The Indirect Continuum-GMM Estimation

Rachidi Kotchoni*
CREA - Université Laval

March 2, 2011

Abstract

A “curse of dimensionality” arises when using the generalized method of moment based on a continuum of moments conditions to estimate a multivariate model with three or more dimensions. The solution proposed consists of turning the multivariate model into a continuum of univariate auxiliary models. The ideal solution of the multivariate model is then estimated indirectly as a weighted average of the solutions to the univariate models. The optimal weighting function is derived and the relative efficiency of the optimal indirect estimator vis-à-vis the maximum likelihood estimator is discussed. Two simulations studies and an empirical application illustrate the effectiveness of the new estimator.

Keywords: Autoregressive Gamma, Bootstrap, Continuum of Moments Conditions, Covariance Operator, Indirect Estimator, Realized Volatility.

JEL Classification: C01, C13, C15, C60.

*Correspondance: Pavillon Paul-Comtois; 2425, rue de l’Agriculture, Local#4424-F; Québec (QC) G1V 0A6; Phone: 418 656 2131 ext. 3940; Fax: 418 656 7821; E-mail: rachidi.kotchoni@fsaa.ulaval.ca.
1 Introduction

When implementing the generalized method of moments based on a continuum of moments conditions (henceforth, CGMM) proposed by Carrasco and Florens (2000), the objective function to minimize takes the form of multiple integrals with respect to continuous variables. In solving this minimization problems, a major difficulty lies in the evaluation of the multiple integrals embedded in the objective function. Indeed, if 10 quadrature points are needed to achieve a given level of precision for a one-dimensional integration, about $10^d$ quadrature points are required to obtain the same level of precision in evaluating the CGMM objective function, where $d$ is the multiplicity of the integrals. This raises a “curse of dimensionality”. To circumvent this problem, a solution may consist of discarding quadrature points that have very low weights, or reducing the number of quadrature points. Unfortunately, none of these solutions provide a substantial numerical efficiency gain without impeding the accuracy of the estimation procedure. We propose a solution that consists of converting the multivariate model of interest into a continuum of univariate auxiliary models that are easy to estimate. The solutions to the univariate models are then optimally combined to recover the ideal solutions of the unfeasible multivariate model.

Although the CGMM has a general scope in econometric theory, it has been mostly used to deal with moments conditions that are based on the characteristic function (henceforth CF). There is a one-to-one link between the CF and the density of a random variable, the former being a Fourier transform of the latter. Hence one might expect an inference method that adequately exploits the continuum of moment conditions given by the difference between the theoretical CF and its empirical counterpart to be as efficient as an alternative likelihood-based approach. The CGMM may be implemented to estimate the parameters of a distribution that admits a closed form expression for its CF, but for which the likelihood function is not known (e.g. fat-tailed stable distributions; discrete samples from a diffusion process). It may also be used when the density of the random variable of interest takes the unfriendly form of an infinite sum or an integral over an unbounded set (e.g. variance gamma models, discrete sample from a square-root process).

The CGMM builds on the same philosophy as the GMM of Hansen (1982). In particular, both are based on the minimization of a quadratic form associated with some scalar product. But the scalar product of the GMM is defined on a finite dimensional vector space while that of the CGMM is defined on an infinite dimensional Hilbert space. The canonical scalar product of two finite dimensional vectors $u$ and $v$ is the sum of the products of their corresponding coordinates, and the quadratic form associated with this scalar product is the Euclidian norm. In a complex Hilbert space, an example of scalar product between two functions $h(\tau)$ and $g(\tau)$ is given by the integral of the function $h(\tau)\overline{g(\tau)}$ against a finite measure $\pi(\tau)d\tau$, where $\overline{g(\tau)}$ is the complex conjugate of $g(\tau)$. The norm of $h(\tau)$ associated with this scalar product is given by the integral of $|h(\tau)|^2 \equiv h(\tau)\overline{h(\tau)}$ against $\pi(\tau)d\tau$. Hence the multiplicity of the integral is determined by the dimensionality of $\tau$. In the context of the CGMM based on the CF, $h(\tau) \equiv h(\tau, \theta_0)$ is a moment function obtained by taking the difference between the theoretical and the empirical CF of a multivariate random variable $x \in \mathbb{R}^d$. The variable $\tau$ is the Fourier transformation index and $\theta_0$ is a finite dimensional parameter that fully characterized the distribution of $x$. The objective function of the CGMM is the norm of $h(\tau, \theta_0)$ and the dimensionality of $\tau$ is equal to the number of coordinates of $x$. If the data are generated by a Markov process of order $p$, then a function of $p$ lagged observations may be used as instruments in the moment function, which would increase the dimensionality of $\tau$ from $d$ to $d(p + 1)$. In this case, if 10 quadrature points are needed to achieve a certain level
of precision in one dimension, about $10^{d(p+1)}$ quadrature points are required to obtain the same level of precision in evaluating the CGMM objective function. Hence the CGMM becomes quickly unfeasible for values of $d$ as low as 3, particularly when a large number of iterations is required for the convergence of the optimization algorithm.

To avoid the curse of dimensionality in the CGMM, one draws a vector $\tau$ from a bounded subset of $\mathbb{R}^{d(p+1)}$, and define the set of all moment functions along the dimension $\tau$ as $h_{\tau,t}(u, \theta_0) \equiv h_t(u, \theta_0)$, $u \in \mathbb{R}$. For a given $\tau$, an estimator can be computed by minimizing the norm of $h_{\tau,t}(u, \theta_0)$ which is a function of a one dimensional index $u$. The norm of $h_{\tau,t}(u, \theta_0)$ may be defined as the integral of $h_{\tau,t}(u, \theta_0)\bar{h}_{\tau,t}(u, \theta_0)$ against the measure $\omega (u) du$. The minimization of this norm delivers an estimator $\hat{\theta}(\tau)$ of $\theta_0$ which is obviously a function of $\tau$. This estimator is consistent for $\theta_0$ under some regularity conditions that are discussed subsequently. An estimator $\hat{\theta}_\pi$ that does not depend on $\tau$ is obtained by integrating $\hat{\theta}(\tau)$ against an arbitrary measure $\pi(\tau) d\tau$ that sums to one. Empirically, the estimator $\hat{\theta}_\pi$ may be approximated by $\frac{1}{S} \sum_{s=1}^{S} \hat{\theta}(\tau_s)$, where $\tau_s, s = 1, ..., S$ are independent draws from the measure $\pi(\tau)$. Note that the solution proposed above may be viewed as a resampling technique, as it consists of aggregating the estimators $\hat{\theta}(\tau)$ obtained from the collections of samples $\{y_t = \tau x, \tau \in \mathbb{R}^d\}$ generated from the frequency domain of the distribution of $x$. It also has the flavor of the indirect inference of Gourieroux, Monfort and Renault (1993) or the Efficient Method of Moment of Gallant and Tauchen (1996), as it consists of converting a high dimensional model into a continuum of auxiliary univariate models. We shall thus refer to $\hat{\theta}_\pi$ as the indirect CGMM estimator (henceforth ICGMM).

Three major theoretical issues are addressed: the identifiability of $\theta_0$ from the objective function based on the reduced information set $h_{\tau,t}(u, \theta_0)$, $u \in \mathbb{R}$, the design of the optimal aggregating weight $\pi^*(\tau)$, and the relative efficiency of $\hat{\theta}_\pi^*$ with respect to the maximum likelihood estimator (MLE). It appears that the optimal weighting scheme is closely related to the inverse of the covariance operator associated with $\hat{\theta}(\tau)$ viewed as a function of $\tau$. Also, a sufficient condition for $\hat{\theta}_\pi^*$ to be as optimal as the unfeasible MLE is that the matrix $\frac{\partial h(\tau)}{\partial \tau}$ be of full rank. If this rank condition is not satisfied, then $\hat{\theta}_\pi^*$ may still be as efficient as the maximum likelihood estimator, but we may not be able to prove this. Intuitively, the rank condition will be satisfied if the manifold $\left\{ \hat{\theta}(\tau), \tau \in \mathbb{R}^d \right\}$ is rich enough to encompassed the MLE.

An approach similar to the one presented above is used in Chen, Jacho-Chavez and Linton (2009) in the context of instrumental variable estimation. These authors face a set of conditional moment restrictions of type $E [\rho (Z_t, \theta_0) | X_t] = 0$, for some scalar function $\rho (Z_t, \theta_0)$. The standard approach in this literature consists of turning these conditional moment restrictions into unconditional ones by using $E [\rho (Z_t, \theta_0) A (X_t)] = 0$, for any vector function $A (X_t)$. One then estimates the optimal instrumental function $A_{oiv} (X_t)$, and the GMM estimator $\hat{\theta}_{oiv}$ based on the unconditional restriction $E [\rho (Z_t, \theta_0) A_{oiv} (X_t)] = 0$ is called the optimal instrumental variable estimator. Chen, Jacho-Chavez and Linton (2009) proposed the alternative estimator $\hat{\theta}_w = \sum_{j=1}^{N} w_j \hat{\theta}_j$, where $N$ is allowed to increase with the sample size, $\hat{\theta}_j$ is the GMM estimator based on the moment restrictions $E [\rho (Z_t, \theta_0) A_j (X_t)] = 0$, and $\{A_j (X_t)\}_{j=1}^{\infty}$ are basis functions chosen by the econometrician. Chen et al. show that $\hat{\theta}_w$ is as efficient as $\hat{\theta}_{oiv}$ for optimally designed $N$ and $w = \{w_j\}_{j=1}^{N}$.

The rest of the paper is organized as follows. In the next section, we present the general framework and introduce some notation. In Section 3, we discuss the properties of classical CGMM estimators. In Section 4, we derive the theoretically optimal aggregating weight $\pi^*(\tau)$ for the
ICGMM estimator. In particular, we compare the performance of the ideal ICGMM estimator \( \hat{\theta}_{\pi^*} \) to that of the MLE. In Section 5, we present the feasible ICGMM estimator and show its asymptotic equivalence with its ideal theoretical counterpart. The feasible ICGMM estimator is the one obtained by plugging an estimator \( \pi^*(\tau) \) into the expression of \( \hat{\theta}_{\pi^*} \). Section 6 presents a Monte Carlo simulation study based on a Gaussian AR(1) model. The MLE is feasible for this case and thus, can be used as benchmark to assess the performance of the ICGMM estimator. Section 7 presents a Monte Carlo study based on the Autoregressive Factor Gamma Model. Section 8 presents an empirical application based on the Autoregressive Variance Gamma model of order \( p \) specified for the joint dynamic of the daily return on Alcoa and its realized variance. Section 9 concludes the paper and the proofs are left in appendix.

2 The General Framework

This section introduces the notation and the definition of the ICGMM estimator. The assumptions underlying the ICGMM procedure are also discussed and illustrated.

2.1 The ICGMM estimator

Let \( x_t \in \mathbb{R}^n \) be an IID random variable, and assume that the distribution of \( x_t \) is fully characterized by a finite dimensional parameter \( \theta_0 \in \mathbb{R}^q \). Let us consider the function \( h_t(\tau, \theta) \) given by:

\[
h_t(\tau, \theta) = \exp(i\tau'x_t) - \varphi(\tau, \theta), \tau \in \mathbb{R}^n,
\]

where \( \varphi(\tau, \theta) = E^\theta[\exp(i\tau'x_t)] \) and \( E^\theta \) is the expectation operator with respect to the data generating process indexed by \( \theta \). Because \( E^{\theta_0}[h_t(\tau, \theta_0)] = 0 \) for all \( \tau \in \mathbb{R}^n \), the function \( h_t(\tau, \theta) \) defines a continuum of valid moment conditions that can be used to estimate \( \theta_0 \) from an observed sample. Let \( \pi(\tau) \) be a probability density function on \( \mathbb{R}^n \) and \( L^2(\pi) \) denote the Hilbert space of complex valued functions that are square integrable with respect to \( \pi \), that is:

\[
L^2(\pi) = \{ f : \mathbb{R}^n \rightarrow \mathbb{C} \text{ such that } \int f(\tau)\overline{f(\tau)}\pi(\tau)d\tau < \infty \}.
\]

A scalar product \( \langle ., . \rangle \) on \( L^2(\pi) \times L^2(\pi) \) is given by:

\[
\langle f, g \rangle = \int f(\tau)\overline{g(\tau)}\pi(\tau)d\tau
\]

where \( \overline{z} \) is the complex conjugate of \( z \) and \( E[\cdot] \) is the expectation with respect to the density \( \pi(\tau) \). It can easily be checked that the moment function \( h_t(\tau, \theta_0) \) is bounded in modulus and hence, belongs to \( L^2(\pi) \) for any finite measure \( \pi \). Taking advantage of this, Carrasco and Florens (2000) defined the objective function of the CGMM by mean of the quadratic form associated with the scalar product above:

\[
Q_T(\theta) = \left\langle K^{-1/2}\hat{h}_T(\cdot, \theta), K^{-1/2}\hat{h}_T(\cdot, \theta) \right\rangle
\]
where \( h_t(\tau, \theta) = \frac{1}{T} \sum_{t=1}^{T} h_t(\tau, \theta) \) and \( K \) is the covariance operator \( K \) associated with the moment function. The CGMM estimator is defined as the particular value of \( \theta \) that minimizes \( Q_T(\theta) \).

When \( x_t \) is not IID but instead Markov of order \( p \), the relevant moment function is defined as:

\[
h_t(\tau, \theta) = \exp \left( i \tau' x_{t+1} - \varphi_t(\tau_1, \theta) \right) \exp \left( \sum_{k=1}^{p} i \tau'_{k+1} x_{t+1-k} \right), \tag{5}
\]

where \( \tau = (\tau_1, ..., \tau_{p+1}) \in \mathbb{R}^{n(p+1)} \) and \( \varphi_t(\tau_1, \theta) = E^\theta \left[ \exp \left( i \tau' x_{t+1} \right) | \{ x_{t+1-k} \}_{k=1}^{p} \right] \) is the CF of \( x_{t+1} \) conditional on \( p \) lags. Accordingly, we would define the scalar product \( \langle ., . \rangle \) on \( L^2(\mathbb{R}) \times L^2(\pi) \) similarly as above but now using a probability measure \( \pi(\tau) \) on \( \mathbb{R}^{n(p+1)} \).

When \( x_t \) is dependent so that its distribution depends on its entire past, a suggestion is to use a moment function based on the joint CF:

\[
h_t(\tau, \theta) = e^{i \tau' Y_t} - E^\theta(e^{i \tau' Y_t}), \tau \in \mathbb{R}^{np}, \tag{6}
\]

where \( Y_t = (x_t, x_{t-1}, ..., x_{t-p+1}) \). In theory, the larger \( p \) the more efficient the CGMM estimator. But in practice, the quest for efficiency must be balanced with the computing cost. In particular, the curse of dimensionality described in the introduction quickly emerges as \( p \) increases. For more discussions on the use of the moment function (6), see Jiang and Knight (2002), Yu (2004) and Carrasco, Chernov, Florens and Ghysels (2007).

Henceforth, we will use the generic notation \( h_t(\tau, \theta), \tau \in \Lambda \) to denote either of the moment functions above, and we let \( d \) denote the dimensionality of \( \tau \). To implement the dimensionality reduction technique outlined in the introduction, let us consider the normalized set \( \Lambda \) given by:

\[
\Lambda = \{ \tau \in \mathbb{R}^{n(p+1)} : ||\tau||_E = 1 \}, \tag{7}
\]

where \( ||\tau||_E \) is the Euclidian norm. From this point on, we let \( \pi(\tau) \) denote a density on \( \Lambda \). For a given draw \( \tau \) from the measure \( \pi() \), we define the set of all moment functions along the dimension \( \tau \) as:

\[
h_{\tau,t}(u, \theta) \equiv h_t(u\tau, \theta), \ u \in \mathbb{R}, \tag{8}
\]

For fixed \( \tau, h_{\tau,t}(u, \theta) \) may be viewed as a function of \( u \) alone, that is, a univariate mapping from \( \mathbb{R} \) to \( \mathbb{C} \). Under certain regularity conditions discussed below, a consistent CGMM estimator is given by:

\[
\hat{\theta}(\tau) = \arg \min_\theta \left\{ Q_{\tau,T}(\theta) = \int h_{\tau,T}(u, \theta) \bar{h}_{\tau,T}(u, \theta) \omega(u) \, du \right\}, \tag{9}
\]

where \( \omega(u) \) is a weighting function on \( \mathbb{R} \) and \( \bar{h}_{\tau,T}(u, \theta) = \frac{1}{T} \sum_{t=1}^{T} h_{\tau,t}(u, \theta) \). To make the overall estimation procedure independent of the particular draw \( \tau \), we define the final estimator as the average:

\[
\hat{\theta}_\pi = \int \hat{\theta}(\tau) \pi(\tau) \, d\tau. \tag{10}
\]

where \( \pi(\tau) \) is a density on \( \Lambda \).

The estimator \( \hat{\theta}(\tau) \) defined in (9) may be used as a first step estimator to build a more efficient
second step CGMM estimator. The second step estimator indexed by $\tau$ is defined as:

$$\hat{\theta}^{(2)} (\tau) = \arg\min_\theta Q^{(2)}_{\tau,T}(\theta),$$

where

$$Q^{(2)}_{\tau,T}(\theta) = \int K_{\tau}^{-1/2} h_{\tau,T}(u, \theta) K_{\tau}^{-1/2} \overline{h}_{\tau,T}(u, \theta) \omega(u) \, du,$$

In IID and Markov models, $K_\tau$ is the linear operator with kernel

$$k_\tau(u_1, u_2) = E^{\theta_0} \left[ h_{\tau,t}(u_1, \theta) \overline{h}_{\tau,t}(u_2, \theta) \right],$$

and $K_\tau f(u_1) = \int k_\tau(u_1, u_2) f(u_2) \, du_2$ for all $f : \mathbb{R} \to \mathbb{R}$. In dependent models, the expression of the kernel $k_\tau(u_1, u_2)$ is given by:

$$k_\tau(u_1, u_2) = E^{\theta_0} \left[ h_{\tau,t}(u_1, \theta) \overline{h}_{\tau,t}(u_2, \theta) \right] \quad + \quad \sum_{j=1}^\infty E^{\theta_0} \left[ h_{\tau,t}(u_1, \theta) \left( \overline{h}_{\tau,t-j}(u_2, \theta) + \overline{h}_{\tau,t+j}(u_2, \theta) \right) \right].$$

Other details on this kernel are given in Carrasco, Chernov, Florens and Ghysels (2007). In practice, $\hat{\theta}(\tau)$ will be used to estimate the covariance operator associated with the moments function $K_\tau$ before implementing the second step estimator $\hat{\theta}^{(2)}(\tau)$.

### 2.2 The Assumptions

The following assumptions are posited.

**Assumption 1:** The pdf $\omega(\cdot)$ is strictly positive on $\mathbb{R}$ and has finite moments at any order.

**Assumption 2:** For all $\tau \in \Lambda \setminus \emptyset$, the equation

$$E^{\theta_0} [h_{\tau,t}(u, \theta)] = 0 \text{ for all } u \in \mathbb{R}, \omega - \text{almost everywhere},$$

has a unique solution $\theta_0$ which is an interior point of a compact set $\Theta$, where $\emptyset$ is a null set with respect to $\pi$, $E^{\theta_0}$ denotes the expectation with respect to the distribution of the data at $\theta = \theta_0$.

**Assumption 3:** For all $\tau \in \Lambda \setminus \emptyset$, $h_{\tau,t}(u, \theta)$ is three times continuously differentiable with respect to $\theta$.

**Assumption 4:** For all $\theta$ and $\tau \in \Lambda \setminus \emptyset$, $E^{\theta_0} [h_{\tau,T}(\cdot, \theta)]$ and its first three derivatives with respect to $\theta$ belong to the range of $K_\tau^\beta$ for $\beta \geq 1/2$, where $K_\tau$ is the covariance operator associated with the moment function $h_{\tau,t}(\cdot, \theta)$.

**Assumption 5:** $h_{\tau,t}(u, \theta)$ is at least twice continuously differentiable with respect to $\tau$ in $\Lambda \setminus \emptyset$.

**Assumption 6:** (i) $\frac{\partial^2 Q_{\tau,T}}{\partial \theta_0 \partial \theta}$ is positive definite and (ii) $\frac{\partial^2 Q_{\tau,T}}{\partial \theta_0 \partial \tau}$ is of full rank in $\Lambda \setminus \emptyset$.

**Assumption 7:** The measure $\pi(\cdot) \tau$ on $\Lambda$ satisfies: $\int \pi(\tau) \, d\tau = 1$.

**Assumption 8:** The random variable $x_t$ is stationary and satisfies $x_t = x(\theta_0, \varepsilon_t, Z_{t-1})$ where $x(\cdot, \varepsilon_t, Z_{t-1})$ is three times continuously differentiable with respect to $\theta$, $\varepsilon_t$ is a IID white noise whose distribution does not depend on $\theta_0$, and $Z_{t-1}$ can only contain lagged values of $x_t$.  

6
The first assumption ensures that \( 0 < E_{\omega(u)} \left[ f(u)f(u) \right] < \infty \) for all \( f \neq 0 \). Assumption 2 stipulates that the set of all \( \tau \) such that \( \theta_0 \) is not identifiable from the moment function \( h_{\tau,t}(u, \theta) \) represents a null set \( \mathcal{N} \) with respect to the continuous measure \( \pi \) on \( \Lambda \). For instance, if \( d = 2 \) and \( x_t = (x_{1,t}, x_{2,t})' \), choosing \( \tau = (\tau_1, 0) \) amounts to rely on the marginal distribution of \( x_{1,t} \) for the estimation of \( \theta_0 \). In this case, the parameters that characterize the dependence between \( x_{1,t} \) and \( x_{2,t} \) cannot be identified from the distribution of \( y_{\tau,t} = \tau_1 x_{1,t} \). But the set of all \( \tau = (\tau_1, \tau_2) \) such that \( \tau_1 = 0 \) or \( \tau_2 = 0 \) is a null set with respect to any continuous measure on \( \mathbb{R}^2 \). Hence Assumption 2 is satisfied by setting \( \pi \) equal to an arbitrary continuous probability distribution function on \( \mathbb{R}^2 \).

The CGMM estimator can be derived under weaker conditions than in Assumption 3, but the derivation of some of the asymptotic properties may become difficult if this assumption is not satisfied. Assumption 4 ensures that the limit of the objective function is well defined as \( T \) goes to infinity. Assumptions 5 and 6 ensure that \( \hat{\theta}(\tau) \) is unique and is a smooth function of \( \tau \). Assumption 2 already ensures the positive definiteness of \( \frac{\partial^2 Q_{c,t}}{\partial \theta \partial \theta'} \) as \( T \) goes to infinity, but we request this to be satisfied in finite sample for simplicity. The measure \( \pi (\tau) \) in Assumption 7 need not be positive for all \( \tau \). Finally, Assumption 8 is used in Carrasco and Kotchoni (2009) to select a regularization parameter that enters in the expression of the feasible optimal second step CGMM estimator \( \hat{\theta}^{(2)}(\tau) \). Actually, this assumption is not necessary for the derivation of the good properties of the ICGMM estimator.

### 2.3 Scope of the Identification Assumption

Assumption 2 is crucial for the consistency of the ICGMM estimator. In this section, we examine the extent by which the requirement of this assumption can be easily met in cases of practical interest. For illustration purposes, let us assume that \( d = 2 \) and consider the bivariate normal process \( x_t = (x_{1,t}, x_{2,t})' \):

\[
x_t \sim N \left( \left[ \begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right], \left[ \begin{array}{cc} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{array} \right] \right)
\]  

(14)

Let \( \theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) \) be the vector of parameters. Note that \((\mu_1, \sigma_1^2)\) can be efficiently estimated from the marginal distribution of \( x_{i,t}, i = 1, 2 \). We have:

\[
\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} x_{i,t} \quad \text{and} \quad \hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^{T} (x_{i,t} - \hat{\mu}_i)^2, i = 1, 2
\]

(15)

Given this, one may focus exclusively on the estimation of the parameter \( \rho \) that governs the dependence between \( x_{1,t} \) and \( x_{2,t} \), which amounts to consider the concentrated likelihood deduced from the distribution:

\[
z_t \sim N \left( \left[ \begin{array}{c} 0 \\ \rho \end{array} \right], \left[ \begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right] \right)
\]

(16)

where \( z_{i,t} = \frac{x_{i,t} - \hat{\mu}_i}{\sqrt{\hat{\sigma}_i^2}}, i = 1, 2 \). It can be showed that the MLE of \( \rho \) based on (16) solves the fixed-point relation:

\[
\hat{\rho} = \frac{\hat{\rho}^3 - \hat{\rho}^2 \frac{1}{T} \sum z_{1,t} z_{2,t} - \frac{1}{T} \sum z_{1,t} z_{2,t}}{1 - \frac{1}{T} \sum z_{1,t}^2 - \frac{1}{T} \sum z_{2,t}^2}
\]

(17)
Now, let \( \tau = (\tau_1, \tau_2) \) and \( y_{r,t} = \tau_1 z_{1,t} + \tau_2 z_{2,t} \). The concentrated likelihood of \( y_{r,t} \) may be deduced from the distribution:
\[
y_{r,t} = \tau_1 z_{1,t} + \tau_2 z_{2,t} \sim N \left( 0, \tau_1^2 + 2\rho \tau_1 \tau_2 + \tau_2^2 \right)
\]  
(18)

It can be showed that the MLE of \( \rho \) based on (18) is given by:
\[
\hat{\rho}(\tau) = \frac{1}{2\tau_1 \tau_2} \left[ \frac{1}{T} \sum y_{r,t}^2 - \tau_1^2 - \tau_2^2 \right]
\]
(19)

This shows that \( \rho \) is identifiable from the reduced information set consisting of \((i)\) the knowledge of the distribution of \( y_{r,t} = \tau_1 z_{1,t} + \tau_2 z_{2,t} \), and \((ii)\) the knowledge of “weights” \( \tau_1 \) and \( \tau_2 \). This reduced information set is strictly included in the joint distribution of \((z_{1,t}, z_{2,t})\), but it is larger than the sole knowledge of the marginal distribution of \( y_{r,t} \). Precisely, this justifies the presence of \( \tau_1 \) and \( \tau_2 \) in the expression of \( \hat{\rho}(\tau) \), which is a disaggregated information compared to \( y_{r,t} \).

If \( x_{1,t} \) is a stationary time series and \( x_{2,t} = x_{1,t-1} \), then we have \( \mu_1 = \mu_2 \equiv \mu \) and \( \sigma_1 = \sigma_2 \equiv \sigma \). In this case, one may define \( y_{r,t} \) as a linear combination of the elements of \( x_t \), that is:
\[
y_{r,t} = \tau_1 x_{1,t} + \tau_2 x_{2,t} \sim N \left( \mu (\tau_1 + \tau_2), \sigma^2 (\tau_1^2 + 2\rho \tau_1 \tau_2 + \tau_2^2) \right),
\]
(20)

The MLE of \( \theta = (\mu, \sigma^2, \rho) \) based on (20) is:
\[
\hat{\mu}(\tau) = \frac{1}{T(\tau_1 + \tau_2)} \sum y_{r,t},
\]
(21)
\[
\hat{\sigma}^2(\tau) = \frac{1}{T(\tau_1^2 + 2\hat{\rho}(\tau) \tau_1 \tau_2 + \tau_2^2)} \sum (y_{r,t} - (\tau_1 + \tau_2) \hat{\mu})^2 \quad \text{and}
\]
(22)
\[
\hat{\rho}(\tau) = \frac{1}{2\tau_1 \tau_2} \left[ \frac{1}{T} \sum \left( \frac{y_{r,t} - \hat{\mu}(\tau)}{\hat{\sigma}(\tau)} \right)^2 - \tau_1^2 - \tau_2^2 \right],
\]
(23)

Note that the system above requires to solve for \( \hat{\sigma}^2(\tau) \) and \( \hat{\rho}(\tau) \) simultaneously. The whole vector of parameters \( \theta \) is identifiable from the reduced information set (20) because all of its elements are common to the marginal distributions \( x_{i,t}, i = 1, 2 \). This shows that Assumption 2 is always satisfied in stationary time series models. Obviously, the conclusions drawn above for the joint distribution of \((x_{1,t}, x_{2,t})\) also apply to the distribution of \( x_{1,t} \) conditional on \( x_{2,t} \).

To generalize the results to an arbitrary multivariate distribution, let \( F(x_{1,t}, ..., x_{n,t}, \theta) \) be the joint distribution of \( x_{i,t}, i = 1, ..., n \). Based on Sklar’s (1959) theorem, we can cast this joint distribution into as a set of univariate distributions that are bound by a copula function, that is:
\[
F(x_{1,t}, ..., x_{n,t}, \theta) = C(F_1(x_{1,t}, \theta_0, \theta_1), ..., F_n(x_{n,t}, \theta_0, \theta_n), \rho)
\]
(24)

where \( \theta = (\theta_0, \theta_1, ..., \theta_n, \rho) \) and \( F_i(x_{i,t}, \theta_0, \theta_i) \) is the marginal distribution of \( x_{i,t}, i = 1, ..., n \). The univariate distributions depend on a set of common parameters \( \theta_0 \) and a set of specific parameters \( \theta_i \), while the dependence between these distributions is characterized by a set of parameters that are gathered in \( \rho \). According to the results derived above, \( \theta_0 \) and \( \rho \) are identifiable from the moment function \( h_{x,t}(u, \theta) \) while the specific parameters \( \theta_i \) may not. Fortunately, these specific parameters can be inferred from the marginal distributions of \( x_{i,t}, i = 1, ..., n \).
3 Properties of the CGMM Estimators

In this section, we review the properties of the CGMM estimators \( \hat{\theta} (\tau) \) and \( \hat{\theta}^{(2)} (\tau) \). Under assumptions 1 to 4, \( \hat{\theta} (\tau) \) is consistent for \( \theta_0 \) (for almost all \( \tau \)) and is asymptotically normal. The proof of this statement can be found in Carrasco and Florens (2000) and Carrasco, Chernov, Florens and Ghysels (2007). The following property also holds for \( \hat{\theta} (\tau) \).

**Proposition 1** Under Assumptions 1 to 6, \( \hat{\theta} (\tau) \) is unique for each \( \tau \) in \( \Lambda \setminus \emptyset \). Moreover, \( \hat{\theta} (\tau) \) is continuously differentiable with respect to \( \tau \).

Proposition 1 is useful for the derivation of the minimum variance ICGMM estimator and for the comparison of the latter with the MLE. The estimator \( \hat{\theta} (\tau) \) can be used to consistently estimate the covariance operator \( K_\tau \) that enters in the computation of the second step CGMM estimator \( \hat{\theta}^{(2)} (\tau) \). In IID and Markov models, a natural estimator of \( K_\tau \) is given by the linear empirical operator \( K;T \) with kernel:

\[
\hat{k}_\tau (u_1, u_2) = \frac{1}{T} \sum_{t=1}^{T} h_{\tau,t}(u_1, \hat{\theta}) h_{\tau,t}(u_2, \hat{\theta}),
\]

where \( \hat{\theta} \equiv \hat{\theta} (\tau) \). In the specific case of IID models, the first step estimator \( \hat{\theta} (\tau) \) may be bypassed by using the formula:

\[
\hat{k}_\tau (u_1, u_2) = \frac{1}{T} \sum_{t=1}^{T} (e^{iu_1 y_{\tau,t}} - \hat{\varphi}_{\tau,T}) \overline{(e^{iu_1 y_{\tau,t}} - \hat{\varphi}_{\tau,T})},
\]

where \( \hat{\varphi}_{\tau,T} = \frac{1}{T} \sum_{t=1}^{T} e^{iu_1 y_{\tau,t}} \). Finally, in the dependent model, \( k_\tau (u_1, u_2) \) is estimated by:

\[
\hat{k}_\tau (u_1, u_2) = \frac{1}{T} \sum_{t=1}^{T} h_{\tau,t}(u_1, \hat{\theta}) h_{\tau,t}(u_2, \hat{\theta})
\]

\[
+ \sum_{j=1}^{J_T} \left( 1 - \frac{j-1}{J_T} \right) \sum_{t=1}^{T} h_{\tau,t}(u_1, \hat{\theta}) \left( h_{\tau,t-j}(u_2, \hat{\theta}) + h_{\tau,t+j}(u_2, \hat{\theta}) \right),
\]

and \( J_T \) is a bandwidth that is increasing in \( T \).

The operator \( K_\tau \) has an infinite and discrete spectrum. By letting \( l_{\tau,i} \) be its eigenvalue associated with the eigenfunction \( \psi_{\tau,i} \) and assuming that \( l_{\tau,i} \) is decreasing in \( i \), we have: (i) \( l_{\tau,1} < \infty \), (ii) \( l_{\tau,i} > l_{\tau,i+1} > 0 \) for all \( i \), and (iii) \( \lim_{i \to \infty} l_{\tau,i} = 0 \). By contrast, \( K_{\tau,T} \) has a degenerate spectrum. More precisely, if we let \( \hat{l}_{\tau,i} \) be an eigenvalue of \( K_{\tau,T} \) associated with the eigenfunction \( \hat{\psi}_{\tau,i} \), then we can label \( \hat{l}_{\tau,i} \) and \( \hat{\psi}_{\tau,i} \) so that: (i) \( \hat{l}_{\tau,1} < \infty \), (ii) \( \hat{l}_{\tau,i} > \hat{l}_{\tau,i+1} \geq 0 \) for all \( i \), and (iii) \( \hat{l}_{\tau,i} = 0 \) for all \( i > T \), where \( T \) is the sample size. As a result, \( K_{\tau,T} \) is not invertible on \( L^2 (\omega) \). For more details on the properties of covariance operators, we refer the reader to Carrasco, Florens and Renault (2007).

To estimate \( K^{-1}_{\tau} \), the following generalized inverse is used:

\[
K^{-1}_{\tau,T,\alpha,T} = (K^2_{\tau,T} + \alpha I)^{-1} K_{\tau,T},
\]
With the same notations as above, it can be checked that $\hat{\psi}_{r,i}$ is an eigenfunction of $K_{r,T}^{-1}$ associated with the eigenvalue $\frac{\lambda_{r,i}}{\lambda_{r,i} + \alpha_T}$.

In IID and Markov models, under Assumptions 1 and 2, the operator $K_{r,T}$ satisfies:

$$\|K_{r,T} - K\| = O_p(T^{-1/2}),$$

where $K$ is the covariance operator defined in equation (4). The regularized inverse $K_{r,T,\alpha_T}^{-1}$ has the property that for any function $f$ in the range of $K_{r,T}^{1/2}$, the function $K_{r,T,\alpha_T}^{-1/2}f$ converges to $K_{r,T}^{-1/2}f$ as $T$ goes to infinity and $\alpha_T$ goes to zero. Assumptions 1 to 4 then ensure that replacing $K_{r,T}^{-1/2}$ by $K_{r,T,\alpha_T}^{-1/2}$ in (11) yields:

$$T^{1/2} \left( \hat{\theta}^{(2)}(\tau) - \theta_0 \right) \overset{L}{\to} N(0, I_{r,\theta_0}^{-1}),$$

as $T$ and $\alpha_T T^{3/2}$ go to infinity and $\alpha_T$ goes to zero, where $I_{r,\theta_0}^{-1}$ denotes the asymptotic variance of the MLE based on the reduced information set.

In dependent models however, only the CGMM efficiency can be attained under some additional technical assumptions discussed in Carrasco, Chernov, Florens and Ghysels (2007). By CGMM efficiency, it is meant that $\hat{\theta}^{(2)}(\tau)$ is optimal among the following class indexed by a linear operator $B$:

$$\arg\min_{\theta} \int B\hat{h}_{r,T}(u, \theta)B\hat{h}_{r,T}(u, \theta)\omega(u)\,du.$$ 

In order for $\hat{\theta}^{(2)}(\tau)$ to be truly optimal in the sense of (28), the regularization parameter $\alpha_T$ needs to be calibrated in practice. Let $\alpha_T(\theta_0)$ be the value of $\alpha_T$ that is optimal in the mean square error sense, that is:

$$\alpha_T(\theta_0) = \arg\min_{\alpha} \mathbb{E} \left[ \left( \hat{\theta}^{(2)}(\tau) - \theta_0 \right) \left( \hat{\theta}^{(2)}(\tau) - \theta_0 \right)' \right].$$

Under Assumptions 1 to 4 and Assumptions 8, Carrasco and Kotchoni (2008) showed the consistency of $\hat{\alpha}_T$ for $\alpha_T(\theta_0)$, where:

$$\hat{\alpha}_T = \arg\min_{\alpha} \frac{1}{M} \sum_{k=1}^{M} \left( \hat{\theta}^{(2,k)}(\tau) - \hat{\theta}(\tau) \right) \left( \hat{\theta}^{(2,k)}(\tau) - \hat{\theta}(\tau) \right)' ,$$

and $\hat{\theta}^{(2,k)}(\tau)$ is the second step CGMM estimator of $\theta_0$ computed using a sample simulated from the data generating process indexed by the point estimate $\hat{\theta}(\tau)$, and $M$ is the total number of simulated samples.

---

1. The consistency and optimality are guaranteed for $\beta \geq 1/2$. The asymptotic normality has been proved in Carrasco, Chernov, Florens and Ghysels (2007) only under $\beta \geq 1$ in Assumption 4, which is satisfied in the characteristic function based CGMM.
4 The Ideal ICGMM Estimator

In Equation (10), we have defined the ICGMM estimator as the weighted sum of a continuum of \( \sqrt{T} \)-consistent estimators indexed by \( \tau \), that is:

\[
\hat{\theta}_x = \int \hat{\theta}(\tau) \pi(\tau) \, d\tau
\]

where \( \pi(\tau) \) is a measure on \( \Lambda \) that sums to one. The continuity of \( \hat{\theta}(\tau) \) as a function of \( \tau \) allows to consider the use of continuous pdfs \( \pi(\tau) \) for the weighting function. Below we discuss the consistency of \( \hat{\theta}_x \) and derive the weighting function \( \pi^* \) that minimizes the variance. The ideal ICGMM estimator \( \hat{\theta}_x^* \) is then compared to the MLE.

4.1 Consistency and Optimal Aggregating Measure

For any \( \lambda \in \mathbb{R}^q \), the covariance of \( \sqrt{T} \hat{\lambda} \hat{\theta}_x \) is given by:

\[
\text{Var} \left( \sqrt{T} \hat{\lambda} \hat{\theta}_x \right) = \int \int g_\lambda(\tau_1, \tau_2) \pi(\tau_1) \pi(\tau_2) \, d\tau_1 \, d\tau_2,
\]

where

\[
g_\lambda(\tau_1, \tau_2) = \lambda \text{Cov} \left( \sqrt{T} \hat{\theta} (\tau_1), \sqrt{T} \hat{\theta} (\tau_2) \right) \lambda.
\]

The potential dependence of \( g_\lambda(\tau_1, \tau_2) \) on \( T \) is hidden for simplicity.

Using the first order Taylor expansion of \( \hat{\theta}(\tau) \) deduced from the first order condition that it solves, it can be showed that:

\[
\lim_{T \to \infty} \text{Cov} \left( \sqrt{T} \hat{\theta} (\tau_1), \sqrt{T} \hat{\theta} (\tau_2) \right) = W_{\tau_1}^{-1} \langle G_{\tau_1}(\cdot, \theta_0), K_{\tau_1, \tau_2} G_{\tau_2}(\cdot, \theta_0) \rangle W_{\tau_2}^{-1},
\]

where \( K_{\tau_1, \tau_2} \) is the operator with kernel:

\[
k_{\tau_1, \tau_2}(u, v) = \lim_{T \to \infty} \text{Cov} \left( \sqrt{T} \hat{h}_{\tau_1,T}(u, \theta_0), \sqrt{T} \hat{h}_{\tau_2,T}(v, \theta_0) \right),
\]

and

\[
G_{\tau}(u, \theta_0) = P \lim_{T \to \infty} \frac{\partial \hat{h}_{\tau,T}(u, \hat{\theta}(\tau))}{\partial \theta} \text{ and } W_{\tau} = \langle G_{\tau}(\cdot, \theta_0), G_{\tau}(\cdot, \theta_0) \rangle.
\]

We have the following consistency result for \( \hat{\theta}_x \):

**Proposition 2** Under Assumptions 1 to 4 and Assumptions 7, the ICGMM estimator satisfies:

\[
\hat{\theta}_x - \theta_0 = O_p \left( T^{-1/2} \right).
\]

According to Proposition 2, the consistency of the ICGMM estimator is obtained by using any arbitrary measure \( \pi \) that sums to one. For example, specifying \( \pi \) as a Gaussian measure would ease the computations, but this may not deliver the ICGMM estimator of \( \lambda \hat{\theta}_x \) with minimum
variance. The ideal measure $\pi^*_\lambda (\tau)$ solves:

$$\pi^*_\lambda = \arg \min_{\pi} \int \int g_\lambda (\tau_1, \tau_2) \pi (\tau_1) \pi (\tau_2) d\tau_1 d\tau_2,$$

subject to $\int \pi (\tau) d\tau = 1$. In practice, $\lambda$ may be set according to some particular hypothesis one which to test on $\hat{\theta}_\pi$.

Let $V_\lambda$ be the linear operator with kernel $g_\lambda (\tau_1, \tau_2)$, that is, the covariance operator associated with $\sqrt{T\lambda} \hat{\theta} (\tau)$. The operator $V_\lambda$ is compact if we have:

$$\int_{\Lambda} \int_{\Lambda} [g_\lambda (\tau_1, \tau_2)]^2 d\tau_1 d\tau_2 < \infty$$

This condition is met here because $\Lambda$ is a bounded set while $g_\lambda (\tau_1, \tau_2)$ is finite and continuous at all $(\tau_1, \tau_2)$. These properties of $g_\lambda (\tau_1, \tau_2)$ follow from the consistency of $\hat{\theta} (\tau)$ and its continuity as a function of $\tau$. The compactness of the covariance operator $V_\lambda$ ensures that it has a discrete spectrum. If we let $\phi_{\lambda,j} (\tau_1)$ denote the eigenfunction of $V_\lambda$ associated with the eigenvalue $\nu_{\lambda,j}$, then we have $\nu_{\lambda,j} \geq 0$ and $\phi_{\lambda,i} (\tau_1)$ and $\phi_{\lambda,j} (\tau_1)$ are orthogonal for all $i \neq j$. The following proposition characterizes the optimal weighting function.

**Proposition 3** The solution of (32) $\pi^*_\lambda (\tau)$ with minimal norm is given by:

$$\pi^*_\lambda (\tau) = \left[ \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left( \int \phi_{\lambda,j} (\tau_1) d\tau_1 \right)^2 \right]^{-1} \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left( \int \phi_{\lambda,j} (\tau_1) d\tau_1 \right) \phi_{\lambda,j} (\tau),$$

At the optimum, the variance of $\sqrt{T\lambda} \hat{\theta}_\pi$ is:

$$\text{Var} \left( \sqrt{T\lambda} \hat{\theta}_\pi \right) = \left[ \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left( \int \phi_{\lambda,j} (\tau) d\tau \right)^2 \right]^{-1}. \quad (34)$$

Note that $\pi^*_\lambda (\tau) + \tilde{f} (\tau)$ is also a solution of (32) for any function $\tilde{f} (\tau)$ in the null set of $V$, and $\pi^*_\lambda (\tau)$ is the unique solution if the null set of $V_\lambda$ reduces to the null function. The expression of the theoretically optimal ICGMM estimator is:

$$\hat{\theta}_{\pi^*} = \int \pi^*_\lambda (\tau) \hat{\theta} (\tau) d\tau$$

$$= \left[ \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left( \int \phi_{\lambda,j} (\tau_1) d\tau_1 \right)^2 \right]^{-1} \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left( \int \phi_{\lambda,j} (\tau_1) d\tau_1 \right) \int \phi_{\lambda,j} (\tau) \hat{\theta} (\tau) d\tau,$$

To obtain a representation of $\hat{\theta}_{\pi^*}$ that is more compact than in Proposition 3, we define the following scalar product on the Hilbert space of bounded real functions on $\Lambda$:

$$\langle f, g \rangle_\Lambda = \int f(\tau) g(\tau) d\tau$$

12
Under this notation, we have:

\[ \pi^*_\lambda (\tau) = \frac{1}{\langle V^{-1}_\lambda, \iota \rangle_\Theta} V^{-1}_\lambda \iota (\tau), \]

\[ \hat{\theta}_{\pi^*} = \frac{1}{\langle V^{-1}_\lambda, \iota \rangle_\Theta} \langle V^{-1}_\lambda, \hat{\theta} \rangle_\Theta \quad \text{and} \quad \text{Var} \left( \sqrt{T} \lambda' \hat{\theta}_{\pi^*_\lambda} \right) = \frac{1}{\langle V^{-1}_\lambda, \iota \rangle_\Theta}. \]

where the relation between \( \hat{\theta}_{\pi^*} \) and \( \hat{\theta} \equiv \hat{\theta} (\tau) \) is defined on an element-by-element basis and \( \iota (\tau) = 1 \) for all \( \tau \). This compact representation brings more intuitions about how to estimate \( \pi^*_\lambda (\tau) \) and \( \hat{\theta}_{\pi^*} \) from finite dimensional objects.

### 4.2 Comparison with the Maximum Likelihood

The CGMM procedure may be used when the MLE is either costly to implement or unfeasible. In turn, the ICGMM is a good alternative that delivers a \( \sqrt{T} \)-consistent estimator of the parameter of interest when the CGMM itself is unfeasible due to the curse of dimensionality outlined in the introduction. In this section, we discuss the conditions under which the ideal ICGMM estimator \( \hat{\theta}_{\pi^*_\lambda} \) is as efficient as the unfeasible MLE.

Let \( \hat{\theta}_{\text{MLE}} \) be the unknown MLE of \( \theta_0 \) and define the linear manifold \( \hat{D}_T (\theta_0) \) by:

\[ \hat{D}_T (\theta_0) = \left\{ \theta \in \mathbb{R}^q \text{ s.t. } \theta = \int \hat{\theta} (\tau) \pi (\tau) \, d\tau \text{ and } \int \pi (\tau) \, d\tau = 1 \right\}. \quad (35) \]

For a given sample, \( \hat{\theta}_{\text{MLE}} \) and \( \hat{D}_T (\theta_0) \) are deterministic functions of the data. Let us assume that for each given sample, there exists \( \pi^* (\tau) \) such that:

\[ \int \hat{\theta} (\tau) \pi^* (\tau) \, d\tau = \hat{\theta}_{\text{MLE}}, \quad (36) \]

In this case, we would have \( \hat{\theta}_{\text{MLE}} \in \hat{D}_T (\theta_0) \) and:

\[ \text{Var} \left( \lambda' \hat{\theta}_{\pi^*_\lambda} \right) = \arg \min_{\hat{\theta} \in \hat{D}_T (\theta_0)} \text{Var} \left( \lambda' \hat{\theta} \right) \leq \text{Var} \left( \lambda' \hat{\theta}_{\text{MLE}} \right). \]

The following proposition gives a condition under which Equation (36) is satisfied.

**Proposition 4** Under Assumptions 1 to 7, the optimal ICGMM estimator \( \lambda' \hat{\theta}_{\pi^*_\lambda} \) is as efficient as the MLE \( \lambda' \hat{\theta}_{\text{MLE}} \) if the rank of \( \frac{\partial \hat{\theta} (\tau)}{\partial \tau} \) is equal to \( q = \text{dim} (\theta_0) \).

The intuition behind this result is as follows. Around a particular \( \tau \), we have:

\[ \hat{\theta} (\tau + \tau_0) = \hat{\theta} (\tau) + \frac{\partial \hat{\theta} (\tau)}{\partial \tau} \tau_0. \quad (37) \]

When the rank of \( \frac{\partial \hat{\theta} (\tau)}{\partial \tau} \) is equal to \( q \), the manifold \( \hat{D}_T (\theta_0) \) replicates the entire parameter space \( \Theta \). This happens because \( \hat{\theta} (\tau + \tau_0) \) replicates the entire neighborhood of \( \hat{\theta} (\tau) \) as \( \tau_0 \) varies in the
neighborhood of the null vector of $\mathbb{R}^d$. In this case, Equation (36) holds and $\lambda \hat{\theta}_{\pi_\Lambda}$ is as efficient as $\lambda \hat{\theta}_{\text{MLE}}$. A necessary condition for this rank condition to be satisfied is $d \geq \max\{q, 2\}$. Indeed, when $d = 1$, the normalized set $\Lambda$ reduces to the singleton $\{\tau = 1\}$ and $\hat{D}_T (\theta_0) = \{\hat{\theta} (1)\}$. In this case, $\lambda \hat{\theta}_{\pi_\Lambda}$ is not as efficient as the $\lambda \hat{\theta}_{\text{MLE}}$, and implementing the second step estimator $\hat{\theta}^{(2)} (1)$ would permit to approach the maximum likelihood efficiency. However, when $d \geq 2$, $\Lambda$ contains a continuum of normalized vectors and the set $\{\hat{\theta} (\tau), \tau \in \Lambda\}$ is a continuum of estimators. The rank condition on $\frac{\partial \hat{\theta} (\tau)}{\partial \tau}$ is satisfied if the dimensionality of this set is equal to that of the parameter space $\Theta$. If this rank condition is not satisfied, then $\hat{D}_T (\theta_0)$ can still encompass the MLE, but there is no simple way to verify this.

Analogue of the rank condition of Proposition 4 have been formulated in the context of previous indirect estimators. For example, in Gourieroux, Monfort and Renault (1993), the indirect inference estimator is consistent for the parameter of interest if the gradient of the function that binds the true model to the auxiliary model is of full rank. In Gallant and Tauchen (1996), the efficient method of moment estimator is as efficient as the MLE if the auxiliary model is rich enough to encompasses the structural model of interest. Intuitively, $\hat{D}_T (\theta_0)$ is more likely to encompass $\hat{\theta}_{\text{MLE}}$ if there is enough variability in the set $\hat{\theta} (\tau)$ across $\tau$. In light of this, using the suboptimal CGMM estimator $\hat{\theta} (\tau)$ in the definition of $\hat{\theta}_\pi$ has two advantages. First of all, $\hat{\theta} (\tau)$ is less efficient than $\hat{\theta}^{(2)} (\tau)$ and thus has more variability than the latter, thus allowing the manifold $\hat{D}_T (\theta_0)$ to have a higher probability of encompassing the MLE. And secondly, by permitting to avoid computing the covariance operator $K^{-1/2}$, the use of $\hat{\theta} (\tau)$ makes the computation of the ICGMM estimator easier.

5 The Feasible Optimal ICGMM

Our goal is to compute the estimator $\hat{\theta}_{\pi^*} = \int \hat{\theta} (\tau) \pi^*_\Lambda (\tau) d\tau$. If we knew how to draw from the measure $\pi^*_\Lambda (\tau)$, then we would approximate $\hat{\theta}_{\pi^*}$ by Monte Carlo using the formula:

$$\hat{\theta}_{\pi^*} \simeq \frac{1}{S} \sum_{i=1}^S \hat{\theta} (\tau_i),$$

where $\tau_i, i = 1, ..., S$ are independent draws from $\pi^*_\Lambda (\tau)$. Unfortunately, the expression of the measure $\pi^*_\Lambda (\tau)$ is not friendly and we do not know how to simulate data from this distribution.

Alternatively, one may consider the approximation:

$$\hat{\theta}_{\pi^*} \simeq \frac{1}{S} \sum_{i=1}^S \pi^*_\Lambda (\tau_i) \hat{\theta} (\tau_i), \quad (38)$$

where $\tau_i, i = 1, ..., S$ are independent draws from the uniform distribution on $\Lambda$. We know how to draw from the uniform distribution on $\Lambda$ and we know how to compute $\hat{\theta} (\tau)$. Hence the formula (38) can be implemented provided we find a way to estimate $\pi^*_\Lambda (\tau)$. This second approach is pursued below.
Thus, let \( \tau_i, i = 1, \ldots, S \) be \( S \) draws from the multivariate uniform distribution on \( \Lambda \), and assume that we can simulate from the data generating process of \( x_t \). Further let \( \left\{ x_t^{(l)} \right\}_{t=1}^T, l = 1, \ldots, L \) be \( L \) independent samples of size \( T \) simulated from the distribution of interest. For each sample indexed by \( l \) and each possible \( \tau \), we compute the univariate samples:

\[
\left\{ y_{\tau, l}^{(l)} \right\} = \left\{ \tau' x_t^{(l)} \right\}, \quad l = 1, \ldots, L.
\]

Finally, let \( \hat{\theta} (\tau, l) \) be the first step CGMM estimator based on the sample \( \left\{ y_{\tau, l}^{(l)} \right\} \), and \( \hat{\Theta}_\lambda \) be the \( L \times S \) matrix with \((l, i)\) element given by \( \lambda' \hat{\theta} (\tau_i, l) \). Note that \( \hat{\theta} (\tau_i, l), l = 1, \ldots, L \) are IID copies of the CGMM estimator \( \hat{\theta} (\tau_i) \). An estimator of the covariance operator associated with \( \hat{\theta} (\tau) \) is given by the \((S \times S)\) empirical covariance matrix of \( \hat{\Theta}_\lambda \):

\[
\hat{V}_\lambda = \frac{T}{SL} \left( \hat{\Theta}_\lambda - \overline{\Theta}_\lambda \right)' \left( \hat{\Theta}_\lambda - \overline{\Theta}_\lambda \right),
\]

where \( \hat{\Theta}_\lambda \) is the matrix with \((l, i)\) element given by \( \lambda' \hat{\theta} (\tau_i, l) \) and \( \overline{\Theta}_\lambda \) contains the means of the columns of \( \hat{\Theta}_\lambda \). Hence the \((l, i)\) element of \( \hat{V}_\lambda \) is given by:

\[
\hat{g}_\lambda (\tau_i, \tau_j) = \frac{T}{SL} \sum_{l=1}^L \lambda' \left( \hat{\theta} (\tau_i, l) - \overline{\theta} (\tau_i, l) \right) \left( \hat{\theta} (\tau_j, l) - \overline{\theta} (\tau_j, l) \right)' \lambda,
\]

where \( \overline{\theta} (\tau_i, l) = \frac{1}{L} \sum_{l=1}^L \hat{\theta} (\tau_i, l) \).

Obviously, \( \theta_0 \) is unknown. However, it can be proxied by the consistent estimator \( \hat{\Theta}_S = \frac{1}{S} \sum_{i=1}^S \hat{\theta} (\tau_i) \) computed from the actual data. We have the following result about \( \hat{V}_\lambda \).

**Proposition 5** Let \( f = (f (\tau_1), \ldots, f (\tau_S))' \) where \( \tau_1, \ldots, \tau_S \) are \( S \) draws from the multivariate uniform distribution on \( \Lambda \) and \( f \) is continuous. Then as \( L \) and \( S \) go to infinity, we have:

\[
\left( \hat{V}_\lambda f \right)_i - V_\lambda f (\tau_i) = O_p \left( L^{-1/2} \right) + O_p \left( S^{-1/2} \right)
\]

for all \( \tau_i \), where \( \left( \hat{V}_\lambda f \right)_i \) is the \( i \)th element of the vector \( \hat{V}_\lambda f \).

Acting on the consistency result of Proposition 5, we use \( \hat{V}_\lambda \) to estimate the optimal aggregating weight \( \pi^*_\lambda \) by:

\[
\hat{\pi}^*_\lambda = \left( \hat{\pi}^*_{\lambda, 1}, \ldots, \hat{\pi}^*_{\lambda, S} \right)' = \left( \frac{l'}{S} \hat{V}^{-1}_{\lambda, 1} \frac{l}{S} \right) \left( \frac{l'}{S} \hat{V}^{-1}_{\lambda, S} \frac{l}{S} \right),
\]

where \( l \) is a vector of ones and \( \hat{V}^{-1}_{\lambda, i} \) is the regularized inverse of \( \hat{V}_\lambda \) defined as:

\[
\hat{V}^{-1}_{\lambda, i} = \left( \hat{V}_\lambda^2 + \alpha I \right)^{-1} \hat{V}_\lambda, \quad \alpha \in (0, 1).
\]

This regularization is necessary because as \( S \) increases, some elements of the set \( \tau_1, \ldots, \tau_S \) eventually become arbitrarily close so that \( \hat{V}_\lambda \) is singular or nearly so. In Equation (42), the ratio \( \frac{1}{S} \) is a discrete
approximation of the Lebesgue measure. The feasible optimal ICGMM estimator is defined as:

\[
\hat{\theta}_{\pi_{\lambda, \alpha}}^* = \frac{1}{S} \sum_{i=1}^{S} \pi_{\lambda, \alpha}^*(\tau_i) \hat{\theta}(\tau_i),
\]

\[
= \left(\frac{1}{T} \sum_{i=1}^{T} \pi_{\lambda, \alpha}^*(\tau_i) \right)^{-1} \left(\frac{1}{T} \sum_{i=1}^{T} \pi_{\lambda, \alpha}^*(\tau_i) \right)^{-1} \hat{\theta}
\]

where \( \hat{\theta}(\tau_i), i = 1, ..., S \) are computed from the actual data and \( \hat{\theta} = \left(\hat{\theta}(\tau_1), ..., \hat{\theta}(\tau_S)\right) \). To prove the consistency of this estimator, the following assumption is needed.

**Assumption 9:** \( \|V_{\lambda}^{-\epsilon} t\| < \infty \) for some \( \epsilon > 1 \), where \( t(\tau) = 1 \) for all \( t \).

Assumption 9 can be shown to hold for \( \epsilon = 1 \). Indeed, the optimal measure \( \pi_{\lambda}^*(\tau) = \frac{1}{\langle V_{\lambda}^{-1} t, \pi_{\lambda} \rangle_{L^2}} V_{\lambda}^{-1} t(\tau) \) is continuous, and is thus square integrable on the bounded set \( \Lambda \). Because the ratio \( \frac{1}{\langle V_{\lambda}^{-1} t, \pi_{\lambda} \rangle_{L^2}} \) gives the variance of the optimal ICGMM estimator of \( \sqrt{T} \lambda \hat{\theta}_{\pi_{\lambda}}^{\ast} \), it is finite. Hence the function \( V_{\lambda}^{-1} t(\tau) \) is square integrable on \( \Lambda \), that is, \( \|V_{\lambda}^{-1} t\| < \infty \). Hence Assumption 9 may not be as strong as it seems at a glance. The following proposition contains the ingredient that will be used in the proof of the consistency of the optimal ICGMM estimator.

**Proposition 6** Under Assumptions 1 to 7 and Assumption 9, we have:

\[
\left\| \frac{1}{\langle V_{\lambda}^{-1} t, \pi_{\lambda} \rangle_{L^2}} (V_{\lambda}^{-1} - \lambda^{-1}) t \right\| = O_p \left( \alpha^{-1} L^{-1/2} + \alpha^{-1} S^{-1/2} \right), \quad \text{and}
\]

\[
\left\| (V_{\lambda}^{-1} - \lambda^{-1}) t \right\| = O \left( \alpha^{\min(1, \infty)} \right) .
\]

According to Proposition 6, we have:

\[
(V_{\lambda}^{-1} - \lambda^{-1}) t = O \left( \alpha^{\min(1, \infty)} \right) + O_p \left( \alpha^{-1} L^{-1/2} + \alpha^{-1} S^{-1/2} \right). \quad (47)
\]

This implies that \( (V_{\lambda}^{-1} - \lambda^{-1}) t \) converges to zero provided that \( L \to \infty, S \to \infty, \alpha \to 0, \alpha^{-1} L^{-1/2} \to 0 \) and \( \alpha^{-1} S^{-1/2} \to 0 \). The next result establishes the consistency of the estimated optimal weighting function and the asymptotic optimality of the feasible ICGMM estimator.

**Proposition 7** Under Assumptions 1 to 7 and Assumption 9, \( \hat{\pi}_{\lambda, \alpha}^* (\tau) \) converges to \( \pi_{\lambda}^* (\tau) \) and we have:

\[
\hat{\pi}_{\lambda, \alpha}^* - \pi_{\lambda}^* = O \left( \alpha^{\min(1, \infty)} \right) + O_p \left( \alpha^{-1} L^{-1/2} + \alpha^{-1} S^{-1/2} \right)
\]

as \( L \to \infty, S \to \infty, \alpha \to 0, \alpha^{-1} L^{-1/2} \to 0 \) and \( \alpha^{-1} S^{-1/2} \to 0 \). Moreover, the asymptotic variances of \( \hat{\theta}_{\pi_{\lambda, \alpha}}^* \) and \( \hat{\theta}_{\pi_{\lambda}}^* \) are the same:

\[
\hat{\theta}_{\pi_{\lambda, \alpha}}^* - \hat{\theta}_{\pi_{\lambda}}^* = o_p \left( T^{-1/2} \right)
\]

In Proposition 7, the larger the number \( \epsilon \), the faster the rate of convergence of \( \hat{\pi}_{\lambda, \alpha}^* \). However the value of \( \epsilon \) does not impact on the optimality of the feasible ICGMM estimator. With the
estimator \( \hat{\pi}_{\lambda,\alpha} \) in hand, we can compute the ICGMM estimator from the \( l \)-th sample as:

\[
\hat{\theta}^{(l)}_{\pi_{\lambda,\alpha}} = \sum_{i=1}^{S} \hat{\pi}_{\lambda,\alpha}^*(\tau_i) \hat{\theta}^{(l)} (\tau_i, l), \quad l = 1, \ldots, L.
\]

(48)

Hence, the distribution of \( \hat{\theta}^{(l)}_{\pi_{\lambda,\alpha}} \) is obtained as a by-product of the ICGMM procedure and it can be used to perform inferences about \( \theta_0 \).

The higher order asymptotics of \( \hat{\theta}^{(l)}_{\pi_{\lambda,\alpha}} \) will depend on the regularization parameter \( \alpha \). The optimal regularization parameter may be estimated by minimizing the following criterion:

\[
\hat{\alpha} = \arg \min_{\alpha \in (0,1)} \frac{T}{L} \sum_{k=1}^{L} \left( \hat{\theta}^{(l)}_{\pi_{\lambda,\alpha}} - \hat{\theta}_S \right)' \left( \hat{\theta}^{(l)}_{\pi_{\lambda,\alpha}} - \hat{\theta}_S \right),
\]

(49)

where \( \hat{\theta}_S = \frac{1}{S} \sum_{i=1}^{S} \hat{\theta}^{(l)} (\tau_i) \) is the proxy used for \( \theta_0 \).

Finally, instead of using Monte Carlo simulations in the ICGMM procedure, one can resort to a classical time domain resampling to generated bootstrap copies \( \hat{\theta}^{(l)} (\tau_i, l), l = 1, \ldots, L \) of \( \hat{\theta} (\tau_i) \). Both approaches are implemented below.

6 Simulation Study 1: The Gaussian AR(1) Model

We base the simulation study of this section on the Gaussian AR(1):

\[
x_t = \rho x_{t-1} + \varepsilon_t,
\]

(50)

where \( \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \). The mean of this process is \( \mu \equiv E(x_t) = 0 \) and its variance is \( \sigma^2 \equiv Var(x_t) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \). If \( \rho = \sigma_\varepsilon = \frac{1}{\sqrt{2}} \), then the marginal distribution of \( x_t \) is standard normal while that of \((x_t, x_{t-1})\)' is given by:

\[
\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).
\]

If \( \mu \) and \( \sigma^2 \) are known, then the only parameter to estimate is \( \theta = \rho \). In this case, we have \( q = 1 \) and \( d = 2 \). The MLE of \( \rho \) obtained from the joint distribution of \((x_t, x_{t-1})\)' is given by:

\[
\hat{\rho} = \frac{\hat{\rho}^3 - \hat{\rho}^2 + \frac{1}{T} \sum x_t x_{t-1} - \frac{1}{T} \sum x_t^2 x_{t-1}}{1 - \frac{1}{T} \sum x_t^2 - \frac{1}{T} \sum x_{t-1}^2}.
\]

(51)

On the other hand, MLE can be computed for \( \rho \) based on the reduced information set:

\[
y_{\tau,t} = \tau_1 x_t + \tau_2 x_{t-1} \sim N \left( 0, \tau_1^2 + 2\rho \tau_1 \tau_2 + \tau_2^2 \right).
\]

This is given by:

\[
\hat{\rho}_{MLE} (\tau) = \frac{1}{2\tau_1 \tau_2} \left[ \frac{1}{T} \sum y_{\tau,t} - \frac{\tau_1^2}{\tau_2} - \frac{\tau_2^2}{\tau_1} \right].
\]

(52)
When $\mu$ and $\sigma^2$ are unknown so that the $\theta = (\mu, \sigma^2, \rho)$, $q = 3$ and $d = 2$, the MLE of $\theta$ based on the joint distribution of $(x_t, x_{t-1})'$ solves the fixed point problem:

$$
\hat{\mu}_{MLE} = \frac{1}{2T} \sum (x_t + x_{t-1}),
$$

$$
\hat{\sigma}_{MLE}^2 = \frac{1}{2T} \sum \frac{(x_t - \hat{\mu})^2 - 2\hat{\rho} (x_t - \hat{\mu})(x_{t-1} - \hat{\mu}) + (x_{t-1} - \hat{\mu})^2}{(1 - \hat{\rho}^2)} \quad \text{and}
$$

$$
\hat{\rho}_{MLE} = \frac{\hat{\rho}^3 - \hat{\sigma}^2 \frac{1}{T} \sum \left( \frac{x_t - \hat{\mu}}{\hat{\sigma}} \right) \left( \frac{x_{t-1} - \hat{\mu}}{\hat{\sigma}} \right) - \frac{1}{T} \sum \left( \frac{x_t - \hat{\mu}}{\hat{\sigma}} \right)^2 - \frac{1}{T} \sum \left( \frac{x_{t-1} - \hat{\mu}}{\hat{\sigma}} \right)^2}{1 - \frac{1}{T} \sum \left( \frac{x_t - \hat{\mu}}{\hat{\sigma}} \right)^2 - \frac{1}{T} \sum \left( \frac{x_{t-1} - \hat{\mu}}{\hat{\sigma}} \right)^2}.
$$

The MLE of $\theta$ based on the reduced information set is given by:

$$
\hat{\mu}_{MLE}(\tau) = \frac{1}{T(\tau_1 + \tau_2)} \sum y_{\tau,t},
$$

$$
\hat{\sigma}_{MLE}^2(\tau) = \frac{1}{T(\tau_1^2 + 2\hat{\rho}(\tau)\tau_1\tau_2 + \tau_2)} \sum (y_{\tau,t} - (\tau_1 + \tau_2) \hat{\mu})^2 \quad \text{and}
$$

$$
\hat{\rho}_{MLE}(\tau) = \frac{1}{2\tau_1\tau_2} \left[ \frac{1}{T} \sum \left( \frac{y_{\tau,t} - (\tau_1 + \tau_2) \hat{\mu}(\tau)}{\hat{\sigma}(\tau)} \right)^2 - \tau_1^2 - \tau_2^2 \right].
$$

To compute the first step CGMM estimator $\hat{\theta}(\tau)$, we consider the following moment function deduced from the joint CF of $(x_t, x_{t-1})'$:

$$
h_{\tau,t}(u, \theta) = \exp(iu y_{\tau,t}) - \exp\left( iu \mu_{\tau} - \frac{1}{2} u^2 \sigma_{\tau}^2 \right), \quad u \in \mathbb{R}.
$$

where

$$
y_{\tau,t} = \tau' x_t,
$$

$$
\mu_{\tau} \equiv E(y_{\tau,t}) = \mu (\tau_1 + \tau_2) \quad \text{and}
$$

$$
\sigma_{\tau}^2 \equiv Var(y_{\tau,t}) = \sigma^2 (\tau_1^2 + 2\rho \tau_1 \tau_2 + \tau_2^2).
$$

We would like to compare the performance of the indirect maximum likelihood (IMLE) and the ICGMM estimators to that of the MLE based on the joint distribution of $(x_t, x_{t-1})'$. To compute the indirect estimators (IMLE and ICGMM), we define $\delta$ as the sum of the unknown parameters. For example, $\delta = \mu + \sigma^2 + \rho$ when all the parameters are estimated while $\delta = \mu + \rho$ when $\sigma^2$ is assumed known. Next, we simulate $L = 100$ samples of size $T = 250$. For each sample indexed by $l = 1, ..., M$, we compute the estimators $\delta^{(l)}(\tau_i)$ of $\delta$, $i = 1, 2, ..., S = 100$. Let $\hat{\Theta}$ denote the $L \times S$ matrix with $(l, s)$ element $\hat{\delta}^{(l)}(\tau_i)$, and $\widehat{V}$ the matrix given by:

$$
\widehat{V} = \frac{T}{SL} \left( \hat{\Theta} - \bar{\Theta} \right)' \left( \hat{\Theta} - \bar{\Theta} \right),
$$

where $\bar{\Theta}$ contains the means of the columns of $\hat{\Theta}$. Then, the optimal indirect estimator from the
\( t^{th} \) sample is given by

\[
\hat{\theta}_{\hat{\pi}_a}^{(l)} = \frac{1}{S} \sum_{i=1}^{S} \hat{\pi}_a^*(\tau_i) \hat{\delta}_{\hat{\pi}_a}^{(l)}(\tau_s)
\]

where

\[
\hat{\pi}_a^* = (\hat{\pi}_a^*(\tau_1), \ldots, \hat{\pi}_a^*(\tau_S))^\prime = S \left( \hat{\nu} \hat{V}_a^{-1} \right)^{-1} \hat{\nu} \hat{V}_a^{-1}, \quad \text{and}
\]

\[
\hat{V}_a^{-1} = \left( \hat{V}^2 + \alpha I \right)^{-1} \hat{V}, \quad \alpha \in (0, 1)
\]

The simulation results are shown in Table 1.
Table 1.
Comparison of the MLE, IMLE and ICGMM by Simulation.

(IC5 and IC95 are respectively the 5th and 95th percentile of the simulated distributions)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>IC5</th>
<th>IC95</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho ) unknown, ( \mu = 0, \sigma^2 = 1 )</td>
<td>MLE</td>
<td>0.4985</td>
<td>0.0411</td>
<td>1.3251</td>
</tr>
<tr>
<td></td>
<td>IMLE</td>
<td>0.4976</td>
<td>0.0419</td>
<td>1.3120</td>
</tr>
<tr>
<td></td>
<td>ICGMM</td>
<td>0.4940</td>
<td>0.0439</td>
<td>1.3922</td>
</tr>
<tr>
<td>Results reported for ( \rho = 0.5 )</td>
<td>MLE</td>
<td>0.4976</td>
<td>0.0705</td>
<td>0.3638</td>
</tr>
<tr>
<td></td>
<td>IMLE</td>
<td>0.4970</td>
<td>0.0705</td>
<td>0.3715</td>
</tr>
<tr>
<td></td>
<td>ICGMM</td>
<td>0.4951</td>
<td>0.0714</td>
<td>0.3557</td>
</tr>
<tr>
<td>( (\mu, \rho) ) unknown, ( \sigma^2 = 1 )</td>
<td>MLE</td>
<td>1.5169</td>
<td>0.0971</td>
<td>1.3612</td>
</tr>
<tr>
<td></td>
<td>IMLE</td>
<td>1.5303</td>
<td>0.1340</td>
<td>1.3376</td>
</tr>
<tr>
<td></td>
<td>ICGMM</td>
<td>1.5211</td>
<td>0.1169</td>
<td>1.3244</td>
</tr>
<tr>
<td>Results reported for ( \sigma^2 + \rho = 1.5 )</td>
<td>MLE</td>
<td>1.5131</td>
<td>0.1169</td>
<td>1.3251</td>
</tr>
<tr>
<td></td>
<td>IMLE</td>
<td>1.5351</td>
<td>0.1496</td>
<td>1.3120</td>
</tr>
<tr>
<td></td>
<td>ICGMM</td>
<td>1.5316</td>
<td>0.0809</td>
<td>1.3922</td>
</tr>
</tbody>
</table>

In all the scenarios, the performances of indirect estimators are quite similar to that of the MLE derived from the joint likelihood of the data. In the last scenario where all the parameters are estimated, the number of unknown parameters is larger than the dimensionality of \( \tau \). Nonetheless, the variance of the ICGMM estimator is slightly smaller than that of the IMLE and the MLE. Because the second step CGMM estimator is asymptotically as efficient as the MLE, the results of Table 1 confirm that one can rely on the first step CGMM estimator \( \hat{\theta}(\tau) \) to compute the efficient ICGMM estimator.

7 Simulation Study 2: The Autoregressive Factor Gamma Model

The expected return of financial assets is positively correlated with the expected risk while the unexpected return is negatively correlated with the unpredictable risk. French, Schwertz and Stambaugh (1987) documented this fact more than two decades ago by performing the regression of the excess return onto estimates of expected and unexpected volatilities. They also found that the excess return is negatively correlated with the unexpected risk. The increase in the expected excess return following an increase in the expected risk is driven by the risk premium while the negative correlation between the excess return and the volatility shocks is often called the leverage effect.

However, it is not clear whether the risk on a financial asset should be solely measured by its volatility. For this simulation study, we consider a latent risk factor model for assets returns. This model assumes that the returns are positively correlated with some latent risk factor while being negatively correlated with the innovations of that factor. Because the considered latent risk factor
is not exactly the variance of the return, this model offers an alternative framework to assess the risk premium and the leverage effect on financial markets.

### 7.1 The Autoregressive Factor Gamma Model

The Autoregressive Factor Gamma Model (henceforth ARFG) is a stochastic volatility model for asset returns. The return $r_t$ is expressed as linear function of lagged realization of some latent risk factor $V_{t-1}$ and its contemporaneous innovation $V_t$, that is:

$$r_t = \mu_0 + \mu_1 V_{t-1} + \delta (V_t - E[V_t|V_{t-1}]) + \sigma \varepsilon_t,$$

where $\varepsilon_t \sim N(0,1)$ is uncorrelated with $V_{t-1}$ and $V_t - E[V_t|V_{t-1}]$. The risk premium is modeled as a positive relationship between the return and the expected risk, while the leverage effect is modeled as a negative relation between the return and the unexpected risk. The latent variable $V_t$ is assumed to follow an Autoregressive Gamma process of order one:

$$f(V_t|V_{t-1}) = \sum_{j=0}^{\infty} \frac{V_{t}^{j+q-1} e^{j+q}}{\Gamma(j+q)} \exp(-cV_t) p_j(V_{t-1}),$$

with $c = \frac{2\kappa}{\sigma^2(1-e^{-\kappa})}$, $q = \frac{2\kappa \beta}{\sigma^2}$, $(\kappa, \beta, \sigma) > 0$ and $p_j(V_{t-1})$ are Poisson weights given by:

$$p_j(V_{t-1}) = \frac{(ce^{-\kappa}V_{t-1})^j}{j!} \exp(-ce^{-\kappa}V_{t-1}).$$

The marginal distribution of $V_t$ is a Gamma with density given by:

$$f(V_t) = \frac{V_t^{q-1}}{\Gamma(q)} \left(\frac{2\kappa}{\sigma^2}\right)^q \exp\left(-\frac{2\kappa}{\sigma^2}V_t\right).$$

Its conditional and unconditional CF are:

$$E[e^{irV_t}|V_{t-1}] = \left(1 - \frac{i\tau}{c}\right)^{-q} \exp\left(\frac{i\tau e^{-\kappa}V_{t-1}}{1 - \frac{i\tau}{c}}\right)$$

By looking at the above conditional CF, we see that the distribution of $V_t$ is nested by the Wishart Autoregressive process of Gourieroux, Jasiak and Sufana (2005). In particular, the series $V_t$ can be thought of as a discrete sample from the CIR diffusion.
The conditional expectation and variance of $V_t$ are given by:
\begin{align*}
E[V_t|V_{t-1}] &= \beta (1 - e^{-\kappa}) + e^{-\kappa}V_{t-1}, \\ Var[V_t|V_{t-1}] &= \frac{\beta \sigma^2}{2\kappa} (1 - e^{-\kappa})^2 + \frac{\sigma^2}{\kappa} e^{-\kappa} (1 - e^{-\kappa}) V_{t-1}.
\end{align*}
(65)\hspace{1cm}(66)
This implies that the conditional mean and variance of $r_t$ are linear in the lagged realization of the risk factor:
\begin{align*}
E[r_t|V_{t-1}] &= \mu_0 + \mu_1 V_{t-1}, \\ Var[r_t|V_{t-1}] &= \delta^2 Var[V_t|V_{t-1}] + \sigma^2.
\end{align*}
(67)
Hence
\begin{align*}
&= \frac{\delta^2 \beta \sigma^2}{2\kappa} (1 - e^{-\kappa})^2 + \frac{\sigma^2}{\kappa} e^{-\kappa} (1 - e^{-\kappa}) V_{t-1}.
\end{align*}
(68)
In particular, $E[r_t|V_{t-1}]$ is a linear function of $Var[r_t|V_{t-1}]$.

We derive similarly the third and fourth conditional moments of $V_t$:
\begin{align*}
E[(V_t - E[V_t|V_{t-1}])^3|V_{t-1}] &= \frac{\beta \sigma^4}{2\kappa^3} (1 - e^{-\kappa})^3 + \frac{3\sigma^4 e^{-\kappa}}{2\kappa^2} (1 - e^{-\kappa})^2 V_{t-1}, \\ E[(V_t - E[V_t|V_{t-1}])^4|V_{t-1}] &= 3Var[V_t|V_{t-1}]^2 + \frac{3\sigma^6}{4\kappa^3} \left( \beta (1 - e^{-\kappa})^2 + 4 e^{-\kappa} (1 - e^{-\kappa})^3 V_{t-1} \right).
\end{align*}
(69)\hspace{1cm}(70)
Equation (70) shows that $V_t$ has a positive excess kurtosis. The third and fourth conditional moments of $r_t$ are linked to those of $V_t$ by:
\begin{align*}
E[(r_t - E[r_t|V_{t-1}])^3|V_{t-1}] &= \delta^3 E[(V_t - E[V_t|V_{t-1}])^3|V_{t-1}], \\ E[(r_t - E[r_t|V_{t-1}])^4|V_{t-1}] &= \delta^4 E[(V_t - E[V_t|V_{t-1}])^4|V_{t-1}] + 3\sigma^4.
\end{align*}
(71)\hspace{1cm}(72)
Hence $r_t$ has a time varying negative skewness whenever $\delta < 0$. To see the implications of the model in terms of kurtosis, we note that the last equality implies:
\begin{align*}
E[(r_t - E[r_t|V_{t-1}])^4|V_{t-1}] &= 3Var[r_t|V_{t-1}]^2 + \frac{3\delta^4 \sigma^6 e^{-\kappa}}{2\kappa^3} (1 - e^{-\kappa})^3 V_{t-1},
&+ \frac{3\delta^2 \sigma^2}{\kappa} \left( \frac{\delta^4 \sigma^4}{4\kappa^2} (1 - e^{-\kappa})^2 - \sigma^2 \right) \beta (1 - e^{-\kappa})^2 \\
&+ \frac{6\delta^2 \sigma^2}{\kappa} \left( \frac{\delta^4 \sigma^4}{4\kappa^2} (1 - e^{-\kappa})^2 - \sigma^2 \right) e^{-\kappa} (1 - e^{-\kappa}) V_{t-1}.
\end{align*}
(73)
Clearly, this model can reproduce fat tailed distributions. In particular, the distribution of $r_t$ given $V_t$ is fat tailed when $\frac{\delta^4 \sigma^4}{4\kappa^2} (1 - e^{-\kappa})^2 - \sigma^2 < 0$ is positive.\footnote{The noncentered conditional moments of $r_t$ are given by $E[(V_t)^n|V_{t-1}] = \frac{1}{\kappa^n} \frac{\partial^n E[e^{\tau V_t}|V_{t-1}]}{\partial \tau^n}|_{\tau=0}$. These derivatives may be computed using a mathematical software.}

\footnote{This is only a sufficient condition, not necessary.}
7.2 Estimation of the ARFG Model from Observed Returns

While the joint process of observed return and latent risk factor \((r_t, V_t)\) is Markov, the process \(r_t\) alone is not. Since only the returns are observed, the estimation strategy will necessarily be based on the joint CF of the returns. Writing \(r_t\) as a linear function of \((V_t, V_{t-1})\) allows to easily integrate out the latent factor.

Proposition 8 The joint CF of the observed returns \((r_t, ..., r_{t+d})\) is given by:

\[
E \left[ \exp \left( \sum_{k=1}^{d} i \tau_k r_{t+1-k} \right) \right] = \exp \left( \left[ \mu_0 - \delta \beta \left( 1 - e^{-\kappa} \right) \right] \sum_{k=1}^{d} i \tau_k - \frac{\sigma_z^2}{2} \sum_{k=1}^{d} \tau_k^2 \right) \\
\times \left( 1 - \frac{i u_{d+1} \sigma_z^2}{2 \kappa} \right)^{-q} \prod_{k=1}^{d} \left( 1 - \frac{i u_k}{c} \right)^{-q},
\]  

where:

\[
\begin{align*}
u_1 &= \tau_1 \delta; \\
u_k &= \frac{u_{k-1} e^{-\kappa}}{1 - i u_{k-1} / c} + \tau_{k-1} \left( \mu_1 - \delta e^{-\kappa} \right) + \tau_k \delta, \quad k = 2, ..., d; \\
u_{d+1} &= \frac{u_d e^{-\kappa}}{1 - i u_d / c} + \tau_d \left( \mu_1 - \delta e^{-\kappa} \right).
\end{align*}
\]

The details of the derivation of this CF are left in Appendix. The moment function we will use in the frequency domain resampling of the indirect estimation procedure is:

\[
h_{r,t} (u, \theta) = \exp \left( i u \sum_{k=1}^{d} \tau_k r_{t+1-k} \right) - E \left[ \exp \left( i u \sum_{k=1}^{d} \tau_k r_{t+1-k} \right) \right],
\]

where \(\tau = (\tau_1, ..., \tau_d) \in \Lambda, \ u \in \mathbb{R}, \ \theta = (\mu_0, \mu_1, \delta, \beta, \kappa, \sigma, \sigma_z^2)\). Note that Equation (75) is a moment function of type (6).

7.3 Monte Carlo Simulations and Results

To generate a return process \(r_t\) from the ARFG model, we need to first generate the latent factor \(V_t\). This is done using the Poisson Mixing Gamma representation (61) as suggested by Devroye (1986). At time \(t = 0\), one draws an initial value \(V_0\) from the stationary Gamma distribution (63). At \(t = 1\), one draws an integer \(j_0\) from the Poisson distribution with parameter \(c e^{-\kappa} V_0\). The current realization \(V_1\) of the state variable is then drawn from the Gamma distribution with density \(f_{j_0} (v)\), where:

\[
f_{j_0} (v) = \frac{v^{j_0 + 1} e^{-v} c^{j_0 + q}}{\Gamma (j_0 + q)} \exp (-cv) .
\]

At \(t = 2\), one draws again an integer \(j_1\) from the Poisson distribution with parameter \(c e^{-\kappa} V_1\). The new realization \(V_2\) of the state variable is now drawn from the Gamma distribution with density \(f_{j_1} (v)\), and so forth. At an arbitrary step \(t\), the realization \(V_t\) is drawn from the Gamma distribution
with density $f_{j-1}(v)$, where $j_{l-1}$ is a draw from the Poisson distribution with parameter $ce^{-c}V_{l-1}$. Having simulated a path $(V_0, V_1, ..., V_T)$ as described above, a sample of returns $(r_1, ..., r_T)$ can be generated using Equation (60).

In this Monte Carlo experiment, we set $d = 10$ so that the estimation of $\theta_0$ is based on the joint CF of the vector $(r_t, ..., r_{t-9})$. We used $T = 500$ for the sample size and $M = 100$ for the number of replications. The optimal weight $\pi_{\lambda,0}^*$ and regularization parameter $\alpha^*$ are estimated by generating $S = 100$ samples in the frequency domain for each Monte Carlo replication. We arbitrarily fixed $\lambda = (1, ..., 1)'$ and compute $\pi_{\lambda,\alpha}^*$ for $\alpha$ on the grid:

$$\alpha \in \{7 \times 10^{-4}, 5 \times 10^{-4}, 3 \times 10^{-4}, 1 \times 10^{-4}, 7 \times 10^{-5}, 5 \times 10^{-5}, 1 \times 10^{-5}, 5 \times 10^{-6}, 1 \times 10^{-6}, 1 \times 10^{-7}\}$$

For each $\pi_{\lambda,\alpha}^*$, we compute the mean square error of the ICGMM estimator using the formula:

$$MSE(\alpha) = \frac{1}{L} \sum_{k=1}^{L} \left( \hat{\theta}_{\pi_{\lambda,\alpha}^*}^{(l)} - \theta_0 \right)' \left( \hat{\theta}_{\pi_{\lambda,\alpha}^*}^{(l)} - \theta_0 \right)$$

(76)

where $\hat{\theta}_{\pi_{\lambda,\alpha}^*}^{(l)}$ is defined in (48) and:

$$\theta_0 = \left( \mu_0, \mu_1, \delta, \beta, \kappa, \sigma, \sigma_c^2 \right)'$$

$$= \left( 0, 10^{-2}, -5 \times 10^{-2}, 1 \times 10^{-4}, 2 \times 10^{-2}, 5 \times 10^{-2}, 2 \times 10^{-4} \right)'$$

The following figure shows the plot of $MSE(\alpha)$ against $\alpha$. For this application, the mean square error is minimized for $\alpha^* = 10^{-4}$. We see that the graph of $MSE(\alpha)$ is L-shaped. The MSE increase faster when $\alpha$ moves from $\alpha^*$ to zero than when $\alpha$ moves in the opposite direction. This suggests that an overestimation of $\alpha$ is preferable to its underestimation.

Figure 1.

The mean square error of the ICGMM estimators of the ARFG model parameters as a function of the regularization parameter $\alpha$.

The following table shows the simulation results. The column labeled "Mean", "Median" and "Std. Dev" contain respectively the empirical mean, median and standard deviations of $\hat{\theta}_{\pi_{\lambda,\alpha}^*}^{(l)}$. IC1
and IC2 are respectively the lower and upper bound of the 90% confidence interval.
8 Empirical Application

The present empirical application is based the Autoregressive Variance Gamma model (ARVG) of order $p$ presented below. Unlike the in ARFG model, the ARVG model assumes that the risk factor is observed. Moreover, this risk factor is assumed to be the integrated volatility.

### 8.1 The Autoregressive Variance Gamma Model or Order $p$

The Autoregressive Variance Gamma model of order $p$ (henceforth $ARVG(p)$) specifies the return process $r_t$ as a function of the expected variance $E[V_t \mid \{V_{t-k}\}_{k=1}^p]$ and the innovation $V_t - E[V_t \mid \{V_{t-k}\}_{k=1}^p]$:

$$r_t = \mu_0 + \mu_1 \sqrt{E[V_t \mid \{V_{t-k}\}_{k=1}^p]} + \delta (V_t - E[V_t \mid \{V_{t-k}\}_{k=1}^p]) + \sqrt{V_t} \epsilon_t$$

(77)

where $\epsilon_t \sim N(0,1)$ is uncorrelated with past, current and future realizations of $V_t$, $\mu_1 \geq 0$ and $\delta \leq 0$. Like in the ARFG model considered in the previous section, the parameter $\mu_1$ captures the premium for bearing the expected risk while $\delta$ is the leverage effect. The variance $V_t$ is assumed to follow an Autoregressive Gamma process of order $p$ whose conditional density is given by:

$$f (V_t \mid \{V_{t-k}\}_{k=1}^\infty) = f (V_t \mid \{V_{t-k}\}_{k=1}^p)$$

$$= \sum_{j=0}^{\infty} \frac{V_t^{j+q-1} e^{j+q}}{\Gamma (j + q)} \exp (-cV_t) p_j (\{V_{t-k}\}_{k=1}^p)$$

The standard deviations of the estimators are small compared to their means, and the 90% confidence intervals contain the true values for all the parameters. Although the number of Monte Carlo replications $L$ and the number of frequency domain samples $S$ are quite moderate, the results of this experiment suggest that the proposed indirect inference method based on frequency domain resampling is a reliable inference method.
where \( p_j (\{V_{t-k}\}_{k=1}^p) \) are Poisson weights given by:

\[
p_j (\{V_{t-k}\}_{k=1}^p) = \left( \frac{c \sum_{k=1}^p \rho_k V_{t-k}}{j!} \right) \exp \left( -c \sum_{k=1}^p \rho_k V_{t-k} \right)
\]

The parameters of the model are \((\kappa, \beta, \sigma, \{\rho_1\}_{k=1}^p, \mu_0, \mu, \delta)\). In addition to \(\mu_1 \geq 0\) and \(\delta \leq 0\), we further have the constraints:

\[
c = \frac{2\kappa}{\sigma^2 \rho_0}, \quad q = \frac{2\beta \kappa}{\sigma^2},
\]

\[(\kappa, \beta, \sigma) > 0, \quad \{\rho_1\}_{k=0}^p \geq 0 \quad \text{and} \quad \sum_{k=1}^p \rho_k = 1.\]

The specified dynamic for \(V_t\) extends the model of Gourieroux and Jasiak (2005) which is an autoregressive Gamma or order one. The conditional CF of \(V_t\) is an exponential affine form given by:

\[
E \left[ e^{i\tau V_t} \mid \{V_{t-k}\}_{k=1}^p \right] = \left( 1 - \frac{i\tau}{c} \right)^{-q} \exp \left( \frac{i\tau}{1 - \frac{i\tau}{c}} \sum_{k=1}^p \rho_k V_{t-k} \right) \quad (78)
\]

This CF shows that the Autoregressive Gamma model of order \(p\) is identical (up to a re-parameterization) to a univariate Wishart autoregressive process of order \(p\) discussed in Gourieroux, Jasiak and Sufana (2005).\(^4\) The following moments can be computed by using the two first derivatives of the above conditional CF evaluated at zero:\(^5\):

\[
E \left[ V_t \mid \{V_{t-k}\}_{k=1}^p \right] = \beta \rho_0 + \sum_{k=1}^p \rho_k V_{t-k} \quad (79)
\]

\[
Var \left[ V_t \mid \{V_{t-k}\}_{k=1}^p \right] = \frac{1}{c} \left[ \beta \rho_0 + 2 \sum_{k=1}^p \rho_k V_{t-k} \right] \quad (80)
\]

Thus, the conditional mean and variance of \(V_t\) are linear in lagged realizations. Moreover, the ARVG model has the potential to generate asymmetry and fat tails. In fact we have:

\[
E \left[ (r_t - E [r_t \mid \{V_{t-k}\}_{k=1}^p])^3 \mid V_{t-1} \right] = \delta^3 E \left[ (V_t - E [V_t \mid V_{t-1}])^3 \mid V_{t-1} \right]
\]

so that the return process has a negative and time varying skewness whenever \(\delta < 0\). In the specific case where \(p = 1\), \(\kappa = -\log \rho_1\) and \(\delta = 0\) the conditional excess kurtosis of \(r_t\) is given by:

\[
\frac{E \left[ \left( r_t - E [r_t \mid V_{t-1}] \right)^4 \mid V_{t-1} \right]}{Var \left[ r_t \mid V_{t-1} \right]^2} - 3 = \frac{3Var \left[ V_t \mid V_{t-1} \right]}{Var \left[ r_t \mid V_{t-1} \right]^2}
\]

Equation (79) provides a good forecasting formula for the volatility.

\(^4\)See Gourieroux, Jasiak and Sufana (2005), Section 2.3, Definition 2.

\(^5\)The noncentered conditional moments of \(V_t\) are given by:

\[
E \left[ (V_t)^n \mid \{V_{t-k}\}_{k=1}^p \right] = \left. \frac{\partial^n E [e^{i\tau \xi} \mid \{V_{t-k}\}_{k=1}^p]}{\partial \tau^n} \right|_{\tau = 0}
\]
Estimation of the ARVG(p) Using High Frequency Data

Unlike the ARFG model previously discussed, the ARVG(p) satisfies:

\[ \text{Var} \left[ r_t \mid \{V_{t-k}\}_{k=1}^p \right] = V_t \]

Hence if we let \( V_t \equiv \int_{t-1}^t \sigma_s^2 ds \) where \( \{\sigma_s\} \) is a spot volatility process, this equation becomes:

\[ \text{Var} \left[ r_t \mid \{\sigma_s\}_{s=-\infty}^\infty \right] = \int_{t-1}^t \sigma_s^2 ds. \]

The above equation is a standard implication of continuous time models of assets (log) prices. We will use this argument to proxy \( V_t \) by a good estimator of the integrated volatility, as in Kotchoni (2010).

The estimation may be done in two steps. In the first step, we estimate by CGMM an Autoregressive Gamma model for \( V_t \), using the moment function:

\[ h_t(\tau, \theta_1) = (\exp(i\tau V_t) - E[\exp(i\tau V_t) \mid \{V_{t-k}\}_{k=1}^p]) \exp \left( \sum_{k=1}^p i\tau_{k+1}V_{t-k} \right), \]

where \( E[\exp(i\tau V_t) \mid \{V_{t-k}\}_{k=1}^p] \) is given by (78), \( \tau = (\tau_1, ..., \tau_{p+1}) \) and \( \theta_1 = (\rho_0, ..., \rho_p, \kappa, \beta, \sigma^2) \). In the estimation process, \( V_t \) is replaced by any good estimator of the integrated volatility, e.g. the realized kernels of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008) or the shrinkage realized kernels of Carrasco and Kotchoni (2011).

Having computed \( \hat{\theta}_1 \), the expected variance \( \hat{V}_t \) is estimated by:

\[ \hat{V}_t = \hat{\beta}\hat{\rho}_0 + \sum_{k=1}^p \hat{\rho}_k V_{t-k} \]

The remaining set of parameters \( \theta_2 = (\mu_0, \mu_1, \delta) \) can then be estimated in the second step by Gaussian maximum likelihood based on the distribution of \( \varepsilon_t \), where the following proxy is used for \( \varepsilon_t \):

\[ \hat{\varepsilon}_t = V_t^{-1/2} \left[ r_t - \mu_0 - \mu_1 \hat{V}_t^{1/2} - \delta \left( V_t - \hat{V}_t \right) \right] \sim N(0, 1) \]

Below, the ARVG(p) is implemented with real data.

An Application with the Alcoa Index

The data used in this section are the transaction prices of Alcoa, an index listed in the Dow Jones Industrials. The prices are observed every minute from January 1st, 2002 to December 31st, 2007 (T = 1510 trading days). In a typical trading day, the market is open from 9:30 am to 4:00 pm, and this results in \( m = 390 \) observations per day. There are a few missing observations (less than 5 missing data per day) which we filled in using the previous tick method. Following Kotchoni (2010), we construct the proxy of \( V_t \) using the shrinkage realized kernels of Carrasco and Kotchoni (2011).

The implementation of the ICGMM is conducted exactly as in the previous section, except
that the Monte Carlo step is replaced by a resampling with replacement from the set of moment functions computed with the actual data, as illustrated by Equation (??). We resample $L = 100$ times in the time domain and $S = 50$ times in the frequency domain. Finally, we set $p = 30$ (six weeks) in order to assess the level of persistence of the volatility process.

To select the regularization parameter $\alpha$, we minimize the following tracking error:

$$MSE(\alpha) = \frac{1}{T} \sum_{t=1}^{T} \left[ V_t - \hat{\beta} \left( 1 - \sum_{k=1}^{p} \hat{\rho}_k \right) - \sum_{k=1}^{30} \hat{\rho}_k V_{t-k} \right]^2$$

The following graph suggests that the optimal regularization parameter is around $\alpha^* = 10^{-4}$ for these data. We compute the optimal weighting function using this value of $\alpha^*$.

Figure 2.
The mean square error of the ICGMM estimators of the ARVG model parameters as a function of the regularization parameter $\alpha$.

Table 3 shows the summary of the results for the parameters $(\kappa, \beta, \sigma^2)$ and $(\mu_0, \mu_1, \delta)$. One important difference between the current results and those of the case $p = 1$ presented in Kotchoni (2010) is that the estimates of $\kappa$ are considerably lower here. The variances of the estimators are relatively high due the the fact that they are estimated using only $L = 100$ frequency domain samples. Accordingly, the confidence intervals are also large.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\theta}_1$</th>
<th>$\hat{\theta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\kappa}$</td>
<td>$\hat{\beta}$</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0528</td>
<td>0.0017</td>
</tr>
<tr>
<td>Median</td>
<td>0.1358</td>
<td>0.0003</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.3656</td>
<td>0.0023</td>
</tr>
<tr>
<td>IC1(95)</td>
<td>-0.8730</td>
<td>-0.0000</td>
</tr>
<tr>
<td>IC2(95)</td>
<td>0.4516</td>
<td>0.0060</td>
</tr>
</tbody>
</table>

Below, we plot the estimators of the autoregressive coefficients $(\rho_0, \ldots, \rho_{30})$. 

29
The volatility strongly responds to its own lags lying within one week, which is consistent with volatility clustering (See Mandelbrot, 1963 and Cont, 2005). There also seems to be some responses of smaller magnitude to lags lying between 20 and 25 days. The following graph shows the volatility and its estimated expectation conditional on past realizations.

The conditional expectation of the volatility is quite smooth compared to the actual series. In fact, the Autoregressive Gamma model ignores the erratic fluctuations and jumps of the volatility and tracks the trend. This suggest that the current model may be used to decompose the volatility into its continuous component and its noise plus jump component.
9 Conclusion

The CGMM introduced by Carrasco and Florens (2000) has the potential to deliver estimators that are as efficient as maximum likelihood estimators. However, the objective function of the characteristic function based CGMM involves as many integrals as the dimensionality of the observations. Unfortunately, the complexity of the numerical integration grows as an exponential function of this dimensionality, and this makes the CGMM unfeasible in multivariate models with three and more dimensions. To circumvent this "curse of dimensionality", we propose a solution that consists of converting the unfeasible multivariate optimization problem into several suitably designed univariate optimization problems. The ideal solutions of the unfeasible multivariate problem is then recovered by computing a weighted average of the solutions to the univariate problems. The overall procedure is termed indirect CGMM estimator (ICGMM).

We derived the optimal aggregation weight for the ICGMM estimator and compare its efficiency to that of the MLE. The conditions under which the optimal ICGMM estimator is as efficient as the MLE are discussed. Three illustrations are then proposed. The first illustration is a Monte Carlo simulation study based on a Gaussian AR(1) model. This model admits closed form expressions for the MLE, and hence, allows to compare the performance of the ICGMM estimator to the MLE. The second illustration is a Monte Carlo study based on the autoregressive factor gamma model. In this model, the return of a financial asset depends linearly in the realizations of a latent autoregressive gamma risk factor. The latent factor is not observed and must be integrated out, and this results in a non Markov model for the observed returns. These two simulation studies are conclusive and demonstrate the effectiveness of the ICGMM procedure. The third illustration is an empirical application based on the Autoregressive Variance Gamma model of order $p$. The results of this application grossly support the results of French and al (1987) stating that the expected return of financial assets is positively correlated with the expected risk while the unexpected return is negatively correlated with the unpredictable risk. The empirical results also provide evidence of volatility clustering.

10 Acknowledgement

The author wish to thanks Marine Carrasco, Russel Davidson, Pierre Duchesne, Marc Henry, Bruno Larue, Pierre Evariste Nguimkeu and Benoît Perron for useful comments.
References


Appendix: Proofs

Proof of Proposition 1: Under Assumptions 1 to 4, \( \hat{\theta}(\tau) \) is a well-defined consistent estimator of \( \theta_0 \). Furthermore, Assumption 6(i) ensures \( \hat{\theta}(\tau) \) is the unique minimizer of \( Q_{\tau,T} \). According to Assumptions 3 and 5, \( Q_{\tau,T}(\theta) \) is twice continuously differentiable with respect to \( \tau \) and \( \theta \). Hence by the implicit function theorem, \( \hat{\theta}(\tau) \) is continuously differentiable with respect to \( \tau \) and we have:

\[
\frac{\partial \hat{\theta}^{(1)}(\tau)}{\partial \tau} = -\left[ \frac{\partial^2 Q_{\tau,T}}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial^2 Q_{\tau,T}}{\partial \theta \partial \tau}.
\]

where the invertibility of \( \frac{\partial^2 Q_{\tau,T}}{\partial \theta \partial \theta'} \) follows from Assumption 6(i).\( \blacksquare \)

Proof of Proposition 2: According to Assumptions 1 to 4, each CGMM estimator \( \hat{\theta}(\tau) \) is consistent for \( \theta_0 \) (See for example Carrasco and Kotchoni, 2009). Hence for any choice of measure \( \pi \) satisfying Assumption 7, we have:

\[
\begin{align*}
Var_{\pi}(X\hat{\theta}) &= \int \int X' Cov\left( \hat{\theta}(\tau), \hat{\theta}(\tau) \right) \lambda \pi(\tau_1, \tau_2) d\tau_1 d\tau_2 \\
&\leq \max_{\tau} \left[ \lambda' Var\left( \hat{\theta}(\tau) \right) \right] \int \int \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2 \\
&= \max_{\tau} \left[ \lambda' Var\left( \hat{\theta}(\tau) \right) \right].
\end{align*}
\]

The result follows from \( \max_{\tau} \left[ \lambda' Var\left( \hat{\theta}(\tau) \right) \right] = O_p(T^{-1}) \) for \( \tau \in \Lambda \setminus \mathbb{N} \).\( \blacksquare \)

Proof of Proposition 3: The ideal measure \( \pi^*_\lambda(\tau) \) solves:

\[
\pi^*_\lambda = \arg \min_{\pi} \int \int g_\lambda(\tau_1, \tau_2) \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2,
\]

subject to:

\[
\int \pi(\tau_1) d\tau_1 = 1
\]

where \( g_\lambda(\tau_1, \tau_2) = X' Cov\left( \hat{\theta}(\tau), \hat{\theta}(\tau) \right) \lambda \). The Lagrangian for this problem is given by:

\[
\mathcal{L}(\pi) = \int \int g_\lambda(\tau_1, \tau_2) \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2 + \mu_\lambda \left( 1 - \int \pi(\tau_1) d\tau_1 \right),
\]

where \( \mu_\lambda \) is a Lagrange multiplier. The first order necessary condition for this problem is obtained by differentiating \( \mathcal{L}(\pi) \) with respect to \( \pi(\tau_1) \):

\[
\int g_\lambda(\tau_1, \tau_2) \pi(\tau_2) d\tau_2 = \mu_\lambda I(\tau_1),
\]

where \( I(\tau_1) = 1 \) for all \( \tau_1 \) in \( \Lambda \setminus \mathbb{N} \).

Let \( V_\lambda \) be the linear operator with kernel \( v_\lambda(\tau_1, \tau_2) \). The first order condition becomes:

\[
V_\lambda \pi(\tau_1) = \mu_\lambda I(\tau_1)
\]

(83)
Because $V_\lambda$ is compact a covariance operator, it has a discrete nonnegative spectrum with orthogonal eigenfunctions. Let $\phi_{\lambda,j}(\tau)$ be the eigenfunction of $V_\lambda$ associated with the eigenvalue $\nu_{\lambda,j}$. For any function $f(\tau), \tau_1 \in \Lambda$, we have:

$$ f(\tau_1) = \sum_{j=1}^{\infty} \left( \int \phi_{\lambda,j}(\tau) f(\tau) d\tau \right) \phi_{\lambda,j}(\tau_1) + \tilde{f}(\tau_1), $$

where $\tilde{f}$ is in the null set of $V_\lambda$ so that $V_\lambda \tilde{f}(\tau_1) = 0$. Hence if $f_0(\tau_1)$ solves (83) for $\pi(\tau_1)$, then $f(\tau_1) = f_0(\tau_1) + \tilde{f}(\tau_1)$ also solves (83) for $\pi(\tau_1)$. The solution of (83) with minimal norm is the one in which $\tilde{f}(\tau_1) = 0$. This is given by:

$$ \pi_\lambda^*(\tau_1) = \mu_\lambda V_\lambda^{-1} I(\tau_1) = \mu_\lambda \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left( \int \phi_{\lambda,j}(\tau) d\tau \right) \phi_{\lambda,j}(\tau_1). $$

The Lagrange multiplier is identified using the constraint $\int \pi(\tau_1) d\tau_1 = 1$. This yields:

$$ \mu_\lambda = \left[ \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left( \int \phi_{\lambda,j}(\tau) d\tau \right)^2 \right]^{-1}. $$

We substitute this for $\mu_\lambda$ in $\pi_\lambda^*(\tau_1)$ to obtain:

$$ \pi_\lambda^*(\tau_1) = \left[ \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left( \int \phi_{\lambda,j}(\tau) d\tau \right)^2 \right]^{-1} \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left( \int \phi_{\lambda,j}(\tau) d\tau \right) \phi_{\lambda,j}(\tau_1). $$

At the solution $\pi_\lambda^*(\tau_1) = \mu_\lambda V_\lambda^{-1} I(\tau_1)$, we have:

$$ Var \left( \lambda \hat{\theta}_{\pi_\lambda^*} \right) = \mu_\lambda \int \left[ \int g_\lambda(\tau_1, \tau_2) \pi_\lambda^*(\tau_1) d\tau_1 \right] \pi_\lambda^*(\tau_2) d\tau_2 = \mu_\lambda \int \pi_\lambda^*(\tau_2) d\tau_2 = \mu_\lambda $$

**Proof of Proposition 4:** By Proposition 1 (which holds under Assumptions 1 to 6), $\hat{\theta}(\tau)$ is continuously differentiable with respect to $\tau$ and we have:

$$ \frac{\partial^2 Q_{\tau,i}}{\partial \tau^2} = - \left[ \frac{\partial^2 Q_{\tau,i}}{\partial \hat{\theta}_{\tau,i} \partial \hat{\theta}_{\tau,j}} \right]^{-1} \frac{\partial^2 Q_{\tau,i}}{\partial \hat{\theta}_{\tau,i} \partial \hat{\theta}_{\tau,j}}. $$

35
Around a particular \( \tau \) in \( \Lambda \setminus \mathcal{N} \), we have:

\[
\hat{\theta}(\tau + \tau_0) = \hat{\theta}(\tau) + \frac{\partial \hat{\theta}(\tau)}{\partial \tau} \tau_0.
\]

By assumption 6(ii), \( \frac{\partial^2 Q_{x,T}}{\partial \theta \partial \tau} \) is of full rank so that \( \frac{\partial \hat{\theta}(\tau)}{\partial \tau} \) is also of full rank. This implies that for \( d \geq \max\{q, 2\} \), \( q \) linearly independent vectors of type \( \hat{\theta}_T(\tau + \tau_0) \) can be constructed by varying \( \tau_0 \). As a consequence, the manifold \( \hat{D}_T(\theta_0) \) defined by:

\[
\hat{D}_T(\theta_0) = \left\{ \theta \in \mathbb{R}^q \text{ s.t } \theta = \int \pi(\tau) \hat{\theta}(\tau) d\tau \text{ and } \int \pi(\tau) d\tau = 1 \right\}
\]

has exactly \( q \) dimensions. In particular, there exist a basis \( \hat{\theta}_{(j)} \), \( j = 1, \ldots, q \) such that \( \hat{\theta}_{MLE} = \sum w_j \hat{\theta}_{(j)} \in \hat{D}_T(\theta_0) \). Hence \( Var \left( \lambda' \hat{\theta}_{MLE} \right) \leq Var \left( \lambda' \hat{\theta}_{MLE} \right) \).

**Proof of Proposition 5:** We recall that:

\[
\hat{V}_\lambda = \frac{T}{SL} \left( \hat{\Theta}_\lambda - \bar{\Theta}_\lambda \right)' \left( \hat{\Theta}_\lambda - \bar{\Theta}_\lambda \right),
\]

with the \((i,j)\) element of \( \hat{V}_\lambda \) given by:

\[
\hat{g}_{\lambda}(\tau_i, \tau_j) = \frac{T}{SL} \sum_{i=1}^L \lambda' \left( \hat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left( \hat{\theta}(\tau_j, l) - \bar{\theta}(\tau_j, l) \right)' \lambda.
\]

and \( \lambda = (f(\tau_1), \ldots, f(\tau_S))' \), where \( \hat{\theta}(\tau_i, l) \) is IID across \( l \). We have:

\[
\begin{aligned}
\left( \hat{V}_\lambda f \right)_i &= \sum_{j=1}^S \hat{g}_{\lambda}(\tau_i, \tau_j) f(\tau_j) \\
&= \frac{1}{S} \sum_{j=1}^S \left( \frac{T}{L} \sum_{i=1}^L \lambda' \left( \hat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left( \hat{\theta}(\tau_j, l) - \bar{\theta}(\tau_j, l) \right)' \lambda \right) f(\tau_j)
\end{aligned}
\]

Note that \( \tau_i, i = 1, \ldots, S \) are drawn from the multivariate uniform distribution of \( \Lambda \). Hence by the Law of Large Numbers, as \( S \) goes to infinity, we have:

\[
\begin{aligned}
\left( \hat{V}_\lambda f \right)_i &= \frac{1}{S} \sum_{j=1}^S \left( \frac{T}{L} \sum_{i=1}^L \lambda' \left( \hat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left( \hat{\theta}(\tau_j, l) - \bar{\theta}(\tau_j, l) \right)' \lambda \right) f(\tau_j) \\
&\to \int_{\Lambda} \left( \frac{T}{L} \sum_{i=1}^L \lambda' \left( \hat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left( \hat{\theta}(\tau, l) - \bar{\theta}(\tau, l) \right)' \lambda \right) f(\tau) d\tau,
\end{aligned}
\]

36
and
\[
\left( \tilde{V}_\lambda f \right)_i - \int_A \left( \frac{T}{L} \sum_{i=1}^L \lambda' \left( \bar{\theta}(\tau_i, l) - \tilde{\theta}(\tau_i, l) \right) \left( \bar{\theta}(\tau, l) - \tilde{\theta}(\tau, l) \right)' \lambda \right) f(\tau) d\tau \\
= O_p \left( S^{-1/2} \right).
\]

Next, \( \tilde{\theta}(\tau_i, l) \) is IID across \( l = 1, \ldots, L \). Hence, as \( L \) goes to infinity, we have:
\[
\int_A \left( \frac{T}{L} \sum_{i=1}^L \lambda' \left( \bar{\theta}(\tau_i, l) - \tilde{\theta}(\tau_i, l) \right) \left( \bar{\theta}(\tau, l) - \tilde{\theta}(\tau, l) \right)' \lambda \right) f(\tau) d\tau \\
- \int_A \lambda' \text{Cov} \left( \sqrt{T} \bar{\theta}(\tau_i, l), \sqrt{T} \tilde{\theta}(\tau, l) \right) \lambda f(\tau) d\tau \\
= O_p \left( L^{-1/2} \right).
\]

This result holds because \( f \) is continuous and hence, \( \int_A f(\tau) d\tau \) is finite. We obtain the desired result by replacing the second asymptotic in the first one.

**Proof of Proposition 6:** We have:
\[
\left\| \left( \tilde{V}_\lambda^2 + \alpha I \right)^{-1} \tilde{V}_\lambda - (V_\lambda^2 + \alpha I)^{-1} V_\lambda \right\| \leq \\
\left\| \left( \tilde{V}_\lambda^2 + \alpha I \right)^{-1} (\tilde{V}_\lambda - V_\lambda) \right\| + \left\| \left( \tilde{V}_\lambda^2 + \alpha I \right)^{-1} V_\lambda - (V_\lambda^2 + \alpha I)^{-1} V_\lambda \right\| \leq \\
\left\| \left( \tilde{V}_\lambda^2 + \alpha I \right)^{-1} \right\| \left\| \tilde{V}_\lambda - V_\lambda \right\| + \left\| \left[ \left( \tilde{V}_\lambda^2 + \alpha I \right)^{-1} - (V_\lambda^2 + \alpha I)^{-1} \right] V_\lambda \right\| \leq \\
\leq \alpha^{-1} = O_p \left( L^{-1/2} \right) + O_p \left( S^{-1/2} \right)
\]

Hence:
\[
\left\| \left[ \tilde{V}_\lambda^2 + \alpha I \right]^{-1} - (V_\lambda^2 + \alpha I)^{-1} V_\lambda \right\| \leq \left\| \left[ \tilde{V}_\lambda^2 + \alpha I \right]^{-1} \right\| \left\| (V_\lambda^2 - \tilde{V}_\lambda^2) \right\| \left\| V_\lambda^2 + \alpha I \right\|^{-1/2} \left\| (V_\lambda^2 + \alpha I)^{-1/2} V_\lambda \right\| \leq \alpha^{-1} = O_p \left( L^{-1/2} \right) + O_p \left( S^{-1/2} \right) \leq \alpha^{-1/2} \left\| (V_\lambda^2 + \alpha I)^{-1/2} V_\lambda \right\|^{-1}
\]

To obtain (45), we use Assumption 9. We rewrite (45) as
\[
\left\| \left( \tilde{V}_{\lambda, \alpha}^{-1} - V_{\lambda, \alpha}^{-1} \right) V_\lambda \right\| = \left\| \left( \tilde{V}_{\lambda, \alpha}^{-1} - V_{\lambda, \alpha}^{-1} \right) V_\lambda \right\| \leq \left\| \left( \tilde{V}_{\lambda, \alpha}^{-1} - V_{\lambda, \alpha}^{-1} \right) V_\lambda \right\| \left\| V_\lambda^{-1} t \right\| = \left\| \left( \tilde{V}_{\lambda, \alpha}^{-1} - V_{\lambda, \alpha}^{-1} \right) V_\lambda \right\| \left\| V_\lambda^{-1} t \right\|.
\]

37
We have
\[
\left( \tilde{V}_{\lambda}^{-1} - V_{\lambda}^{-1} \right) V_{\lambda} \\
= \left( \tilde{V}_{\lambda}^2 + \alpha I \right)^{-1} \tilde{V}_{\lambda} V_{\lambda} - \left( V_{\lambda}^2 + \alpha I \right)^{-1} V_{\lambda}^2 \\
= \left( \tilde{V}_{\lambda}^2 + \alpha I \right)^{-1} \left( \tilde{V}_{\lambda} - V_{\lambda} \right) V_{\lambda} \\
+ \left[ \left( \tilde{V}_{\lambda}^2 + \alpha I \right)^{-1} - \left( V_{\lambda}^2 + \alpha I \right)^{-1} \right] V_{\lambda}^2. 
\]
(84)

The term (84) can be bounded in the following manner
\[
\left\| \left( \tilde{V}_{\lambda}^2 + \alpha I \right)^{-1} \left( \tilde{V}_{\lambda} - V_{\lambda} \right) V_{\lambda} \right\| \leq \left\| \left( \tilde{V}_{\lambda}^2 + \alpha I \right)^{-1} \right\| \left\| \tilde{V}_{\lambda} - V_{\lambda} \right\| \left\| V_{\lambda} \right\| \\
= O_p \left( \alpha^{-1} L^{-1/2} \right) + O_p \left( \alpha^{-1} S^{-1/2} \right). 
\]

For the term (85), we use the fact that $A^{-1/2} - B^{-1/2} = A^{-1/2} \left( B^{1/2} - A^{1/2} \right) B^{-1/2}$. It follows that
\[
\left\| \left[ \left( \tilde{V}_{\lambda}^2 + \alpha I \right)^{-1} - \left( V_{\lambda}^2 + \alpha I \right)^{-1} \right] V_{\lambda}^2 \right\| \\
= \left\| \left( \tilde{V}_{\lambda}^2 + \alpha I \right)^{-1} \left( V_{\lambda}^2 - \tilde{V}_{\lambda}^2 \right) \left( V_{\lambda}^2 + \alpha I \right)^{-1} V_{\lambda}^2 \right\| \\
\leq \left\| \left( \tilde{V}_{\lambda}^2 + \alpha I \right)^{-1} \right\| \left\| V_{\lambda}^2 - \tilde{V}_{\lambda}^2 \right\| \left\| \left( V_{\lambda}^2 + \alpha I \right)^{-1} V_{\lambda}^2 \right\| \\
\leq O_p \left( \alpha^{-1} L^{-1/2} \right) + O_p \left( \alpha^{-1} S^{-1/2} \right). 
\]

Now we turn our attention to the equation (46). We can write
\[
(V_{\lambda}^2 + \alpha I)^{-1} V_{\lambda} t - V_{\lambda}^{-1} t = \sum_{j=1}^{\infty} \left[ \frac{v_j}{\alpha + v_j^2} - \frac{1}{v_j^2} \right] \left< t, \phi_j \right> \phi_j \\
= \sum_{j=1}^{\infty} \left( \frac{v_j}{\alpha + v_j^2} - 1 \right) \frac{\left< t, \phi_j \right>}{v_j} \phi_j. 
\]

We now take the norm:
\[
(46) = \left\| \left( V_{\lambda}^2 + \alpha I \right)^{-1} V_{\lambda} t - V_{\lambda}^{-1} t \right\| \\
= \left( \sum_{j=1}^{\infty} \left( \frac{v_j}{\alpha + v_j^2} - 1 \right) \frac{\left< t, \phi_j \right>}{v_j^2} \right)^{1/2} \\
= \left( \sum_{j=1}^{\infty} \frac{v_j^{2\epsilon -2}}{\alpha + v_j^2} \left( \frac{v_j^2}{\alpha + v_j^2} - 1 \right) \frac{\left< t, \phi_j \right>}{v_j^{2\epsilon}} \right)^{1/2} \\
\leq \left( \sum_{j=1}^{\infty} \left( \frac{\left< t, \phi_j \right>}{v_j^{2\epsilon}} \right)^2 \sup_{1 \leq j \leq \infty} v_j^{2\epsilon -1} \left| \frac{v_j^2}{\alpha + v_j^2} - 1 \right| \right)^{1/2}. 
\]

Recall that as $K$ is a compact operator, its largest eigenvalue $\nu_1$ is bounded. We need to find an equivalent to
\[ \sup_{0 \leq \nu \leq \nu_1} \nu^{\epsilon - 1} \left( 1 - \frac{\nu^2}{\alpha + \nu^2} \right) = \sup_{0 \leq \lambda \leq \nu_1^2} \lambda^{\frac{\epsilon - 1}{2}} \left( 1 - \frac{1}{\alpha/\lambda + 1} \right) \] (86)

Case where $\epsilon \leq 3/2$

We apply another change of variables to the objective function (86), $x = \alpha/\lambda$ and obtain
\[ \sup_{x \geq 0} x^{\epsilon/2 - 1/2} \left( 1 - \frac{1}{x + 1} \right). \]

We see that an equivalent to (86) is $\alpha^{\epsilon/2 - 1/2}$ provided that
\[ \sup_{x \geq 0} x^{\epsilon/2 - 1/2} \left( 1 - \frac{1}{x + 1} \right) \]
is bounded. We study the properties of $g(x) \equiv \frac{1}{x^{\epsilon/2 - 1/2}} \left( 1 - \frac{1}{x + 1} \right).$ Note that $g(x)$ is continuous and therefore bounded on any interval of $(0, +\infty)$. It remains to study its behavior at 0 and $+\infty$. It goes to 0 at $+\infty$ (for any $\epsilon > 1$). For the limit at 0, we apply l’Hopital’s rule and obtain
\[ g(x) \sim \frac{1}{0} \left( \frac{\epsilon - 1}{2} \right) x^{\frac{\epsilon - 3}{2}} = 0, \]
provided $\epsilon < 3/2$. For $\epsilon = 3/2$, we have $g(x) \sim \frac{1}{0} \left( \frac{\epsilon - 1}{2} \right)$. Hence $g(x)$ is bounded on $\mathbb{R}^+$ for all $\epsilon \leq 3/2$.

Case where $\epsilon > 3/2$

We rewrite (86) as
\[ \sup_{0 \leq \lambda \leq \nu_1^2} \alpha \lambda^{\frac{\epsilon - 3}{2}} \left[ \frac{\left( 1 - \frac{1}{\alpha/\lambda + 1} \right)}{\left( \frac{1}{\alpha/\lambda} \right)} \right]. \]
The term $\lambda^{\frac{\epsilon - 3}{2}} \left[ \frac{\left( 1 - \frac{1}{\alpha/\lambda + 1} \right)}{\left( \frac{1}{\alpha/\lambda} \right)} \right]$ is the product of an increasing function of $\lambda$, namely $\lambda^{\frac{\epsilon - 3}{2}}$ (which is bounded because $\lambda$ is bounded) and a function of the form $\frac{1 - \frac{1}{x + 1}}{x}$. It is easy to show using the l’Hopital’s rule that $\frac{1 - \frac{1}{x + 1}}{x}$ is bounded on $\mathbb{R}^+$:
\[ \left( 1 - \frac{1}{x + 1} \right) / x \sim 0 \text{ and } \left( 1 - \frac{1}{x + 1} \right) / x \sim 1 \]

Hence the rate of (86) is given by $\alpha$.

Finally, $f \in \Phi_\epsilon$:
\[ (46) = O \left( \alpha^{\min(1, \frac{\epsilon - 1}{2})} \right). \]
Proof of Proposition 7: Let \( \hat{\phi}_{\lambda,j} \) be the eigenfunctions of \( \hat{V}_\lambda \) associated with eigenvalues \( \hat{\nu}_{\lambda,j} \). Then we have:

\[
\tilde{V}^{-1}_\alpha = \left( \hat{V}^2 + \alpha \mathbf{I} \right)^{-1} \hat{V}_t = \sum_{i=1}^{S} \frac{\hat{\nu}_{\lambda,j}}{\alpha + \hat{\nu}_{\lambda,j}^2} \left( \sum_{k=1}^{S} \hat{\phi}_{\lambda,j,k} \right) \hat{\phi}_{\lambda,i}
\]

Hence

\[
\tilde{\pi}_\alpha^* = S \left( t' \tilde{V}_\alpha^{-1} t \right)^{-1} t' \tilde{V}_\alpha^{-1}
\]

\[
= S \left[ \sum_{i=1}^{S} \frac{\hat{\nu}_{\lambda,j}}{\alpha + \hat{\nu}_{\lambda,j}^2} \left( \sum_{k=1}^{S} \hat{\phi}_{\lambda,j,k} \right) \right]^{-1} \sum_{i=1}^{S} \frac{\hat{\nu}_{\lambda,j}}{\alpha + \hat{\nu}_{\lambda,j}^2} \left( \sum_{k=1}^{S} \hat{\phi}_{\lambda,j,k} \right) \hat{\phi}_{\lambda,i}
\]

According to proposition 5, \( \hat{V} \) is consistent for \( V \) as \( L \) and \( S \) go to infinity. Hence the eigenfunctions \( \hat{\phi}_{\lambda,i} \) and eigenvalues \( \hat{\nu}_{\lambda,j} \) are consistent for their theoretical counterpart of \( \phi_{\lambda,i} \) and \( \nu_{\lambda,j} \), and the rate of convergence are the same as that of \( \hat{V} \). See Carrasco and Florens (2000).

Thus, as \( L \) and \( S \) go to infinity and \( \alpha \) is fixed, we have \( \tilde{\pi}_\alpha^* - \pi_{\lambda,\alpha}^* (\tau) = O_p \left( L^{-1/2} \right) + O_p \left( S^{-1/2} \right) \), where

\[
\pi_{\lambda,\alpha}^* (\tau) = \sum_{j=1}^{\infty} \frac{\nu_{\lambda,j}}{\alpha + \nu_{\lambda,j}^2} \left( \int \phi_{\lambda,j} (\tau_1) d\tau_1 \right)^{-1} \int \frac{\nu_{\lambda,j}}{\alpha + \nu_{\lambda,j}^2} \left( \int \phi_{\lambda,j} (\tau_1) d\tau_1 \right) \phi_{\lambda,j} (\tau).
\]

If \( \alpha \) is not fixed, the one should note that \( \tilde{\pi}_\alpha^* \) is a function of \( \tilde{V}_{\lambda,\alpha}^{-1} \), and hence, that the rate of convergence of \( \tilde{\pi}_\alpha^* \) cannot be faster than the rate of \( \left( \tilde{V}_{\lambda,\alpha}^{-1} - V_{\lambda}^{-1} \right) t \) given in (47). This implies:

\[
\tilde{\pi}_\alpha^* - \pi_{\lambda,\alpha}^* (\tau) = O \left( \alpha^{\min\left(1, \frac{1-t}{2} \right)} \right) + O_p \left( \alpha^{-1} L^{-1/2} \right) + O_p \left( \alpha^{-1} S^{-1/2} \right)
\]

Now, we show the asymptotic equivalence of \( \hat{\theta}_{\tilde{\pi}_{\lambda,\alpha}}^* \) to \( \theta_{\tilde{\pi}_{\lambda}}^* \). We have:

\[
\hat{\theta}_{\tilde{\pi}_{\lambda,\alpha}} = \frac{1}{S} \sum_{s=1}^{S} \tilde{\pi}_{\lambda,\alpha}^* (\tau_s) \hat{\theta} (\tau_s)
\]

\[
= \frac{1}{S} \sum_{s=1}^{S} \pi_{\lambda}^* (\tau_s) \hat{\theta} (\tau_s) + \frac{1}{S} \sum_{s=1}^{S} \left( \tilde{\pi}_{\lambda,\alpha}^* (\tau_s) - \pi_{\lambda}^* (\tau_s) \right) \hat{\theta} (\tau_s)
\]

\[
= \hat{\theta}_{\pi_{\lambda}} + \frac{1}{S} \sum_{s=1}^{S} \left( \tilde{\pi}_{\lambda,\alpha}^* (\tau_s) - \pi_{\lambda}^* (\tau_s) \right) \hat{\theta} (\tau_s)
\]
Hence:

\[
\hat{\theta}_{\tilde{\pi}_{\lambda,\alpha}} - \hat{\theta}_{\pi_{\lambda}} = \frac{1}{S} \sum_{s=1}^{S} \left( \tilde{\pi}_{\lambda,\alpha}^* (\tau_s) - \pi_{\lambda}^* (\tau_s) \right) \hat{\theta} (\tau_s)
\]

\[
= \theta_0 \frac{1}{S} \sum_{s=1}^{S} \left( \tilde{\pi}_{\lambda,\alpha}^* (\tau_s) - \pi_{\lambda}^* (\tau_s) \right) + \frac{1}{S} \sum_{s=1}^{S} \left( \tilde{\pi}_{\lambda,\alpha}^* (\tau_s) - \pi_{\lambda}^* (\tau_s) \right) \left( \hat{\theta} (\tau_s) - \theta_0 \right)
\]

\[
\rightarrow \int_{\Lambda} \left( \tilde{\pi}_{\lambda,\alpha}^* (\tau) - \pi_{\lambda}^* (\tau) \right) \left( \hat{\theta} (\tau) - \theta_0 \right) d\tau,
\]

by the Law of Large Numbers. By the Cauchy-Schartz Inequality, we have:

\[
\left( \hat{\theta}_{\tilde{\pi}_{\lambda,\alpha}} - \hat{\theta}_{\pi_{\lambda}} \right)_j \leq \left( \int_{\Lambda} \left( \tilde{\pi}_{\lambda,\alpha}^* (\tau) - \pi_{\lambda}^* (\tau) \right)^2 d\tau \right)^{1/2} \left( \int_{\Lambda} \left( \hat{\theta}_j (\tau) - \theta_{0,j} \right)^2 d\tau \right)^{1/2}
\]

\[
= o_p (T^{-1/2})
\]

\[\Box\]

**Proof of Proposition 8:** The joint CF of \((r_t, r_{t-1}, \ldots, r_{t+1-d})\) is derived as follows:

\[
E \left[ \exp \left( \sum_{k=1}^{d} i \tau_k r_{t+1-k} \right) \mid \{V_{t-k}\}_{k=1}^{d} \right]
\]

\[
= E \left[ \exp \left( \sum_{k=1}^{d} i \tau_k \left[ \mu_0 + \mu_1 V_{t-k} + \delta \left( V_{t+1-k} - \beta (1 - e^{-\kappa}) - e^{-\kappa} V_{t-k} \right) + \sigma_{z} \varepsilon_{t} \right] \right) \mid \{V_{t-k}\}_{k=1}^{d} \right]
\]

\[
= E \left[ \exp \left( \sum_{k=1}^{d} i \tau_k \left( \mu_0 - \delta \beta (1 - e^{-\kappa}) + (\mu_1 - \delta e^{-\kappa}) V_{t-k} + \delta V_{t+1-k} \right) \right) \mid \{V_{t-k}\}_{k=1}^{d} \right]
\]

\[
= \exp \left[ \sum_{k=1}^{d} \left( i \tau_k \left( \mu_0 - \delta \beta (1 - e^{-\kappa}) \right) - \frac{\sigma_{z}^2}{2} \tau_k^2 \right) \right]
\]

\[
\times E \left[ \exp \left( i \tau_d \left( \mu_1 - \delta e^{-\kappa} \right) V_{t-d} + \sum_{k=1}^{d-1} \left( i \tau_k \left( \mu_1 - \delta e^{-\kappa} \right) + i \tau_{k+1} \delta \right) V_{t-k} + i \tau_1 \delta V_t \right) \mid \{V_{t-k}\}_{k=1}^{d} \right]
\]
Hence:

\[
\begin{align*}
&= \exp \left[ \sum_{k=1}^{d} \left( i\tau_k \left[ \mu_0 - \delta \beta (1 - e^{-\kappa}) \right] - \frac{\sigma_e^2}{2} \tau_k^2 \right) \right] \\
&\quad \times \exp \left[ i\tau_d (\mu_1 - \delta e^{-\kappa}) V_{t-d} + \sum_{k=2}^{d-1} \left[ i\tau_k (\mu_1 - \delta e^{-\kappa}) + i\tau_k \delta \right] V_{t-k} \right] \\
&\quad \times \left( 1 - \frac{i\tau_1 \delta}{c} \right)^{-q} \exp \left[ \left( i\tau_1 \delta e^{-\kappa} \frac{1}{1 - \frac{i\tau_1 \delta}{c}} + i\tau_1 (\mu_1 - \delta e^{-\kappa}) \right) V_{t-1} \right]
\end{align*}
\]

Let \( u_1 = \tau_1 \delta \) and \( u_2 = \frac{\tau_1 \delta e^{-\kappa}}{1 - \frac{i\tau_1 \delta}{c}} + \tau_1 (\mu_1 - \delta e^{-\kappa}) + \tau_2 \delta \). Taking the expectation with respect to \( V_{t-1} \) yields:

\[
\begin{align*}
&= \exp \left[ \sum_{k=1}^{d} \left( i\tau_k \left[ \mu_0 - \delta \beta (1 - e^{-\kappa}) \right] - \frac{\sigma_e^2}{2} \tau_k^2 \right) \right] \\
&\quad \times \exp \left[ i\tau_d (\mu_1 - \delta e^{-\kappa}) V_{t-d} + \sum_{k=2}^{d-1} \left[ i\tau_k (\mu_1 - \delta e^{-\kappa}) + i\tau_k \delta \right] V_{t-k} \right] \\
&\quad \times \left( 1 - \frac{i u_1}{c} \right)^{-q} \exp \left[ i u_2 V_{t-1} \right]
\end{align*}
\]

\[
\begin{align*}
&= \exp \left[ \sum_{k=1}^{d} \left( i\tau_k \left[ \mu_0 - \delta \beta (1 - e^{-\kappa}) \right] - \frac{\sigma_e^2}{2} \tau_k^2 \right) \right] \\
&\quad \times \exp \left[ i\tau_d (\mu_1 - \delta e^{-\kappa}) V_{t-d} + \sum_{k=3}^{d-1} \left[ i\tau_k (\mu_1 - \delta e^{-\kappa}) + i\tau_k \delta \right] V_{t-k} \right] \\
&\quad \times \left( 1 - \frac{i u_1}{c} \right)^{-q} \left( 1 - \frac{i u_2}{c} \right)^{-q} \exp \left[ i u_3 V_{t-2} \right]
\end{align*}
\]

where \( u_3 = \frac{u_2 \delta e^{-\kappa}}{1 - \frac{i u_2 \delta}{c}} + \tau_2 (\mu_1 - \delta e^{-\kappa}) + \tau_3 \delta \). Integrating out recursively \( V_{t-2} \) conditional on \( V_{t-3}, V_{t-3} \) conditional on \( V_{t-4} \) and so forth, we get:

\[
\begin{align*}
&= \exp \left[ \sum_{k=1}^{d} \left( i\tau_k \left[ \mu_0 - \delta \beta (1 - e^{-\kappa}) \right] - \frac{\sigma_e^2}{2} \tau_k^2 \right) \right] \\
&\quad \times \exp \left[ \prod_{k=1}^{d-1} \exp \left( i\tau_k \left[ \mu_0 - \delta \beta (1 - e^{-\kappa}) \right] - \frac{\sigma_e^2}{2} \tau_k^2 \right) \left( 1 - \frac{i u_k}{c} \right)^{-q} \right] \\
&\quad \times \exp \left[ i u_{d+1} V_{t-d} \right]
\end{align*}
\]
where:

\[
\begin{align*}
    u_1 & = \tau_1 \delta \\
    u_k & = \frac{u_{k-1} e^{-\kappa}}{1 - \frac{i u_{k-1}}{c}} + \tau_{k-1} (\mu_1 - \delta e^{-\kappa}) + \tau_k \delta, \quad k = 2, \ldots, d \\
    u_{d+1} & = \frac{i u_d e^{-\kappa}}{1 - \frac{i u_d}{c}} + i \tau_d (\mu_1 - \delta e^{-\kappa})
\end{align*}
\]

Finally, integrating out \( V_{t-d} \) yields the joint CF of \((r_t, \ldots, r_{t+1-d})\):

\[
E \left[ \exp \left( \sum_{k=1}^{d} i \tau_k r_{t+1-k} \right) \right] = \exp \left( \mu_0 - \delta \beta (1 - e^{-\kappa}) \sum_{k=1}^{d} i \tau_k - \frac{\sigma^2}{2} \sum_{k=1}^{d} \tau^2_k \right) \times \left( 1 - \frac{i u_{d+1} \sigma^2}{2 \kappa} \right)^{-q} \prod_{k=1}^{d} \left( 1 - \frac{i u_k}{c} \right)^{-q}
\]