are uncorrelated, the former formulas simplify: $Uh_t = \bar{h}_t$, $v_t = \overline{w}_t$, $c_{tl} = \left\langle h_l\left(\tau; \hat{\theta}_T^1\right), h_t\left(\tau; \hat{\theta}_T^1\right)\right\rangle$. Hence we obtain the Proposition 2.1.

Similarly, an estimator of the asymptotic variance of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ given in (3.4) can be computed in a simple way.

Proposition 3.5 *Suppose that the assumptions of Proposition 3.3 hold and T, $T^{\nu/(2\nu+1)}\alpha_T^{3/4}$ go to infinity and α_T goes to zero. Then a consistent estimator of the $q \times q-$matrix $\left\langle E^{\theta_0}(\nabla_\theta h), E^{\theta_0}(\nabla_\theta h)\right\rangle_K$ is given by*

$$\begin{aligned} &\left\langle \nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right), \left(K_T^{\alpha_T}\right)^{-1} \nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right)\right\rangle \\ = &\frac{1}{\alpha_T(T-q)} \underline{w}'(\theta) \left[I_T - C \left[\alpha_T I_T + C^2 \right]^{-1} C \right] \underline{v}(\theta) \end{aligned}$$

where C is the $T \times T-$matrix defined in Proposition 3.4, I_T is the $T \times T$ identity matrix, $\underline{v} = [v_1, ..., v_T]'$ and $\underline{w} = [w_1, ..., w_T]'$ are $T \times q-$matrices with (t,j) element

$$\begin{aligned} (v_t(\theta))_j &= \int U h_t\left(\tau; \hat{\theta}_T^1\right) \nabla_{\theta_j} \hat{h}_T\left(\hat{\theta}_T\right) \pi(\tau) d\tau, \\ (w_t(\theta))_j &= \left\langle h_t\left(\tau; \hat{\theta}_T^1\right), \nabla_{\theta_j} \hat{h}_T\left(\hat{\theta}_T\right)\right\rangle. \end{aligned}$$

3.4 Efficiency

In Proposition 3.2, we saw that the asymptotic variance of $\hat{\theta}_T$ is $\left(\left\langle E^{\theta_0}(\nabla_\theta h), E^{\theta_0}(\nabla_\theta h)\right\rangle_K\right)^{-1}$. Using results on RKHS (see Appendix B), it is possible to compute this term and hence to establish conditions under which this variance coincides with the Cramer Rao efficiency bound. We consider arbitrary functions $h(\tau, Y_t; \theta_0)$ that satisfy the identification Assumption A.3 and where, as usual, Y_t is the $(L+1)-$ vector of r.v.: $Y_t = (X_t, X_{t+1}, \ldots, X_{t+L})'$. Let $L^2(Y_t)$ be the set of random variables of the form $g(Y_t)$ with $E^{\theta_0}\left[|g(Y_t)|^2\right] < \infty$. It is assumed that $h(\tau, Y_t; \theta_0)$ belongs to $L^2(Y_t)$. Let S be the set of all random variables that may be written as $\sum_{j=1}^{n} c_j h(\tau_j, Y_t; \theta_0)$ for arbitrary integer n, real constants $c_1, c_2, ..., c_n$ and points $\tau_1, ..., \tau_n$ of I. Denote $\overline{S}$ its closure, $\overline{S}$ contains all the elements of S and their limits in $L^2(Y_t)-$norm.

Proposition 3.6 *Assume that (i) X_t is stationary, $\alpha-$ mixing, and Markov of order L, (ii) $\int \sup_{\theta\in\Theta} |h(\tau, y_0; \theta) f_\theta(x_0, x_1, ..., x_L)| dx_0 dx_1 ... dx_L < \infty$. (iii) the conditional pdf of x_t given $x_{t-1}, .., x_{t-L}$ is differentiable w.r. to θ. (iv) the optimal GMM estimator, $\hat{\theta}_T$, based on $h(\tau, Y_t; \theta)$, $\tau \in \mathbb{R}^d$ is tractable and its asymptotic distribution is as given in (3.4).*

Then, $\hat{\theta}_T$ is asymptotically as efficient as MLE if and only if

$$\nabla_\theta \ln f_\theta(x_{t+L}|x_{t+L-1}, .., x_t; \theta)|_{\theta=\theta_0} \in \overline{S}.$$

A proof of this proposition is given in Carrasco and Florens (2002). It states that the GMM estimator is efficient if and only if the score belongs to the span of the moment conditions. This result is close to that of Gallant and Long (1997) who show that if the auxiliary model is rich enough to encompass the DGP, then the efficient method of moments estimator is asymptotically efficient. Note that the condition (ii) is trivially satisfied when $\{h_t\}$ is bounded. It is important to remark that π does not affect the efficiency as long as $\pi > 0$ on $\mathbb{R}^d$. In small samples however, the choice of π might play a role.

4 GMM estimators based on the characteristic function

This section studies the properties of moment conditions (3.1) based on the conditional or joint characteristic. The first subsection will focus on Markov processes while the second subsection will discuss mainly the nonmarkovian case.

4.1 Using the conditional characteristic function

In this subsection, we consider moment conditions based on the conditional CF (CCF) and discuss the choice of the instruments so that the estimator based on these moment conditions achieves ML efficiency.

Suppose an econometrician observes realizations of a Markov process $X \in \mathbb{R}^p$. The conditional characteristic function of X_{t+1}

$$\psi_\theta\left(s|X_t;\theta\right) = E^\theta\left[e^{isX_{t+1}}|X_t\right]$$

is assumed to be known. We denote $\psi_\theta\left(s|X_t;\theta_0\right)$ by $\psi_\theta\left(s|X_t\right)$. Let $Y_t = \left(X_t, X_{t+1}\right)'$.

The CCF permits to construct unconditional moment conditions for an arbitrary choice of instrument, m, which is a function of X_t. Besides being a function of X_t, m may be a function of an index r either equal to or different from s. The following two types of unconditional moment functions are of particular interest:

SI – the **S**ingle **I**ndex moment functions: $h\left(s, Y_t;\theta\right) = m(s, X_t)\left(e^{isX_{t+1}} - \psi_\theta\left(s|X_t\right)\right)$ where $s \in \mathbb{R}^p$ and $m\left(s, X_t\right) = \overline{m\left(-s, X_t\right)}$

DI – the **D**ouble **I**ndex moment functions: $h\left(\tau, Y_t;\theta\right) = m(r, X_t)\left(e^{isX_{t+1}} - \psi_\theta\left(s|X_t\right)\right)$ where $\tau = (r,s)' \in \mathbb{R}^{2p}$ and $m\left(r, X_t\right) = \overline{m\left(-r, X_t\right)}$

Note that in either case, the sequence of moment functions $\left\{h\left(., Y_t;\theta\right)\right\}$ is a martingale difference sequence with respect to the filtration $I_t = \left\{X_t, X_{t-1}, ..., X_1\right\}$, hence it is uncorrelated, which simplifies the estimation of the covariance operator K.

We now discuss which choice of instruments m used in the unconditional moment functions is optimal, i.e. yields an efficient CF-GMM estimator, where "efficient" means as efficient as the MLE. One set of optimal instruments is known since Feuerverger and McDunnough (1981) in the i.i.d. setting, and was extended by Singleton (2001) for affine diffusions. Singleton (2001) shows that the optimal **SI** instrument is

$$m(s, X_t) = \frac{1}{(2\pi)^p} \int e^{-isx} \nabla_\theta \ln f_\theta \left(x|X_t; \theta_0\right) dx. \tag{4.1}$$

He proves, that under standard regularity assumptions, the solution of

$$\int m(s, X_t) \left(e^{isX_{t+1}} - \psi_\theta \left(s|X_t\right)\right) ds = 0$$

for this choice of instrument is asymptotically efficient. As noted in Section 1, the drawback of this instrument is that it depends on the unknown probability density function.

When $r = s$ is not imposed, there is a choice of instrument that does not depend on the unknown p.d.f., while attaining the ML-efficiency. The optimal **DI** instrument is

$$m(r, X_t) = e^{irX_t}. \tag{4.2}$$

For this choice of instrument, the GMM estimator is asymptotically efficient.

Proposition 4.1 *Consider*

$$h\left(\tau, Y_t; \theta\right) = e^{irX_t} \left(e^{isX_{t+1}} - \psi_\theta \left(s|X_t\right)\right), \tag{4.3}$$

with $\tau = (r, s)' \in \mathbb{R}^{2p}$ *and denote* K *the covariance operator of* $\{h\left(., Y_t; \theta\right)\}$. *Assume that* $\mathcal{N}\left(K\right) = \{f \in \mathbb{L}^2\left(\pi\right) : Kf = 0\} = \{0\}$ *and Assumptions A.2, A.3, A.7, and A.8 hold. Then the optimal GMM estimator based on (4.3) satisfies*

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, I_{\theta_0}^{-1}\right)$$

as T, $T\alpha_T^{5/2}$ *go to infinity and* α_T *goes to zero.* I_{θ_0} *denotes the Information matrix.*

The condition $\mathcal{N}\left(K\right) = \{0\}$ is equivalent to require that the covariance matrix be non singular in the usual GMM. The efficiency resulting from moment functions (4.3) can be proved from Proposition 3.6. Indeed $\bar{S}$ the closure of the span of $\{h_t\}$ includes all functions in $\mathbb{L}^2\left(Y_t\right)$ hence it also includes the score function. Alternatively, one can prove this result directly by computing the asymptotic variance of the GMM estimator and comparing it with the information matrix, see Equation (D.14) in Appendix.

The intuition for the efficiency result is as follows. For the GMM estimator to be as efficient as the MLE, the moment conditions need to be sufficiently rich to permit to recover the score. By the Fourier inversion formula, the **SI** moment functions correspond to the score, indeed we have

$$\frac{1}{(2\pi)^p}\int e^{isX_{t+1}}\left[\int e^{-isx}\nabla_\theta \ln f_\theta\left(x|X_t;\theta_0\right)dx\right]ds = \nabla_\theta \ln f_\theta\left(X_{t+1}|X_t;\theta_0\right).$$

The advantage of **SI** moment functions is that they depend on a single index. This is done at a cost: the optimal instrument is a function of the unknown score. If one does not want to rely on instruments of type (4.1), one needs to use a double index. The **DI** moment functions with instruments defined in (4.2) span all functions in $L^2\left(Y_t\right)$ and the unknown score in particular.

Singleton (2001) addresses the problem of the unknown score by replacing the integral in (4.1) by a sum over a finite grid and computing the respective m as an optimal instrument in the Hansen (1985) framework. This estimator approaches ML efficiency as the grid becomes finer and finer. However, for too fine a grid, the covariance matrix of the resulting moment functions becomes singular. Hence, one has to know the optimal rate of convergence of the discretization interval to be able to implement such an estimator. The second caveat is that optimal instruments depend on the selected grid, i.e. as one refines the grid, new instruments have to be selected. Therefore, it is not clear how it is going to impact the estimator in practice. In our approach the counterpart of the discretization grid size is the penalization term α_T whose convergence rate was discussed in the previous section. Moreover, we choose our instrument (4.2) prior to choosing the discretization grid to compute the integrals. Therefore, we are facing a pure numerical error, which can be controlled via standard numerical techniques.

Finally, we notice that since the moment functions are uncorrelated and the optimal instrument is known to have an exponential form, the computation of the terms C and v in the objective function (3.10) is simplified. Appendix C outlines these computations. Note that all elements involving the index r can be computed analytically. Therefore, using the

DI instrument does not introduce computational complications.

4.2 Using the joint characteristic function

Many important models in finance involve latent factors, the most prominent example being the stochastic volatility (SV) model. In this case, the full system can be described by a Markov vector $(X_t, \mathcal{X}_t)'$ consisting of observable and latent components. As a result, X_t is most likely not Markov.[19]

For non-Markovian processes, the conditional characteristic function is usually unknown and difficult to estimate. On the other hand, the joint characteristic function (JCF), if not known, can be computed by simulations.[20] [21] Denote the JCF as:

$$\psi_\theta^L(\tau) = E^\theta\left(e^{i\tau' Y_t}\right) \tag{4.4}$$

where $\tau = (\tau_0, \tau_1, ..., \tau_L)'$, and $Y_t = (X_t, X_{t+1}, .., X_{t+L})'$.

Feuerverger (1990) has considered this problem. His estimator is the solution to

$$\int \left(\psi_\theta^L(\tau) - \psi_T^L(\tau)\right) m(\tau)\, d\tau = 0. \tag{4.5}$$

where $\psi_T^L(\tau)$ denotes the empirical JCF. For a special weighting function m, which is very similar to (4.1), Feuerverger shows that the estimator is as efficient as the estimator which solves

$$\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta \ln f_\theta\left(X_{t+L}|X_{t+L-1}, ..., X_t; \theta\right) = 0 \tag{4.6}$$

where $f_\theta(X_{t+L}|X_{t+L-1}, ..., X_t)$ is the true distribution of X_{t+L} conditional on $X_t, ..., X_{t+L-1}$. This result holds even if the process X_t is not Markovian of order L (or less).

If X_t is Markovian of order L then the variance of the resulting estimator is $I_\theta^{-1}(L)$ with

$$I_\theta(L) = E_\theta\left(\nabla_\theta \ln f_\theta\left(X_{t+L}|X_{t+L-1}, ..., X_t; \theta\right)^2\right) \tag{4.7}$$

[19] Florens, Mouchard, and Rolin (1993) give necessary and sufficient conditions for the marginal of a jointly Markov process to be itself Markov.

[20] Jiang and Knight (2002) discuss examples of diffusion models for which JCF is available in analytical form. Yu (2001) derives JCF of the Merton model generalization to self-exciting jump component.

[21] Simulations are discussed in Section 5.

which is the Cramér-Rao efficiency bound. If X_t is not Markovian of order L then the variance of the estimator has the usual sandwich form because $\nabla_\theta \ln f_\theta (X_{t+L}|X_{t+L-1}, ..., X_t; \theta_0)$ is not a martingale difference sequence with respect to $\{X_{t+L}, ..., X_1\}$. This variance differs from $I_\theta^{-1}(L)$ and is greater than the Cramér-Rao efficiency bound. Note that (4.6) should not be confused with quasi-maximum likelihood estimation because $f_\theta (X_{t+L}|X_{t+L-1}, ..., X_t; \theta)$, is the exact distribution conditional on a restricted information set.

Feuerverger (1990) notes that the estimator based on the JCF can be made arbitrarily efficient provided that "L (fixed) is sufficiently large" although no proof is provided. This argument is clearly valid when the process is Markovian of order L. However, in the non-Markovian case, the only feasible way to achieve the efficiency would be to let L go to infinity with the sample size at a certain (slow) rate, the question of the optimal speed of convergence has to the best of our knowledge not been addressed in the literature. The implementation of such approach might be problematic since for L too large, the lack of data to estimate consistently the characteristic function might result in an $\hat{\theta}_T$ with undesirable properties.

The approach of Feuerverger based on the joint characteristic function of basically the full vector $(X_1, X_2, ..., X_T)$ is not realistic because only one observation of this vector is available. Instead, we can avoid using the unknown instrument m in (4.5) by considering a moment condition based on the JCF of Y_t :

$$h(\tau, Y_t, \theta) = e^{i\tau Y_t} - \psi_\theta^L(\tau). \tag{4.8}$$

for some small $L = 0, 1, 2, \ldots$[22] Assume that the JCF is sufficient to identify the parameters. Now the moments $h(\tau, Y_t, \theta_0)$ are not a martingale difference sequence (even if X_t is Markovian) and the kernel of K is given by

$$k(\tau_1, \tau_2) = \sum_{j=-\infty}^{\infty} E^{\theta_0}\left[h(\tau_1, Y_t, \theta_0)\overline{h(\tau_2, Y_{t-j}, \theta_0)}\right].$$

When X_t is Markov of order L, the optimal GMM estimator is efficient as stated below.

Proposition 4.2 *Assume that X_t is Markov of order L and that the assumptions of Proposition 3.3 hold and T, $T^{\nu/(2\nu+1)}\alpha_T^{5/4}$ go to infinity and α_T goes to zero. Then the optimal GMM estimator using the moments (4.8) is as efficient as the MLE.*

[22] Jiang and Knight (2002), in a particular case of an affine stochastic volatility model, arbitrary base the instrument m on the normal density and experiment with values of L from 1 to 5.

As the closure of the span of $\{h_t\}$ contains the score $\nabla_\theta \ln f_\theta \left(X_{t+L}|X_{t+L-1}, ..., X_t; \theta_0\right)$, the efficiency follows from Proposition 3.6.

Note that if X_t is Markov, it makes more sense to use moment conditions based on the CCF because the resulting estimator, while being efficient, is easier to implement (as $\{h_t\}$ are m.d.s.). If X_t is not Markov, the JCF-GMM estimator will not be efficient. However, it might still have some good properties if the temporal dependence dies out quickly. As the computation of the optimal K_T may be burdensome (it involves two smoothing parameters S_T and α_T), one may decide to use a suboptimal weighting operator obtained by inverting the covariance operator without the autocorrelations.

One interesting question is then: What is the resulting loss of efficiency? We can answer this question only partially because we are not able to compute the variance of the optimal JCF-GMM estimator when X_t is not Markov. However, we have a full characterization of the variance of the suboptimal JCF-GMM estimator.

Assume that one ignores the autocorrelations and uses as weighting operator the inverse of the operator $\widetilde{K}$ associated with the kernel:

$$\widetilde{k}\left(\tau_1, \tau_2\right) = E^{\theta_0}\left[h\left(\tau_1, Y_t, \theta_0\right) \overline{h\left(\tau_2, Y_t, \theta_0\right)}\right]. \tag{4.9}$$

Proposition 4.3 *Assume that the assumptions of Proposition 4.2 hold. The asymptotic variance of the suboptimal JCF-GMM estimator $\hat{\theta}_T$ using (4.8) and (4.9) is the same as that of the estimator $\widetilde{\theta}_T$ which is the solution of*

$$\frac{1}{T}\sum_t \nabla_\theta \ln f_\theta \left(Y_t; \theta\right) = 0 \tag{4.10}$$

where $\ln f_\theta \left(Y_t; \theta\right)$ is the exact joint distribution of Y_t.

Since using the efficient weighting matrix should result in a gain of efficiency, the asymptotic variance of $\widetilde{\theta}_T$ (given in Appendix D) can be considered as an upper bound for the variance of the estimator obtained by using the optimal weighting operator that is K^{-1}. To illustrate the results of Proposition 4.3, consider first the case where $\{X_t\}$ is i.i.d. and $L = 1$. Then solving (4.10) is basically (for T large) equivalent to solving

$$2\frac{1}{T}\sum_t \nabla_\theta \ln f_\theta \left(X_t; \theta\right) = 0$$

so that the resulting estimator $\hat{\theta}_T$ is efficient. Now turn to the case where $\{X_t\}$ is Markov of order 1 and again $L = 1$ then (4.10) is equivalent to

$$\frac{1}{T}\sum_t \nabla_\theta \ln f_\theta \left(X_t|X_{t-1};\theta\right) + \frac{1}{T}\sum_t \nabla_\theta \ln f_\theta \left(X_{t-1};\theta\right) = 0$$

which will not deliver an efficient estimator in general.

5 Simulated Method of Moments

This section introduces the Simulated Method of Moments extensions of CF-based estimators. Such estimators are of interest for applications involving processes that do not have analytical expressions for the CF.

In this section we assume that X_t is a Markov process satisfying

$$X_{t+1} = H\left(X_t, \varepsilon_t, \theta\right) \tag{5.1}$$

where ε_t is an i.i.d. sequence independent of X_t with known distribution.

For instance, X_t may be the solution of a dynamic asset pricing model as that presented by Duffie and Singleton (1993). If X_t is a discretely sampled diffusion process then H in (5.1) can be obtained from an Euler discretization.[23] Moments based on the unknown conditional or joint characteristic function are used to estimate θ. Simulation methods are required in this case. Consider two cases of interest:

- Assume X_t is fully observable. Then (5.1) permits to draw data from the conditional distribution of X_{t+1} given X_t and to estimate the CCF. This simulation scheme will be called conditional simulation.

[23] However, there is a pitfall with this approach. When the number of discretization intervals per unit of time, N, is fixed, none of the J simulated paths, $\tilde{X}_t^j$, is distributed as X_t and the estimator $\hat{\theta}_T$ is biased. Broze, Scaillet and Zakoian (1998) document the discretization bias of the Indirect Inference estimator and show that it vanishes when $N \to \infty$ and J is fixed. In a recent paper, Detemple, Garcia, and Rindisbacher (2002) study estimators of the conditional expectation of diffusions. They show that if J is allowed to diverge too fast relative to N, then the bias of their estimator blows up. The same is likely to be true here. However as there is no limitation on how fine we can discretize (besides the computer precision), we assume that N is chosen sufficiently large for the discretization bias to vanish.

- Assume $X_t = (Z_t, Z_t^*)$ where Z_t^* is a latent variable, e.g. the volatility in a stochastic volatility model. Z_t only is observable. In such cases, it is usually unknown how to draw from the conditional distribution of Z_t. Moreover using the CCF of Z_t will not deliver an efficient estimator because Z_t is not Markovian. On the other hand, it is easy to simulate a path of Z_t and to construct an estimator of the joint characteristic function.

The main difference in the properties of the two estimators is that in the first case, the estimator is as efficient as the MLE when the number of simulated paths, J, goes to infinity while in the second case the estimator will, in general, never reach the efficiency bound even if J goes to infinity. A subsection will be devoted to each case.

5.1 Conditional simulation

Assume X_t is observable. For a given θ and conditionally on X_t, we generate a sequence $\left\{\tilde{X}_{t+1|t}^{\theta,j}, j = 1, 2, ..., J\right\}$ from

$$\tilde{X}_{t+1|t}^{\theta,j} = H\left(X_t, \tilde{\varepsilon}_{j,t+1}, \theta\right)$$

where $\left\{\tilde{\varepsilon}_{j,t}\right\}_{j,t}$ are identically and independently distributed as $\{\varepsilon_t\}$. Note that $\left\{\tilde{X}_{t+1|t}^{\theta,j}\right\}_j$ are i.i.d. conditionally on X_t and distributed as $X_{t+1}|X_t$ when $\theta = \theta_0$. The moment conditions become

$$\tilde{h}_T^J(\tau;\theta) = \frac{1}{T}\sum_{t=1}^{T} e^{irX_t}\left(e^{isX_{t+1}} - \frac{1}{J}\sum_{j=1}^{J} e^{is\tilde{X}_{t+1|t}^{\theta,j}}\right)$$

where $\tau = (r, s)$. To facilitate the discussion, we introduce the following notations:

$$\begin{aligned}
Y_t &= (X_t, X_{t+1})' \\
h_t(\tau;\theta) &= e^{irX_t}[e^{isX_{t+1}} - \psi_\theta(s|X_t)] \\
\tilde{h}\left(Y_t, \tilde{X}_{t+1|t}^{\theta,j}, \tau\right) &= e^{irX_t}[e^{isX_{t+1}} - e^{is\tilde{X}_{t+1|t}^{\theta,j}}] \\
\tilde{h}_t^J(\tau;\theta) &= \frac{1}{J}\sum_{j=1}^{J}\tilde{h}\left(Y_t, \tilde{X}_{t+1|t}^{\theta,j}, \tau\right) = e^{irX_t}\left(e^{isX_{t+1}} - \frac{1}{J}\sum_{j=1}^{J} e^{is\tilde{X}_{t+1|t}^{\theta,j}}\right)
\end{aligned}$$

The resulting $\left\{\tilde{h}_t^J(\tau;\theta_0)\right\}$ are martingale difference sequences with respect to $\{X_t, X_{t-1}, ..., X_1\}$ and therefore are uncorrelated. Moreover,

$$E^{\theta}\left[\tilde{h}\left(Y_t, \tilde{X}_{t+1|t}^{\theta,j}, \tau\right)|Y_t\right] = h_t(\tau;\theta).$$

Let K be the covariance operator associated with the kernel

$$k(\tau_1, \tau_2) = E^{\theta_0}\left[h_t(\tau_1;\theta_0)\overline{h_t(\tau_2;\theta_0)}\right] \tag{5.2}$$

and let U be the operator with kernel

$$u(\tau_1, \tau_2) = E^{\theta_0}\left[\left(\tilde{h}_t^J - h_t\right)(\tau_1;\theta_0)\overline{\left(\tilde{h}_t^J - h_t\right)}(\tau_2;\theta_0)\right].$$

Denote $\tilde{K}$ the covariance operator of $\left\{\tilde{h}_t^J\right\}$. Let $\tilde{K}_T^{\alpha_T}$ be the regularized estimator of $\tilde{K}$. The GMM estimator associated with the moments $\tilde{h}^J$ is defined as

$$\tilde{\theta}_T = \underset{\theta}{\text{argmin}} \left\|\tilde{h}_T^J(.,\theta)\right\|_{\tilde{K}_T^{\alpha_T}}^2.$$

Now we can state the efficiency result:

Proposition 5.1 *Suppose that Assumptions A.2, A.3, A.7, A.8(i), A.9, and A.10 hold for* $\tilde{h}_t^J$ *and a fixed* J. *We have*

$$\sqrt{T}\left(\tilde{\theta}_T - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(\left\langle E^{\theta_0}(\bigtriangledown_\theta h), E^{\theta_0}(\bigtriangledown_\theta h)\right\rangle_{\tilde{K}}\right)^{-1}\right)$$

as T *and* $T\alpha_T^{5/2}$ *go to infinity and* α_T *goes to zero. Moreover,* $\tilde{K} = K + \frac{1}{J}U$ *and we have the inequality*

$$\left(\left\langle E^{\theta_0}(\bigtriangledown_\theta h), E^{\theta_0}(\bigtriangledown_\theta h)\right\rangle_{\left(K+\frac{1}{J}U\right)}\right) \leq \left(\left\langle E^{\theta_0}(\bigtriangledown_\theta h), E^{\theta_0}(\bigtriangledown_\theta h)\right\rangle_K\right). \tag{5.3}$$

For J large, the SMM estimator will be as efficient as the CCF-GMM estimator which itself has been shown to reach the Cramér-Rao Efficiency bound because we have

$$\left\langle E^{\theta_0}(\bigtriangledown_\theta h), E^{\theta_0}(\bigtriangledown_\theta h)\right\rangle_K = I_{\theta_0}.$$

5.2 Path simulation

Assume now that observable S_t is only a subset of a larger system $X_t = (Z_t, Z_t^*)$. The JCF of $Y_t = (Z_t, Z_{t+1}, .., Z_{t+L})'$, as defined in (4.4), is assumed to be unknown and will be estimated via simulations. First, note that because only Z_t is observable, we might not be able to identify all the parameters characterizing the distribution of the full system X_t. The identification Assumption A.3 needs to be checked on a case by case basis. Second, note also that even if the full system is Markov, Z_t is usually not Markov. Therefore there is no hope to reach the Cramér-Rao efficiency bound when L is fixed, as discussed in Section 4.2.

For a given θ, we generate a sequence $\left\{\tilde{X}_j^\theta, j = 1, 2, ..., J(T)\right\}$ from

$$\begin{aligned} \tilde{X}_{j+1}^\theta &= H\left(\tilde{X}_j^\theta, \tilde{\varepsilon}_{j+1}, \theta\right) \\ \tilde{X}_0^\theta &= \tilde{X}_0 \end{aligned} \tag{5.4}$$

where $\{\tilde{\varepsilon}_j\}$ are identically and independently distributed as $\{\varepsilon_t\}$, $\tilde{X}_0$ is some arbitrary starting value, and the number of simulations $J(T)$ goes to infinity with T. This path simulation scheme was suggested by Duffie and Singleton (1993). Contrary to the simulation scheme in the previous section, the sequence $\left\{\tilde{X}_j^\theta\right\}$ is completely independent of the observations $\{X_t\}$. Note that, as the starting value $\tilde{X}_0$ is not drawn from the stationary distribution of X_t, the sequence $\left\{\tilde{X}_j^\theta\right\}$ is not stationary. We assume that X_t and consequently $\left\{\tilde{X}_j^\theta\right\}$ are geometric ergodic, which guarantees that $\tilde{X}_j^\theta$ becomes stationary exponentially fast. Hence the initial starting value will not affect the distribution of our estimator. A possibility that is not exploited here is to draw a sequence of values from (5.4) and to use as $\tilde{X}_0$ the say nth value drawn which is basically stationary for n sufficiently large.Denote $\tilde{Z}_j^\theta$ the component of $\tilde{X}_j^\theta$ corresponding to Z_j and $\tilde{Y}_j = \left(\tilde{Z}_j^\theta, ..., \tilde{Z}_{j+L}^\theta\right)'$.

The estimation procedure is based on

$$\tilde{h}_T(\tau;\theta) = \frac{1}{T}\sum_{t=1}^{T} e^{i\tau Y_t} - \frac{1}{J(T)}\sum_{j=1}^{J(T)} e^{i\tau\tilde{Y}_j} \equiv \frac{1}{T}\sum_{t=1}^{T}\tilde{h}_t(\tau;\theta).$$

If ψ_θ^L were known, the following moment conditions would be used

$$h_T(\tau;\theta) = \frac{1}{T}\sum_{t=1}^{T} e^{i\tau Y_t} - \psi_\theta^L(\tau) \equiv \frac{1}{T}\sum_{t=1}^{T} h_t(\tau;\theta)$$

Note that $\{h_t(\tau;\theta)\}$ are not a martingale difference sequence and are autocorrelated. Therefore, K, the covariance operator associated with $\{h_T(\tau;\theta)\}$, has a more complicated expression than in the previous subsection:

$$k(\tau_1,\tau_2) = \sum_{i=-\infty}^{\infty} E^{\theta_0}\left[\left(e^{i\tau_1' Y_t} - \psi_\theta^L(\tau_1)\right)\overline{\left(e^{i\tau_2' Y_{t-i}} - \psi_\theta^L(\tau_2)\right)}\right]$$

We estimate K using the kernel estimator K_T described in 3.7 and 3.8 where $\psi_\theta^L(\tau_1)$ can be estimated using the observations Y_t. Let $K_T^{\alpha_T}$ be the regularized version of K_T. The GMM estimator associated with moments $\tilde{h}$ is defined as

$$\tilde{\theta}_T = \underset{\theta}{\text{argmin}} \left\|\tilde{h}_T(.,\theta)\right\|^2_{K_T^{\alpha_T}}.$$

Note that $\tilde{X}_{j+1}^\theta$ depends on θ through the past history of $\left\{\tilde{X}_j^\theta\right\}$. Sufficient conditions for the uniform weak law of large numbers of $\tilde{h}_T(.,\theta)$ are discussed in Duffie and Singleton (1993). Let $T/J(T)$ converge to ζ as T goes to infinity. Then, under the additional assumption of geometric ergodicity of X_t (Assumption A.11) we have the following result:

Proposition 5.2 *Suppose that Assumptions A.2 to A.6 (for $\tilde{h}_t$ replacing h_t and E^{θ_0} denotes the expectation with respect to the stationary distribution of Y_t), A.9, and A.11 hold. Let K_T be the kernel estimator of K with kernel ω and bandwidth S_T satisfying the conditions of Proposition 3.3. Then,*

$$\sqrt{T}\left(\tilde{\theta}_T - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, (1+\zeta)\left(\left\langle E^{\theta_0}(\nabla_\theta h), E^{\theta_0}(\nabla_\theta h)\right\rangle_K\right)^{-1}\right)$$

as T and $T^{\nu/(2\nu+1)}\alpha_T^{5/4}$ go to infinity and α_T goes to zero.

It should be noted that the variance of $\tilde{\theta}_T$ can be made as close as possible to that of $\hat{\theta}_T$ in Proposition 3.2 by letting $T/J(T)$ go to 0. Because of the autocorrelations, the estimation of the optimal weighting operator K is burdensome. To simplify this computation we could use the covariance operator that ignores the autocorrelations but the resulting estimator would be less efficient. Its variance is given by Proposition 4.3 for the non-simulated case. The variance of the C-SMM estimator is again equal to $(1+\zeta)$ times the variance obtained in the non-simulated context.

6 Monte-Carlo Study

The purpose of this section is to evaluate the performance of the CF-GMM estimator via Monte-Carlo analysis. We consider several term-structure models, which have conditional a CF in closed-form. First, in order to show that our method is on par with others, we compare its performance with that of MLE, QMLE, and EMM on the example of the CIR, or square-root, process. We then proceed with an example of a scalar jump-diffusion, for which MLE is not available, and EMM implementation would involve considerable numerical difficulties. Finally, we report performance of the estimator for a three-factor affine diffusions model, which was previously handled by EMM (Dai and Singleton, 2000) and QMLE (Duffee, 2002).

In all three cases, when we perform computations of the objective function, we select the standard normal density as the p.d.f. π, which defines the inner product. Numerical integration is performed over the real line truncated from -5 to +5. The value of α_T has been chosen equal to 0.02 in all the experiments and sample sizes.

6.1 A Scalar Diffusion (CIR)

The work horse of term structure models is the CIR square-root process:

$$dr_t = (\theta - \kappa r_t)\, dt + \sigma\sqrt{r_t} dW_t \tag{6.1}$$

which has the following conditional characteristic function (see e.g. Singleton, 2001):

$$\begin{aligned} \psi(\tau|r_t) &= \left(1 - \frac{i\tau}{c}\right)^{-2\theta/\sigma^2} \exp\left\{\frac{i\tau\sigma^2 e^{-\kappa}}{1 - \frac{i\tau}{c}} r_t\right\} \\ c &= \frac{2\kappa}{\sigma^2(1 - e^{-\kappa})} \end{aligned} \tag{6.2}$$

This specification is rejected by many studies for term structure pricing, but found to model the short rate quite reasonably. Therefore, we will use this model as an example. When κ, θ and σ are all strictly positive and $\sigma^2 \leq 2\theta$ then the square root process has a unique fundamental solution and its marginal density is *Gamma* and its transition density is a type I Bessel function distribution or noncentral χ^2 with a fractional order (see e.g. Cox et al., 1985). Hence the ML is well defined as a closed form solution for the transition density is available.

The simulation design is identical to Zhou (2001). This is done on purpose as it allows us to compare our results with the MLE, QMLE and EMM results reported in Zhou (2001). Two sample sizes are considered, namely $T = 500$, and $T = 1500$ with a weekly sampling frequency in mind.

We consider the two scenarios Zhou reports in detail. The first picks an empirically plausible process and considers the parameter estimates obtained from Gallant and Tauchen (1998), namely $dr_t = (0.02491 - 0.00285 r_t)dt + 0.0275\sqrt{r_t}dW_t$. This process exhibits the near-unit root behavior that matches the data and has a small unconditional variance. As a result, the conditional density of the process looks almost Gaussian.

Zhou (2001) considers other cases where one would expect moment-based estimators to do relatively poorly, among them we consider the worst scenario (i.e. "Scenario 8" in Zhou (2001)). It is a scenario with rich conditional volatility dynamics and non-Gaussian innovations with the process $dr_t = (2.491 - 0.285 r_t)dt + 1.1\sqrt{r_t}dW_t$. It is not an empirically plausible process, but a challenge for moment-based procedures.

Table 1 reports the results for 1000 iterations for the two scenarios with sample sizes $T = 500, 1500$. We report the Mean Bias, Median Bias and Root Mean Squared Error of the following estimators: MLE, QMLE, EMM , CF-GMM without instruments, i.e. $m(\tau, X_t) \equiv 1$, and finally CF-GMM using optimal **DI**-instruments (4.2). The first three estimators appeared in Zhou (2001) and we report his results only for the purpose of comparison.

Panel A of Table 1 covers the first parameter setting taken from Gallant and Tauchen. The performance of CF-GMM with and without optimal instruments for θ and κ is comparable to MLE and vastly better than QMLE and EMM. However, performance of CF-GMM is worse for σ, especially when compared to MLE and QMLE. Still, the underperformance of CF-GMM for σ is marginal compared to that of EMM. In Panel B of Table 1 we report the more challenging scenario. The CF-GMM with optimal instruments is again close to MLE, particularly in the larger sample. The same patterns as in Panel A reappear.

6.2 A Scalar Jump-Diffusion (Vasicek with exponential jumps)

Even if one considers affine jump-diffusion models, there are not many of them which possess a closed-form characteristic function: most of them require numerical solution of the ODEs. One notable exception is a model proposed by Das and Foresi (1996). The diffusion part of their model is represented by the Vasicek, or Ornstein-Uhlenbeck, process and the jump component has an exponentially distributed absolute value of jump size with a sign of the jump determined by a Bernoulli variable:

$$\begin{aligned} dr_t &= (\theta - \kappa r_t)dt + \sigma dW_t + J_t dN_t \\ |J_t| &\sim EXP(\alpha) \\ sign(J_t) &\sim BIN(\beta) \\ N_t &\sim POI(\lambda) \end{aligned} \tag{6.3}$$

Das and Foresi (1996) derive the characteristic function for this process:

$$\begin{aligned} \psi(\tau|r_t) &= \exp\left(A(\tau) + B(\tau)r_t\right) \\ A(\tau) &= \frac{i\tau\theta}{\kappa}\left(1 - e^{-\kappa}\right) - \frac{\tau^2\sigma^2}{4\kappa}\left(1 - e^{-2\kappa}\right) \\ &+ \frac{i\lambda(1-2\beta)}{\kappa}\left[\arctan\left(\tau\alpha e^{-\kappa}\right) - \arctan\left(\tau\alpha\right)\right] + \frac{\lambda}{2\kappa}\log\left(\frac{1+\tau^2\alpha^2 e^{-2\kappa}}{1+\tau^2\alpha^2}\right) \\ B(\tau) &= i\tau e^{-\kappa} \end{aligned} \tag{6.4}$$

The presence of parameter β allows for negative and positive jumps and controls skewness of the interest rates. For simplicity we set $\beta = 1$, i.e. we evaluate a simplified version of the model with positive jumps only.

In order to be consistent with the previous subsection, we take parameter values from Gallant and Tauchen (1998). They estimate a Vasicek model $dr_t = (0.02949 - 0.00283 r_t)dt + 0.09802 dW_t$. We make up parameters for the jump part: $\lambda = 0.09615$, which corresponds to five jumps a year, and $\alpha = 0.005$, which corresponds to five basis points change in interest rate.

This model does not have an analytic likelihood, so we can not compare it with this efficiency benchmark. We are not aware of any frequentist methodology, which would be able to handle this model. The exceptions are method of moments based approaches, such as GMM, EMM, and our CF-based method. Clearly, GMM with a finite set of moments conditions will be less efficient than our method. EMM produces an objective function, which is very jagged in the jump intensity parameter. As a result estimation is very difficult: one either has to approximate a jump component by a diffusive component as in Andersen, Benzoni, and Lund (2002) or profile the intensity parameter as in Chernov, Gallant, Ghysels, and Tauchen (2002). As a result, we will not be able to contrast our methodology with any benchmark.

Table 2 reports the results for 1000 iterations with sample sizes $T = 500, 1500$. As in the previous section, we report the Mean Bias, Median Bias and Root Mean Squared Error. The results are fairly straightforward to summarize. Without optimal instruments, the CF-GMM estimator has considerable bias and is also inefficient. For example, the parameter κ with a true value of 0.00283, has a mean bias of 7.3483. When the optimal instruments are used this bias is completely eliminated. The other parameters have less severe bias without optimal instruments, which is again eliminated with optimal instruments. This includes the jump intensity, a parameter typically difficult to estimate. The same observations apply to efficiency. The $RMSE$ of κ is disastrous without instruments. Optimal instruments reduce the $RMSE$ dramatically for all the parameters.

6.3 A Vector Diffusion (Three-factor affine model)

In this section we consider an affine model estimated by Duffee (2002) who uses a Gaussian density involving analytic expressions for the first and second conditional moments. In particular, we consider the three factor affine model appearing in equations (19a-d) of Duffee (2002) with empirical results reported in his Table V. For the purpose of the exercise, we simulate the state process that determines zero coupon bonds via a linear transformation, i.e., we simulate equation (using Duffee's notation):

$$\begin{pmatrix} dX_{t,1} \\ dX_{t,2} \\ dX_{t,3} \end{pmatrix} = \left[\begin{pmatrix} 0 \\ (K\theta)_2 \\ (K\theta)_3 \end{pmatrix} + \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} \right] dt + S_t dW_t$$

where S_t is a diagonal matrix with elements $S_{t(ii)} = \sqrt{\alpha_i + (\beta_{i1} \quad \beta_{i2} \quad \beta_{i3})X_t}$ for $i = 1, 2$ and 3. The parameter values corresponding to the empirical estimates of Duffee (2002) are: $K\theta_2$ = 0.222, $K\theta_3$ = -2.299, k_{11} = 0.172, k_{12} = -0.295, k_{21} = -0.197, k_{22} = 0.406, k_{31} = 0.564, k_{32} = -1.669, k_{33} = 1.721 and all elements of the diagonal matrices S_t zero, except $\alpha_3 = \beta_{11} = \beta_{22} = \beta_{32} = 1$.

The results are reported in Table 3, again considering the two sample sizes $T = 500, 1500$. We observe a picture qualitatively very similar to the one in the previous section. It is clear, despite the absence of the efficiency benchmark, that CF-GMM with optimal instrument provides a very accurate estimator even in fairly small ($T = 500$) samples.

Conclusion

This paper showed how to construct maximum likelihood efficient estimators in the settings where the maximum likelihood estimation itself is not feasible. The solution is to use GMM and to select moment functions, which are based on characteristic functions, and optimal instruments, which form a basis spanning the unknown likelihood score. The efficiency is achieved by using the whole continuum of possible moment conditions resulting from this approach. We provide practical results allowing to construct such an estimator as well as auxiliary results pertaining to the cases when data are not Markov (estimation based on the joint characteristic function) and when characteristic functions are not available in analytical form (simulated method of moments estimation). The method is especially attractive for term structure models, where typically there are more data than underlying factors, and for mixture models, such as jump-diffusions, where likelihood could be unbounded even if available in analytic form. Our Monte-Carlo study shows that the method indeed performs on par with MLE, and fares better than other methods. We also provide favorable finite-sample evidence for the cases where MLE is not feasible.

The methodology is applicable to estimation of a wide range of non-linear time series models. It has particular relevance for empirical work in finance. Asset pricing models are frequently formulated in terms of stochastic differential equations, which have no closed form solution for the conditional density based on discrete-time observations. Motivated by these avenues of application, the future work will have to refine our results on estimation of non-Markovian processes and latent states as well as develop statistical inference in the framework of characteristic function based continuum of moment conditions.

References

[1] Aït-Sahalia (1996), "Nonparametric Pricing of Interest Rate Derivative Securities", *Econometrica*, 64, 527-560.

[2] Aït-Sahalia , Y. (2002a) "Maximum Likelihood estimation of discretely sampled diffusions: A closed-form approach", *Econometrica*, 70, 223-262.

[3] Aït-Sahalia , Y. (2002b) "Closed-Form Expansions for Multivariate Diffusions", Discussion paper Princeton University.

[4] Aït-Sahalia , Y. and R. Kimmel (2002) "Maximum Likelihood Estimation of Affine Yield Models", Discussion paper Princeton University.

[5] Aït-Sahalia , Y. and P Mykland (2002) "The Effects of Random and Discrete Sampling when Estimating Continuous-Time Diffusions", *Econometrica*, forthcoming.

[6] Andrews, D. (1991) "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, **59**, No. 3, 817-858.

[7] Brandt, M. and P. Santa-Clara (2002) "Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets", *Journal of Financial Economics*, 63, 161-210.

[8] Broze, L., O. Scaillet, and J-M. Zakoian (1998) "Quasi-Indirect Inference for Diffusion Processes", *Econometric Theory*, 14, 161-186.

[9] Carr, P., H. Geman, D. Madan and M. Yor (2001) "Stochastic volatility for Lévy processes", *Mathematical Finance*, forthcoming.

[10] Carr, P. and L. Wu (2002) "Time-Changed Lévy Processes and Option Pricing", *Journal of Financial Economics*, forthcoming.

[11] Carrasco, M. and J. P. Florens (2000a) "Generalization of GMM to a continuum of moment conditions", *Econometric Theory*, 16, 797-834.

[12] Carrasco, M. and J. P. Florens (2000b) "Efficient GMM Estimation Using the Empirical Characteristic Function", working paper, CREST, Paris.

[13] Carrasco, M. and J. P. Florens (2002) "On the Asymptotic Efficiency of GMM", mimeo, University of Rochester.

[14] Carrasco, M., L. P. Hansen, and X. Chen (1999) "Time Deformation and Dependence", mimeo, University of Chicago.

[15] Chacko, G. and L. Viceira (1999) "Spectral GMM estimation of continuous-time processes", Discussion paper, Harvard Business School.

[16] Chen, X., L. P. Hansen, and M. Carrasco (1999) "Nonlinearity and Temporal Dependence", mimeo, University of Chicago.

[17] Chen, X., L. P. Hansen, and J. Scheinkman (1998) "Shape Preserving Estimation of Diffusions", Discussion paper, University of Chicago.

[18] Chen, X. and H. White (1998) "Central Limit and Functional Central Limit Theorems for Hilbert Space-Valued Dependent Processes", *Econometric Theory,* **14**, 260-284.

[19] Chernov, M., A.R. Gallant, E. Ghysels and G. Tauchen (2002) "Alternative Models of Stock Price Dynamics", *Journal of Econometrics*, forthcoming.

[20] Clark, P. (1973) "A Subordinated Stochastic Process Model with Finite Variance and Speculative Price", *Econometrica*, 41, 135-156.

[21] Conley, T., L. P. Hansen, E., Luttmer, and J. Scheinkman (1997), "Short Term Interest Rates as Subordinated Diffusions", *Review of Financial Studies*, 10, 525-578.

[22] Cox, J.C., J. Ingersoll and S.A. Ross (1985) "A theory of the term structure of interest rates", *Econometrica*, 53, 385-408.

[23] Dai, Q. and K. Singleton (2000) "Specification Analysis of Affine Term Structure Models", *Journal of Finance*, 55, 1943-1978.

[24] Das, S. and S. Foresi (1996) "Exact Solutions for Bond and Option Prices with Systematic Jumpr Risk", *Review of Derivatives Research*, 1, 7-24.

[25] Darolles, S., J.P. Florens, and E. Renault (2002), "Nonparametric Instrumental Regression", Working paper 05-2002, CRDE.

[26] Dautray, R. and J.-L. Lions (1988) *Analyse mathématique et calcul numérique pour les sciences et les techniques.* Vol. 5. Spectre des opérateurs, Masson, Paris.

[27] Debnath, L. and L. Mikunsinski (1999) "Introduction to Hilbert Spaces with Applications", Academic Press, San Diego.

[28] Detemple, J., R. Garcia, and M. Rindisbacher (2002) "Asymptotic Properties of Monte Carlo Estimators of Diffusion Processes", mimeo, University of Montreal.

[29] Duffee, G. (2002) "Term Premia and Interest Rate Forecasts in Affine Models", *Journal of Finance*, 57, 405-443.

[30] Duffie, D., D. Filipovic, and W. Schachermayer (2002) "Affine Processes and Applications in Finance", forthcoming in *Annals of Applied Probability*

[31] Duffie, D. and P. Glynn (2001) "Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals", *Working paper*, Stanford University

[32] Duffie, D. and R. Kan (1996) "A Yield-Factor Model of Interest Rates", *Mathematical Finance*, 6, 379-406

[33] Duffie, D., J. Pan and K. Singleton (2000) "Transform Analysis and Option Pricing for Affine Jump-Diffusions", *Econometrica*, 68, 1343-1376.

[34] Duffie, D. and K. Singleton (1993) "Simulated Moments Estimation of Markov Models of Asset Prices", *Econometrica*, 61, 929-952.

[35] Dunford, N. and J. Schwartz (1988) *Linear Operators, Part II: Spectral Theory*, Wiley, New York.

[36] Durham, G. and A.R. Gallant (2002) "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion", *Journal of Business and Economic Statistics*, 20, 297-316.

[37] Elerian, O., S. Chib and N. Shephard (2001) "Likelihood Inference for Discretely Observed Nonlinear Diffusions", *Econometrica*, 69, 959-1012.

[38] Eraker, B., M. Johannes and N. Polson (2001) "The Impact of Jumps in Equity Index Volatility and Returns", Discussion paper Columbia University.

[39] Feuerverger, A. (1990) "An efficiency result for the empirical characteristic function in stationary time-series models", *Canadian Journal of Statistics*, 18, 155-161.

[40] Feuerverger, A. and P. McDunnough (1981) "On the Efficiency of Empirical Characteristic Function Procedures", *J. R. Statist. Soc.* B, 43, 20-27.

[41] Florens, J.-P., Mouchard, and Rolin (1993) "Noncausality and Marginalization of Markov Processes", *Econometric Theory*, 9, 239-260.

[42] Gallant, A.R. and J. R.Long (1997) "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square", *Biometrika*, 84, 125-141.

[43] Gallant, A.R. and G. Tauchen (1996) "Which Moments to Match?" *Econometric Theory*, 12, 657-681.

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

[45] Geman, H. and M. Yor (1993) "Bessel Processes, Asian Options and Perpetuities", *Mathematical Finance*, 3, 349-375.

[46] Genz, A. and B.D. Keister (1996) "Fully Symmetric Interpolatory Rules for Multiple Integrals over Infinite Regions with Gaussian Weight", *Journal of Computational and Apllied Mathematics*, 71, 299-309

[47] Ghysels, E., C. Gouriéroux and J. Jasiak (1997) "Market Time and Asset Price Movements: Theory and Estimation", in *Statistics in Finance*, D. Hand and S. Jacka (eds.), Edward Arnold, London, Chapter 15, 307-332.

[48] Ghysels, E. and J. Jasiak (1996): "Stochastic Volatility and Time Deformation: An Application to Trading Volume and Leverage Effects", *Discussion Paper* CIRANO.

[49] Gouriéroux, C., A. Monfort and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics*, 8, S85-S118.

[50] Hansen, L. (1982) "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

[51] Hansen, L. (1985) "A Method of calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators", *Journal of Econometrics*, 30, 203-238

[52] Hansen, L. and J. Scheinkman (1995) "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes", *Econometrica*, 63, 767-804.

[53] Honoré, P. (1998): "Pifalls in Estimation Jump-Diffusion Models", *Working paper*, University of Aarhus

[54] Jiang, G. and J. Knight (2002) "Estimation of Continuous Time Processes via the Empirical Characteristic Function", *Journal of Business and Economic Statistics*, 20, 198-212.

[55] Johannes, M. (2000) "A Nonparametric View of the Role of Jumps in Interest Rates", Discussion Paper, Columbia University.

[56] Lo, A. W. (1988) "Maximum Likelihood estimation of generalized Itô processes with discretely sampled data", *Econometric Theory* 4, 231-247.

[57] Merton, R. (1976) "Option pricing when underlying stock retruns are Discontinuous", *Journal of Financial Economics*, 3,

[58] Pan, J. (2002) The jump-risk premia implicit in options: evidence from an integrated time-series study, *Journal Of Financial Economics*, 63, 3-50.

[59] Parzen, E. (1957) “On Consistent Estimates of the Spectrum of a Stationary Time Series”, *Annals of Mathematical Statistics*, 28, 329-348.

[60] Parzen, E. (1959) “Statistical Inference on Time Series by Hilbert Space Methods”, Technical report N. 23, Applied Mathematics and Statistics Laboratory, Stanford.

[61] Parzen, E. (1960) “On the estimation of a probability density function and mode”, *The Annals of Mathematical Statistics* 33, 1065-1076.

[62] Parzen, E. (1970) “Statistical Inference on time series by RKHS methods”, *12th Biennial Seminar Canadian Mathematical Congress Proc.*, R. Pyke, ed., Canadian Mathematical Society, Montreal.

[63] Piazzesi, M. (2000), “An Econometric Model of the Yield Curve with Macroeconomic Jump Effect”, NBER Working Paper 8246.

[64] Politis, D. and J. Romano (1994) “Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap”, Statistica Sinica, 4, 451-476.

[65] Ranga Rao, R. (1962) “Relations Between Weak and Uniform Convergence of Measures With Applications”, *Annals of Mathematical Statistics*, 33, 659-680.

[66] Singleton, K. (2001) “Estimation of Affine Pricing Models Using the Empirical Characteristic Function”, *Journal of Econometrics*, 102, 111-141.

[67] Tauchen, G. and M. Pitts (1983) “The Price Variability-Volume Relationship on Speculative Markets”, *Econometrica*, 51, 485-506.

[68] Van der Vaart, A. and J. Wellner (1996) *Weak Convergence and Empirical Processes*, Springer Verlag, New York.

[69] Vasicek, O. (1977), “An equilibrium characterization of the term structure”, *Journal of Financial Economics*, 5, 177-188.

[70] White, H. (1994) *Estimation, Inference and Specification Analysis,* Cambridge University Press, Cambridge, UK.

[71] Yu, J. (2001) "Empirical Characteristic Function Estimation and Its Applications", *Working paper*, University of Auckland

[72] Zhou, H. (2001) "Finite Sample Properties fo EMM, GMM, QMLE, and MLE for a Square-Root Interest Rate Diffusion Model", *Journal of Computational Finance* (forthcoming).

A Regularity Conditions

Assumption A.1 *The stochastic process X_t is a $p \times 1$-vector of random variables. X_t is stationary and $\alpha-$mixing with coefficients α_j that satisfy $\sum_{j=1}^{\infty} j^2 \alpha_j < \infty$. The distribution of $(X_1, X_2, X_3, ...)$ is indexed by a finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^q$ and Θ is compact.*

The condition on the mixing numbers is satisfied if X_t is $\alpha-$mixing of size -3.[24] Sufficient conditions for $\rho-$ and $\beta-$ mixing (and, therefore, $\alpha-$mixing) of univariate diffusions can be found in Chen et al. (1999). For subordinated diffusions, they can be found in Carrasco et al. (1999) with many examples.

For illustration, consider a stationary scalar diffusion process with drift coefficient, μ, and diffusion coefficient σ^2. Then a sufficient condition for x_t to be $\beta-$mixing with geometric decay (and therefore $\alpha-$mixing with geometric decay) is that $(\mu/\sigma + 0.5\,(\partial\sigma/\partial x))$ is negative at the right boundary and positive at the left boundary. Weaker conditions on μ and σ permit to establish $\beta-$mixing with polynomial decay rate. The condition in Assumption A.1 is relatively weak and is expected to be satisfied for a large class of processes.

The following assumption introduces the Hilbert space of reference.

Assumption A.2 *π is the p.d.f. of a distribution that is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$. $\pi(\tau) > 0$ for all $\tau \in \mathbb{R}^d$. $\mathbb{L}^2(\pi)$ is the Hilbert space of complex-valued functions that are square integrable with respect to π :*

$$\mathbb{L}^2(\pi) = \left\{ g : \mathbb{R}^d \to \mathbb{C} \mid \int |g(\tau)|^2 \pi(\tau)\, d\tau < \infty \right\}$$

Denote $\langle .,. \rangle$ and $\| \, . \, \|$ the inner product and the norm defined on $\mathbb{L}^2(\pi)$. The inner product is $\langle f, g \rangle = \int f(\tau) \overline{g(\tau)} \pi(\tau)\, d\tau$ where $\overline{g(\tau)}$ denotes the complex conjugate of $g(\tau)$. If $f = (f_1, ..., f_m)'$ and $g = (g_1, ..., g_m)'$ are vectors of functions of $\mathbb{L}^2(\pi)$, we denote $\langle f, g \rangle$ the $m \times m-$matrix with (i,j) element $\int f_i(\tau) \overline{g_j(\tau)} \pi(\tau)\, d\tau$.

We also have to define a Hilbert-space analogue of a random variable:

Definition A.1 *An $\mathbb{L}^2(\pi)-$ valued random element g has a Gaussian distribution on $\mathbb{L}^2(\pi)$ with covariance operator K if, for all $f \in \mathbb{L}^2(\pi)$, the real-valued random variable $\langle g, f \rangle$ has a Gaussian distribution*

[24] Note that a size -2 (instead of -3) is sufficient to establish the asymptotic normality of the estimator (Proposition 3.2). However we need a stronger condition (weaker dependency structure) to show the consistency of covariance operator estimate, K_T (Proposition 3.3).

on $\mathbb{R}$ *with variance* $\langle Kf, f\rangle$.[25]

We assume that the moment conditions (3.1) identify all the parameters of interest:

Assumption A.3 *The equation*

$$E^{\theta_0}\left(h_t\left(\tau;\theta\right)\right)=0 \textit{ for all } \tau\in\mathbb{R}^d,\ \pi-\textit{almost everywhere}$$

has a unique solution θ_0 *which is an interior point of* Θ. E^{θ_0} *denotes the expectation with respect to the distribution of* Y_t *for* $\theta=\theta_0$.

$\{h_t(\theta_0)\}$ is supposed to satisfy the set of assumptions:

Assumption A.4 *(i)* h *is a measurable function from* $\mathbb{R}^d\times\mathbb{R}^{\dim(Y)}\times\Theta$ *into* $\mathbb{C}$.

(ii) $h_t\left(\tau;\theta\right)$ *is continuously differentiable with respect to* θ *and* $h_t\left(\tau;\theta\right)\in L_\infty\left(\pi\otimes P^\theta\right)$ *where* $L_\infty\left(\pi\otimes P^\theta\right)$ *is the set of measurable bounded functions of* $\left(\tau, Y_t\right)$.

(iii) $\sup_{\theta\in\Theta}\left\|\hat{h}_T\left(\theta\right)-E^{\theta_0}h_t\left(\theta\right)\right\|=O_p\left(\frac{1}{\sqrt{T}}\right)$

$\sup_{\theta\in\Theta}\left\|\nabla_\theta\hat{h}_T\left(\theta\right)-E^{\theta_0}\nabla_\theta h_t\left(\theta\right)\right\|=O_p\left(\frac{1}{\sqrt{T}}\right)$, *where* ∇_θ *denotes the derivative with respect to* θ.

Note that we do not try to provide minimal assumptions and that A.4(ii) could certainly be relaxed. However, as our moment conditions are based on the conditional CF and on the joint CF, they will be necessarily bounded. We will check later that all our assumptions are satisfied in the context of CF-GMM.

The following assumption about the moment function h_t is required for establishing the properties of the optimal C-GMM estimator:

Assumption A.5 *Let* K *be the asymptotic covariance operator of* $\sqrt{T}\hat{h}_T\left(\theta_0\right)$. *The null space of* K : $\mathcal{N}\left(K\right)=\left\{f\in\mathbb{L}^2\left(\pi\right):Kf=0\right\}=\{0\}$. $E^{\theta_0}h_t\left(\theta\right)\in\mathcal{H}\left(K\right)$ *for all* $\theta\in\Theta$ *and* $E^{\theta_0}\nabla_\theta h_t\left(\theta\right)\in\mathcal{H}\left(K\right)$ *for all* $\theta\in\Theta$.

The kernel ω, used in construction of the estimator of the covariance kernel, satisfies the following conditions:

[25]Background material on the Hilbert space - valued random elements can be found in, for instance, Chen and White (1998).

Assumption A.6 *(i) The kernel* ω *satisfies* $\omega : \mathbb{R} \to [-1,1]$, $\omega(0) = 1$, $\omega(x) = \omega(-x)$, $x \in \mathbb{R}$, $\int \omega^2(x)\,dx < \infty$, $\int |\omega(x)|\,dx < \infty$. ω *is continuous at 0 and at all but a finite number of points.*

(ii) For each $\tau \in \mathbb{R}^d$, $E^{\theta_0} \sup_{\theta\in\Theta} \|\nabla_\theta h_t(\tau;\theta)\|_E^2 < \infty$ *and* $E^{\theta_0} \sup_{\theta\in\Theta} \|\nabla_{\theta\theta} h_t(\tau;\theta)\|_E^2 < \infty$ *where* $\nabla_{\theta\theta} h_t(\tau;\theta)$ *denotes the* $q \times q$ *matrix of second derivatives of* $h_t(\tau;\theta)$ *and* $\|.\|_E$ *denotes the Euclidean norm.*

The following assumption is needed in Section 4.1. to use the condition characteristic function in the markovian case.

Assumption A.7 *The stochastic process* X_t, *is a* $p \times 1-$*vector of random variables.* X_t*is stationary, Markov, and* $\alpha-$*mixing with* $\sum_{j=1}^{\infty} j^2\alpha_j < \infty$. *The conditional pdf of* X_{t+1} *given* X_t, $f_\theta(x_{t+1}|x_t;\theta)$, *is indexed by a parameter* $\theta \in \Theta \subset R^q$ *and* Θ, *is compact.* $f_\theta(x_{t+1}|x_t;\theta)$ *is continuously differentiable w.r. to* θ. $f_\theta(x_{t+1}|x_t;\theta_0)$ *is denoted* $f_\theta(x_{t+1}|x_t)$.

Now, we elaborate on the conditions to implement the efficient C-GMM estimator. Some of the assumptions, e.g. Assumption A.5, might seem to be difficult to verify. We can check these conditions using the properties of the RKHS (see Appendix B). Below, we give a set of primitive assumptions under which the general Assumptions A.1, A.5, A.4 to A.6(ii) are satisfied.

Assumption A.8 *(i)* $f_\theta(x_{t+1}|x_t;\theta) < \infty$ *for all* $\theta \in \Theta$ *and*

$$E^{\theta_0}\left[\frac{\nabla_\theta f_\theta(x_{t+1}|x_t)\,\nabla_\theta f_\theta(x_{t+1}|x_t)'}{f_{\theta_0}(x_{t+1}|x_t)}\right] < \infty$$

for all $\theta \in \Theta$. ψ_θ *is differentiable and* $\int \sup_{\theta\in\Theta} |\nabla_\theta \psi_\theta(s|x_t)|\,ds < \infty$.

(ii) $\psi_\theta(s|X_t;\theta)$ *is twice continuously differentiable in* θ. $E^{\theta_0}\|\nabla_\theta\psi_\theta(.|X_t;\theta)\|^{2+\delta} < \infty$ *for some* $\delta > 0$ *and* $\sum_{j=1}^{\infty} \alpha_j^{\delta/(2+\delta)} < \infty$.

(iii) For each $s \in \mathbb{R}^{\prime}$, $E^{\theta_0} \sup_{\theta\in\Theta} \|\nabla_\theta \psi_\theta(s|X_t;\theta)\|_E^2 < \infty$ *and* $E^{\theta_0} \sup_{\theta\in\Theta} \|\nabla_{\theta\theta} \psi_\theta(s|X_t;\theta)\|_E^2 < \infty$.

Proposition A.1 *Assumption A.7 implies Assumption A.1. If Assumption A.7 is satisfied and* h_t *is defined by*

$$h(\tau, Y_t;\theta) = e^{irX_t}\left(e^{isX_{t+1}} - \psi_\theta(s|X_t)\right), \tag{A.1}$$

with $\tau = (r,s)' \in \mathbb{R}^{2p}$ *then Assumption A.8 implies Assumptions A.4, A.5, and A.6(ii).*

The proof is provided in Appendix D.

In the section on the simulated method of moments, our starting point is the following model.

Assumption A.9 *X_t satisfies*

$$X_{t+1} = H(X_t, \varepsilon_t, \theta)$$

for some measurable transition function $H : R^p \times R^N \times \Theta$ for some $N > 0$. ε_t is an i.i.d. sequence of R^N-valued random variables independent of X_t with a known distribution that does not depend on θ.

We will need an assumption, which corresponds to Assumption A.8(ii) and (iii) for the particular moments $\tilde{h}_t^J$ used in the conditional simulation case of the simulated method of moments (section 5.1).

Assumption A.10 *(i) H is twice continuously differentiable in θ.*

(ii) $E^{\theta_0} E_\varepsilon \|\nabla_\theta H(X_t, \varepsilon_t, \theta)\|_E^{2+\delta} < \infty$ for some $\delta > 0$ and $\sum_{j=1}^{\infty} \alpha_j^{\delta/(2+\delta)} < \infty$. E_ε denotes the expectation with respect to the distribution of ε_t.

(iii) $E^{\theta_0} E_\varepsilon \sup_{\theta \in \Theta} \|\nabla_\theta H(X_t, \varepsilon_t, \theta)\|_E^2 < \infty$ and $E^{\theta_0} E_\varepsilon \sup_{\theta \in \Theta} \|\nabla_{\theta\theta} H(X_t, \varepsilon_t, \theta)\|_E^2 < \infty$.

The following assumption is required for the proof of asymptotic properties of the simulated estimator in case of path simulation (section 5.2).

Assumption A.11 *X_t is geometrically ergodic and $E^{\theta_0} \sup_{\theta \in \Theta} \left[\nabla_\theta e^{i\tau \tilde{Y}_j}\right] < \infty$. E^{θ_0} denotes the expectation with respect to the stationary distribution of Y_t.*

B Norm in a RKHS

To verify whether Assumption A.5 is satisfies, we need to be able to compute $\|g\|_K^2$, the norm of g in the RKHS associated with K. This section gives results on the norm in a RKHS that appear in Parzen (1970) and are further discussed in Carrasco and Florens (2002).

Let K be the covariance operator

$$\begin{aligned} K &: \mathbb{L}^2(\pi) \to \mathbb{L}^2(\pi) \\ f &\to g(\tau) = \int k(\tau, \tau_2) f(\tau_2) \pi(d\tau_2). \end{aligned}$$

$k(.,.)$ defines an inner product:

$$k(\tau_1, \tau_2) = \langle h(\tau_1), h(\tau_2) \rangle_{H^0}.$$

If $\{h_t\}$ is a martingale difference sequence, we have

$$\langle h(\tau_1), h(\tau_2) \rangle_{H^0} = E^{\theta_0} \left(h_t(\tau_1) \overline{h_t(\tau_2)} \right)$$

where $\overline{h(\tau_2)}$ denotes the complex conjugate of $h(\tau_2)$. If $\{h_t\}$ depends on some $Y_t = (X_t, X_{t+1}, ..., X_{t+L})'$ which is autocorrelated, we may have

$$\langle h(\tau_1), h(\tau_2)\rangle_{H^0} = \sum_{j=-\infty}^{\infty} E^{\theta_0}\left(h(\tau_1, Y_0)\overline{h(\tau_2, Y_j)}\right).$$

The question of interest is to compute $\|g\|_K$ for a specific $g \in \mathcal{H}(K)$. Let

$$C(g) = \left\{G : g(\tau) = \langle G, h(\tau)\rangle_{H^0} \quad \forall \tau \in R^d\right\}. \tag{B.1}$$

Proposition B.1 *The norm in $\mathcal{H}(K)$ satisfies*

$$\|g\|_K^2 = \min_{G \in C(g)} \|G\|_{H^0}^2.$$

Note that for the purpose of computing $\|g\|_K^2$ and because $\|g\|_K^2 = \|\bar{g}\|_K^2$, it is same whether one defines $C(g)$ as in (B.1) or as follows

$$C(g) = \left\{G : g(\tau) = \langle h(\tau), G\rangle_{H^0} \quad \forall \tau \in R^d\right\}. \tag{B.2}$$

If g is a $L-$vector, then G is also a $L-$vector and the equation is $C(g)$ becomes $g_j(\tau) = \langle h(\tau), G_j\rangle_{H^0}$ for $j = 1, ..., L$.

C On the Computation of the C-GMM Objective Function

This appendix discusses the computation of the matrix C and the vector $\underline{v}$ in the simplified C-GMM objective function (3.10). Let $y_t = (x_{t+1}, x_t)$, $\tau = (r, s)$, and $\hat{\pi}$ is the Fourier transform of π defined as

$$\hat{\pi}(x_t, x_{t+1}) = \int e^{i(rx_t + sx_{t+1})}\pi(\tau)\, d\tau. \tag{C.1}$$

Assuming r and s to be independent,

$$\pi(\tau) = \pi(r, s) = \pi_r(r)\,\pi_s(s) \tag{C.2}$$

If π is the p.d.f. of the bivariate normal variable y with zero mean and variance Σ, then $\hat{\pi}(y) = \exp[-(y'\Sigma y/2)]$ where Σ is diagonal. Consider the moments of the type:

$$h(y_{t+1}, \tau; \theta) = \left(e^{isx_{t+1}} - \psi_\theta(s|x_t)\right)e^{irx_t}. \tag{C.3}$$

An element of $\underline{v}$ is computed as follows:

$$\begin{aligned}
v_t(\theta) &= \frac{1}{T}\sum_j \int \overline{h}\left(y_t,\tau;\hat{\theta}_T^1\right) h\left(y_j,\tau;\theta\right)\pi(\tau)\,d\tau \\
&= \frac{1}{T}\sum_j \int \overline{\left(e^{isx_{t+1}} - \psi_{\hat{\theta}_T^1}(s|x_t)\right)e^{irx_t}}\left(e^{isx_{j+1}} - \psi_\theta(s|x_j)\right)e^{irx_j}\pi(\tau)\,d\tau \\
&= \frac{1}{T}\sum_j \int e^{is(x_{j+1}-x_{t+1})}e^{ir(x_j-x_t)}\pi(\tau)\,d\tau \\
&\quad -\frac{1}{T}\sum_j \int e^{i(sx_{j+1}+r(x_j-x_t))}\psi_{\hat{\theta}_T^1}(-s|x_t)\,\pi(\tau)\,d\tau \\
&\quad -\frac{1}{T}\sum_j \int e^{i(-sx_{t+1}+r(x_j-x_t))}\psi_\theta(s|x_j)\,\pi(\tau)\,d\tau \\
&\quad +\frac{1}{T}\sum_j \int \psi_{\hat{\theta}_T^1}(-s|x_t)\,\psi_\theta(s|x_j)\,e^{ir(x_j-x_t)}\pi(\tau)\,d\tau.
\end{aligned}$$

The first term is equal to $1/T\sum_j \hat{\pi}(x_j - x_t, x_{j+1} - x_{t+1})$. Given (C.2), the other terms involve:

$$I_r \equiv \int e^{ir(x_j-x_t)}\pi_r(r)dr = \hat{\pi}(x_j - x_t, 0). \tag{C.4}$$

Therefore, the second and third terms have the form:

$$I_1 = I_r \cdot \int e^{-isv}\psi_\theta(s|w)\,\pi_s(s)\,ds. \tag{C.5}$$

with opposite signs, and the last term is equal to:

$$I_2 = I_r \cdot \int \psi_{\hat{\theta}_T^1}(-s|x_t)\,\psi_\theta(s|x_j)\,\pi_s(s)\,ds \tag{C.6}$$

The remaining integrals, which have to be evaluated numerically, can be characterized as multidimensional integrals over infinite integration regions with a Gaussian weight function π. Evaluation of such integrals represents an important problem in the evaluation of quantum-mechanical matrix elements with gaussian wave functions in atomic and molecular, nuclear, and particle physics as well as in other fields. Hence a plethora of fast and accurate numerical methods have been developed, see e.g. Genz and Keister (1996).

Note that integral I_1 in (C.5) evaluated at $(v, w) = (x_{t+1}, x_t)$ looks very similar to the Fourier inverse of the CF used in Singleton (2001, Equation (14)) to construct conditional density for MLE estimation. Presence of the density π turns out to be critical in the simplification of the numerical integration task. Figure 1 compares the integrand used in Singleton with I_1 and I_2. It is clear that π dampens off all the oscillating behavior of the integrand needed for MLE.

The elements of the matrix C can be computed similarly by replacing θ by $\hat{\theta}_T^1$.

D Proofs

Proof of Lemma 3.1. To prove this result, we need a functional central limit theorem for weakly dependent process. We use the results of Politis and Romano (1994). By Assumptions A.1 and A.4(i), $\{h_t\}$ is stationary $\alpha-$mixing with $\sum_{j=1}^{\infty} j^2\alpha_j < \infty$. Moreover by Assumption A.4(ii), $\{h_t\}$ is bounded with probability one. The result follows directly from Theorem 2.2 of Politis and Romano (1994). Note that Politis and Romano require that the α coefficient of $\{h_t\}$ satisfies $\sum_{i=1}^{j} i^2\alpha(i) \leq Kj^{\mu}$ for all $1 \leq j \leq T$ and some $\mu< 3/2$ which is satisfied.

Note that K is an integral operator with kernel k defined in Equation (3.6). An operator $K : \mathbb{L}^2(\pi) \rightarrow \mathbb{L}^2(\pi)$ with kernel k is an operator of Hilbert Schmidt if

$$\int\int |k(\tau_1,\tau_2)|^2 \pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2 < \infty.$$

As π is a pdf, it is enough to show that $k(\tau_1,\tau_2) < \infty$. As $k(\tau_1,\tau_2)$ is the long-run covariance of $\{h_t\}$, it is well-known (see e.g. Politis and Romano, 1994) that a sufficient condition for k to be finite is that $\{h_t\}$ is bounded with probability one and the α-coefficients of $\{h_t\}$ are summable i.e. $\sum_j \alpha(j) < \infty$. These two conditions are satisfied under our assumptions. Hence K is a Hilbert-Schmidt operator.

Proof of Proposition 3.1. The proof of Proposition 3.1(1) is similar to that of Theorem 2 in Carrasco and Florens (2000a) and is not repeated here. The optimality argument follows from the proof of Theorem 8 in Carrasco and Florens (2000a).

Proof of Proposition 3.2. We need as preliminary result the following lemma. It generalizes Theorem 7 of Carrasco and Florens (2000a) to the case where K_T has typically a slower rate of convergence than $T^{-1/2}$. Its proof is given after that of Proposition 3.2.

Lemma D.1 *Assume K_T is such that $\|K_T - K\| = O_p(T^{-a})$, $(K_T^{\alpha_T})^{-1} = (K_T^2 + \alpha_T I)^{-1} K_T$, and α_T goes to zero. We have*

$$\left\|(K_T^{\alpha_T})^{-1/2} - (K^{\alpha_T})^{-1/2}\right\| = O_p\left(\frac{1}{T^a \alpha_T^{3/4}}\right).$$

Let f and f_T be such that $\|f_T - f\| = O\left(1/\sqrt{T}\right)$. Then, for $f \in \mathcal{H}(K) + \mathcal{H}(K)^{\perp}$, we have

$$\left\|(K_T^{\alpha_T})^{-1/2} f_T - K^{-1/2} f\right\| = O_p\left(\frac{1}{T^a \alpha_T^{3/4}}\right).$$

Although in the standard GMM framework, the proof of the rate of convergence is not needed prior to proving asymptotic normality, here the $\sqrt{T}-$rate of convergence of the estimator is required for a reason that will become clear later. First we prove consistency, second we compute the rate of convergence, third we prove normality.

Consistency. The consistency follows from Theorem 3.4. of White (1994) under the following three conditions.

(a) $Q_T(\theta) = -\left\| (K_T^{\alpha_T})^{-1/2} \hat{h}_T(\theta) \right\|^2$ is a continuous function of θ for all finite T.

(b) $Q_T(\theta) \xrightarrow{P} Q(\theta) = -\left\| (K)^{-1/2} E^{\theta_0} h_t(\theta) \right\|^2$ uniformly on Θ.

(c) $Q(\theta)$ has a unique maximizer θ_0 on Θ.

We check successively (a), (b), and (c). (a) $h_T(\theta)$ is continuous in θ by Assumption A.4 (ii). For T finite, $(K_T^{\alpha_T})^{-1/2}$ is a bounded operator (because $\alpha_T > 0$) and therefore $\left\| (K_T^{\alpha_T})^{-1/2} \hat{h}_T(\theta) \right\|^2$ is a continuous function of θ.

(b) The uniform convergence as T and $T^a \alpha_T^{3/4}$ go to infinity follows from A.4 and Lemma D.1.

(c) Assumption A.5 implies that K is a positive definite operator.. By the property of the norm, we have $\left\| E^{\theta_0} h(\theta) \right\|_K^2 = 0 \Rightarrow E^{\theta_0} h(\theta) = 0$ which implies $\theta = \theta_0$ by Assumption A.3.

Rate of convergence. To establish that $\sqrt{T}\hat{\theta}_T = O_p(1)$, we apply Theorem 3.2.5 of van der Vaart and Wellner (1996). Note that this theorem does not assume that the data are iid and does not impose any specific form on Q. We need to check that the following two conditions hold:

(a) $Q(\theta) - Q(\theta_0) \leq -C \|\theta - \theta_0\|^2$.

(b) For every T and sufficiently small δ, we have

$$E^{\theta_0} \sup_{\|\theta - \theta_0\| < \delta} |(Q_T - Q)(\theta) - (Q_T - Q)(\theta_0)| \leq \frac{C\delta}{\sqrt{T}}$$

where C is some positive constant.

We prove successively (a) and (b).

(a) Using the mean value theorem around θ_0 of $E^{\theta_0} h(\theta)$, we get

$$\begin{aligned} Q(\theta) &= Q(\theta_0) \\ &\quad -2(\theta - \theta_0)' \left\langle E^{\theta_0} \nabla_\theta h\left(\overline{\theta}\right), E^{\theta_0} h(\theta_0) \right\rangle_K \\ &\quad - (\theta - \theta_0)' \left\langle E^{\theta_0} \nabla_\theta h\left(\overline{\theta}\right), E^{\theta_0} \nabla_\theta h_t\left(\overline{\theta}\right) \right\rangle_K (\theta - \theta_0) \end{aligned}$$

for some mean value $\overline{\theta}$. The result of (a) follows from $E^{\theta_0} h(\theta_0) = 0$ and $\left\langle E^{\theta_0} \nabla_\theta h_t\left(\overline{\theta}\right), E^{\theta_0} \nabla_\theta h_t\left(\overline{\theta}\right) \right\rangle_K < \infty$ and nonsingular by Assumption A.5.

(b) Here again we use the mean value theorem around θ_0 of $E^{\theta_0}h(\theta)$.

$$
\begin{aligned}
& (Q_T - Q)(\theta) - (Q_T - Q)(\theta_0) \\
= \; & -2(\theta-\theta_0)' \left\langle \frac{1}{T}\sum \nabla_\theta h_t\left(\overline{\theta}\right), \hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}} \\
& -(\theta-\theta_0)' \left\{ \left\langle \frac{1}{T}\sum \nabla_\theta h_t\left(\overline{\theta}\right), \frac{1}{T}\sum \nabla_\theta h_t\left(\overline{\theta}\right) \right\rangle_{K_T^{\alpha_T}} - \left\langle E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right), E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right) \right\rangle_K \right\} (\theta-\theta_0)
\end{aligned}
$$

for some mean value $\overline{\theta}$. We have

$$
\begin{aligned}
\left\langle \frac{1}{T}\sum \nabla_\theta h_t\left(\overline{\theta}\right), \hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}} &= \left\langle \frac{1}{T}\sum \nabla_\theta h_t\left(\overline{\theta}\right) - E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right), \hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}} \\
& \quad + \left\langle E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right), \hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}} \\
&\leq \left\| \frac{1}{T}\sum \nabla_\theta h_t\left(\overline{\theta}\right) - E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right) \right\|_{K_T^{\alpha_T}} \left\| \hat{h}_T(\theta_0) \right\|_{K_T^{\alpha_T}} \\
& \quad + \left\langle E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right), \hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}}.
\end{aligned}
$$

By Lemma D.1, $\left\| \frac{1}{T}\sum \nabla_\theta h_t\left(\overline{\theta}\right) - E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right) \right\|_{K_T^{\alpha_T}} = O_p\left(T^a \alpha_T^{3/4}\right)$, and by Lemma 3.1, $\left\| \hat{h}_T(\theta_0) \right\|_{K_T^{\alpha_T}} = O_p\left(1/\sqrt{T}\right)$ and $\left\langle E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right), \hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}} = O_p\left(1/\sqrt{T}\right)$.Moreover Lemma D.1 implies

$$
\left\langle \frac{1}{T}\sum \nabla_\theta h_t\left(\overline{\theta}\right), \frac{1}{T}\sum \nabla_\theta h_t\left(\overline{\theta}\right) \right\rangle_{K_T^{\alpha_T}} - \left\langle E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right), E^{\theta_0}\nabla_\theta h\left(\overline{\theta}\right) \right\rangle_K = O_p\left(T^a \alpha_T^{3/4}\right).
$$

Hence

$$
\sup_{\|\theta-\theta_0\|<\delta} \left| (Q_T - Q)(\theta) - (Q_T - Q)(\theta_0) \right| \leq \frac{C_1 \delta}{\sqrt{T}} + \frac{C_2 \delta^2}{T^a \alpha_T^{3/4}}.
$$

For δ small enough (take $\delta < T^a \alpha_T^{3/4}/\sqrt{T}$), the second term is negligeable with respect to the first term on the right hand side. The result of (b) follows.

Asymptotic Normality. Using a Taylor expansion of the first order condition

$$
\left\langle \nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right), \hat{h}_T\left(\hat{\theta}_T\right) \right\rangle_{K_T^{\alpha_T}} = 0
$$

around $\theta_{0,}$ we obtain

$$
\begin{aligned}
\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) &= -\left[\left\langle \nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right), \nabla_\theta \hat{h}_T\left(\bar{\theta}\right) \right\rangle_{K_T^{\alpha_T}} \right]^{-1} \\
& \quad \times \left\langle \nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right), \sqrt{T}\hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}}
\end{aligned}
$$

where $\bar{\theta}$ is a mean value. We need to establish:

N1 - $\left\langle \nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right), \nabla_\theta \hat{h}_T\left(\bar{\theta}\right) \right\rangle_{K_T^{\alpha_T}} \xrightarrow{P} \left\langle E^{\theta_0}\nabla_\theta h_t(\theta_0), E^{\theta_0}\nabla_\theta h_t(\theta_0) \right\rangle_K$.

N2 - $\left\langle \nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right), \sqrt{T}\hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left\langle E^{\theta_0}\nabla_\theta h_t(\theta_0), E^{\theta_0}\nabla_\theta h_t(\theta_0) \right\rangle_K\right)$.

N1 - First we have $E^{\theta_0}\nabla_\theta h_t(\theta) \in \mathcal{H}(K)$ by Assumption A.5. Second, we have $\left\|\nabla_\theta \hat{h}_T(\theta)\right\|^2_{K_T^{\alpha_T}}$ converges in probability to $\left\|E^{\theta_0}\nabla_\theta h_t(\theta)\right\|^2_K$ uniformly in θ as T and $T^a\alpha_T^{3/4}$ go to infinity by A.4 and Lemma D.1. Hence the result N1 follows.

N2 - We have

$$\begin{aligned}
&\left\langle (K_T^{\alpha_T})^{-1/2}\nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right), (K_T^{\alpha_T})^{-1/2}\sqrt{T}\hat{h}_T(\theta_0)\right\rangle \\
=\ &\left\langle (K_T^{\alpha_T})^{-1/2}\nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right) - K^{-1/2}E^{\theta_0}\nabla_\theta h_t(\theta_0), (K_T^{\alpha_T})^{-1/2}\sqrt{T}\hat{h}_T(\theta_0)\right\rangle && \text{(D.1)}\\
&+\left\langle K^{-1/2}E^{\theta_0}\nabla_\theta h_t(\theta_0), (K_T^{\alpha_T})^{-1/2}\sqrt{T}\hat{h}_T(\theta_0)\right\rangle && \text{(D.2)}
\end{aligned}$$

$$(D.1) \le \left\|(K_T^{\alpha_T})^{-1/2}\nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right) - K^{-1/2}E^{\theta_0}\nabla_\theta h_t(\theta_0)\right\| \left\|(K_T^{\alpha_T})^{-1/2}\right\| \left\|\sqrt{T}\hat{h}_T(\theta_0)\right\|.$$

Using the fact that $\hat{\theta}_T$ converges at the $\sqrt{T}-$rate, we have $\left\|\nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right) - E^{\theta_0}\nabla_\theta h_t(\theta_0)\right\| = O_p\left(1/\sqrt{T}\right)$ and by Lemma D.1, we have

$$\begin{aligned}
\left\|(K_T^{\alpha_T})^{-1/2}\nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right) - K^{-1/2}E^{\theta_0}\nabla_\theta h_t(\theta_0)\right\| &= O_p\left(1/\left(T^a\alpha_T^{3/4}\right)\right),\\
\left\|(K_T^{\alpha_T})^{-1/2}\right\| &= O_p\left(1/\alpha_T^{1/2}\right).
\end{aligned}$$

The term $(D.1)$ is $O_p\left(1/\left(T^a\alpha_T^{5/4}\right)\right) = o_p(1)$ as $T^a\alpha_T^{5/4}$ goes to infinity by assumption.

The term $(D.2)$ can be decomposed as

$$\begin{aligned}
&\left\langle K^{-1/2}E^{\theta_0}\nabla_\theta h_t(\theta_0), (K_T^{\alpha_T})^{-1/2}\sqrt{T}\hat{h}_T(\theta_0)\right\rangle \\
=\ &\left\langle K^{-1/2}E^{\theta_0}\nabla_\theta h_t(\theta_0), \left((K_T^{\alpha_T})^{-1/2} - (K^{\alpha_T})^{-1/2}\right)\sqrt{T}\hat{h}_T(\theta_0)\right\rangle && \text{(D.3)}\\
&+\left\langle K^{-1/2}E^{\theta_0}\nabla_\theta h_t(\theta_0), (K^{\alpha_T})^{-1/2}\sqrt{T}\hat{h}_T(\theta_0)\right\rangle. && \text{(D.4)}
\end{aligned}$$

We have

$$\begin{aligned}
(D.3) &\le \left\|K^{-1/2}E^{\theta_0}\nabla_\theta h_t(\theta_0)\right\| \left\|(K_T^{\alpha_T})^{-1/2} - (K^{\alpha_T})^{-1/2}\right\| \left\|\sqrt{T}\hat{h}_T(\theta_0)\right\|\\
&= O_p\left(\frac{1}{T^a\alpha_T^{3/4}}\right)
\end{aligned}$$

by Lemma D.1. It remains to show that (D.4) is asymptotically normal. Denote $(\lambda_j, \phi_j : j = 1, 2...)$ the eigenvalues and eigenfunctions of K.

$$\begin{aligned}
(D.4) &= \sum_{t=1}^T \sum_{j=1}^\infty \frac{1}{\sqrt{T}} \frac{1}{\sqrt{\lambda_j^2 + \alpha_T}} \left\langle E^{\theta_0}\nabla_\theta h_t, \phi_j\right\rangle \overline{\langle h_t, \phi_j\rangle}\\
&\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^T Z_{Tt}.
\end{aligned}$$

where θ_0 is dropped to simplify the notation. By Theorem A.3.7. of White (1994), we have

$$\frac{1}{\sigma_T}\sum_{t=1}^{T} Z_{Tt} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

if the following assumptions are satisfied:

(a) $E^{\theta_0}\left(|Z_{Tt}|^{\mu}\right) \leq \Delta < \infty$ for some $\mu > 2$

(b) Z_{Tt} is near epoch dependent on $\{V_t\}$ of size -1 where $\{V_t\}$ is mixing of size -2$\mu/(\mu-2)$.

(c) $\sigma_T^2 \equiv var\left(\sum_{t=1}^{T} Z_{Tt}\right)$ satisfies $\sigma_T^{-2} = O\left(T^{-1}\right)$.

We verify Conditions (a) to (c) successively. (a) is satisfied for all μ because Z_{Tt} is bounded with probability one. Indeed, we have

$$\begin{aligned} Z_{Tt} &= \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j^2+\alpha_T}} \left\langle E^{\theta_0}\nabla_\theta h, \phi_j \right\rangle \overline{\langle h_t, \phi_j\rangle} \\ &\leq \left(\sum_{j=1}^{\infty} \frac{1}{\lambda_j^2+\alpha_T} \left|\left\langle E^{\theta_0}\nabla_\theta h, \phi_j \right\rangle\right|^2\right)^{1/2} \left(\sum_{j=1}^{\infty} |\langle h_t, \phi_j\rangle|^2\right)^{1/2} \end{aligned}$$

by Cauchy-Schwartz inequality. As $\mu_j^2+\alpha_T \geq \mu_j^2$ and

$$\begin{aligned} \sum_{j=1}^{\infty} \frac{1}{\lambda_j^2} \left|\left\langle E^{\theta_0}\nabla_\theta h, \phi_j \right\rangle\right|^2 &= \left\langle E^{\theta_0}\nabla_\theta h, E^{\theta_0}\nabla_\theta h \right\rangle_K, \\ \sum_{j=1}^{\infty} |\langle h_t, \phi_j\rangle|^2 &= \|h_t\|^2. \end{aligned}$$

We have

$$Z_{Tt} \leq \left\langle E^{\theta_0}\nabla_\theta h, E^{\theta_0}\nabla_\theta h \right\rangle_K \|h_t\|^2 \leq \Delta < \infty$$

with probability one. (b) It is easy to verify that Z_{Tt} is near epoch dependent on $\{h_t\}$ of arbitrary size. (c) We have

$$\begin{aligned} &\frac{1}{T} var\left(\sum_{t=1}^{T} Z_{Tt}\right) \\ &= var\left\{\sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j^2+\alpha_T}} \left\langle E^{\theta_0}\nabla_\theta h, \phi_j \right\rangle \overline{\left\langle \sqrt{T}\hat{h}_T, \phi_j \right\rangle}\right\} \\ &= \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j^2+\alpha_T}} \left|\left\langle E^{\theta_0}\nabla_\theta h, \phi_j \right\rangle\right|^2 var\left(\left\langle \sqrt{T}\hat{h}_T, \phi_j \right\rangle\right) \\ &\quad + \sum_{i\neq j} \frac{1}{\sqrt{\lambda_i^2+\alpha_T}\sqrt{\lambda_j^2+\alpha_T}} \left\langle E^{\theta_0}\nabla_\theta h, \phi_i \right\rangle \overline{\left\langle E^{\theta_0}\nabla_\theta h, \phi_j \right\rangle} cov\left(\left\langle \sqrt{T}\hat{h}_T, \phi_i \right\rangle, \left\langle \sqrt{T}\hat{h}_T, \phi_j \right\rangle\right). \end{aligned}$$

Using as before $\lambda_j^2+\alpha_T \geq \lambda_j^2$, both sums can be bounded by a term that does not depend on T, therefore we may, in passing at the limit as $T\to\infty$, interchange the limit and the summation. By Lemma 3.1, we have $\sqrt{T}\hat{h}_T \xrightarrow{L} \mathcal{N}(0, K)$ and hence

$$\lim_{T\to\infty} cov\left(\left\langle \sqrt{T}\hat{h}_T, \phi_i \right\rangle, \left\langle \sqrt{T}\hat{h}_T, \phi_j \right\rangle\right) = \langle K\phi_i, \phi_j \rangle = \begin{cases} \lambda_j \text{ if } i = j \\ 0 \text{ otherwise} \end{cases}.$$

Therefore

$$\frac{1}{T} var\left(\sum_{t=1}^{T} Z_{Tt}\right) \to \sum_j \frac{1}{\lambda_j} \left|\left\langle E^{\theta_0} \bigtriangledown_\theta h, \phi_j \right\rangle\right|^2 = \left\langle E^{\theta_0} \bigtriangledown_\theta h, E^{\theta_0} \bigtriangledown_\theta h \right\rangle_K$$

as $T \to \infty$ proving that $\sigma_T^{-2} = O\left(T^{-1}\right)$.

Hence, we have

$$(D.4) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left\langle E^{\theta_0} \bigtriangledown_\theta h, E^{\theta_0} \bigtriangledown_\theta h \right\rangle_K\right).$$

This completes the proof.

Proof of Lemma D.1. Note that

$$\begin{aligned}
& \left\| (K_T^{\alpha_T})^{-1/2} - (K^{\alpha_T})^{-1/2} \right\| \\
= & \left\| \left(\alpha_T + K_T^2\right)^{-1/2} K_T^{1/2} - \left(\alpha_T + K^2\right)^{-1/2} K^{1/2} \right\| \\
= & \left\| \left(\alpha_T + K_T^2\right)^{-1/2} K_T^{1/2} - \left(\alpha_T + K_T^2\right)^{-1/2} K^{1/2} \right\| \\
& + \left\| \left(\alpha_T + K_T^2\right)^{-1/2} K^{1/2} - \left(\alpha_T + K^2\right)^{-1/2} K^{1/2} \right\| \\
\leq & \underbrace{\left\| \left(\alpha_T + K_T^2\right)^{-1/2} \right\|}_{\leq \alpha_T^{-1/2}} \underbrace{\left\| K_T^{1/2} - K^{1/2} \right\|}_{=O_p(T^{-a})} \qquad \text{(D.5)} \\
& + \left\| \left[\left(\alpha_T + K_T^2\right)^{-1/2} - \left(\alpha_T + K^2\right)^{-1/2} \right] K^{1/2} \right\| \qquad \text{(D.6)}
\end{aligned}$$

Using $A^{-1/2} - B^{-1/2} = A^{-1/2}\left[B^{1/2} - A^{1/2}\right]B^{-1/2}$, we get

$$\begin{aligned}
& (D.6) \\
= & \left\| \left(\alpha_T + K_T^2\right)^{-1/2} \left[\left(\alpha_T + K_T^2\right)^{1/2} - \left(\alpha_T + K^2\right)^{1/2} \right] \left(\alpha_T + K^2\right)^{-1/2} K^{1/2} \right\| \\
\leq & \underbrace{\left\| \left(\alpha_T + K_T^2\right)^{-1/2} \right\|}_{\leq \alpha_T^{-1/2}} \underbrace{\left\| \left(\alpha_T + K_T^2\right)^{1/2} - \left(\alpha_T + K^2\right)^{1/2} \right\|}_{=O_p(T^{-a})} \underbrace{\left\| \left(\alpha_T + K^2\right)^{-1/4} \right\|}_{\leq \alpha_T^{-1/4}} \underbrace{\left\| \left(\alpha_T + K^2\right)^{-1/4} K^{1/2} \right\|}_{\leq 1}.
\end{aligned}$$

Hence $(D.6) = O_p\left(T^{-a}\alpha_T^{-3/4}\right)$. The first equality of Lemma D.1 follows from the fact that $(D.5)$ is negligeable with respect to $(D.6)$. The second equality can be proved similarly using Theorem 7 of Carrasco and Florens (2000a).

Proof of Proposition 3.3. Let $\|A\|_{HS}$ denote the Hilbert Schmidt norm of the operator A (see Dautray and Lions, 1988, for a definition and the properties of the Hilbert-Schmidt norm). If $\|A\|$ denotes the usual operator norm, $\|A\| \leq \|A\|_{HS}$. We have

$$\|K_T - K\|_{HS}^2 = \int\int \left|\hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2)\right|^2 \pi(\tau_1)\pi(\tau_2)\, d\tau_1 d\tau_2.$$

Next we use the following result. If $X_T \geq 0$ is such that $EX_T = O(1)$ then $X_T = O_p(1)$. This result is proved in Darolles, Florens, and Renault (2000, footnote 12). We have

$$\begin{aligned} E^{\theta_0} \|K_T - K\|_{HS}^2 &= \int\int E^{\theta_0} \left| \hat{k}_T(\tau_1, \tau_2) - E^{\theta_0} \hat{k}_T(\tau_1, \tau_2) \right|^2 \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2 \\ &+ \int\int \left| E_T^{\theta_0} \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right|^2 \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2. \end{aligned}$$

Parzen (1957) and Andrews (1991) consider kernel estimators of the covariance of real-valued random variables. Here, we have complex-valued h_t but their results remain valid. From Parzen (1957) and Andrews (1991), we have

$$\begin{aligned} S_T^q \left(E_T^{\theta_0} \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right) &= -2\pi \omega_\nu f^{(\nu)}, \\ \frac{T}{S_T} E^{\theta_0} \left| \hat{k}_T(\tau_1, \tau_2) - E_T^{\theta_0} \hat{k}_T(\tau_1, \tau_2) \right|^2 &= \pi^2 f^2 \int \omega^2(x) \, dx. \end{aligned}$$

Hence if $S_T^{2\nu+1}/T \to \gamma$, we have:

$$\frac{T}{S_T} E^{\theta_0} \|K_T - K\|_{HS}^2 = 4\pi^2 \left(\frac{\omega_\nu^2 f^{(\nu)2}}{\gamma} + 2f^2 \int \omega^2(x) \, dx \right).$$

and therefore, $E^{\theta_0} \|K_T - K\|_{HS}^2 = O\left(T^{-2\nu/(2\nu+1)}\right)$. This yields the result.

Proof of Proposition 3.4. The C-GMM estimator is solution of:

$$\begin{aligned} \hat{\theta}_T &= \arg\min_\theta \left\| \left(K_T^{\alpha_T}\right)^{-1/2} h_T(\theta) \right\|^2 \\ &\iff \hat{\theta}_T = \arg\min_\theta \left\langle \left(K_T^{\alpha_T}\right)^{-1} \hat{h}_T(.;\theta), \hat{h}_T(.;\theta) \right\rangle \end{aligned} \tag{D.7}$$

Let $g = \left(K_T^{\alpha_T}\right)^{-1} \hat{h}_T(\theta)$ so that g satisfies:

$$\begin{aligned} \left(\alpha_T I_T + K_T^2\right) g &= K_T \hat{h}_T(\theta) \\ \alpha_T g(\tau) + \frac{1}{(T-q)^2} \sum_{t,l=1}^T h_t\left(\tau; \hat{\theta}^1\right) c_{tl} \beta_l &= \frac{1}{T-q} \sum_{t=1}^T h_t\left(\tau; \hat{\theta}^1\right) v_t(\theta) \end{aligned} \tag{D.8}$$

with

$$\beta_l = \int U h_l\left(\tau; \hat{\theta}^1\right) g(\tau) \pi(\tau) \, d\tau.$$

First, we compute $\beta_l, l = 1, ..., T$. We premultiply (D.8) by $Uh_k\left(\tau; \hat{\theta}^1\right) \pi(\tau)$ and integrate with respect to τ to obtain:

$$\alpha_T \beta_k + \frac{1}{(T-q)^2} \sum_{t,l=1}^T c_{kt} c_{tl} \beta_l = \frac{1}{T-q} \sum_{t=1}^T c_{kt} v_t(\theta). \tag{D.9}$$

Using the matrix notation, (D.9) can be rewritten

$$\left[\alpha_T I_T + C^2\right] \underline{\beta} = C \underline{v}(\theta).$$

where $\underline{\beta} = [\beta_1, ..., \beta_T]'$. Solving in $\underline{\beta}$, we get

$$\underline{\beta} = \left[\alpha_T I_T + C^2\right]^{-1} C \underline{v}(\theta). \tag{D.10}$$

Now we want to compute $\langle g, h_T(\theta) \rangle$ that appears in (D.7). To do so, we multiply all terms of (D.8) with $\overline{\hat{h}_T(\tau;\theta)}\pi(\tau)$ and integrate with respect to τ :

$$\alpha_T \left\langle g, \hat{h}_T(\theta) \right\rangle + \frac{1}{(T-q)^2} \sum_{t,l=1}^{T} w_t(\theta) c_{tl} \beta_l = \frac{1}{T-q} \sum_{t=1}^{T} w_t(\theta) v_t(\theta).$$

So that

$$\langle g, h_T(\theta) \rangle = \frac{1}{\alpha_T (T-q)} \left[\underline{w}'(\theta) \underline{v}(\theta) - \underline{w}'(\theta) C \underline{\beta}\right]$$

and using (D.10), we obtain

$$\langle g, h_T(\theta) \rangle = \frac{1}{\alpha_T (T-q)} \underline{w}'(\theta) \left[I_T - C \left[\alpha_T I_T + C^2\right]^{-1} C\right] \underline{v}(\theta).$$

This yields the result.

Proof of Proposition 3.5. The proof is very similar to that of 3.4 and is not repeated here. The consistency follows from Lemma D.1**.**

Proof of Proposition A.1. Assumption A.7 $\Rightarrow$ Assumption A.1 is obvious. We check successively the conditions of Assumption A.4.

A.4(i) and (ii):

$$\begin{aligned} |h_t| &= \left| e^{i(sx_{t+1}+rx_t)} - \psi_\theta(s|x_t) e^{irx_{t+1}} \right| \\ &\leq \left| e^{i(sx_{t+1}+rx_t)} \right| + \left| \psi_\theta(s|x_t) e^{irx_{t+1}} \right| \\ &\leq 2 \end{aligned}$$

as $|\psi_\theta(s|x_t)| \leq 1$ for all s. h_t is continuously differentiable by A.8(ii).

A.4(iii): The same way as we proved Lemma 3.1, we can prove that $\sqrt{T}\left(\hat{h}_T(\theta) - E^{\theta_0} h_t(\theta)\right)$ converges weakly to a Gaussian process with mean zero. Hence $\left\| \hat{h}_T(\theta) - E^{\theta_0} h_t(\theta) \right\| = O_p\left(1/\sqrt{T}\right)$. As $h_t(\theta)$ is bounded, the convergence is uniform by Ranga Rao (1962). Now, we turn to the term involving $\nabla_\theta \hat{h}_T(\theta)$. By Politis and Romano (1994, Theorem 2.3(i)), $\sqrt{T}\left(\nabla_\theta \hat{h}_T(\theta) - E^{\theta_0} \nabla_\theta h_t(\theta)\right)$ converges weakly to a Gaussian process with mean zero under the Assumptions A.8(ii). Hence $\left\| \nabla_\theta \hat{h}_T(\theta) - E^{\theta_0} \nabla_\theta h_t(\theta) \right\| = O_p\left(1/\sqrt{T}\right)$. By Ranga Rao (1962), the convergence is uniform under the extra assumption: there is some function $b(s, x_t) > 0$ such that $|\nabla_\theta \psi_\theta(s|x_t)| \leq b(s, x_t)$ for all $\theta \in \Theta$ and $E^{\theta_0} b(s, X_t) < \infty$. This last assumption is satisfied under the stronger condition A.8(iii).

A.5: First we check that $E^{\theta_0} h_t(\theta) \in \mathcal{H}(K)$ for all $\theta \in \Theta$. Note that

$$E^{\theta_0} h_t(\theta) = E^{\theta_0}\left[e^{irX_t}\left(\psi_{\theta_0}(s|X_t) - \psi_\theta(s|X_t)\right)\right] \equiv g(r,s) \tag{D.11}$$

We apply Proposition B.1 to compute $\|g\|_K^2$. We need to find G such that

$$\begin{aligned} g(r,s) &= E^{\theta_0}\left[\left(e^{i(sX_{t+1}+rX_t)} - \psi_\theta(s|X_t)\, e^{irX_{t+1}}\right) G(X_t, X_{t+1})\right] \\ &= E^{\theta_0}\left[e^{i(sX_{t+1}+rX_t)}\left\{G(X_t, X_{t+1}) - E^{\theta_0}\left[G(X_t, X_{t+1})|X_t\right]\right\}\right]. \end{aligned}$$

Let us denote $\tilde{G} = G - E^{\theta_0}[G|X_t]$. We want to solve in $\tilde{G}$ the equation

$$g(\tau) = \int e^{i(sx_{t+1}+rx_t)} \tilde{G}(x_t, x_{t+1})\, f_{\theta_0}(x_{t+1}|x_t)\, f_{\theta_0}(x_t)\, dx_{t+1} dx_t.$$

Applying twice the Fourier inversion formula, we obtain a unique solution

$$\tilde{G}(x_t, x_{t+1}) = \frac{1}{(2\pi)^2} \int\int \frac{g(r,s)\, e^{-i(sx_{t+1}+rx_t)}}{f_{\theta_0}(x_{t+1}|x_t)\, f_{\theta_0}(x_t)} ds dr. \tag{D.12}$$

We now replace $g(r,s)$ by its expression (D.11) into (D.12) to calculate $\tilde{G}$. Applying the Fourier inversion formula, we have

$$\begin{aligned} \frac{1}{2\pi}\int e^{-irx_t}\left(\int e^{iru}\psi_\theta(s|u)\, f_{\theta_0}(u)\, du\right) dr &= \psi_\theta(s|x_t)\, f_{\theta_0}(x_t), \\ \frac{1}{2\pi}\int \psi_\theta(s|x_t)\, e^{-isx_{t+1}} ds &= f_\theta(x_{t+1}|x_t). \end{aligned}$$

Hence we have

$$\tilde{G}(x_t, x_{t+1}) = \frac{f_{\theta_0}(x_{t+1}|x_t) - f_\theta(x_{t+1}|x_t)}{f_{\theta_0}(x_{t+1}|x_t)}$$

and $\|g\|_K^2 = E^{\theta_0}\tilde{G}^2 < \infty$ if and only if

$$\int\int \left[\frac{f_{\theta_0}(x_{t+1}|x_t) - f_\theta(x_{t+1}|x_t)}{f_{\theta_0}(x_{t+1}|x_t)}\right]^2 f_{\theta_0}(x_t, x_{t+1})\, dx_t dx_{t+1} < \infty$$

for all $\theta \in \Theta$. We recognize the chi-square distance, it is finite as long as $f_\theta(x_{t+1}|x_t) < \infty$ for all $\theta \in \Theta$. Now, we check that $E^{\theta_0}\nabla_\theta h_t(\theta) \in \mathcal{H}(K)$ for all $\theta \in \Theta$. We replace $g(r,s)$ by

$$g(r,s) \equiv E^{\theta_0}\nabla_\theta h_t(\theta) = -E^{\theta_0}\left[e^{irX_t}\nabla_\theta\psi_\theta(s|X_t)\right]$$

in Equation (D.12) to calculate $\tilde{G}$. We again apply the Fourier inversion formula to obtain

$$\begin{aligned} \frac{1}{2\pi}\int e^{-irx_t}\left(\int e^{iru}\nabla_\theta\psi_\theta(s|u)\, f_{\theta_0}(u)\, du\right) dr &= \nabla_\theta\psi_\theta(s|x_t)\, f_{\theta_0}(x_t), \\ \frac{1}{2\pi}\int \nabla_\theta\psi_\theta(s|x_t)\, e^{-isx_{t+1}} ds &= \frac{1}{2\pi}\nabla_\theta \int \psi_\theta(s|x_t)\, e^{-isx_{t+1}} ds \qquad \text{(D.13)} \\ &= \nabla_\theta f_\theta(x_{t+1}|x_t). \end{aligned}$$

We are allowed to interchange the order of integration and derivation in (D.13) because of $\int \sup_{\theta\in\Theta} |\nabla_\theta \psi_\theta (s|x_t)| ds < \infty$ and by Lemma 3.6 of Newey and McFadden (1994). Hence we have

$$\tilde{G} = -\frac{\nabla_\theta f_\theta (x_{t+1}|x_t)}{f_{\theta_0} (x_{t+1}|x_t)}$$

and

$$\begin{aligned} \left\| E^{\theta_0} \nabla_\theta h_t (\theta) \right\|_K^2 &= E^{\theta_0} \tilde{G}\tilde{G}' \\ &= E^{\theta_0} \left[\frac{\nabla_\theta f_\theta (x_{t+1}|x_t)}{f_{\theta_0} (x_{t+1}|x_t)} \left(\frac{\nabla_\theta f_\theta (x_{t+1}|x_t)}{f_{\theta_0} (x_{t+1}|x_t)} \right)' \right] \end{aligned} \quad \text{(D.14)}$$

which is finite by assumption. When $\theta = \theta_0$, the term in (D.14) coincides with the information matrix I_{θ_0} which proves the ML-efficiency without using Proposition 3.6.

Assumption A.6(ii) follows from A.8(iii) because

$$\begin{aligned} \|\nabla_\theta h_t (\theta)\|^2 &= \int \left| e^{irX_t} \nabla_\theta \psi_\theta (s|X_t) \right|^2 ds \\ &\leq \int |\nabla_\theta \psi_\theta (s|X_t)|^2 ds \\ &= \|\nabla_\theta \psi_\theta (.|X_t)\|^2 \end{aligned}$$

and similarly $\|\nabla_{\theta\theta} h_t (\theta)\|^2 \leq \|\nabla_{\theta\theta} \psi_\theta (.|X_t)\|^2$.

Proof of Proposition 4.1. The asymptotic distribution of $\hat{\theta}_T$ follows from Propositions 3.2 and A.1. The asymptotic efficiency follows from Equation (D.14).

Proof of Proposition 4.3. To simplify the notation, we omit θ_0 also all the terms in this proof are taken at θ_0. Recall that the variance of $\widetilde{\theta}_T$ is given by $J^{-1}\Sigma J^{-1}$ with

$$\begin{aligned} J &= E^{\theta_0} (\nabla_{\theta\theta} \ln f_\theta (Y_0)) = -E^{\theta_0} \left[\nabla_\theta \ln f_\theta (Y_0) (\nabla_\theta \ln f_\theta (Y_0))' \right], \\ \Sigma &= \sum_{j=-\infty}^{\infty} E^{\theta_0} \left[\nabla_\theta \ln f_\theta (Y_0) (\nabla_\theta \ln f_\theta (Y_j))' \right]. \end{aligned}$$

The asymptotic variance of $\hat{\theta}_T$ is given by Theorem 2 in Carrasco and Florens (2000a) by replacing B by $\tilde{K}^{-1/2}$:

$$V = \left(\|g\|_{\tilde{K}}^2 \right)^{-1} \left(\tilde{K}^{-1} g, K\tilde{K}^{-1} g \right) \left(\|g\|_{\tilde{K}}^2 \right)^{-1}$$

where $g = E^{\theta_0} (\nabla_\theta h)$. Theorem 2 assumes that B is a bounded operator, here B is not bounded but a proof similar to that of Theorem 8 of Carrasco and Florens (2000a) would show that the result is also valid for $\tilde{K}^{-1/2}$.

a - Calculation of $\|g\|_{\tilde{K}}^2$:

We apply results from Proposition B.1. First we check that

$$G_0 = \nabla_\theta \ln f_\theta (Y_t)$$

belongs to $C(g)$ that is

$$\begin{aligned} \nabla_\theta \psi_\theta (\tau) &= \int \nabla_\theta \ln f_\theta (y_t) \left(e^{i\tau y_t} - \psi_\theta (\tau) \right) f_\theta (y_t) \, dy_t \\ &= \int \nabla_\theta f_\theta (y_t) \left(e^{i\tau y_t} - \psi_\theta (\tau) \right) dy_t \\ &= \int \nabla_\theta f_\theta (y_t) \, e^{i\tau y_t} dy_t \\ &= \nabla_\theta \int f_\theta (y_t) \, e^{i\tau y_t} dy_t. \end{aligned}$$

Now consider a general solution $G = G_0 + G_1$. The condition $G \in C(g)$ implies

$$\begin{aligned} \int G_1 (y_t) \left(e^{i\tau y_t} - \psi_\theta (\tau) \right) f_\theta (y_t) \, dy_t &= 0 \quad \forall \tau \\ \Leftrightarrow \int (G_1 (y_t) - EG_1) \, e^{i\tau y_t} f_\theta (Y_t) \, dY_t &= 0 \quad \forall \tau \\ \Rightarrow G_1 - EG_1 &= 0 \\ \Rightarrow E^{\theta_0} (G_0 G_1) &= 0. \end{aligned}$$

This shows that the element of $C(g)$ with minimal norm is G_0. Hence we have

$$\|g\|_{\tilde{K}}^2 = E^{\theta_0} (G_0 G_0') = E^{\theta_0} \left[(\nabla_\theta \ln f_\theta (Y_t)) (\nabla_\theta \ln f_\theta (Y_t))' \right].$$

b - Calculation of $\tilde{K}^{-1} g$: We verify that $g = \tilde{K} \omega$ with

$$\omega (\tau) = \int e^{-i\tau v} \nabla_\theta \ln f_\theta (v) \, dv$$

where v is a $L-$vector and f_θ denotes the joint likelihood of Y_t. Because Y_t is assumed to be stationary, f_θ does not depend on t. We have

$$\begin{aligned} \left(\tilde{K} \omega \right) (\tau_1) &= \int (\psi_\theta (\tau_1 + \tau_2) - \psi_\theta(\tau_1)\psi_\theta(\tau_2)) \int e^{-i\tau_2 v} \nabla_\theta \ln f_\theta (v) \, dv d\tau_2 \\ &= \int \psi_\theta (\tau_1 + \tau_2) \int e^{-i\tau_2 v} \nabla_\theta \ln f_\theta (v) \, dv d\tau_2 - \psi_\theta(\tau_1) \int \nabla_\theta f_\theta (v) \, dv \\ &= \int e^{i\tau_1 y} \left[\int\int e^{i\tau_2 y} e^{-i\tau_2 v} \nabla_\theta \ln f_\theta (v) \, dv d\tau_2 \right] f_\theta (y) \, dy \\ &= \int e^{i\tau_1 y} \nabla_\theta \ln f_\theta (y) \, f_\theta (y) \, dy = g (\tau_1). \end{aligned}$$

The fourth equality follows from a property of the Fourier transform, see Theorem 4.11.12. in Debnath and Mikusinsky (1999).

c - Calculation of $\left(\tilde{K}^{-1} g, K \tilde{K}^{-1} g \right)$:

Note that $\left(\tilde{K}^{-1}g, K\tilde{K}^{-1}g\right) = (\omega, K\omega)$. The kernel of K is given by

$$k(\tau_1,\tau_2) = \sum_{j=-\infty}^{\infty}\left[E^{\theta_0}\left(e^{i(\tau_1 Y_0+\tau_2 Y_j)}\right) - \psi_\theta(\tau_1)\psi_\theta(\tau_2)\right] \equiv \sum_{j=-\infty}^{\infty} k_j(\tau_1,\tau_2).$$

Let us denote K_j the operator with kernel $k_j(\tau_1,\tau_2)$.

$$(K_j\omega)(\tau_1) = \int E^{\theta_0}\left(e^{i(\tau_1 Y_0+\tau_2 Y_j)}\right)\int e^{-i\tau_2 v}\nabla_\theta \ln f_\theta(v)\, dv d\tau_2$$

because the second term equals zero.

$$\begin{aligned}(K_j\omega)(\tau_1) &= \int e^{i\tau_1 y_0}\left[\int\int e^{i\tau_2 y_j}e^{-i\tau_2 v}\nabla_\theta \ln f_\theta(v)\, dv d\tau_2\right] f_\theta(y_0,y_j)\, dy_0 dy_j\\ &= \int e^{i\tau_1 y_0}\nabla_\theta \ln f_\theta(y_j) f_\theta(y_0,y_j)\, dy_0 dy_j.\end{aligned}$$

We have

$$\begin{aligned}(\omega, K_j\omega) &= \int\left[\int\int e^{i\tau_1 y_0}e^{-i\tau_1 v}\nabla_\theta \ln f_\theta(v)\, dv d\tau_1\right]\nabla_\theta \ln f_\theta(y_j)' f_\theta(y_0,y_j)\, dy_0 dy_j\\ &= \int \nabla_\theta \ln f_\theta(y_0)\nabla_\theta \ln f_\theta(y_j)' f_\theta(y_0,y_j)\, dy_0 dy_j\\ &= E^{\theta_0}\left[\nabla_\theta \ln f_\theta(Y_0)(\nabla_\theta \ln f_\theta(Y_j))'\right].\end{aligned}$$

It follows that

$$(\omega, K\omega) = \sum_{j=-\infty}^{\infty}(\omega, K_j\omega) = \sum_{j=-\infty}^{\infty} E^{\theta_0}\left[\nabla_\theta \ln f_\theta(Y_0)(\nabla_\theta \ln f_\theta(Y_j))'\right]$$

which finishes the proof.

Proof of Proposition 5.1. We wish to apply Proposition 3.2 on $\left\{\tilde{h}_t^J\right\}$. To do this, we need to check that the conditions of this proposition are satisfied for $\left\{\tilde{h}_t^J\right\}$. The mixing properties of $\left\{\tilde{h}_t^J\right\}$ are the same as those of $\{X_t\}$, moreover $\left\{\tilde{h}_t^J\right\}$ is a martingale difference sequence. Hence by Assumption A.7 and Politis and Romano (1994), we have

$$\sqrt{T}\tilde{h}_t^J \Rightarrow \mathcal{N}\left(0,\tilde{K}\right)$$

as $T\to\infty$ in $\mathbb{L}^2(\pi)$ where $\tilde{K}$ is the operator with kernel $\tilde{k}$ satisfying

$$\begin{aligned}\tilde{k}(\tau_1,\tau_2) &= cov\left(\tilde{h}_t^J(\tau_1), \tilde{h}_t^J(\tau_2)\right)\\ &= E\left[cov\left(\tilde{h}_t^J(\tau_1), \tilde{h}_t^J(\tau_2)|Y_t\right)\right]\\ &+ cov\left[E\left(\tilde{h}_t^J(\tau_1)|Y_t\right), E\left(\tilde{h}_t^J(\tau_2)|Y_t\right)\right]\\ &= \frac{1}{J}E_Y E_\varepsilon\left[\left(\tilde{h}^J(\tau_1)-h(\tau_1)\right)\overline{\left(\tilde{h}^J(\tau_2)-h(\tau_2)\right)}|Y_t\right] + cov\left(h(\tau_1), h(\tau_2)\right)\\ &= \frac{1}{J}u(\tau_1,\tau_2) + k(\tau_1,\tau_2).\end{aligned}$$

Note that we use E and cov for the expectation and covariance with respect to both ε_t and Y_t. Therefore $\tilde{K} = K + U/J$. Note that U is a positive definite operator. Assumption A.5 is satisfied under Assumption A.8(i) because

$$\begin{aligned} Eh_t(\theta) &= E\tilde{h}_t^J(\theta), \\ E\nabla_\theta h_t(\theta) &= E\nabla_\theta \tilde{h}_t^J(\theta). \end{aligned}$$

The second equality follows from

$$\begin{aligned} E\left(\triangledown_\theta \tilde{h}_t\right) &= E_Y E_\varepsilon\left[\triangledown_\theta \tilde{h}_t | Y_t\right] \\ &= E_Y\left[\triangledown_\theta E_\varepsilon\left(\tilde{h}_t | Y_t\right)\right]. \end{aligned} \tag{D.15}$$

The order of integration and differentiation in D.15 can be exchanged because the distribution of $\tilde{\varepsilon}_{j,t}$ does not depend on θ and $E\sup_{\theta\in\Theta}|\nabla_\theta H| < \infty$ which is true under A.10(iii). Therefore $E\left(\triangledown_\theta \tilde{h}_t^J\right) = E\left(\triangledown_\theta h\right).$ Finally, using a proof very similar to that of Proposition A.1, we see that Assumptions A.4(iii) and A.6(ii) are satisfied under Assumption A.10. It is enough to notice that

$$\begin{aligned} \left|\triangledown_\theta \tilde{h}_t^J\right| &= \left|\frac{1}{J}\sum_{j=1}^J is\triangledown_\theta H\left(X_t, \varepsilon_{j,t+1}, \theta\right) e^{isX_{t+1|t}^{\theta j}}\right| \\ &\leq \frac{1}{J}\sum_{j=1}^J \left|\triangledown_\theta H\left(X_t, \varepsilon_{j,t+1}, \theta\right)\right| \end{aligned}$$

and $\left|\triangledown_{\theta\theta}\tilde{h}_t^J\right| \leq \frac{1}{J}\sum_{j=1}^J \left|\triangledown_{\theta\theta} H\left(X_t, \varepsilon_{j,t+1}, \theta\right)\right|$. Hence, from Proposition 3.2, we have

$$\sqrt{T}\left(\tilde{\theta}_T - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(\left\langle E^{\theta_0}\left(\nabla_\theta \tilde{h}^J\right), E^{\theta_0}\left(\nabla_\theta \tilde{h}^J\right)\right\rangle_{\tilde{K}}\right)^{-1}\right).$$

We can rewrite the variance by using $E\left(\triangledown_\theta \tilde{h}_t^J\right) = E\left(\triangledown_\theta h\right).$

Now, we show the inequality $\|g\|_{\tilde{K}}^2 \leq \|g\|_K^2$ for any function g in the range of K. For sake of simplicity, we assume g scalar, the proof for g vector is very similar. Denote

$$\begin{aligned} f &= \left(K + \frac{1}{J}U\right)^{-1} g \\ l &= K^{-1}g. \end{aligned}$$

We have $\|g\|_{\tilde{K}}^2 = \langle f, g\rangle$ and $\|g\|_K^2 = \langle l, g\rangle$. We want to show $\langle l - f, g\rangle \geq 0$.

$$\begin{aligned} & \langle l - f, g\rangle \geq 0 \\ \Leftrightarrow\ & \langle K(l-f), g\rangle_K \geq 0 \\ \Leftrightarrow\ & \left\langle \frac{1}{J}Uf, Kf + \frac{1}{J}Uf\right\rangle_K \geq 0 \\ \Leftrightarrow\ & \frac{1}{J}\langle Uf, f\rangle + \|Uf\|_K^2 \geq 0 \end{aligned}$$

This last inequality is true because U is definite positive.

Proof of Proposition 5.2 The consistency holds under Assumptions A.2-A.4. By the geometric ergodicity and the boundedness of $e^{i\tau Y_t}$ and $e^{i\tau \widetilde{Y}_j}$, the functional CLT of Chen and White (1998, Theorem 3.9) gives:

$$\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\left(e^{i\tau Y_t}-\psi_{\theta}^{L}(\tau)\right)\Rightarrow\mathcal{N}(0,K),$$
$$\frac{\sqrt{J(T)}}{J(T)}\sum_{j=1}^{J(T)}\left(e^{i\tau \widetilde{Y}_j}-\psi_{\theta}^{L}(\tau)\right)\Rightarrow\mathcal{N}(0,K).$$

as $T\rightarrow\infty$ in $\mathbb{L}^2(\pi)$. The asymptotic normality follows from

$$\begin{aligned}\sqrt{T}\tilde{h}_T(\tau;\theta_0) &= \frac{\sqrt{T}}{T}\sum_{t=1}^{T}\left(e^{i\tau Y_t}-\psi_{\theta}^{L}(\tau)\right)-\frac{\sqrt{T}}{\sqrt{J(T)}}\frac{\sqrt{J(T)}}{J(T)}\sum_{j=1}^{J(T)}\left(e^{i\tau \widetilde{Y}_j}-\psi_{\theta}^{L}(\tau)\right)\\ &\xrightarrow{\mathcal{L}}\mathcal{N}(0,(1+\zeta)K)\end{aligned}$$

because Y_t and $\tilde{Y}_j$ are independent. Let $\tilde{K}=(1+\zeta)K$. Minimizing $\left\|\tilde{h}_T\right\|_{K_T^{\alpha_T}}$ is equivalent to minimizing $\left\|\tilde{h}_T\right\|_{\tilde{K}_T^{\alpha_T}}$ where $\tilde{K}_T^{\alpha_T}$ denote a regularized estimator of $\tilde{K}$. By Proposition 3.2, $\tilde{\theta}_T$ is asymptotically normal and the inverse of its variance is equal to

$$\left\langle E^{\theta_0}\left(\bigtriangledown_\theta\tilde{h}_t\right),E^{\theta_0}\left(\bigtriangledown_\theta\tilde{h}_t\right)\right\rangle_{\tilde{K}}=\frac{1}{(1+\zeta)}\left\langle E^{\theta_0}\left(\bigtriangledown_\theta\tilde{h}_t\right),E^{\theta_0}\left(\bigtriangledown_\theta\tilde{h}_t\right)\right\rangle_{K}.$$

Now, we compute $E^{\theta_0}\left(\bigtriangledown_\theta\tilde{h}_t\right)$. By Assumption A.11, we have:

$$\begin{aligned}E^{\theta_0}\left(\bigtriangledown_\theta\tilde{h}_t\right) &= -E^{\theta_0}\left(\bigtriangledown_\theta e^{i\tau\widetilde{Y}_j}\right)\\ &= -\bigtriangledown_\theta E^{\theta_0}\left(e^{i\tau\widetilde{Y}_j}\right)\\ &= -\bigtriangledown_\theta \psi_{\theta}^{L}(\tau)\\ &= E^{\theta_0}\left(\bigtriangledown_\theta h\right).\end{aligned}$$

Table 1 : Monte Carlo Comparison of Estimation Methods based on the CIR model of interest rates

We report three measures of estimation method performance – Mean Bias, Median Bias, and Root Mean Squared Error (RMSE) – for five different estimation methods: CF-GMM with the optimal **DI** instrument, CF-GMM without an instrument (unity instrument), MLE, QMLE, and EMM (the results for the former three methods are taken from Zhou, 2001). The simulations are performed based on the CIR model:

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

with two sets of parameter values. Panel A uses the estimated values from Gallant and Tauchen (1998). Panel B uses the Zhou (2001) parameters, which represent a particularly challenging for estimation case. All results are based on 1000 replications of samples with 500 and 1500 observations.

Panel A

	Mean Bias		Median Bias		RMSE	
True Value	$T = 500$	$T = 1500$	$T = 500$	$T = 1500$	$T = 500$	$T = 1500$
	CF-GMM with optimal instrument					
$\theta = 0.02491$	0.0012	0.0011	0.0009	0.0009	0.0101	0.0100
$\kappa = 0.00285$	0.0000	0.0001	0.0002	0.0001	0.0010	0.0008
$\sigma = 0.02750$	-0.0005	0.0004	-0.0006	-0.0005	0.0079	0.0077
	CF-GMM without an instrument					
$\theta = 0.02491$	0.0009	-0.0014	-0.0005	-0.0018	0.0192	0.0210
$\kappa = 0.00285$	0.0000	0.0003	0.0002	0.0003	0.0017	0.0020
$\sigma = 0.02750$	-0.0016	-0.0021	-0.0017	-0.0019	0.0085	0.0084
	MLE					
$\theta = 0.02491$	-0.0123	-0.0130	-0.0119	0.0215	0.0125	0.0131
$\kappa = 0.00285$	-0.0014	0.0015	-0.0014	0.0014	0.0014	0.0015
$\sigma = 0.02750$	-4.4e-5	2.5e-6	-4.6e-5	2.1e-5	0.0009	0.0005
	QMLE					
$\theta = 0.02491$	0.0994	0.285	0.0803	0.0209	0.1343	0.0437
$\kappa = 0.00285$	-0.0113	-0.0033	-0.0091	-0.0025	0.0153	0.0050
$\sigma = 0.02750$	3.0e-5	1.2e-5	4.1e-5	1.9e-5	0.0009	0.0005
	EMM					
$\theta = 0.02491$	0.0451	0.0407	2.6e-4	0.0085	0.1252	0.0944
$\kappa = 0.00285$	-0.0054	-0.0048	-8.1e-5	-0.0012	0.0149	0.0112
$\sigma = 0.02750$	-0.0015	-0.0003	-4.8e-6	-4.3e-7	0.0076	0.0041

Table 1 (continued)

Panel B						
	Mean Bias		Median Bias		RMSE	
True Value	$T = 500$	$T = 1500$	$T = 500$	$T = 1500$	$T = 500$	$T = 1500$
			CF-GMM with optimal instrument			
$\theta = 2.491$	-0.0928	-0.0788	-0.0813	-0.0689	0.2106	0.1014
$\kappa = 0.285$	0.0085	0.0066	0.0071	0.0013	0.0311	0.0188
$\sigma = 1.1$	0.0026	-0.0088	0.0065	0.0001	0.0448	0.0261
			CF-GMM without an instrument			
$\theta = 2.491$	-0.1028	-0.0988	-0.0991	-0.0910	0.4008	0.1188
$\kappa = 0.285$	0.0111	0.0071	0.0063	0.0010	0.0507	0.0202
$\sigma = 1.1$	0.0098	0.0001	0.0059	0.0009	0.0488	0.0233
			MLE			
$\theta = 2.491$	-0.0832	-0.0663	-0.0679	-0.0524	0.1337	0.0923
$\kappa = 0.285$	0.0085	0.0058	0.0029	0.0010	0.0251	0.0161
$\sigma = 1.1$	0.0024	-0.0016	0.0060	0.0000	0.0432	0.0263
			QMLE			
$\theta = 2.491$	0.0742	0.0224	0.0022	0.0006	0.3613	0.0923
$\kappa = 0.285$	-0.0100	-0.0020	-0.0071	-0.0011	0.0448	0.0258
$\sigma = 1.1$	0.0023	0.0003	0.0015	-0.0001	0.0430	0.0246
			EMM			
$\theta = 2.491$	0.1323	-0.0067	0.0433	-0.0173	0.4891	0.2000
$\kappa = 0.285$	-0.0310	-0.0022	-0.0199	-0.0000	0.0694	0.0257
$\sigma = 1.1$	-0.0218	-0.0137	-0.0091	-0.0122	0.0618	0.0296

Table 2 : Monte Carlo Study of a Jump-Diffusion Model

This table focuses on the performance of CF-GMM in estimation of jump-diffusion models. The results are based on the Vasicek model augmented by the exponential jump component:

$$\begin{aligned} dr_t &= (\theta - \kappa r_t)dt + \sigma dW_t + J_t dN_t \\ J_t &\sim EXP(\alpha) \\ N_t &\sim POI(\lambda) \end{aligned}$$

Parameter values for the diffusion part are taken from Gallant and Tauchen (1998). All results are based on 1000 replications of samples with 500 and 1500 observations.

	Mean Bias		Median Bias		RMSE	
True Value	$T = 500$	$T = 1500$	$T = 500$	$T = 1500$	$T = 500$	$T = 1500$
			CF-GMM with optimal instrument			
$\theta = 0.02949$	0.0000	0.0000	0.0001	-0.0001	0.0008	0.0008
$\kappa = 0.00283$	-0.0006	0.0007	-0.0022	-0.0011	0.0122	0.0124
$\sigma = 0.09802$	0.0026	-0.0001	-0.0002	-0.0007	0.0283	0.0286
$\alpha = 0.00500$	0.0002	0.0000	0.0004	0.0000	0.0014	0.0014
$\lambda = 0.09615$	0.0018	0.0004	0.0047	0.0012	0.0289	0.0281
			CF-GMM without an instrument			
$\theta = 0.02949$	-0.4389	-0.3459	-0.0941	-0.0609	1.6010	4.2000
$\kappa = 0.00283$	7.3483	-23.0590	0.6440	0.4871	76.4663	1582.0000
$\sigma = 0.09802$	0.7026	0.6884	0.5781	0.6469	0.7683	0.8000
$\alpha = 0.00500$	-0.0395	-0.0550	-0.0083	-0.0077	0.5248	0.6804
$\lambda = 0.09615$	-0.0216	-0.0418	-0.0014	-0.0015	1.0657	1.0006

Table 3 : Empirical Estimation of Three-factor Affine Term Structure Models

This table evaluates the properties of CF-GMM in the estimation of multivariate diffusions. We use the Duffee (2002) preferred specification with corresponding parameter values:

$$\begin{pmatrix} dX_{t,1} \\ dX_{t,2} \\ dX_{t,3} \end{pmatrix} = \left[\begin{pmatrix} 0 \\ (K\theta)_2 \\ (K\theta)_3 \end{pmatrix} + \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \begin{pmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{pmatrix} \right] dt + S_t dW_t$$

where S_t is a diagonal matrix with elements $S_{t(ii)} = \sqrt{\alpha_i + (\beta_{i1} \quad \beta_{i2} \quad \beta_{i3}) X_t}$ for $i = 1, 2$ and 3 and all elements of the diagonal matrices S_t zero, except $\alpha_3 = \beta_{11} = \beta_{22} = \beta_{32} = 1$. All results are based on 1000 replications of samples with 500 and 1500 observations.

	Mean Bias		Median Bias		RMSE	
True Value	$T = 500$	$T = 1500$	$T = 500$	$T = 1500$	$T = 500$	$T = 1500$
			CF-GMM with optimal instrument			
$K\theta_2 = 0.222$	0.0009	0.0002	0.0010	0.0004	0.0415	0.0091
$K\theta_3 = -2.299$	0.0003	-0.0001	0.0005	0.0000	0.0167	0.0056
$k_{11} = 0.172$	0.0011	0.0000	0.0008	0.0001	0.0189	0.0109
$k_{12} = -0.295$	0.0014	0.0004	0.0011	0.0004	0.0216	0.0089
$k_{21} = -0.197$	0.0006	-0.0003	-0.0001	-0.0001	0.0196	0.0045
$k_{22} = 0.406$	0.0013	0.0007	0.0009	0.0002	0.0219	0.0111
$k_{31} = 0.564$	-0.0001	0.0001	0.0001	-0.0002	0.0089	0.0023
$k_{32} = -1.669$	0.0004	0.0002	0.0003	0.0000	0.0077	0.0012
$k_{33} = 1.721$	0.0007	-0.0001	0.0006	0.0002	0.0163	0.0071
			CF-GMM without an instrument			
$K\theta_2 = 0.222$	0.0931	0.0337	0.0595	0.0541	0.1740	0.0682
$K\theta_3 = -2.299$	0.1511	0.1065	0.1167	0.1072	0.1471	0.0416
$k_{11} = 0.172$	1.8919	0.1065	1.8218	1.6966	0.7074	0.3890
$k_{12} = -0.295$	1.1900	1.0555	1.1577	1.1006	0.3630	0.1723
$k_{21} = -0.197$	0.6225	0.7586	0.7574	0.7879	0.4299	0.0918
$k_{22} = 0.406$	1.4090	1.2634	1.3885	1.3043	0.4120	0.2323
$k_{31} = 0.564$	-0.0886	0.0299	0.0004	0.0268	0.3640	0.0385
$k_{32} = -1.669$	0.6022	0.5368	0.6018	0.5516	0.1822	0.1163
$k_{33} = 1.721$	0.6394	0.5306	0.5808	0.5615	0.3746	0.1402

Figure 1. Plot of real parts of integrands for computing the MLE and C-GMM estimators

We illustrate the degree of the numerical effort involved in computing the integrals necessary for the MLE estimation based on the Fourier inverse technique described in Singleton (2001) and the C-GMM estimation. The integrand is computed for the CIR model, studied in Section 6.1:

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

and evaluated at the point $(r_{t+1}, r_t) = (2\theta, \theta) = (0.04982, 0.02491)$. CGMM1 (CGMM2) denotes the integrand I_1 in (C.5) (I_2 in (C.6)).