Université de Montréal

Efficient Estimation Using the Characteristic Function : Theory and Applications with High Frequency Data

par

Rachidi KOTCHONI

Département de Sciences Économiques

Faculté des Arts et des Sciences

Thèse présentée à la Faculté des Arts et des Sciences

en vue de l'obtention du grade de

Philosophiae Doctor (Ph.D.)

en Sciences Économiques

Mai, 2010

Université de Montréal

Faculté des Arts et des Sciences

Cette thèse intitulée :

Efficient Estimation Using the Characteristic Function : Theory and Applications with High Frequency Data

et présentée par :

Rachidi KOTCHONI

a été évaluée par un jury composé des personnes suivantes :

Marc HENRY

Président-Rapporteur

Marine CARRASCO

Directeur de Recherche

Benoit PERRON

Membre du Jury

Russel DAVIDSON (McGill University)

Examinateur Externe

Pierre DUCHESNE (D.M.S., U. de M)

Représentant du Doyen de la Faculté des Arts et Sciences

Thèse acceptée le : 19 Mai 2010

Résumé de la thèse

Nous abordons deux sujets distincts dans cette thèse : l'estimation de la volatilité des prix d'actifs financiers à partir des données à haute fréquence, et l'estimation des paramétres d'un processus aléatoire à partir de sa fonction caractéristique.

Le chapitre 1 s'intéresse à l'estimation de la volatilité des prix d'actifs. Nous supposons que les données à haute fréquence disponibles sont entachées de bruit de microstructure. Les propriétés que l'on prête au bruit sont déterminantes dans le choix de l'estimateur de la volatilité. Dans ce chapitre, nous spécifions un nouveau modèle dynamique pour le bruit de microstructure qui intègre trois propriétés importantes : (i) le bruit peut être autocorrélé, (ii) le retard maximal au delà duquel l'autocorrélation est nulle peut être une fonction croissante de la fréquence journalière d'observations ; (iii) le bruit peut avoir une composante correlée avec le rendement efficient. Cette dernière composante est alors dite endogène. Ce modèle se différencie de ceux existant en ceci qu'il implique que l'autocorrélation d'ordre 1 du bruit converge vers 1 lorsque la fréquence journalière d'observation tend vers l'infini.

Nous utilisons le cadre semi-paramétrique ainsi défini pour dériver un nouvel estimateur de la volatilité intégrée baptisée "estimateur shrinkage". Cet estimateur se présente sous la forme d'une combinaison linéaire optimale de deux estimateurs aux propriétés différentes, l'optimalité étant défini en termes de minimisation de la variance. Les simulations indiquent que l'estimateur shrinkage a une variance plus petite que le meilleur des deux estimateurs initiaux. Des estimateurs sont également proposés pour les paramètres du modèle de microstructure. Nous clôturons ce chapitre par une application empirique basée sur des actifs du Dow Jones Industrials. Les résultats indiquent qu'il est pertinent de tenir compte de la dépendance temporelle du bruit de microstructure dans le processus d'estimation de la volatilité.

Les chapitres 2, 3 et 4 s'inscrivent dans la littérature économétrique qui traite de la méthode des moments généralisés. En effet, on rencontre en finance des modèles dont la fonction de vraisemblance n'est pas connue. On peut citer en guise d'exemple la loi stable ainsi que les modèles de diffusion observés en temps discrets. Les méthodes d'inférence

basées sur la fonction caractéristique peuvent être envisagées dans ces cas. Typiquement, on spécifie une condition de moment $h_t(\tau, \theta_0)$ basée sur la différence entre la fonction caractéristique (conditionnelle) théorique et sa contrepartie empirique. Dans l'expression de cette condition de moment, θ_0 est la vraie valeur du paramètre d'intérêt et τ est la variable de transformation de Fourier. Le défit ici est d'exploiter au mieux le continuum de conditions de moment $\left\{h_t(\tau, \theta_0), \tau \in \mathbb{R}^d\right\}$ pour atteindre la même efficacité que le maximum de vraisemblance.

Ce défit a été relevé par Carrasco et Florens (2000) qui ont proposé la procédure CGMM (continuum GMM). La fonction objectif que ces auteurs proposent est de la forme :

$$\widehat{Q}_T(\alpha, \theta) = \int_{\mathbb{R}^d} K_{\alpha T}^{-1} h_T(\tau, \theta) \overline{h_T(\tau, \theta)} \pi(\tau) \, d\tau$$

où $\pi(\tau)$ est une mesure finie absolument continue sur $\mathbb{R}^d$, $\overline{h_T(\tau, \theta_0)}$ est le complexe conjugué de $h_T(\tau, \theta) = \frac{1}{T}\sum_{t=1}^{T} h_t(\tau, \theta)$ et $K_{\alpha T}^{-1}$ est l'inverse régularisé de l'operateur de covariance empirique associé à la fonction de moment $h_t(\tau, \theta)$. Le paramètre de régularisation α assure à la fois l'existence et la continuité de $\widehat{Q}_T(\alpha, \theta)$. Carrasco et Florens (2000) ont montré que l'estimateur de θ_0 obtenu en minimisant $\widehat{Q}_T(\alpha, \theta)$ est asymptotiquement aussi efficace que l'estimateur du maximum de vraisemblance si α tend vers zéro lorsque la taille de l'échatillon T tend vers l'infini. La nature de la fonction objectif du CGMM soulève deux questions importantes. La première est celle de la calibration de α en pratique, et la seconde est liée à la présence d'intégrales multiples dans l'expression de $\widehat{Q}_T(\alpha, \theta)$. C'est à ces deux problématiques qu'essayent de répondent les trois derniers chapitres de la présente thèse.

Dans le chapitre 2, nous proposons une méthode de calibration de α basée sur la minimisation de l'erreur quadratique moyenne (EQM) de l'estimateur. Nous suivons une approche similaire à celle de Newey et Smith (2004) pour calculer un développement d'ordre supérieur de l'EQM de l'estimateur CGMM de sorte à pouvoir examiner sa dépendance en α en échantillon fini. Nous proposons ensuite deux méthodes pour choisir α en pratique. La première se base sur le développement de l'EQM, et la seconde se base sur des simulations Monte Carlo. Nous montrons que la méthode Monte Carlo délivre un estimateur convergent de α optimal. Nos simulations confirment la pertinence de la calibration de α en pratique.

Le chapitre 3 essaye de vulgariser la théorie du chapitre 2 pour les modèles avec $d \leq 2$. Nous commençons par passer en revue les propriétés de convergence et de normalité asymptotique de l'estimateur CGMM. Nous proposons ensuite des recettes numériques pour l'implémentation. Enfin, nous conduisons des simulations Monte Carlo basée sur la loi stable. Ces simulations démontrent que le CGMM est une méthode fiable d'inférence. En guise d'application empirique, nous estimons par CGMM un modèle de variance autorégressif Gamma. Les résultats d'estimation confirment un résultat bien connu en finance : le rendement est positivement corrélé au risque espéré et négativement corrélé au choc sur la volatilité.

Lorsqu'on implémente le CGMM, une difficulté majeure réside dans l'évaluation numérique itérative des intégrales multiples présentes dans la fonction objectif. Les méthodes de quadrature sont en principe parmi les plus précises que l'on puisse utiliser dans le présent contexte. Malheureusement, le nombre de points de quadrature augmente exponentiellement en fonction de d. L'utilisation du CGMM devient pratiquement impossible dans les modèles multivariés et non markoviens où $d \geq 3$. Dans le chapitre 4, nous proposons une procédure alternative baptisée "*reéchantillonnage dans le domaine fréquentielle*" qui consiste à fabriquer des échantillons univariés en prenant une combinaison linéaire des éléments du vecteur initial, les poids de la combinaison linéaire étant tirés aléatoirement dans un sous-espace normalisé de $\mathbb{R}^d$. Chaque échantillon ainsi généré est utilisé pour produire un estimateur du paramètre d'intérêt. L'estimateur final que nous proposons est une combinaison linéaire optimale de tous les estimateurs ainsi obtenus. Finalement, nous proposons une étude par simulation et une application empirique basées sur des modèles autorégressifs Gamma.

Dans l'ensemble, nous faisons une utilisation intensive du bootstrap, une technique selon laquelle les propriétés statistiques d'une distribution inconnue peuvent être estimées à partir d'un estimé de cette distribution. Nos résultats empiriques peuvent donc en principe être améliorés en faisant appel aux connaissances les plus récentes dans le domaine du bootstrap.

Summary of the Thesis

In estimating the integrated volatility of financial assets using noisy high frequency data, the time series properties assumed for the microstructure noise determines the proper choice of the volatility estimator. In the first chapter of the current thesis, we propose a new model for the microstructure noise with three important features. First of all, our model assumes that the noise is L-dependent. Secondly, the memory lag L is allowed to increase with the sampling frequency. And thirdly, the noise may include an endogenous part, that is, a piece that is correlated with the latent returns. The main difference between this microstructure model and existing ones is that it implies a first order autocorrelation that converges to 1 as the sampling frequency goes to infinity.

We use this semi-parametric model to derive a new shrinkage estimator for the integrated volatility. The proposed estimator makes an optimal signal-to-noise trade-off by combining a consistent estimators with an inconsistent one. Simulation results show that the shrinkage estimator behaves better than the best of the two combined ones. We also propose some estimators for the parameters of the noise model. An empirical study based on stocks listed in the Dow Jones Industrials shows the relevance of accounting for possible time dependence in the noise process.

Chapters 2, 3 and 4 pertain to the generalized method of moments based on the characteristic function. In fact, the likelihood functions of many financial econometrics models are not known in close form. For example, this is the case for the stable distribution and a discretely observed continuous time model. In these cases, one may estimate the parameter of interest θ_0 by specifying a moment condition $h_t(\tau, \theta_0)$ based on the difference between the theoretical (conditional) characteristic function and its empirical counterpart, where $\tau \in \mathbb{R}^d$ is the Fourier transformation variable. The challenge is then to exploit the whole continuum of moment conditions $\{h_t(\tau, \theta_0), \tau \in \mathbb{R}^p\}$ to achieve the maximum likelihood efficiency.

This problem has been solved in Carrasco and Florens (2000) who propose the CGMM procedure. The objective function of the CGMM is given by :

$$\widehat{Q}_T(\alpha,\theta)=\int_{\mathbb{R}^d}K_{\alpha T}^{-1}h_T(\tau,\theta)\,\overline{h_T(\tau,\theta)}\pi(\tau)\,d\tau$$

where $\pi(\tau)$ is an absolutely continuous finite measure, $\overline{h_T(\tau,\theta)}$ is the complex conjugate of $h_T(\tau,\theta)=\frac{1}{T}\sum_{t=1}^{T}h_t(\tau,\theta)$ and $K_{\alpha T}^{-1}$ is the regularized inverse of the empirical covariance operator associated with the moment function. The parameter α ensures the existence as well as the continuity of $\widehat{Q}_T(\alpha,\theta)$. Carrasco and Florens (2000) have shown that the estimator obtained by minimizing $\widehat{Q}_T(\alpha,\theta)$ is asymptotically as efficient as the maximum likelihood estimator provided that α converges to zero as the sample size T goes to infinity. However, the nature of the objective function $\widehat{Q}_T(\alpha,\theta)$ raises two important questions. First of all, how do we select α in practice ? And secondly, how do we implement the CGMM when the multiplicity of the integrals embedded in the objective-function d is large. These questions are tackled in the last three chapters of the thesis.

In Chapter 2, we propose to choose α by minimizing the approximate mean square error (MSE) of the estimator. Following an approach similar to Newey and Smith (2004), we derive a higher-order expansion of the estimator from which we characterize the finite sample dependence of the MSE on α. We provide two data-driven methods for selecting the regularization parameter in practice. The first one relies on the higher-order expansion of the MSE whereas the second one uses only simulations. We show that our simulation technique delivers a consistent estimator of α. Our Monte Carlo simulations confirm the importance of the optimal selection of α.

The goal of Chapter 3 is to illustrate how to efficiently implement the CGMM for $d\leq 2$. To start with, we review the consistency and asymptotic normality properties of the CGMM estimator. Next we suggest some numerical recipes for its implementation. Finally, we carry out a simulation study with the stable distribution that confirms the accuracy of the CGMM as an inference method. An empirical application based on the autoregressive variance Gamma model led to a well-known conclusion : investors require a positive premium for bearing the expected risk while a negative premium is attached to the unexpected risk.

In implementing the characteristic function based CGMM, a major difficulty lies in the evaluation of the multiple integrals embedded in the objective function. Numerical quadratures are among the most accurate methods that can be used in the present context. Unfortunately, the number of quadrature points grows exponentially with d. When the data generating process is Markov or dependent, the accurate implementation of the CGMM becomes roughly unfeasible when $d \geq 3$. In Chapter 4, we propose a strategy that consists in creating univariate samples by taking a linear combination of the elements of the original vector process. The weights of the linear combinations are drawn from a normalized set of $\mathbb{R}^d$. Each univariate index generated in this way is called a frequency domain bootstrap sample that can be used to compute an estimator of the parameter of interest. Finally, all the possible estimators obtained in this fashion can be aggregated to obtain the final estimator. The optimal aggregation rule is discussed in the paper. The overall method is illustrated by a simulation study and an empirical application based on autoregressive Gamma models.

This thesis makes an extensive use of the bootstrap, a technique according to which the statistical properties of an unknown distribution can be estimated from an estimate of that distribution. It is thus possible to improve our simulations and empirical results by using the state-of-the-art refinements of the bootstrap methodology.

Contents

List of Tables

List of Figures

Sigles et abréviations

ARCH	: Autoregressive Conditional Heteroscedasticity
ARFG	: Autoregressive Factor Gamma
ARVG	: Autoregressive Variance Gamma
CF	: Characteristic Function
CIR	: Cox Ingersol and Ross
CGMM	: Continuum Generalized Method of Moments
ECF	: Empirical Characteristic Function
GARCH	: Generalized Autoregressive Conditional Heteroscedasticity
GMM	: Generalized Method of Moments
FB	: Frequency domain Bootstrap
IID	: Independent and Identically Distributed
MSE	: Mean Square Error
IV	: Integrated Volatility
RV	: Realized Volatility

A mes parents, Alimatou Aboudou et Bio Séidou Kotchoni,

La flamme qu'ils ont allumé brûle encore, contre toutes les attentes.

A ma tante, Salamatou Aboudou Touré,

Qui continue de vivre dans mon coeur.

Remerciements

Tout d'abord, je remercie une excellente directrice de recherche, Marine Carrasco, pour sa disponibilité et sa grande patience. Son support a été très déterminant dans la réussite de ce projet de thèse.

Ensuite, j'adresse un merci spécial à toutes les personnes qui, dans les moments de doutes, m'ont encouragé à continuer. Je citerais Gérard Gaudet, Prosper Dovonon et son épouse Olivia, Firmin Doko et son épouse Carole. Pour tout le soutien qu'ils m'ont apporté, je remercie toute ma famille et plus particulièrement mes parents Alimatou Aboudou et Séidou Kotchoni, mes frères et soeurs, mes oncles Toussaint Kegnidé, Michel Biaou ainsi que leurs épouses. C'est le lieu de mentionner ma gratitude à Chakirou Razaki dont la contribution à cette thèse date de bien avant le début de la thèse. A lui et à toute sa famille, je souhaite beaucoup de joie et de bonheur.

Pour la qualité de leurs conseils, suggestions ou soutiens, je remercie les professeurs Marcel Boyer, Benoit Perron, Marc Henry, Rui Castro, Sílvia Gonçalves. Je remercie tous les professeurs du Département des Sciences Économiques de l'Université de Montréal pour la qualité de leurs enseignements. Je rends hommage à Mme Larouche-Sidoti, Mme Lyne Racine, Mme Jocelyne Demers, Mme Josée Lafontaine et tout le personnel du CIREQ pour leur précieuse contribution dans la bonne marche du département.

J'éprouve une profonde gratitude envers les organismes qui m'ont apporté directement ou indirectement un soutien matériel et financier tout au long de ma formation: la Banque Laurentienne, le Centre Interuniversitaire de Recherche en Analyse des Organisations (CIRANO), le Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ), le Département des Sciences Économiques ainsi que la Faculté des Études Supérieures de l'Université de Montréal.

Pour toutes sortes de choses qu'il serait trop long d'énumérer ici, mes sincères remerciements vont également à mes amis Constant Lonkeng, son épouse Judith et sa fille Abigael, Octave Keutiben et son épouse Sylvie, Bruno Feuno, Constantin Hèdiblè et son épouse Gracia, Clément Yélou et son épouse Jocelyne, Messan Agbaglah, Modeste Somé, Sali Sanou, Nerée Noumon, Cédric Okou, Bertrand Hounkannounon, Jonhson Kakeu, Anderson Walter Nzabandora, Moudjib Moussiliou Coles et son épouse Semirath, sans oublier mes amis

Kenneth Tanimomo et Francis Biaou.

Aux nombreuses personnes que je ne cite pas et qui ont contribué à la réussite de ce projet de thèse de près ou de loin, j'adresse un merci qui vient du coeur.

Chapter 1

Shrinkage Realized Kernels

Note: Cet article rédigé en collaboration avec Marine Carrasco est actuellement sous évaluation pour publication dans "*Journal of Financial Econometrics*"

Mots-Clés: Integrated Volatility, Method of Moment, Microstructure Noise, Realized Kernel, Shrinkage.

1.1 Introduction

Since the theoretical works by Jacod (1994), Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002), it is well established that the realized volatility (RV) is a consistent estimator of the integrated volatility (IV) when prices are observed without error (see for example Ait-Sahalia, Mykland and Zhang, 2005). However, it is commonly admitted that recorded stock prices are contaminated with pricing errors known in the literature as the "market microstructure noise" (henceforth "noise"). The causes of this noise are discussed for example in Stoll (1989, 2000) or Hasbrouck (1993,1996). In the words of Hasbrouck (1993), the pricing errors are mainly due to "... *discreteness, inventory control, the non-information based component of the bid-ask spread, the transient component of the price response to a block trade, etc.*". Its presence in measured prices causes the RV computed with very high frequency data to be a severely biased estimator of the IV.

Many approaches have been proposed in the literature to deal with this curse. One of them consists in choosing in an ad-hoc manner a moderate sampling frequency at which the impact of the noise is sufficiently mitigated[1]. Zhou (1996) and Hansen and Lunde (2006) proposed a bias correction approach while Bollen and Inder (2002) and Andreou and Ghysels (2002) advocated filtering techniques. Under the assumption that the volatility of the high frequency returns are constant within the day, Ait-Sahalia, Mykland and Zhang (2005) derived a highly efficient maximum likelihood estimator of the IV that is robust to both IID noise and distributional mispecification. Zhang, Mykland, and Ait-Sahalia (2005) proposed another consistent estimator in the presence of IID noise which they called the two scale realized volatility. This estimator has been adapted in Ait-Sahalia, Mykland and Zhang (2006) to deal with dependent noise. Since then, other consistent estimators have become available among which the realized kernels of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a) and the pre-averaging estimator of Podolskij and Vetter (2006)[2]. An alternative line of research pursued by Corradi, Distaso and Swanson (2008) advocates the nonparametric estimation of the predictive density and confidence intervals for the IV rather than focusing on point estimates.

In a simulation study, Gatheral and Oomen (2007) showed that consistent estimators often perform poorly at the sampling frequencies commonly encountered in practice. One can explain this result by saying that an inconsistent estimator necessarily delivers its best performance at moderate frequency[3] while a consistent estimator may require quite high frequency data in order to perform well. It turns out from our simulations study of Section 8 that the conclusion of Gatheral and Oomen (2007) strongly depends on the size of the variance of the microstructure noise relative to the discretization error. In fact, the inconsistent estimator tends to perform better than the consistent one only when the variance of the microstructure noise is small. The main idea of the current paper is that even when the variance of the inconsistent estimator is higher, it can still be optimally combined with the consistent estimator to obtain a new one that performs better than both. The weight of the linear combination is selected in order to minimize the variance and the resulting estimator is called "shrinkage realized kernels".

However, a good estimator of the IV must be designed in accordance with the dependence

[1] See Andersen, Bollerslev, Diebold and Labys (2000); Andersen, Bollerslev, Diebold and Ebens (2001).

[2] See also Jacod, Li, Mykland, Podolskij and Vetter (2009).

[3] See Ait-Sahalia, Mykland and Zhang (2005 and Bandi and Russell (2008).

properties of the noise. Awartani, Corradi and Distaso (2007) proposed an hypothesis test to assess the impact of certain features of the noise on realized measures. Hansen and Lunde (2006) construct a Haussman-type test to detect time dependence in the noise process. After applying their test to real data, they concluded that the noise process is time dependent, correlated with latent return, and possibly heteroscedastic[4]. More recently, Ubukata and Oya (2009) proposed some procedures to test for dependence in the noise process in a bivariate framework with irregularly spaced and asynchronous data.

In this paper, we restrict our attention to regularly spaced univariate data and advocate a semi-parametric model for the noise. More precisely, we specify at the highest frequency a parsimonious relation between the microstructure noise on the one side, and the efficient return and the latent volatility process on the other side. We assume a general and flexible type of noise that includes an independent endogenous part ε_s^* and an L-dependent exogenous part ε_s. Contrary to an $AR(1)$ model with constant autoregressive root, the new model has the implication that the correlation between two consecutive realizations of ε_t converges to one as the frequency at which the prices are recorded goes to infinity. This model captures the fact that two consecutive observations of ε_t become arbitrarily close in calendar time and eventually coincide at the limit as the sampling frequency increases.

We derive the properties of common realized measures under the model and propose new unbiased estimators for the IV. One unbiased estimator uses the samples available at all days to estimate the IV of each day, while the second only uses within day data. The shrinkage realized kernels are finally designed as the optimal linear combination of the standard realized kernels of Barndorff-Nielsen and al. (2008a) with the unbiased within day estimator. We also propose a framework to estimate the exogenous noise parameters. Unfortunately, the endogeneity parameters are not identifiable.

We illustrate by simulation the good performance of the shrinkage realized kernels proposed estimator. An empirical application based on fifteen stocks listed in the Dow Jones Industrials shows strong evidences of correlation in the noise process and between the noise and the latent returns. Indeed, the empirical results suggest that the memory parameter L grows slower than $\sqrt{m}$ in general. It should be mentioned that this result is derived under

[4]Kalnina and Linton (2008) propose a consistent estimator in the presence of a noise that exhibits diurnal heteroskedasticity.

the assumption that our model for the microstructure noise is true.

The rest of the paper is organized as follows. The next section motivates the use of shrinkage estimators for the IV when the noise is IID. In Section 3, we present our theoretical model in light of which we study the properties of three standard IV estimators in section 4. Inference procedures about the noise parameters are presented in section 5. In Section 6, we design the shrinkage realized kernels for the dependent noise case. Sections 7 and 8 present respectively a simulation study and an empirical application based on twelve stocks listed in the Dow Jones Industrials. Section 9 concludes. The mathematical proofs and the summaries of the results of Sections 7 and 8 are gathered in appendix.

1.2 Motivating Shrinkage Estimators for the IV

Basically, a shrinkage estimator is an optimal linear combination of several estimators. Here, optimal means that shrinkage estimator minimizes the mean square error (MSE). To motivate shrinkage estimators for the IV, we examine the contribution of the discretization error and the microstructure noise to the MSEs of three estimators. More precisely, we try to understand the trade-offs at play as one moves from a biased estimator to an unbiased estimator on the one hand, and from an unbiased estimator to a consistent estimator on the other hand.

To start with, we introduce some basic notation and concepts.

1.2.1 The Efficient Price and the Microstructure Noise

Let p_s^* denote a latent (or efficient) log-price of an asset and p_s its observable counterpart. Assume that the latent log-price obeys the following stochastic differential equation:

$$dp_s^* = \sigma_s dW_s; \quad p_0^* = 0 \tag{1.1}$$

where W_s is a standard Brownian motion independent of σ_s.

Keeping in mind that we are working with high frequency data, the omitted drift is proportional to ds which is negligible in front of the $O_p\left(\sqrt{ds}\right)$ volatility term. We assume that the volatility process $\{\sigma_s\}_{s=0}^{T}$ is *càdlàg*, implying that all powers of the volatility process are locally integrable with respect to the Lebesgue Measure[5]. Without loss of generality, we condition all our analysis on the whole volatility path but the conditioning is removed from the notations for simplicity. Therefore, all deterministic transformations of the volatility process are treated as constant objects. In particular, the integrated volatilities $IV_t = \int_{t-1}^{t} \sigma_s^2 ds$, $t = 1, 2, 3, ...T$ are fixed parameters we aim to estimate. We will consider a sampling scheme where the unit period is normalized to one in calendar time.

It is maintained throughout the paper that there is neither jump nor leverage effect in our model for the efficient price. If jumps that are uncorrelated with all other randomness are added in the model, the estimators considered for the IV in the sequel are now designed for the quadratic variation[6]. If leverage effect is assumed in (1.1), some of our results can be derived with a few more technical complications. We will check the robustness of our conclusions to the presence of leverage effect in simulation.

By definition, the noise equals $u_s = p_s - p_s^*$, that is, the difference between the observed log-price and the efficient log-price. Let r_t^* denote the latent log-return at period t, and r_t its observable counterpart. Under the above conditions, the process $\{r_t^*\}$ is a semimartingale and we have:

$$r_t \equiv p_t - p_{t-1} = r_t^* + u_t - u_{t-1} \tag{1.2}$$

$$r_t^* = \int_{t-1}^{t} \sigma_s dW_s | \{\sigma_s\}_{s=0}^{T} \sim N(0, IV_t) \tag{1.3}$$

Suppose that we have access to a large number m of intra-period returns $r_{t,1}, r_{t,2}, ..., r_{t,m}$, where $t = 1, ..., T$ are the period labels, m is the number of recorded prices in each period and $r_{t,j}$ is the j^{th} observed return during the period $[t-1,\ t]$. In the sequel, we sometimes use the expression "record frequency" to refer to the frequency m at which the data has been recorded. The noise-contaminated (observed) and true realized volatility (latent) computed at frequency m are:

$$RV_t^{(m)} = \sum_{j=1}^{m} r_{t,j}^2 \text{ and } RV_t^{*(m)} = \sum_{j=1}^{m} r_{t,j}^{*2} \tag{1.4}$$

[5]See e.g Barndorff-Nielsen, Graversen, Jacod and Shephard (2006).

[6]Separating the IV from the contribution of the jumps in the quadratic variation would then be the new issue.

For simplicity, we assume that these observations are equidistant in calendar time. We have:

$$\begin{aligned} r_{t,j}^{*} &\equiv p_{t-1+j/m}^{*} - p_{t-1+(j-1)/m}^{*} = \int_{t-1+(j-1)/m}^{t-1+j/m} \sigma_s dW_s \\ r_{t,j} &= r_{t,j}^{*} + u_{t,j} - u_{t,j-1} \\ u_{t,j} &\equiv u_{t-1+j/m} \end{aligned}$$

Barndorff-Nielsen and Shephard (2002) show that $RV_t^{*(m)}$ converges to IV_t and derived the asymptotic distribution:

$$\frac{RV_t^{*(m)} - IV_t}{\sqrt{\frac{2}{3}\sum_{j=1}^{m} r_{t,j}^{*4}}} \to N(0,1)$$

as m goes to infinity. Meddahi (2002) studied the *finite frequency* behavior of the discretization error $RV_t^{*(m)} - IV_t$ with a focus on the specific case where the true model belongs to the Eigenfunction Stochastic Volatility family. Gonçalves and Meddahi (2009) proposed some bootstrap procedures as alternative inference tools to analyze the asymptotic behavior of realized measures. In both papers, no microstructure noise is assumed.

In the presence of microstructure noise, $RV_t^{*(m)}$ is no longer feasible. We review three feasible estimators below.

1.2.2 Three Standard Estimators of the IV

In this subsection, we consider estimators $\widehat{IV}_t$ of IV_t such that:

$$\widehat{IV}_t = f_{r^*}\left(\left\{r_{t,j}^{*}\right\}_{j=1}^{m}\right) + f_{r^*,u}\left(\left\{r_{t,j}^{*}, u_{t,j}\right\}_{j=1}^{m}\right) + f_u\left(\left\{u_{t,j}\right\}_{j=1}^{m}\right) \tag{1.5}$$

with

$$E\left[f_{r^*}\left(\left\{r_{t,j}^{*}\right\}_{j=1}^{m}\right)\right] = IV_t \tag{1.6}$$

$$E\left[f_{r^*,u}\left(\left\{r_{t,j}^{*}, u_{t,j}\right\}_{j=1}^{m}\right)\right] = 0 \tag{1.7}$$

and

$$f_{r^*,u}\left(\left\{r_{t,j}^{*}, 0\right\}_{j=1}^{m}\right) = f_u\left(\left\{0\right\}_{j=1}^{m}\right) = 0 \tag{1.8}$$

It is further required that the three terms in (1.5) be uncorrelated.

Below are three examples of estimators that can be decomposed as above.

Example 1 *The realized volatility $RV_t^{(m)}$ defined in (1.4) satisfies (1.5) with:*

$$
\begin{aligned}
f_{r^*}\left(\left\{r_{t,j}^*\right\}_{j=1}^m\right) &= \sum_{j=1}^m r_{t,j}^{*2} \\
f_{r^*,u}\left(\left\{r_{t,j}^*, u_{t,j}\right\}_{j=1}^m\right) &= 2\sum_{j=1}^m \left(u_{t,j} - u_{t,j-1}\right) r_{t,j}^* \\
f_u\left(\left\{u_{t,j}\right\}_{j=1}^m\right) &= \sum_{j=1}^m \left(u_{t,j} - u_{t,j-1}\right)^2
\end{aligned}
$$

Under IID noise, $RV_t^{(m)}$ is biased and inconsistent and its bias and variance are linearly increasing in m. See for example Zhang, Mykland and Ait-Sahalia (2005) and Hansen and Lunde (2006).

Example 2 *The first order autocorrelation bias corrected estimator of Zhou (1996) given by*

$$
RV_t^{(AC,m,1)} = \sum_{j=1}^m r_{t,j}^2 + \sum_{j=1}^m r_{t,j}\left(r_{t,j+1} + r_{t,j-1}\right) \tag{1.9}
$$

satisfies (1.5) with:

$$
\begin{aligned}
f_{r^*}\left(\left\{r_{t,j}^*\right\}_{j=1}^m\right) &= \sum_{j=1}^m r_{t,j}^{*2} + \sum_{j=1}^m r_{t,j}^*\left(r_{t,j+1}^* + r_{t,j-1}^*\right) \\
f_{r^*,u}\left(\left\{r_{t,j}^*, u_{t,j}\right\}_{j=1}^m\right) &= 2\sum_{j=1}^m \Delta u_{t,j} r_{t,j}^* + \sum_{j=1}^m \Delta u_{t,j}\left(r_{t,j+1}^* + r_{t,j-1}^*\right) \\
&\quad + \sum_{j=1}^m r_{t,j}^*\left(\Delta u_{t,j+1} + \Delta u_{t,j-1}\right) \\
f_u\left(\left\{u_{t,j}\right\}_{j=1}^m\right) &= \sum_{j=1}^m \Delta u_{t,j}^2 + \sum_{j=1}^m \Delta u_{t,j}\left(\Delta u_{t,j+1} + \Delta u_{t,j-1}\right)
\end{aligned}
$$

and $\Delta u_{t,j} = u_{t,j} - u_{t,j-1}$.

Under IID noise, it is shown in Hansen and Lunde (2006) that $RV_t^{(AC,m,1)}$ is unbiased for IV while its variance is linearly increasing in m. Bandi and Russell (2006) and Hansen and Lunde (2006) derived optimal sampling frequencies for $RV_t^{(m)}$ and $RV_t^{(AC,m,1)}$ based on a signal-to-noise ratio maximization.

Example 3 *The realized Kernel of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a) is given by:*

$$K_t^{BNHLS} = \gamma_{t,0}(r) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}(r) + \gamma_{t,-h}(r)\right) \tag{1.10}$$

for a kernel function $k\left(\frac{h-1}{H}\right)$ *such that* $k(0) = 1$ *and* $k(1) = 0$*. where:*

$$\gamma_{t,h}(x) = \sum_{j=1}^{m} x_{t,j}x_{t,j-h} \tag{1.11}$$

for all variable x *and* h*. If we further define:*

$$\begin{aligned} \gamma_{t,h}(x,y) &= \sum_{j=1}^{m} x_{t,j}y_{t,j-h} \\ K_t(x) &= \gamma_{t,0}(x) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\left[\gamma_{t,h}(x) + \gamma_{t,-h}(x)\right] \\ K_t(x,y) &= \gamma_{t,0}(x,y) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\left[\gamma_{t,h}(x,y) + \gamma_{t,-h}(x,y)\right] \end{aligned}$$

then K_t^{BNHLS} *satisfies (1.5) with:*

$$\begin{aligned} f_{r^*}\left(\left\{r_{t,j}^*\right\}_{j=1}^{m}\right) &= K_t(r^*) \\ f_{r^*,u}\left(\left\{r_{t,j}^*, u_{t,j}\right\}_{j=1}^{m}\right) &= K_t(r^*, \Delta u) + K_t(\Delta u, r^*) \\ f_u\left(\left\{u_{t,j}\right\}_{j=1}^{m}\right) &= K_t(\Delta u) \end{aligned}$$

and $\Delta u_{t,j} = u_{t,j} - u_{t,j-1}$*.*

Barndorff-Nielsen and al. (2008a) show that K_t^{BNHLS} is consistent for IV_t in the presence of microstructure noise under various choice of kernel function. For example, setting $k(x) = 1 - x$ (the Bartlett kernel) and $H \propto m^{2/3}$ leads to $K_t^{BNHLS} - IV_t = O_p\left(m^{-1/6}\right)$ under IID noise. Furthermore, this estimator is robust to special forms of endogeneity and serial correlation in the microstructure noise process[7]. As we can see, the expression of K_t^{BNHLS} is reminiscent of the long run variance estimators of Newey and West (1987) and Andrews and Monahan (1992).

[7]In the current context, an endogenous noise is a noise that is correlated with the efficient price or return.

1.2.3 Discretization Error versus Microstructure Noise

In this subsection, we examine the relative contribution of the discretization error and the microstructure noise to the MSE of $\widehat{IV}_t$. We first consider the bias term:

$$E\left[\widehat{IV}_t\right] - IV_t = E\left[f_u\left(\{u_{t,j}\}_{j=1}^m\right)\right] \tag{1.12}$$

Because the additive terms in (1.5) are uncorrelated, the variance of $\widehat{IV}_t$ is given by:

$$\begin{aligned} Var\left[\widehat{IV}_t\right] &= Var\left[f_{r^*}\left(\left\{r_{t,j}^*\right\}_{j=1}^m\right)\right] + Var\left[f_{r^*,u}\left(\left\{r_{t,j}^*, u_{t,j}\right\}_{j=1}^m\right)\right] \\ &\quad + Var\left[f_u\left(\{u_{t,j}\}_{j=1}^m\right)\right] \end{aligned}$$

Hence the overall MSE is:

$$\begin{aligned} MSE\left[\widehat{IV}_t\right] &= Var\left[f_{r^*}\left(\left\{r_{t,j}^*\right\}_{j=1}^m\right)\right] + Var\left[f_{r^*,u}\left(\left\{r_{t,j}^*, u_{t,j}\right\}_{j=1}^m\right)\right] \\ &\quad + Var\left[f_u\left(\{u_{t,j}\}_{j=1}^m\right)\right] + E\left[f_u\left(\{u_{t,j}\}_{j=1}^m\right)\right]^2 \end{aligned} \tag{1.13}$$

Because $f_{r^*,u}\left(\left\{r_{t,j}^*, 0\right\}_{j=1}^m\right) = f_u\left(\{0\}_{j=1}^m\right) = 0$, the above MSE reduces to the variance of $f_{r^*}\left(\left\{r_{t,j}^*\right\}_{j=1}^m\right)$ when there is no noise in the data. Based on this argument, we adopt the following definition.

Definition 4 *The contribution of the microstructure noise to the MSE of* $\widehat{IV}_t$ *is:*

$$\begin{aligned} MSE_u\left[\widehat{IV}_t\right] &= Var\left[f_{r^*,u}\left(\left\{r_{t,j}^*, u_{t,j}\right\}_{j=1}^m\right)\right] + Var\left[f_u\left(\{u_{t,j}\}_{j=1}^m\right)\right] \\ &\quad + E\left[f_u\left(\{u_{t,j}\}_{j=1}^m\right)\right]^2 \end{aligned} \tag{1.14}$$

Accordingly, we define the MSE due to discretization as:

$$MSE_{r^*}\left[\widehat{IV}_t\right] = Var\left[f_{r^*}\left(\left\{r_{t,j}^*\right\}_{j=1}^m\right)\right]. \tag{1.15}$$

This definition imputes to the microstructure noise the part of the MSE of $\widehat{IV}_t$ that vanishes when there is actually no microstructure noise in the data. In the following table, we examine the expression of $f_{r^*}\left(\left\{r_{t,j}^*\right\}_{j=1}^m\right)$ for the three estimators listed in the examples.

It is seen that this expression includes more and more terms as one moves from the top to the bottom of the table. In fact, $RV_t^{(AC,m,1)}$ kills of the bias of its ancestor $RV_t^{(m)}$ at the expense of a higher discretization error. Likewise, K_t^{BNHLS} brings consistency upon conceding a higher discretization error with respect to the unbiased estimator $RV_t^{(AC,m,1)}$.

	$f_{r^*}\left(\left\{r^*_{t,j}\right\}_{j=1}^m\right)$	$Var\left[f_{r^*}\left(\left\{r^*_{t,j}\right\}_{j=1}^m\right)\right]$
$RV_t^{(m)}$	$\sum_{j=1}^m r^{*2}_{t,j}$	$2\sum_{j=1}^m \sigma^{*4}_{t,j}$
$RV_t^{(AC,m,1)}$	$\sum_{j=1}^m r^{*2}_{t,j} + \sum_{j=1}^m r^*_{t,j}\left(r^*_{t,j+1} + r^*_{t,j-1}\right)$	$2\sum_{j=1}^m \sigma^{*4}_{t,j}$ $+4\sum_{j=1}^m \sigma^{*2}_{t,j}\sigma^{*2}_{t,j-1} + O(m^{-2})$
K_t^{BNHLS}	$\sum_{j=1}^m r^{*2}_{t,j} + \sum_{j=1}^m r^*_{t,j}\left(r^*_{t,j+1} + r^*_{t,j-1}\right)$ $+\sum_{h=2}^H k\left(\frac{h-1}{H}\right)\sum_{j=1}^m r^*_{t,j}\left(r^*_{t,j+h} + r^*_{t,j-h}\right)$	$2\sum_{j=1}^m \sigma^{*4}_{t,j} + 4\sum_{j=1}^m \sigma^{*2}_{t,j}\sigma^{*2}_{t,j-1}$ $+4\sum_{h=2}^H k\left(\frac{h-1}{H}\right)\sum_{j=1}^m \sigma^{*2}_{t,j}\sigma^{*2}_{t,j-h}$ $+O(Hm^{-2})$

Table 1: Part of the MSE due to discretization

We now turn to discuss the MSE due to IID microstructure noise. Unlike $RV_t^{(m)}$ whose bias equals $2mE\left[u^2_{t,j}\right]$, the estimators $RV_t^{(AC,m,1)}$ and K_t^{BNHLS} are unbiased. As a consequence, the MSE of $RV_t^{(m)}$ increases at rate m^2 while those of $RV_t^{(AC,m,1)}$ and K_t^{BNHLS} are only linear in m. On the other hand, the consistency of K_t^{BNHLS} ensures that its variance eventually becomes smaller than that of $RV_t^{(AC,m,1)}$ as m goes to infinity. But there is at least two situations where $RV_t^{(AC,m,1)}$ can have lower variance than K_t^{BNHLS}. The first situation is the one in which the sampling frequency m is not large enough to make the asymptotic results for K_t^{BNHLS} reliable. In fact, the variance of K_t^{BNHLS} can be arbitrarily high in fixed frequency although it vanishes as m goes to infinity. The second situation is the case where the variance of the microstructure noise is so small that it contributes very little to the MSE. In this case, the MSE of the estimators basically reduces to the variance of $f_{r^*}\left(\left\{r^*_{t,j}\right\}_{j=1}^m\right)$ which happens to be larger for K_t^{BNHLS}.

Our intuitions are supported by a simulation study by Gatheral and Oomen (2007). These authors implemented twenty realized measures that aim to estimate the IV. Their main finding is that even though inconsistent, kernel-type estimators like $RV_t^{(AC,m,1)}$ often deliver good performances in term of MSE at sampling frequencies commonly encountered in practice. Unfortunately, there is no clear rule indicating the minimum sampling frequency required for the asymptotic theory of K_t^{BNHLS} to be usable. Moreover, the microstructure noise is not observed so that it is difficult to tell whether or not its size is small compared to the efficient returns.

It turns out that one can construct a linear combination of $RV_t^{(AC,m,1)}$ and K_t^{BNHLS} that outperforms either of the individual estimators. Let us define:

$$K_t^{\varpi} = \varpi K_t^{BNHLS} + (1-\varpi)\, RV_t^{(AC,m,1)} \tag{1.16}$$

Because both estimators are unbiased, the weight ϖ that minimizes the variance of K_t^{ϖ} *conditional on the volatility path* is given:

$$\varpi_t^* = \frac{Cov\left[RV_t^{(AC,m,1)}, RV_t^{(AC,m,1)} - K_t^{BNHLS}\right]}{Var\left[RV_t^{(AC,m,1)} - K_t^{BNHLS}\right]}. \tag{1.17}$$

The resulting $K_t^{\varpi^*}$ is called shrinkage estimator of IV_t. By construction, it satisfies:

$$Var\left(K_t^{\varpi^*}\right) \leq \min\left(Var\left(K_t^{BNHLS}\right), Var\left(RV_t^{(AC,m,1)}\right)\right)$$

Hence the shrinkage estimator inherits the consistency of K_t^{BNHLS} while performing better than $RV_t^{(AC,m,1)}$ in the problematic situations listed above. Although we are using a quadratic loss function, other types of loss functions could have been considered. See Hansen (2007) and references therein.

The estimator $K_t^{\varpi^*}$ is related to the estimator proposed in Ghysels, Mykland and Renault (2008). In fact, Ghysels, Mykland and Renault (2008) observe that the volatility is quite persistent in practice. Based on this, they propose a new estimator of IV_t which is a linear combination of a current period estimator and an optimal forecast of IV_t from the previous period. This can be thought of as shrinking the current period estimator toward the forecast.

We analyze below the asymptotic behavior of the optimal weight.

1.2.4 Asymptotic Theory

It is maintained in this section that the microstructure noise is IID. This assumption is relaxed later in Section 6. Let us assume that K_t^{BNHLS} satisfies:

$$K_t^{BNHLS} - IV_t = O_p\left(m^{-\eta}\right), \eta \geq 0.$$

To ease the readability, let us define:

$$\begin{aligned}
\theta_{1,t} &= \sum_{j=1}^{m} r_{t,j}^2 + \sum_{j=1}^{m} r_{t,j}\left(r_{t,j+1} + r_{t,j-1}\right), \\
\theta_{2,t} &= \sum_{h=2}^{H} k\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}(r) + \gamma_{t,-h}(r)\right).
\end{aligned}$$

With these notations, we have:

$$\begin{aligned}
RV_t^{(AC,m,1)} &= \theta_{1,t}, \\
K_t^{BNHLS} &= \theta_{1,t} + \theta_{2,t}, \\
K_t^{\varpi} &= \theta_{1,t} + \varpi\theta_{2,t}.
\end{aligned}$$

It turns out that K_t^{ϖ} is also a realized kernels with kernel function given by:

$$\begin{aligned}
g_{\varpi}(0) &= k(0) = 1, \\
g_{\varpi}(x) &= \varpi k(x), \ \forall\, x \in (0,1].
\end{aligned}$$

It is seen that $\lim_{x \to 0^+} g_{\varpi}(x) = \varpi$ while $g_{\varpi}(0) = 1$ so that $g_{\varpi}(x)$ is discontinuous at $x = 0$ whenever $\varpi \neq 1$. We may thus refer to the minimum variance estimator $K_t^{\varpi^*}$ as the "*shrinkage realized kernels*". The optimal shrinkage weight is given by:

$$\varpi_t^* = -\frac{Cov\left(\theta_{1,t}, \theta_{2,t}\right)}{Var\left(\theta_{2,t}\right)} = -\rho_{1,2}\sqrt{\frac{Var\left(\theta_{1,t}\right)}{Var\left(\theta_{2,t}\right)}}$$

where $\rho_{1,2}$ is the conditional correlation between $\theta_{1,t}$ and $\theta_{2,t}$. Note that ϖ_t^* is equal to minus the regression slope of $\theta_{1,t}$ onto $\theta_{2,t}$.

This optimal weight results in the following variance for $K_t^{\varpi^*}$:

$$Var\left(K_t^{\varpi^*}\right) = Var\left(\theta_{1,t}\right)\left[1-\rho_{1,2}^2\right] \tag{1.18}$$

It is seen that in comparison with $RV_t^{(AC,m,1)}$, the variance of $K_t^{\varpi^*}$ is smaller by a factor $\left(1-\rho_{1,2}^2\right)$. The variance reduction with respect to K_t^{BNHLS} is given by:

$$\begin{aligned} Var\left(K_t^{BNHLS}\right) - Var\left(K_t^{\varpi^*}\right) &= \rho_{1,2}^2 Var\left(\theta_{1,t}\right) + 2Cov\left(\theta_{1,t},\theta_{2,t}\right) + Var\left(\theta_{2,t}\right) \\ &= \left(\rho_{1,2}\sqrt{Var\left(\theta_{1,t}\right)} + \sqrt{Var\left(\theta_{2,t}\right)}\right)^2 \geq 0 \end{aligned} \tag{1.19}$$

To derive a rate for $\rho_{1,2}$, we use Equations (1.18) and (1.19):

$$\begin{aligned} Var\left(K_t^{\varpi^*}\right) &\leq Var\left(K_t^{BNHLS}\right) = O\left(m^{-2\eta}\right) \Leftrightarrow \\ 1-\rho_{1,2}^2 &\leq \frac{Var\left(K_t^{BNHLS}\right)}{Var\left(\theta_{1,t}\right)} = O\left(m^{-2\eta-1}\right) \end{aligned}$$

We obtain the rate for $\rho_{1,2}$ by applying the following rule:

$$\rho_{1,2} = -\left[1-O\left(m^{-2\eta-1}\right)\right]^{1/2} \approx -1 + \frac{1}{2}O\left(m^{-2\eta-1}\right), \tag{1.20}$$

that is, $\rho_{1,2}$ converges to minus one from above at rate $O\left(m^{-2\eta-1}\right)$. The sign follows from the prior knowledge that $Cov\left(\theta_{1,t},\theta_{2,t}\right)$ is negative.

Note that the consistency of $K_t^{\varpi^*}$ implies that for large enough m, we have:

$$\theta_{1,t} \approx -\varpi_t^*\theta_{2,t} + IV_t,$$

This has two implications. Firstly, because $Var\left(\theta_{1,t}\right) = O(m)$, we have:

$$\frac{1}{m}Var\left(\theta_{2,t}\right) \to O(1)$$

And secondly, we have:

$$Var\left(K_t^{BNHLS}\right) \simeq \left(1-\varpi_t^*\right)^2 Var\left(\theta_{2,t}\right) = O\left(m^{-2\eta}\right).$$

These two implications allow us to conclude that:

$$1 - \varpi_t^* = O\left(m^{-\eta-\frac{1}{2}}\right). \tag{1.21}$$

In summary, (1.19) shows that the shrinkage estimator outperforms the estimator of Barndorff-Nielsen and al. (2008a) in finite frequency while (1.21) shows that the two estimators are asymptotically equivalent. The optimality of the weight ϖ_t^* relies on the unbiasedness of $RV_t^{(AC,m,1)}$ and K_t^{BNHLS}. However, the estimator $RV_t^{(AC,m,1)}$ is biased when the microstructure is time dependent. In the following section, we specify a dependent semiparametric model for the microstructure noise within which the suitable shrinkage estimator will be derived.

1.3 A Semiparametric Model for the Microstructure Noise

Our modeling approach is based on the assumption that the time series properties of the microstructure noise are tied to the frequency at which the prices have been recorded. With this in mind, we specify a link between the noise $u_{t,j}$ and the latent return $r_{t,j}^*$ at the highest frequency and then deduce the properties of the realized volatility computed at lower frequencies. In a second step, we will study the properties of the kernel based estimators of Hansen and Lunde (2006) and Barndorff-Nielsen and al (2008a) when the record frequency m goes to infinity.

We assume that the noise process evolves in calendar time according to:

$$u_{t,j} = a_{t,j} r_{t,j}^* + \varepsilon_{t,j}, \;\; j = 1, 2, ..., m, \text{ for all } t \tag{1.22}$$

where $a_{t,j}$ is a time varying coefficient and $\varepsilon_{t,j}$ is independent of the efficient high frequency return $r_{t,j}^*$. In the words of Hasbrouck (1993), $\varepsilon_{t,j}$ is the information uncorrelated or exogenous *pricing error* and $a_{t,j} r_{t,j}^*$ is the information correlated or endogenous pricing error. For sake of parsimony, our model assumes that time dependence in the noise process can only be due to its information uncorrelated part. We discuss more specifically the assumptions below.

1.3.1 Assumptions

The following assumptions are maintained throughout the paper.

E0. $a_{t,j} = \beta_0 + \frac{\beta_1}{\sqrt{m\sigma_{t,j}^{*2}}}$, where β_0 and β_1 are constants and

$$\sigma_{t,j}^{*2} = Var\left(r_{t,j}^*\right) \equiv \int_{t-1+(j-1)/m}^{t-1+j/m} \sigma_s^2 ds.$$

E1. For fixed m, $\varepsilon_{t,j}$ is a discrete time stationary process with zero mean and finite fourth moments, and independent of $\{\sigma_s\}$ and $r_{t,j}^*$.

E2. $E(\varepsilon_{t,j}\varepsilon_{t,j-h}) = \omega\left(\frac{h}{m}\right) \equiv \omega_{m,h}$, $0 \leq \frac{h}{m} \leq \frac{L}{m} < 1$ and $\omega_{m,h} = 0$ for all $h > L$.

E3. $\omega(0) \equiv \omega_{m,0} = \omega_0$ for all m, $\omega_{m,h} - \omega_{m,h+1} = \omega_0 O(m^{-\alpha})$ for some $\alpha < 2/3$, $h = 0, ..., L-1$.

E4. $L \propto m^{\delta}$ for some $\delta \leq \alpha$.

The aim of Assumption E0 is to introduce endogeneity in the microstructure noise process in such a way that both homoscedasticity ($\beta_0 = 0$) and heteroscedaticity ($\beta_1 = 0$) are allowed. This assumption implies that the variance of the endogenous part of the noise goes to zero at rate m since:

$$Var\left(a_{t,j} r_{t,j}^*\right) = \beta_0 \sigma_{t,j}^{*2} + 2\beta_0\beta_1 \sqrt{\frac{\sigma_{t,j}^{*2}}{m}} + \frac{\beta_1^2}{m}.$$

Assumption E1 is quite standard in the literature. The semi-parametric nature of the model comes from Assumption E2 which only stipulates that $\varepsilon_{t,j}$ is L-dependent. In fact, this assumption does not specify a parametric family for the distribution of $\varepsilon_{t,j}$. Furthermore, L may grow with the record frequency m according to E4. Assumption E3 implies that:

$$Cov\left(\varepsilon_{t,0}, \varepsilon_{t,j}\right) = \omega_0 - \sum_{h=0}^{j-1} \left(\omega_{m,h} - \omega_{m,h+1}\right) = \omega_0 - j\omega_0 O(m^{-\alpha}). \tag{1.23}$$

Hence for any fixed j, $Cov\left(\varepsilon_{t,0}, \varepsilon_{t,j}\right)$ converges to the constant variance ω_0 as m goes to infinity. Intuitively, the time length between $\varepsilon_{t,0}$ and $\varepsilon_{t,j}$ goes to zero as m increases to infinity and these two realizations should coincide at the limit.

We now highlight an important link between assumptions E3 and E4. To this end, let us assume that $j = \lfloor Lc \rfloor$ for some constant $c \in (0,1)$, where $\lfloor x \rfloor$ denote the largest integer that is smaller than x. According to E3 and E4, we have:

$$
\begin{aligned}
Cov\left(\varepsilon_{t,0}, \varepsilon_{t,j}\right) &= \omega_0 - j\omega_0 O(m^{-\alpha}) \\
&= \omega_0 - \omega_0 c O(m^{\delta-\alpha})
\end{aligned}
$$

It is seen that the condition $\delta - \alpha \leq 0$ is necessary in order for $Cov\left(\varepsilon_{t,0}, \varepsilon_{t,j}\right)$ to be bounded.

The requirement that $\alpha < 2/3$ in Assumption E3 is only technical and is imposed to ensure that $\delta < 2/3$. That condition turns out to be crucial for the convergence of the realized kernels with Bartlett kernel. Indeed, the parameters α and δ play important roles in the asymptotics. Note that the memory of noise as measure by L is longer for larger δ, but the time length $\frac{L}{m}$ after which the correlation dies out converges to zero as m goes to infinity.

In summary, the proposed model for the microstructure noise has the implication that the covariance between two consecutive realizations of ε_t converges to its variance as the frequency at which the prices are recorded m goes to infinity. This model aims to capture the fact that two consecutive observations of ε_t become arbitrarily close in calendar time and must thus coincide at the limit. Consequently, the first order autocorrelation of ε_t must converge to one contrary to what is implied for instance by an $AR(1)$ model with constant autoregressive root. The introduction of this feature comes at the cost that the memory parameter L must not grow too fast as a function of m for the realized kernels to continue to deliver their best performance at the largest available frequency.

Below, we compare our models with other specifications.

1.3.2 Nested and Related Models

Imposing $\beta_0 = \beta_1 = \delta = 0$ in our model leads to $u_{t,j} = \varepsilon_{t,j}$ where $\varepsilon_{t,j}$ is a moving average of fix order L. This case has been considered in Hansen and Lunde (2006). Further imposing $L = 0$ leads to the IID noise considered by Ait-Sahalia, Mykland and Zhang (2005) among others. One gets a version of Roll's model (1984) from our specification by setting $\beta_0 = \beta_1 = 0$ and $\varepsilon_{t,j} = \pm Q_{t,j}/2$, where $Q_{t,j}$ is the bid-ask spread. The model of Roll can thus be regarded as nested within our specification with the difference that $\varepsilon_{t,j}$ is now observable. Hasbrouck (1993) used the restriction $\beta_1 = 0$ with $\varepsilon_{t,j}$ IID to model the microstructure noise contaminating daily returns. This particular case results in an $MA(1)$ representation for

$u_{t,j}$ which, as a function of the original parameters, is identifiable if one further imposes the restriction $\varepsilon_{t,j} = 0$ used in Beveridge and Nelson (1981) or the restriction $\beta_0 = 0$ used by Watson (1986).

Ait-Sahalia, Mykland and Zhang (2006) considered an exogenous noise with general mixing properties. Kalnina and Linton (2008) advocated a microstructure noise model that features endogeneity and diurnal heteroscedaticity. These two models cannot be nested within our specification.

We now examine the continuous time limit of our model. As m goes to infinity, we have:

$$\begin{aligned} \int_{t-1/m}^{t} \sigma_s dW_s &\approx \sigma_t dW_t, \\ m\int_{t-1/m}^{t} \sigma_s^2 ds &\approx \sigma_t^2 \end{aligned}$$

for all t. From this, we see that the limit of (1.22) as m becomes very large may be defined as:

$$u_s = \beta_0 \sigma_s dW_s + \beta_1 dW_s + \varepsilon_s, \tag{1.24}$$

where W_s is the same Brownian motion as in (1.1). Equation (1.24) specialized to the case $\beta_0 = \beta_1 = 0$ is reminiscent of a case covered in Section 4.1 of Hansen and Lunde (2006). Note that we have:

$$\begin{aligned} var\left(\beta_0 \sigma_s dW_s\right) &= \beta_0^2 \sigma_s^2 ds = O\left(ds\right), \\ var\left(\beta_1 dW_s\right) &= \frac{\beta_1^2}{m} = O\left(ds\right), \end{aligned}$$

so that the noise basically reduces to its information uncorrelated part ε_s at the limit. Also, we do not suggest that (1.22) can be deduced from (1.24) by aggregation.

The vanishing information correlated noise may happen to be a theoretical weakness in some situations. In this regard, perhaps a more interesting specification is:

$$u_{t,j} = \beta_0 r_{t,j}^* + \beta_1 \sum_{k=0}^{N} \phi_k \frac{r_{t,j-k}^*}{\sqrt{m}\sigma_{t,j-k}^*} + \varepsilon_{t,j}, \ j = 1, 2, ..., m, \text{ for all } t. \tag{1.25}$$

The above specification is suitable if there is a reason to think that the pricing errors are correlated with past information even at the limit. In this case, the limit as m becomes very large may be cast as:

$$u_s = \beta_0 \sigma_s dW_s + \beta_1 \int_{-\infty}^{s} \phi(s,t)\, dW_t + \varepsilon_s \tag{1.26}$$

where $\phi(s,t)$ is a continuous function scaled in such a way that the variance of $\int_{-\infty}^{s} \phi(s,t)\, dW_t$ is constant for all s. In studying model (1.25), the most challenging task will be to identify the coefficients ϕ_k from the autocovariances of $\varepsilon_{t,j}$. Equation (1.26) specialized to the case $\beta_0 = 0$ and $\varepsilon_s = 0$ is reminiscent of a case discussed in a comment of Hansen and Lunde (2006) by Garcia and Meddahi (2006).

In the next subsection, we discuss some implications of model (1.22).

1.3.3 Some Implications of the Model

We first consider the unobservable implications of our postulated model. When $\beta_0 = 0$, the microstructure noise $u_{t,j}$ is identically distributed conditional on the volatility path:

$$u_{t,j} | \{\sigma_s\} \sim \varepsilon_{t,j} + N\left(0, \frac{\beta_1^2}{m}\right) \text{ for all } (t,j)$$

Also, the variance of $u_{t,j}$ and the correlation between $u_{t,j}$ and $r_{t,j}^*$ are given by:

$$\begin{aligned} Var(u_{t,j}) &= \frac{\beta_1^2}{m} + \omega_0 \\ Corr\left(u_{t,j}, r_{t,j}^*\right) &= \frac{\beta_1}{\sqrt{\beta_1^2 + m\omega_0}}. \end{aligned}$$

It is seen that $Corr\left(u_{t,j}, r_{t,j}^*\right)$ goes to zero as m goes to infinity.

When $\beta_1 = 0$ and $\beta_0 \neq 0$, the noise process is no longer identically distributed. We have:

$$\begin{aligned} u_{t,j} | \{\sigma_s^2\} &\sim \varepsilon_{t,j} + N\left(0, \beta_0^2 \sigma_{t,j}^{*2}\right) \\ Corr\left(u_{t,j}, r_{t,j}^*\right) &= \frac{\beta_0 \sigma_{t,j}^*}{\sqrt{\beta_0^2 \sigma_{t,j}^{*2} + \omega_0}} \end{aligned}$$

It is seen that $Corr\left(u_{t,j}, r_{t,j}^*\right)$ is no longer constant.

We now turn to examine the observable implications. The expression of the observed log-returns at the highest frequency m takes on the form:

$$r_{t,j} = \left(1 + \beta_0 + \frac{\beta_1}{\sqrt{m}\sigma^*_{t,j}}\right) r^*_{t,j} - \left(\beta_0 + \frac{\beta_1}{\sqrt{m}\sigma^*_{t,j-1}}\right) r^*_{t,j-1} + (\varepsilon_{t,j} - \varepsilon_{t,j-1}). \tag{1.27}$$

The covariance between two consecutive returns is given by:

$$E(r_{t,j}r_{t,j-1}) = -\left(\beta_0 + \frac{\beta_1}{\sqrt{m}\sigma^*_{t,j-1}}\right)\left(1 + \beta_0 + \frac{\beta_1}{\sqrt{m}\sigma^*_{t,j-1}}\right)\sigma^{*2}_{t,j-1} - \omega_0 + 2\omega_{m,1} - \omega_{m,2} \tag{1.28}$$

where we recall that $E(\varepsilon_{t,j}\varepsilon_{t,j-h}) = \omega_{m,h}$. This covariance is time varying and may be positive or negative depending on the local variance $\sigma^{*2}_{t,j-1}$ and the values of the parameters. The covariance between two non consecutive returns is:

$$E(r_{t,j}r_{t,j-h}) = -\omega_{m,h-1} + 2\omega_{m,h} - \omega_{m,h+1};\ h \geq 2. \tag{1.29}$$

Hence $E(r_{t,j}r_{t,j-L-1}) = 0$ from the L-dependence of $\varepsilon_{t,j}$. Note that if $\varepsilon_{t,j}$ is IID, these formulas reduce to:

$$\begin{aligned} E(r_{t,j}r_{t,j-1}) &= -\left(\beta_0 + \frac{\beta_1}{\sqrt{m}\sigma^*_{t,j-1}}\right)\left(1 + \beta_0 + \frac{\beta_1}{\sqrt{m}\sigma^*_{t,j-1}}\right)\sigma^{*2}_{t,j-1} - \omega_0 && (1.30) \\ E(r_{t,j}r_{t,j-h}) &= 0;\ h \geq 2. && (1.31) \end{aligned}$$

In what follows, we examine the implications of the postulated microstructure noise model for the traditional realized variance.

1.4 Properties of Three IV Estimators

We study successively the traditional realized variance, the kernel estimator of Hansen and Lunde (2006) and the realized kernels of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a). The bias of the realized provides one of the moment conditions that will be used in Section 5 to estimate the correlogram of the microstructure noise. The properties of the kernel estimators are mainly used in Section 6 to design the shrinkage estimator of the IV.

1.4.1 The Realized Volatility

The estimator of interest here is the sparsely sampled realized variance given by:

$$RV_t^{(m_q)} = \sum_{k=1}^{m_q} \widetilde{r}_{t,k}^2 \tag{1.32}$$

where $\widetilde{r}_{t,k}$ is the sum of q consecutive returns, that is:

$$\widetilde{r}_{t,k} = \sum_{j=qk-q+1}^{qk} r_{t,j}, k = 1, ..., m_q = \frac{m}{q}, q \geq 1 \tag{1.33}$$

Hence if $r_{t,j}^*$ is a series of one minute returns for instance, then $\widetilde{r}_{t,k}$ would be a sequence of q minutes return. The following picture illustrates the corresponding subsampling scheme which is quite standard in this literature.

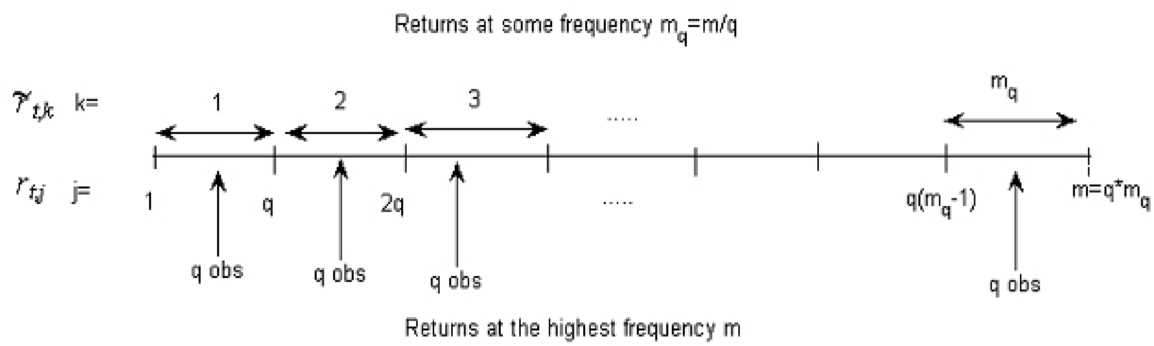


Figure 1: The subsampling scheme

If the noise process is correctly described at the highest frequency by equation (1.22), then the expression of $\widetilde{r}_{t,k}$ is given by:

$$\begin{aligned} \widetilde{r}_{t,k} &= \left(1 + \beta_0 + \frac{\beta_1}{\sigma_{t,qk}^*}\right) r_{t,qk}^* + \sum_{j=qk-q+1}^{qk-1} r_{t,j}^* - \left(\beta_0 + \frac{\beta_1}{\sigma_{t,qk-q}^*}\right) r_{t,qk-q}^* \\ &\quad + \left(\varepsilon_{t,qk} - \varepsilon_{t,qk-q}\right), \end{aligned} \tag{1.34}$$

for $k = 1, ..., m_q$ and for all t, with the convention that $\sum_{j=qk-q+1}^{qk-1} r_{t,j}^* = 0$ when $q = 1$. The covariance between $\widetilde{r}_{t,k}$ and $\widetilde{r}_{t,k-1}$ is given by:

$$\begin{aligned} cov(\widetilde{r}_{t,k}, \widetilde{r}_{t,k-1}) &= -\left(\beta_0 + \frac{\beta_1}{\sigma_{t,qk-q}^*}\right)\left(1 + \beta_0 + \frac{\beta_1}{\sigma_{t,qk-q}^*}\right)\sigma_{t,qk-q}^{*2} \\ &\quad -\omega_0 + 2\omega_{m,q} - \omega_{m,2q} \end{aligned} \tag{1.35}$$

The next theorem gives the bias and variance of $RV_t^{(m_q)}$. The expression of the bias will be useful for the estimation of the correlogram of the microstructure noise in Section 6.

Theorem 5 *Assume that the noise process evolves according to equation (1.22), and let* $RV_t^{(m_q)} = \sum_{k=1}^{m_q} \widetilde{r}_{t,k}^2$ *with* $m_q = \frac{m}{q}$, $q \geq 1$ *and* m *the record frequency. Then we have:*

$$\begin{aligned} E\left[RV_t^{(m_q)}\right] &= IV_t + \underbrace{2m_q\left(\omega_0 - \omega_{m,q}\right)}_{\text{bias due to exogenous noise}} \\ &+ \underbrace{\frac{2\beta_1^2}{q} + \frac{2\beta_1\left(2\beta_0+1\right)}{\sqrt{m}}\sum_{k=1}^{m_q}\sigma_{t,qk}^* + 2\beta_0\left(\beta_0+1\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*2}}_{\text{bias due to endogenous noise}} \\ &+ \underbrace{\beta_0^2\left(\sigma_{t,0}^{*2} - \sigma_{t,m}^{*2}\right) + \frac{2\beta_0\beta_1}{\sqrt{m}}\left(\sigma_{t,0}^* - \sigma_{t,m}^*\right)}_{\text{end effects}}. \end{aligned}$$

$$\begin{aligned} Var\left[RV_t^{(m_q)}\right] &= m_q\kappa + \tfrac{16\beta_1^2}{q}\left(\omega_0 - \omega_{m,q}\right) + \tfrac{12\beta_1^4}{qm} \\ &+8\left[\tfrac{(3+5\beta_0)\beta_1^3}{m\sqrt{m}} + \tfrac{2(1+\beta_0)\beta_1}{\sqrt{m}} + \tfrac{2\beta_0\beta_1}{\sqrt{m}}\left(\omega_0 - \omega_{m,q}\right) + \tfrac{\beta_0\beta_1^3}{m\sqrt{m}}\right]\textstyle\sum_{k=1}^{m_q}\sigma_{t,qk}^* \\ &+4\left(1 + 2\beta_0 + 2\beta_0^2\right)\left[\tfrac{7\left(1+2\beta_0+2\beta_0^2\right)\beta_1^2}{m} + 2\left(\omega_0 - \omega_{m,q}\right)\right]\textstyle\sum_{k=1}^{m_q}\sigma_{t,qk}^{*2} \\ &+8\tfrac{\left(1+4\beta_0+6\beta_0^2+4\beta_0^3\right)\beta_1}{\sqrt{m}}\textstyle\sum_{k=1}^{m_q}\sigma_{t,qk}^3 + 2\sum_{k=1}^{m_q}\left(\sum_{j=qk-q+1}^{qk}\sigma_{t,j}^{*2}\right)^2 \\ &+\tfrac{16\beta_0(1+\beta_0)\beta_1^2}{m}\textstyle\sum_{k=1}^{m_q}\sigma_{t,qk-q}^*\sigma_{t,qk}^* + \tfrac{8(1+\beta_0)\beta_1}{\sqrt{m}}\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,qk}^* \\ &+\tfrac{8\beta_0^2(1+\beta_0)\beta_1}{\sqrt{m}}\textstyle\sum_{k=1}^{m_q}\sigma_{t,qk-q}^{*2}\sigma_{t,qk}^* + \tfrac{8\beta_0(1+\beta_0)^2\beta_1}{\sqrt{m}}\sum_{k=1}^{m_q}\sigma_{t,qk-q}^*\sigma_{t,qk}^{*2} \\ &+\tfrac{8\beta_0\beta_1}{\sqrt{m}}\textstyle\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,qk-q}^* + 2\left(4\beta_0 + 8\beta_0^2 + 8\beta_0^3 + 4\beta_0^4\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^4 \\ &+4\left(2\beta_0 + \beta_0^2\right)\textstyle\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,qk}^{*2} + 4\beta_0^2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,qk-q}^{*2} \\ &+4\beta_0^2\left(1+\beta_0\right)^2\textstyle\sum_{k=1}^{m_q}\sigma_{t,qk-q}^{*2}\sigma_{t,qk}^{*2} + 8\left(\omega_0 - \omega_{m,q}\right)\left(\beta_0^2 + \tfrac{2\beta_0\beta_1}{\sqrt{m}}\right) + O(m^{-1}), \end{aligned}$$

where $\kappa = \frac{1}{m_q} Var\left[\sum_{k=1}^{m_q}\left(\varepsilon_{t,kq} - \varepsilon_{t,kq-q}\right)^2\right]$.

Computing explicitly the exact expression of κ is not of direct interest in our analysis. Note that the dominant terms of the bias and of the variance of $RV^{(m_q)}$ are $O(m_q)$. In the case where $\varepsilon_{t,j}$ is IID, replacing $\beta_0 = \beta_1 = 0$ in the above expressions yields the result of Lemma 4 of Hansen and Lunde (2006) up to some changes in notations:

$$\begin{aligned} E\left[RV_t^{(m_q)}\right] &= IV_t + 2m_q\omega_0 \\ Var\left[RV_t^{(m_q)}\right] &= m_q\kappa + 8\omega_0 IV_t + 2\sum_{k=1}^{m_q}\left(\sum_{j=qk-q+1}^{qk}\sigma_{t,j}^{*2}\right)^2 \end{aligned} \tag{1.36}$$

where $m_q\kappa = 4m_qE\left[\varepsilon_{t,j}^4\right] + 2\left(\omega_0^2 - E\left[\varepsilon_{t,j}^4\right]\right)$ when $\varepsilon_{t,j}$ is IID.

We see that the volatility signature plot may not be able to reveal the presence of the noise in the data if $\varepsilon_{t,j} = 0$, since in this case the bias is equal to:

$$\frac{2\beta_1^2}{q} + 2\beta_1\left(2\beta_0 + 1\right)\frac{1}{\sqrt{m}}\sum_{k=1}^{m_q}\sigma_{t,qk}^{*} + 2q\beta_0\left(\beta_0 + 1\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*2} = O(1) \text{ for all } m_q$$

Moreover, this bias can be negative at some sampling frequencies provided that $\beta_1 < 0$ or $\beta_0 < 0$. Finally, note that the total bias of the RV sampled at the highest frequency may diverge at a lower rate than m, since:

$$2m\left(\omega_0 - \omega_{m,1}\right) = O(\omega_0 m^{1-\alpha}).$$

In the next section, we pursue with the examination of the implication of the microstructure noise model for two kernel-based estimators.

We examine successively the estimators of Hansen and Lunde (2006) and Barndorff-Nielsen and al (2008) under our microstructure noise model. This exercise if a useful step in the process of designing a good shrinkage estimator for the IV.

1.4.2 Hansen and Lunde (2006)

Hansen and Lunde (2006) proposed the following flat kernel estimator:

$$RV_t^{(AC,m,L+1)} = \gamma_{t,0}\left(r\right) + \sum_{h=1}^{L+1}\left(\gamma_{t,h}\left(r\right) + \gamma_{t,-h}\left(r\right)\right), \tag{1.37}$$

where L is the memory of the noise as defined in E2 and $\gamma_{t,h}(r)$ is defined in (1.11). Note that when $L=0$ so that $\varepsilon_{t,j}$ is IID, $RV_t^{(AC,m,L+1)}$ coincides with the estimator of French and al. (1987) and Zhou (1996):

$$RV_t^{(AC,m,1)} = \gamma_{t,0}(r) + 2\sum_{j=1}^{m}\gamma_{t,1}(r) + \underbrace{(r_{t,m+1}r_{t,m} - r_{t,1}r_{t,0})}_{\text{end effects}}. \tag{1.38}$$

The variance of $RV_t^{(AC,m,L+1)}$ is hard to derive in the general case. However, assuming that $\varepsilon_{t,j}$ is IID and neglecting the end effects in (1.38) leads to the following result for $RV_t^{(AC,m,1)}$.

Theorem 6 *Assume that the noise process evolves according to Equation (1.22). If $\varepsilon_{t,j}$ is IID, we have:*

$$\begin{aligned}
&E\left[RV_t^{(AC,m,1)}\right] = IV_t + \left(\beta_0^2 + 2\beta_0\right)\left(\sigma_{t,m}^{*2} - \sigma_{t,0}^{*2}\right) - \tfrac{2\beta_1(1+\beta_0)}{\sqrt{m}}\left(\sigma_{t,m}^{*} - \sigma_{t,0}^{*}\right) \\
&Var\left[RV_t^{(AC,m,1)}\right] = 8m\omega_0^2 + 2\textstyle\sum_{j=1}^m \sigma_{t,j}^{*4} + 2\left(E\left[\varepsilon_{t,j}^4\right] - \omega_0^2\right) \\
&\quad + \tfrac{\beta_1^4 + 6\beta_1^2\omega_0}{m} + \tfrac{8\beta_1^4}{m^2} + \tfrac{8\beta_1}{\sqrt{m}}\left[\tfrac{(\beta_0+1)^2\beta_1^2}{m} + \tfrac{\beta_1}{\sqrt{m}} + 2\omega_0\left(1 + 2\beta_0\right)\right]\textstyle\sum_{j=1}^m \sigma_{t,j}^{*} \\
&\quad + 8\left[\tfrac{\beta_0^2\beta_1^2}{m} + \left(\tfrac{\beta_1^2}{m} + \omega_0\right)\left(1+\beta_0\right)^2 + 2\omega_0\beta_0^2\right]\textstyle\sum_{j=1}^m \sigma_{t,j}^{*2} \\
&\quad + \tfrac{8\beta_1^2}{m}\left(1 + 2\beta_0 + 2\beta_0^2\right)\textstyle\sum_{j=1}^m \sigma_{t,j}^{*}\sigma_{t,j-1}^{*} + \tfrac{16\beta_0\beta_1^2}{m}\left(1+\beta_0\right)\textstyle\sum_{j=1}^m \sigma_{t,j}^{*}\sigma_{t,j-2}^{*} \\
&\quad + \tfrac{8\beta_0\beta_1}{\sqrt{m}}\left(1 + \beta_0 + \beta_0^3\right)\textstyle\sum_{j=1}^m \sigma_{t,j}^{*2}\sigma_{t,j-1}^{*} + \tfrac{8\beta_1}{\sqrt{m}}\left(1 + 2\beta_0 + 2\beta_0^2 + \beta_0^3\right)\textstyle\sum_{j=1}^m \sigma_{t,j-1}^{*2}\sigma_{t,j}^{*} \\
&\quad + \tfrac{8\beta_1(1+\beta_0)\beta_0^2}{\sqrt{m}}\textstyle\sum_{j=1}^m \sigma_{t,j-2}^{*2}\sigma_{t,j}^{*} + \tfrac{8\beta_1\beta_0(1+\beta_0)^2}{\sqrt{m}}\textstyle\sum_{j=1}^m \sigma_{t,j}^{*2}\sigma_{t,j-2}^{*} \\
&\quad + 4\left(1 + 2\beta_0 + 3\beta_0^2 + 2\beta_0^3 + \beta_0^4\right)\textstyle\sum_{j=1}^m \sigma_{t,j-1}^{*2}\sigma_{t,j}^{*2} \\
&\quad + 4\beta_0^2\left(1+\beta_0\right)^2\textstyle\sum_{j=1}^m \sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2} + \beta_0 O\left(m^{-1/2}\right) \blacksquare
\end{aligned}$$

Replacing $\beta_0 = \beta_1 = 0$ in this theorem yields a known result (Lemma 5 of Hansen and Lunde, 2006):

$$\begin{aligned}
E\left[RV_t^{(AC,m,1)}\right] &= IV_t, \\
Var\left[RV_t^{(AC,m,1)}\right] &\simeq 8m\omega_0^2 + 8\omega_0 IV_t - 6\omega_{m,0}^2 + 2\sum_{j=1}^m \sigma_{t,j}^{*4} + 4\sum_{j=1}^m \sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2}.
\end{aligned}$$

When the exogenous noise is absent ($\varepsilon_{t,j} = 0$) and $\beta_0 \neq 0$ or $\beta_1 \neq 0$, the estimator $RV_t^{(AC,m,1)}$ is slightly biased and the bias vanishes at rate $O\left(m^{-1}\right)$.

$$E\left[RV_t^{(AC,m,1)}\right] - IV_t = \left(\beta_0^2 + 2\beta_0\right)\left(\sigma_{t,m}^{*2} - \sigma_{t,0}^{*2}\right) - \frac{2\beta_1\left(1+\beta_0\right)}{\sqrt{m}}\left(\sigma_{t,m}^{*} - \sigma_{t,0}^{*}\right)$$

By examining each of the individual terms in the expression of the variance of $RV_t^{(AC,m,1)}$, it is seen that $RV_t^{(AC,m,1)}$ converges to IV_t at rate $\sqrt{m}$ when $\varepsilon_{t,j} = 0$.

1.4.3 Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a)

The expression of this estimator denoted $K_{t,Lead}^{BNHLS}$ has already been introduced in Equation (1.10). For practical purpose, we shall rewrite it as:

$$\begin{aligned} K_t^{BNHLS} &= \frac{1}{2}\left(K_{t,Lead}^{BNHLS} + K_{t,Lag}^{BNHLS}\right) \\ &= K_{t,Lead}^{BNHLS} + \frac{1}{2}\left(K_{t,Lag}^{BNHLS} - K_{t,Lead}^{BNHLS}\right), \end{aligned}$$

where

$$\begin{aligned} K_{t,Lead}^{BNHLS} &= \gamma_{t,0}(r) + 2\sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\gamma_{s,h}(r) \\ K_{t,Lag}^{BNHLS} &= \gamma_{t,0}(r) + 2\sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\gamma_{s,-h}(r) \end{aligned}$$

In studying the asymptotic properties of K_t^{BNHLS}, the end effects $\frac{1}{2}\left(K_{t,Lag}^{BNHLS} - K_{t,Lead}^{BNHLS}\right)$ are difficult to handle. However, $K_{t,Lead}^{BNHLS}$ and $K_{t,Lag}^{BNHLS}$ have the same expectation and similar asymptotic variances while being imperfectly uncorrelated. This translates into the following equations:

$$\begin{aligned} E\left[K_{t,Lag}^{BNHLS} - K_{t,Lead}^{BNHLS}\right] &= 0 \\ Var\left(\frac{1}{2}K_{t,Lead}^{BNHLS} + \frac{1}{2}K_{t,Lag}^{BNHLS}\right) &\leq Var\left(K_{t,Lead}^{BNHLS}\right) \end{aligned}$$

For simplicity, we shall thus ignore these end effects. In the worse case, this restriction will give an upper bound for the true variance of K_t^{BNHLS}. Accordingly, we introduce the "Lead" versions of $K_t(x)$ and $K_t(x,y)$ defined under Equation (1.10) in order to be able to write:

$$K_t^{BNHLS} = K_t(r^*) + K_t(r^*, \Delta u) + K_t(\Delta u, r^*) + K_t(\Delta u)$$

where

$$\begin{aligned} r_{t,j} &= r_{t,j}^{*} + \Delta u_{t,j} \\ \Delta u_{t,j} &= \left(\beta_0 + \frac{\beta_1}{\sqrt{m}\sigma_{t,j}^{*}}\right) r_{t,j}^{*} - \left(\beta_0 + \frac{\beta_1}{\sqrt{m}\sigma_{t,j-1}^{*}}\right) r_{t,j-1}^{*} + \left(\varepsilon_{t,j} - \varepsilon_{t,j-1}\right). \end{aligned}$$

Barndorff-Nielsen and al (2008a) show that their estimator is robust to an endogenous noise with $\beta_1 = 0$. Also, we have seen in the previous subsection that $RV_t^{(AC,m,1)}$ is consistent for IV_t when there is no exogenous noise in the data. Interestingly, K_t^{BNHLS} has the following representation:

$$K_t^{BNHLS} = RV_t^{(AC,m,1)} + \sum_{h=2}^{H} k\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}(r) + \gamma_{t,-h}(r)\right),$$

where $\sum_{h=2}^{H} k\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}(r) + \gamma_{t,-h}(r)\right)$ is unbiased and consistent for zero when $\varepsilon_{t,j} = 0$. In fact, the observed log-return $r_{t,j}$ is not autocorrelated beyond lag one in this case while $Var\left(r_{t,j}\right) = O(m^{-1})$. As a result, K_t^{BNHLS} is robust to the type of endogenous noise assumed here. For simplicity, we shall thus focus below on the asymptotic behavior of K_t^{BNHLS} under $\beta_0 = \beta_1 = 0$. We have the following theorem.

Theorem 7 *Assume $\beta_0 = \beta_1 = 0$ and that E1 to E4 are satisfied with $\delta \neq 0$. Further let $k(x) = 1 - x$ (the Bartlett kernel). Then for sufficiently large H and m, we have:*

$$\begin{aligned} K_t(r^*) - IV_t &= O_p(H^{1/2}m^{-1/2}), \\ Var\left[K_t(r^*, \Delta u)\right] &\approx \frac{2\omega_0}{H} + 4\sum_{h=1}^{L}\left(\omega_{m,h} - \omega_{m,h+1}\right)\left[1 - \frac{(h+1)^2}{H^2}\right], \\ K_t(\Delta u) &= -\varepsilon_{t,0}^2 + \varepsilon_{t,m}^2 - \frac{4}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H} - \frac{2}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H-1} \\ &\quad -\frac{2}{H}\sum_{h=2}^{H-1}\left(\varepsilon_{t,0}\varepsilon_{t,-h} - \varepsilon_{t,m}\varepsilon_{t,m-h}\right) + \frac{2}{H}\left(\varepsilon_{t,0}\varepsilon_{t,-H} - \varepsilon_{t,m}\varepsilon_{t,m-H}\right). \end{aligned}$$

where we recall that $\omega_{m,L+1} = 0$ in the expression of $Var\left[K_t(r^*, \Delta u)\right]$.

In the IID noise case, we have $\omega_{m,h} = 0$ for all $h \geq 1$. Hence setting $H \propto m^{2/3}$ yields immediately the same result as in Barndorff-Nielsen and al (2008a) up to the end effects:

$$K_t^{BNHLS} - IV_t = -\varepsilon_{t,0}^2 + \varepsilon_{t,m}^2 + O_p(m^{-1/6}).$$

The estimator K_t^{BNHLS} is thus consistent for IV_t if we are willing to neglect the end effects $-\varepsilon_{t,0}^2 + \varepsilon_{t,m}^2$[8].

In the dependent case, we have:

$$\begin{aligned} Var\left[K_t\left(r^*, \Delta u\right)\right] &\simeq 4\omega_{m,L}\left[1 - \frac{(L+1)^2}{H^2}\right] \\ &\quad + \omega_0 O\left(m^{-\alpha}\sum_{h=1}^{L}\left[1 - \frac{(h+1)^2}{H^2}\right]\right) \\ &= 4\omega_{m,L} + \omega_0 O\left(m^{-(\alpha-\delta)}\right) \end{aligned} \tag{1.39}$$

where $\sum_{h=1}^{L}\left[1 - \frac{(h+1)^2}{H^2}\right] = O(L) = O(m^\delta)$ and we recall that $\delta \leq \alpha$ by construction. Here we have two sub-cases:

Case $\delta < \alpha$**:** The term $\omega_0 O\left(m^{-(\alpha-\delta)}\right)$ vanishes so that:

$$\lim_{m\to\infty} Var\left[K_t\left(r^*, \Delta u\right)\right] = 4\lim_{m\to\infty}\omega_{m,L}$$

On the other hand, Equation (1.23), indicates that $\omega_{m,L} = \omega_0 - L\omega_0 O(m^{-\alpha})$. Finally,

$$\lim_{m\to\infty} Var\left[K_t\left(r^*, \Delta u\right)\right] = 4\lim_{m\to\infty}\left(\omega_0 - \omega_0 O(m^{-\alpha+\delta})\right) = 4\omega_0$$

Case $\delta = \alpha$**:** Here the term $\omega_0 O\left(m^{-(\alpha-\delta)}\right) \propto \omega_0$ no longer vanish and:

$$\lim_{m\to\infty}\omega_{m,L} = \lim_{m\to\infty}\left(\omega_0 - \omega_0 O(1)\right) \propto \omega_0$$

This leads to:

$$\lim_{m\to\infty} Var\left[K_t\left(r^*, \Delta u\right)\right] \propto \omega_0$$

Strictly speaking, K_t^{BNHLS} is not consistent in the dependent case. But overall, the estimator

[8] See BNHLS (2007) for the treatment of these end effects in practice.

K_t^{BNHLS} delivers its best performance for large m, no matter whether the noise is IID or not.

In the next section, we study the properties of the microstructure noise.

1.5 Inference on the Microstructure Noise Parameters

From now one, the notation $\gamma_{t,h}$ is used for $\gamma_{t,h}(r)$ where the latter is defined in (1.11). We note from (1.28) that:

$$\begin{aligned} E\left[\gamma_{t,1}\right] &= -\sum_{j=1}^{m}\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma_{t,j-1}^{*}}\right)\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma_{t,j-1}^{*}}\right)\sigma_{t,j-1}^{*2} \\ &\quad +m\left(-\omega_0+2\omega_{m,1}-\omega_{m,2}\right) \end{aligned}$$

where we recall that $\omega_{m,h}$ is the h^{th} autocovariance of $\varepsilon_{t,j}$ when observed at frequency m.

Let $b_t^{(m)} = E\left[RV_t^{(m)} - IV_t\right]$ denote the bias of the realized volatility computed at the record frequency. It follows from Lemma 8 in appendix that when $q=1$, we have:

$$\begin{aligned} b_t^{(m)} &= 2\sum_{j=1}^{m}\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma_{t,j-1}^{*}}\right)\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma_{t,j-1}^{*}}\right)\sigma_{t,j-1}^{*2} \\ &\quad +2m\left(\omega_0-\omega_{m,1}\right)+\beta_0^2\left(\sigma_{t,0}^{*2}-\sigma_{t,m}^{*2}\right)+\frac{2\beta_0\beta_1}{\sqrt{m}}\left(\sigma_{t,0}^{*}-\sigma_{t,m}^{*}\right) \end{aligned}$$

The endogenous parameters β_0 and β_1 hidden in the expression of the bias $b_t^{(m)}$ are unfortunately unidentified. We shall thus focus on the estimation of the bias as a whole rather that tackling β_0 and β_1 individually. In subsection 5.1, we discuss the estimation of $b_t^{(m)}$ and $\{\omega_{m,h}\}_{h=0}^{L}$ while in subsection 5.2 we deal with the memory parameters (L, α, δ).

1.5.1 Estimation of the Correlogram

By neglecting the $O(m^{-1})$ end terms in the expression of the bias $b_t^{(m)}$, we obtain the following moment conditions:

$$E\left[RV_t^{(m)} - b_t^{(m)} - IV_t\right] = 0 \qquad (1.40)$$

$$E\left[b_t^{(m)} + \left(\gamma_{t,1}+\gamma_{t,-1}\right) - 2m\left(\omega_{m,1}-\omega_{m,2}\right)\right] = 0 \qquad (1.41)$$

In addition, we also have:

$$E\left[\left(\gamma_{t,h+1}+\gamma_{t,-h-1}\right)-2m\left(-\omega_{m,h}+2\omega_{m,h+1}-\omega_{m,h+2}\right)\right]=0,\ 1\leq h\leq L \tag{1.42}$$

Given that $\omega_{m,h}=0$ for $h>L$, we have $L+2T$ moment conditions to estimate $L+2T$ parameters. Estimating these parameters by the method of moments is straightforward. Solving first for $\omega_{m,L}$ and then proceeding by backward substitution yields:

$$\begin{aligned}
\widehat{\omega}_{m,h} &= -\frac{1}{2Tm}\sum_{s=1}^{T}\sum_{l=1}^{L-h+1} l\left(\gamma_{s,h+l}+\gamma_{s,-h-l}\right),\ h=1,...L && (1.43)\\
\widehat{b}_t^{(m)} &= -\gamma_{t,1}-\gamma_{t,-1}-\frac{1}{T}\sum_{s=1}^{T}\sum_{l=2}^{L+1}\left(\gamma_{s,l}+\gamma_{s,-l}\right), && (1.44)\\
\overline{RV}_t^{(AC,m,L+1)} &= \gamma_{t,0}+\gamma_{t,1}+\gamma_{t,-1}+\frac{1}{T}\sum_{s=1}^{T}\sum_{l=2}^{L+1}\left(\gamma_{s,l}+\gamma_{s,-l}\right) && (1.45)
\end{aligned}$$

where $\widehat{\omega}_{m,h}$, $\widehat{b}_t^{(m)}$ and $\overline{RV}_t^{(AC,m,L+1)}$ are unbiased estimators of $\omega_{m,h}$, $b_t^{(m)}$ and IV_t respectively[9]. It is seen that $\overline{RV}_t^{(AC,m,L+1)}$ is a bias corrected version of the standard realized variance which uses the data available at all periods to estimate the IV of each period. To estimate the variance ω_0, we use the expression of the bias of the RV sampled at the highest frequency. We have:

$$\begin{aligned}
\widehat{\omega}_0 &= \frac{1}{2mT}\sum_{t=1}^{T}\widehat{b}_t^{(m)}+\widehat{\omega}_{m,1} && (1.46)\\
&= -\frac{1}{2mT}\sum_{t=1}^{T}\sum_{l=1}^{L+1}\left(\gamma_{t,l}+\gamma_{t,-l}\right)+\widehat{\omega}_{m,1}.
\end{aligned}$$

To estimate the covariance matrix of $\widehat{\omega}_m=\left(\widehat{\omega}_{m,0},\widehat{\omega}_{m,1},...,\widehat{\omega}_{m,L}\right)'$, let us define:

$$\begin{aligned}
\widehat{\omega}_{m,(1-L)} &= \left(\widehat{\omega}_{m,1},...,\widehat{\omega}_{m,L}\right)'\\
\overline{\gamma}_{t,(2-L)} &= \left(\overline{\gamma}_{t,2},...,\overline{\gamma}_{t,L+1}\right)'
\end{aligned}$$

where $\overline{\gamma}_{t,h}=\frac{1}{2m}\sum_{j=1}^{m}r_{t,j}\left(r_{t,j-h}+r_{t,j+h}\right)$ for all t and h. Then we have the following relation

[9] When the data are non equally spaced, the expressions of the autocorrelation estimators are more tedious. See for example Ubukata and Oya (2009).

between the autocovariances of the noise and those of the observed returns:

$$\widehat{\omega}_{m,(1-L)} = \frac{1}{T}\sum_{t=1}^{T} P^{-1}\overline{\gamma}_{t,(2-L)}$$

where P is the $L \times L$ matrix with elements: $P_{i,i} = -1$, $P_{i,i+1} = 2$, $P_{i,i+2} = -1$, and $P_{i,j} = 0$ otherwise $1 \leq i, j \leq L$.

If we further define:

$$\widehat{\omega}_{t,m} = \left[-\frac{1}{2m}\sum_{l=1}^{L+1}\left(\gamma_{t,l} + \gamma_{t,-l}\right) + \left(P^{-1}\overline{\gamma}_{t,(2-L)}\right)_1, \left(P^{-1}\overline{\gamma}_{t,(2-L)}\right)' \right]',$$

with $\left(P^{-1}\overline{\gamma}_{t,(2-L)}\right)_1$ being the first element of $P^{-1}\overline{\gamma}_{t,(2-L)}$, then we are able to write:

$$\widehat{\omega}_m = \frac{1}{T}\sum_{t=1}^{T}\widehat{\omega}_{t,m} \tag{1.47}$$

It is seen that $\widehat{\omega}_{t,m}$ depends on only time t observations. Because $r_{t,j}$ is stationary with finite fourth moments under our assumptions on the efficient returns and the microstructure noise, the vector process $\widehat{\omega}_{t,m}$ admits a finite covariance matrix and we have:

$$\sqrt{T}\left(\widehat{\omega}_m - \omega_m\right) \to N\left(0, Avar\left(\widehat{\omega}_{t,m}\right)\right)$$

as T goes to infinity and m is fixed. The long run covariance matrix $Avar\left(\widehat{\omega}_{t,m}\right)$ may be estimated as in Newey and West (1987). For example:

$$\begin{aligned} \widehat{Avar}\left(\widehat{\omega}_{t,m}\right) &= \frac{1}{T}\sum_{t=1}^{T}\widehat{\omega}_{t,m}\widehat{\omega}'_{t,m} \\ &+ \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{q}\left(1 - \frac{k-1}{q}\right)\widehat{\omega}_{t,m}\left(\widehat{\omega}'_{t+k,m} + \widehat{\omega}'_{t-k,m}\right) \end{aligned} \tag{1.48}$$

where q is the bandwidth.

1.5.2 Assessing the true values of L, α and δ

The knowledge of the memory parameter L is required to estimate the correlogram of the microstructure noise. We suggest the following information criterion for its estimation:

$$\widehat{L} = \underset{0 \leq l \leq H-1}{\arg\min} \left\{ \Delta(l) = \frac{1}{T} \sum_{t=1}^{T} \left(K_t^{H,T} - \overline{RV}_t^{(AC,m,l+1)} \right)^2 \right\}, H \propto m^{2/3} \tag{1.49}$$

where $\overline{RV}_t^{(AC,m,l+1)}$ is defined as in (1.45) and:

$$K_t^{H,T} = RV_t^{(Ac,m,1)} + \frac{1}{T} \sum_{s=1}^{T} \sum_{h=2}^{H} \left(1 - \frac{h-1}{H} \right) \left(\gamma_{s,h} + \gamma_{s,-h} \right)$$

To see the intuition underlying this information criterion, note that $\Delta(l)$ satisfies:

$$E\left[\Delta(l)\right] = Var\left(K_t^{H,T} - \overline{RV}_t^{(AC,m,l+1)} \right) + \left[E\left(K_t^{H,T} - \overline{RV}_t^{(AC,m,l+1)} \right) \right]^2$$

where the moments are taken unconditionally. On the one hand, $\overline{RV}_t^{(AC,m,l+1)}$ is obtained by truncating the expression of $\widehat{IV}_t$ to l autocovariance terms and is thus unbiased for IV_t when $l \geq L$. On the other hand, $K_t^{H,T}$ is a smoothed version of $\overline{RV}_t^{(AC,m,H)}$ which is also unbiased for IV_t due to $L < H \propto m^{2/3}$. Hence $E\left(K_t^{H,T} - \overline{RV}_t^{(AC,m,l+1)} \right)$ is decreasing in l in the area $l < L$ and equals zero in the area $l \geq L$. Because the term $Var\left(K_t^{H,T} - \overline{RV}_t^{(AC,m,l+1)} \right)$ is increasing in l, there is a trade-off between bias and variance that results in a L-shaped curve $\Delta(l)$. The power of this information criterion comes from the fact each of the statistics $K_t^{H,T}$ and $\widehat{IV}_t^{(L)}$ fully exploits the stationarity of the exogenous noise across days. An illustration is provided in the simulation study of Section 8.

We now discuss the estimation of α and δ. Assumption E3 stipulates that ω_0 is constant for all m while $\omega_{m,h} - \omega_{m,h+1} = \omega_0 O(m^{-\alpha})$ for some $\alpha < 2/3$, $h = 0, ..., L-1$. More precisely, we write:

$$\frac{\omega_{m,h} - \omega_{m,h+1}}{\omega_0} \simeq C_h m^{-\alpha}, \text{ with } C_h \in \left[\underline{C}, \overline{C} \right]$$

Taking the logs of both side of the equality and averaging over h yields:

$$\alpha \simeq \frac{-1}{L \log m} \sum_{h=0}^{L-1} \log \left(\frac{\omega_{m,h} - \omega_{m,h+1}}{\omega_0} \right) + \frac{1}{L \log m} \sum_{h=0}^{L-1} \log C_h,$$

where $\frac{1}{L \log m} \sum_{h=1}^{L-1} \log C_h \in \left[\frac{\log \underline{C}}{\log m}, \frac{\log \overline{C}}{\log m}\right]$ so that this term shrinks to zero as m goes to infinity. A consistent estimator of α is thus given by:

$$\widehat{\alpha} = \frac{1}{\log m}\left[\log \widehat{\omega}_0 - \frac{1}{L}\sum_{h=0}^{L-1} \log\left(\widehat{\omega}_{m,h} - \widehat{\omega}_{m,h+1}\right)\right] \tag{1.50}$$

Note that $\widehat{\alpha}$ is a Hill (1975) type estimator for the tail index of a distribution. It is difficult to analyze the asymptotic behavior of this estimator as m goes to infinity. However, we can exploit the functional relationship between $\widehat{\alpha}$ and $\{\widehat{\omega}_{m,h}\}_{h=1}^{m}$ to obtain an asymptotic distribution as T goes to infinity and m is fixed. Using the Delta method, we obtain the following Central Limit result:

$$\sqrt{T}\left(\widehat{\alpha} - \alpha\right) \to N\left(0, \frac{\left(\nabla\alpha\right)' Avar\left(\widehat{\omega}_{t,m}\right)\left(\nabla\alpha\right)}{\left(\log m\right)^2}\right)$$

as T goes to infinity, where the elements of the vector $\nabla\alpha$ are given by:

$$\begin{aligned}
\left(\nabla\alpha\right)_1 &= \frac{1}{\omega_0} - \frac{1}{L\left(\omega_0 - \omega_1\right)}; \\
\left(\nabla\alpha\right)_h &= \frac{1}{L\left(\omega_{m,h-2} - \omega_{m,h-1}\right)} - \frac{1}{L\left(\omega_{m,h-1} - \omega_{m,h}\right)}; \; 2 \leq h \leq L \\
\left(\nabla\alpha\right)_{L+1} &= \frac{1}{L\left(\omega_{L-1} - \omega_L\right)}.
\end{aligned}$$

To estimate the parameter δ, we note that $L \propto m^{\delta}$ according to Assumption E4. Assuming again $L \approx Cm^{\delta}$ leads to:

$$\widehat{\delta} = \frac{\log \widehat{L}}{\log m} \tag{1.51}$$

where we neglect the bias $\frac{\log C}{\log m}$.

Our next step is to propose a good estimator of the IV.

1.6 Shrinkage Realized Kernels

In subsections 2.3 and 2.4, we have designed and studied the asymptotic properties of a shrinkage estimator of IV that is suitable for the IID noise case. In this section, we extend the analysis to the case where the noise is dependent.

On the one hand, we have the unbiased method of moment estimators $\overline{RV}_t^{(AC,m,L+1)}$ introduced in (1.45) which makes an efficient use of the data under the assumption that the properties of the noise do not change over time. On the other hand, we have the realized kernels K_t^{BNHLS} of Barndorff-Nielsen and al (2008a) which delivers its best performance at the highest possible frequency. We have:

$$K_t^{BNHLS} - IV_t = O_p\left(m^{-\eta}\right), \text{ for some } \eta \geq 0$$

It is thus tempting to design the shrinkage estimator as an optimal linear combination of K_t^{BNHLS} and $\overline{RV}_t^{(AC,m,L+1)}$. However, we prefer not to use $\overline{RV}_t^{(AC,m,L+1)}$ for two reasons. Firstly, the properties of the noise may be changing over time contrary to what we assumed. And secondly, an estimator that combines K_t^{BNHLS} and $\overline{RV}_t^{(AC,m,L+1)}$ cannot be written as a proper realized kernel that uses only within day observations. So instead, of combining K_t^{BNHLS} and $\overline{RV}_t^{(AC,m,L+1)}$, we combine K_t^{BNHLS} and a smoother version of $\overline{RV}_t^{(AC,m,L+1)}$ denoted $\theta_{1,t}^{(L)}$:

$$\begin{aligned}
\theta_{1,t}^{(L)} &= \gamma_{t,0}\left(r\right) + \sum_{h=1}^{L+1} k\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}\left(r\right) + \gamma_{t,-h}\left(r\right)\right) && (1.52)\\
\theta_{2,t}^{(L)} &= \sum_{h=L+2}^{H} k\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}\left(r\right) + \gamma_{t,-h}\left(r\right)\right) && (1.53)
\end{aligned}$$

so that

$$K_t^{BNHLS} = \theta_{1,t}^{(L)} + \theta_{2,t}^{(L)}$$

Also, $\theta_{1,t}^{(L)}$ is a smoothed version of $RV_t^{(AC,m,L+1)}$ and is unbiased for the IV when $k\left(x\right) = 1 - x$.

We can thus take linear combinations of the form:

$$\begin{aligned}
K_t^{\varpi} &= \varpi K_t^{BNHLS} + (1-\varpi)\theta_{1,t}^{(L)}, \ \varpi \in \mathbb{R}. && (1.54)\\
&= \theta_{1,t}^{(L)} + \varpi\theta_{2,t}^{(L)}
\end{aligned}$$

Indeed, K_t^{ϖ} is a realized kernels with kernel function given by:

$$
\begin{aligned}
g(x) &= k(x), \ 0 \leq x \leq \frac{L}{H} \\
g(x) &= \varpi k(x), \ \frac{L}{H} < x \leq 1
\end{aligned}
$$

The function $g(x)$ is discontinuous at $x = \frac{L}{H}$ unless $\varpi = 1$.

As in the IID case, the weight ϖ is selected to minimize the quadratic loss function:

$$
\varpi_t^* = \arg\min_{\varpi} E\left[\left(K_t^{\varpi} - IV_t\right)^2 \mid \{\sigma\}\right] \tag{1.55}
$$

The optimal shrinkage weight is:

$$
\varpi_t^* = -\frac{Cov\left[\theta_{1,t}^{(L)}, \theta_{2,t}^{(L)} \mid \{\sigma\}\right]}{Var\left[\theta_{2,t}^{(L)} \mid \{\sigma\}\right]}. \tag{1.56}
$$

However, these conditional second moments are not easy to compute. A simple strategy is to look for a constant shrinkage weight ϖ^* that minimizes the marginal variance of K_t^{ϖ}. By the law of total variance, we have:

$$
\begin{aligned}
Var_{Total}\left(K_t^{\varpi}\right) &= Var\left[E\left(K_t^{\varpi} \mid \{\sigma\}\right)\right] + E\left[Var\left(K_t^{\varpi} \mid \{\sigma\}\right)\right] \\
&= Var\left[IV_t\right] + E\left[Var\left(K_t^{\varpi} \mid \{\sigma\}\right)\right].
\end{aligned}
$$

Therefore, choosing ϖ to minimize the marginal variance of K_t^{ϖ} is equivalent to choosing ϖ to minimize the expected conditional variance of K_t^{ϖ}. While ϖ^* achieves on average the same goal as the ideal weight ϖ_t^*, it is easier to estimate:

$$
\widehat{\varpi}^* = -\frac{\frac{1}{T}\sum_{t=1}^{T}\left(\theta_{1,t}^{(L)} - \overline{\theta_{1,T}^{(L)}}\right)\theta_{2,t}^{(L)}}{\frac{1}{T}\sum_{t=1}^{T}\left(\theta_{2,t}^{(L)}\right)^2}. \tag{1.57}
$$

where $\overline{\theta_{1,T}^{(L)}} = \sum_{t=1}^{T}\theta_{1,t}^{(L)}$.

The marginal variance of the shrinkage estimator with constant weight ϖ^* is:

$$Var_{Total}\left(K_t^{\varpi^*}\right) = Var_{Total}\left(\theta_{1,t}^{(L)}\right)\left[1-\rho_{1,2}^2\right]$$

where $\rho_{1,2}$ now denotes the marginal correlation between $\theta_{1,t}^{(L)}$ and $\theta_{2,t}^{(L)}$. This implies that the rate at which $1-\rho_{1,2}^2$ goes to zero is slower compared to the IID case:

$$1-\rho_{1,2}^2 = \frac{Var_{Total}\left(K_t^{\varpi^*}\right)}{Var_{Total}\left(\theta_{1,t}^{(L)}\right)} = O(m^{-1}). \tag{1.58}$$

In fact, this follows from:

$$\begin{aligned} Var_{Total}\left(K_t^{\varpi^*}\right) &\leq \underbrace{Var\left[IV_t\right]}_{O(1)} + \underbrace{E\left[Var\left(K_t^{BNHLS}\middle|\{\sigma\}\right)\right]}_{O(m^{-2\eta})} = O(1) \\ Var_{Total}\left(\theta_{1,t}^{(L)}\right) &= Var\left[IV_t\right] + \underbrace{E\left[Var\left(\theta_{1,t}^{(L)}\middle|\{\sigma\}\right)\right]}_{O(m)} = O(m) \end{aligned}$$

The efficiency gain of the shrinkage estimator with respect to K_t^{BNHLS} is:

$$\begin{aligned} Var_{Total}\left(K_t^{BNHLS}\right) - Var_{Total}\left(K_t^{\varpi^*}\right) &= \\ \left(\rho_{1,2}\sqrt{Var_{Total}\left(\theta_{1,t}\right)} + \sqrt{Var_{Total}\left(\theta_{2,t}\right)}\right)^2 &\geq 0 \end{aligned}$$

The consistency of $K_t^{\varpi^*}$ implies at large frequency the approximation $\theta_{1,t}^{(L)} = -\varpi^*\theta_{2,t}^{(L)} + IV_t$. On the one hand, we have:

$$\begin{aligned} Var_{Total}\left(K_t^{BNHLS}\right) &= Var_{Total}\left(\theta_{1,t}^{(L)} + \theta_{2,t}^{(L)}\right) \\ &= \left(1-\varpi^*\right)^2 Var\left(\theta_{2,t}\right) \end{aligned}$$

On the other hand,

$$Var_{Total}\left(K_t^{BNHLS}\right) = \underbrace{Var\left[IV_t\right]}_{O(1)} + \underbrace{E\left[Var\left(K_t^{BNHLS}\middle|\{\sigma\}\right)\right]}_{O(m^{-2\eta})} = O(1)$$

Putting the two together yields $\left(1-\varpi^{*}\right)^{2} Var\left(\theta_{2,t}\right)=O(1)$ so that finally:

$$1-\varpi^{*}=O(m^{-1/2}) \tag{1.59}$$

Hence the rate of the unconditionally optimal shrinkage weight does not depend of η.

This shrinkage method can be used independently of the postulated microstructure noise model. In particular, it can be adapted to the two scale realized volatility of Ait-Sahalia, Mykland and Zhang (2006).

1.7 Monte Carlo Evidence

The aim in this subsection is to assess the performance of the shrinkage estimator of IV and the quality of the estimators of $\{\omega_{m,l}\}_{l=0}^{\widehat{L}}$ by simulations.

1.7.1 Simulation Design

We assumed that the efficient log-price process evolves according to the model of Heston (1993):

$$\begin{aligned} dp_t^* &= \sigma_t dW_{1,t} && (1.60)\\ d\sigma_t^2 &= \kappa\left(\alpha-\sigma_t^2\right)dt+\gamma\sigma_t\left[\rho dW_{1,t}+\sqrt{1-\rho^2}dW_{2,t}\right] && (1.61)\end{aligned}$$

where $W_{1,t}$ and $W_{2,t}$ are independent Brownian motions and the parameter ρ captures the so-called leverage effect. Following Zhang and al. (2005), we set the parameters values as follows:

$$\kappa=5;\alpha=0.04;\gamma=0.5;\rho\in\{0,-0.5\}$$

where $\rho=0$ corresponds to the no leverage assumption made in deriving our analytical results. The case $\rho=-0.5$ is used to check the robustness of our conclusions. The unit period in this specification is one year.

We simulated data for $T=1000$ days using Euler discretization scheme at one second. Assuming that the market opens from 9:30 am to 4:00 pm, this yields 23400 discretization points within each day. We then consider four frequencies at which the price can be observed:

30 seconds, one minute, two minutes and five minutes. This yields four data sets with respectively $m = 780$, 390, 195 and 78 observations per day. Each data set is contaminated with a microstructure noise process simulated according to the following model:

$$u_{t,j} = \left(\beta_0 + \frac{\beta_1}{\sqrt{m}\sigma^*_{t,j}}\right) r^*_{t,j} + \varepsilon_{t,j},\ j = 1, ..., m$$

where the exogenous noise $\varepsilon_{t,j}$ is an MA(3).

$$\begin{aligned} \varepsilon_{t,j} &= v_{t,j} + \alpha_1 v_{t,j-1} + \alpha_2 v_{t,j-2} + \alpha_3 v_{t,j-3} \\ & \quad v_{t,j} \overset{IID}{\sim} N(0,\ \alpha_0) \end{aligned}$$

We set the following values for the noise parameter:

$$\begin{aligned} \beta_0 &= 0.5;\ \beta_1 = 0.5; \\ \alpha_1 &= 0.5;\ \alpha_2 = 0.2;\ \alpha_3 = 0.05. \end{aligned}$$

In order to make this simulation design less arbitrary, we will vary α_0 in order to increase or decrease the autocovariances of $\varepsilon_{t,j}$. In fact, we have:

$$\begin{aligned} \omega_0 &\equiv E\left(\varepsilon^2_{t,j}\right) = \alpha_0\left(1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2\right) = 1.2925\alpha_0, \\ \omega_{m,1} &\equiv E\left(\varepsilon_{t,j}\varepsilon_{t,j-1}\right) = \alpha_0\left(\alpha_1 + \alpha_1\alpha_2 + \alpha_2\alpha_3\right) = 0.61\alpha_0, \\ \omega_{m,2} &\equiv E\left(\varepsilon_{t,j}\varepsilon_{t,j-2}\right) = \alpha_0\left(\alpha_2 + \alpha_1\alpha_3\right) = 0.225\alpha_0, \\ \omega_{m,3} &\equiv E\left(\varepsilon_{t,j}\varepsilon_{t,j-2}\right) = \alpha_0\alpha_3 = 0.05\alpha_0, \\ \omega_{m,h} &\equiv E\left(\varepsilon_{t,j}\varepsilon_{t,j-h}\right) = 0 \text{ for all } h \geq 4. \end{aligned}$$

Because the link between ω_0 and α_0 is one-to-one, we will directly vary ω_0 within the range:

$$\omega_0 \in \left\{2.25 \times 10^{-8}; 2.5 \times 10^{-7}; 2.25 \times 10^{-6}; 2.5 \times 10^{-5}\right\}.$$

The value $\omega_0 = 2.5 \times 10^{-7}$ has been used in Zhang and al. (2005) at five minute sampling frequency while $\omega_0 = 2.25 \times 10^{-6}$ has served in Ait-Sahalia and al. (2005) at frequencies that range from one minute to thirty minutes.

We consider three IV estimators: the unbiased estimator $\theta_{1,t}^{(L)}$ (Equation (1.52)), the consistent estimator K_t^{BNHLS} (Equation (1.10) with Bartlett kernel) and the shrinkage estimator $K_t^{\varpi^*}$ (Equation (1.54) with ϖ^* given by (1.56)). After several trials, the bandwidth $H = \lfloor 0.4m^{2/3} \rfloor$ seems to work well for K_t^{BNHLS}.

First, we consider the volatility signature plots, that is, the curve of $\frac{1}{T}\sum_{t=1}^{T} RV_t^{(m_q)}$ plotted against $q = \frac{m}{m_q}$. In Figure 2 of Appendix C, the left hand side graphs describe one simulated sample without noise while the right hand side graphs describe a noisy version of the same data. It is seen that the volatility signature plots (at the top) are quite informative about the presence of the noise. Secondly, we estimate L from the first sample using the plot of $\Delta(l)$ against l as suggested in Equation (1.49). The curve of $\Delta(l)$ at the bottom of Figure 2 is L-shaped with the bend located around $L = 3$. Because a slight overestimation of L still results in an unbiased estimator $\theta_{1,t}^{(L)}$, we will thus set once and for all $\widehat{L} = 4$ in the subsequent computations for robustness check. Also, this is an economic choice that speeds up the simulations.

1.7.2 Simulation Results

For any arbitrary estimator $\widehat{IV}_t$ of IV_t, the empirical MSE of $\widehat{IV}_t$ is given by:

$$MSE(\widehat{IV}_t) = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{IV}_t - IV_t\right)^2, \tag{1.62}$$

Note that this MSE converges to the marginal variance of $\widehat{IV}_t$ for the three unbiased estimators considered here. In Appendix B, table 2 displays the MSE of $\theta_{1,t}^{(L)}$, K_t^{BNHLS} and $K_t^{\varpi^*}$ for the efficient price model with no leverage while Table 3 shows the results when leverage is included. It is interesting to note that the introduction of leverage slightly reduces the variance in all the scenarios. Otherwise, the two tables display qualitatively similar results.

Before analyzing the results, we recall that the efficient return data has been contaminated with a Gaussian MA(3) microstructure noise driven by the same parameters regardless of the sampling frequency. Hence for a given ω_0, the signal-to-noise ratio deteriorates as the sampling frequency increases. Likewise for a fixed sampling frequency, the signal-to-noise ratio deteriorates as ω_0 increases.

It turns out that the shrinkage weight allocated to the consistent estimator ϖ^* heavily depends on the variance of the microstructure noise. In general, ϖ^* is increasing in ω_0. In large ω_0 scenarios, the weight is close to one and tends to decrease very slowly as m increases. By contrast, the weight is smaller in small ω_0 scenarios and increases quite fast as m decreases. Overall, the relative efficiency gain of the shrinkage estimator over the consistent estimator is actually large when m is large and ω_0 is small. Note that compared to the consistent estimator K_t^{BNHLS}, the MSE of $K_t^{\varpi^*}$ is smaller by more than one half in the scenario ($\omega_0 = 2.25 \times 10^{-7}$, $m = 780$) and by about one third for ($\omega_0 = 2.25 \times 10^{-7}$, $m = 390$).

Not surprisingly, the unbiased estimator $\theta_{1,t}^{(L)}$ performs better than the consistent estimator in small ω_0 scenarios ($\omega_0 = 2.25 \times 10^{-7}$). In the large ω_0 scenarios ($\omega_0 > 2.25 \times 10^{-7}$), $\theta_{1,t}^{(L)}$ is worse than K_t^{BNHLS} at all the sampling frequencies considered while the best performance of $\theta_{1,t}^{(L)}$ is achieved at lower frequencies. This is consistent with the fact that the optimal signal-to-noise ratio of $\theta_{1,t}^{(L)}$ is attained at lower frequencies for larger ω_0. For a discussion on optimal sampling frequencies in the IID noise context, see for example Bandi and Russell (2006).

Tables 3 and 4 show the estimation results for the correlogram of the noise in the scenarios ($\omega_0 = 2.25 \times 10^{-7}$; $m = 780$) and ($\omega_0 = 2.25 \times 10^{-7}$; $m = 390$) respectively. In the two tables, the column labeled "True" contains the true values of the parameters. The estimates are computed using the Equation (1.47) while the standard deviations are computed from Equation (1.48) with ten lags. Firstly, we note that the estimator of ω_0 is biased upward and the bias decreases as the record frequency increases. In fact, the bias of $\widehat{\omega}_0$ is due to the presence of endogenous noise. Under the assumption that $\sigma_{t,qk}^*$ is stationary, the unconditional bias of $\widehat{\omega}_0$ is given by:

$$E\left[\widehat{\omega}_0\right] - \omega_0 = \frac{\beta_1^2}{m} + \frac{\beta_1\left(2\beta_0 + 1\right)}{\sqrt{m}} E\left[\sigma_{t,qk}^*\right] + \beta_0\left(\beta_0 + 1\right) E\left[\sigma_{t,qk}^{*2}\right]$$

Hence while $\widehat{\omega}_0$ is biased for the variance of the exogenous noise, it does reflect the actual size of the overall noise contaminating the asset prices.

The results suggest that the higher order autocovariances estimators $\{\widehat{\omega}_l\}_{l=1}^4$ are unbiased. The Student-t statistics displayed in the last column indicate that the null hypothesis $\omega_4 = 0$

cannot be rejected at level 5%.

We present an empirical application in the next section.

1.8 Empirical Application

We have shown in the simulation study that the shrinkage estimator $K_t^{\varpi^*}$ performs well relatively to the benchmarks K_t^{BNHLS} and $\theta_{1,t}^{(L)}$. We have also seen that the endogenous parameters of the noise model are not identified although this raises no problem for estimating the IV. Our focus in this empirical investigation will essentially be to test the assumptions E1-E4 made on the exogenous noise. We describe the data in the first subsection below while the empirical results are presented in the second subsection.

1.8.1 Data and Preprocessing

For this application, we used the data on twelve stocks listed in the Dow Jones Industrial[10]. The prices are observed every one minute from January 1^{st}, 2002 to December 31^{th}, 2007 (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[11].

While Equation (1.1) assumes no jumps in the efficient price process, the conclusions of many studies strongly suggest the presence of a jump component in real world prices (see e.g Eraker (2004)). Thus following the same intuitions as in Barndorff-Nielsen and al (2008b) for quote data[12], we applied the following cleaning rule to the initial data which we denoted $r_{t,j}^{OLD}$:

$$r_{t,j}^{NEW} = \begin{cases} r_{t,j}^{OLD} \text{ if } \left|r_{t,j}^{OLD}\right| \leq Q \times \underline{r}^{OLD} \\ sign\left(r_{t,j}^{OLD}\right) \times Q \times \underline{r}^{OLD} \text{ otherwise} \end{cases}$$

where $\underline{r}^{OLD}$ is the empirical median of $\left|r_{t,j}^{OLD}\right|$ across t and j and. We use $Q = 50$, and the resulting $r_{t,j}^{NEW}$ is treated as our initial observed return $r_{t,j} = r_{t,j}^{NEW}$. Our cleaning rule treat

[10]The data we use in this paper have been purchased from a private provider who has ensured its accuracy by comparision with three other independent financial data providers. Please see Section 9 for the preprocessing details.

[11]This amounts to replace a missing data point by the most recent observation available.

[12]For quote data, BNHLS (2008b) suggest to delete entries for which the spread is more that 50 times the median spread on that day.

the jumps and outliers due to recording errors as the same. We advocate this approach for three reasons. Firstly, we want to preserve the structure of dependence of the noise which is of interest in our analysis. Secondly, the process $\left|r_{t,j}^{OLD}\right|$ obviously contains substantial information about the range of the data. And finally, the median is robust to the extreme values that arises in the series $r_{t,j}^{OLD}$ due to the presence of outliers. Figure 3 in Appendix C show examples of the impact of this preprocessing on the data.

1.8.2 Empirical Results

We follow four basic steps in conducting this empirical study. The first step consists in assessing the memory parameter L by mean of the criterion $\Delta(l)$ given in (1.49). In the second step, the estimator $\widehat{L}$ of L is used to compute the estimators of $\{\omega_{m,l}\}_{l=1}^{L}$, α and δ given by Equations (1.43), (1.50) and (1.51). In the third step, we estimate the variance of $\{\widehat{\omega}_{m,l}\}_{l=1}^{L}$ using Equation (1.48) and compute the Student-t statistics. Finally in the last step, we compute the shrinkage estimator $K_t^{\varpi^*}$ for the IV. The empirical results are shown in Appendix C.

For all the stocks considered, we found that the noise is L-dependent with the value of L between 5 minutes (American Express) and 14 minutes (AIG and General Electric). The top graphs of Figure 4 in Appendix C show the curves of $\Delta(l)$ for 3M Co, Alcoa and AIG. The finding that the noise is autocorrelated is not new in the literature (see for example Hansen and Lunde, 2006). What is new here is that we will use the estimates of L and $\{\omega_{m,l}\}_{l=1}^{L}$ to assess Assumptions E3 and E4 in the second step.

Table 5 of Appendix C shows the estimates $\widehat{\alpha}$ and $\widehat{\delta}$. It is seen that $\widehat{\delta} < \widehat{\alpha} < 2/3$ for all the assets. The fact that $\widehat{\delta} < \widehat{\alpha}$ is not surprising because this equality should hold by construction. However, $\widehat{\alpha} < 2/3$ is an interesting result because it indicates that the estimator of Barndorff-Nielsen and al. (2008a) delivers its best performance at the highest available frequency.

The graphs in the second row Figure 4 in Appendix C show the estimated correlogram for 3M Co, Alcoa and AIG along with the estimated Student-t statistics. The t-stats are essentially consistent with the choice of L based on the information criterion $\Delta(l)$. But the Student-t test could have not been used to select L at the beginning because the prior knowledge of L is required in order to implement the whole procedure.

To compute the realized kernels in the last step, we set $H = 30$ for the bandwidth except for the American Express index (AXP) which has necessitate $H = 10$. These bandwidth values appear to produce better results than $\left\lfloor (390)^{2/3} \right\rfloor = 53$. We have also estimated the bias of the RV by the alternative formula $\widetilde{b}_t = RV^{(m)} - K_t^{\varpi^*}$ which tends to have less variance than the natural method of moment estimator $\widehat{b}_t^{(m)}$ given in (1.44). The time series plots of $K_t^{\varpi^*}$ and $\widetilde{b}_t$ are displayed respectively in the third and fourth row of Figure 4. Our results suggest that the sign of $\widetilde{b}_t$ is not constant through time. It turns out that when the correlogram is positive as we found for 3M Co, Alcoa and AIG, a negative bias can only be due to a negative correlation between the noise and the latent return. In light of this, these empirical results suggest that either β_0 or β_1 is negative.

1.9 Conclusion

This paper proposes a flexible semi-parametric model for the market microstructure noise. We specify the microstructure noise as the sum of an information correlated process and an information uncorrelated process. The information uncorrelated noise is modeled as an L-dependent process, where L is allowed to increase with the frequency at which the prices are recorded. In light of this model, we study the properties of common realized measures that aim to estimate the integrated volatility.

We propose a new shrinkage realized kernels which is an optimal linear combination of the consistent realized kernels of Barndorff-Nielsen and al (2008a) and an unbiased estimator constructed for this purpose. It is shown theoretically that the shrinkage estimator has lower variance than the consistent estimator in small samples while both estimators are asymptotically equivalent in large samples. The Monte Carlo simulations show that the relative efficiency gain of the shrinkage realized kernels over the standard realized kernel is substantial in situations where the variance of the microstructure noise is small. When the variance of the noise is large, the inconsistent estimator is markedly dominated.

Finally, we propose a framework to assess the true values of the noise parameters via the observed returns. Unfortunately, the endogenous parameters are not identified. Our empirical findings about the noise confirm the conclusions of Hansen and Lunde (2006): there is strong evidence that the noise is autocorrelated and correlated with the latent returns. Our contribution here is to show how to estimate the rate at which L increases with the sampling frequency. We found that in general, L increases slower than $\sqrt{m}$.

Bibliography

[1] Ait-Sahalia, Y., Mykland, P.A., Zhang, L. (2005) "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise". *The Review of Financial Studies* 18, No 2.

[2] Ait-Sahalia, Y., Mykland, P.A., Zhang, L. (2006) "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise", *NBER Working Paper* No. W11380.

[3] Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2002) "Parametric and nonparametric measurement of volatility", In Y. Aït-Sahalia and LP Hansen (Eds) *Handbook of Financial Econometrics*, Elsevier, North-Holland.

[4] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001) "The distribution of realized stock return volatility", *Journal of Financial Economics* 61, 43-76.

[5] Andersen, T. G., Bollerslev, T. and Diebold, F. X. and Labys, P. (2000) "Great realizations", *Risk Magazine* 13, 105-108.

[6] Andersen, T. G., Bollerslev, T. and Diebold, F. X. and Labys, P. (2003) "Modelling and forecasting Realized Volatility", *Econometrica* 71, 579-625.

[7] Andreou, E. and Ghysels, E. (2002) "Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation and Empirical Results" *Journal of Business and Economic Statistics*, Forthcoming.

[8] Andrew, D.W.K. and Monahan, J.C. (1992) "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator", *Econometrica* 60, 953-966.

[9] Awartani, B., Corradi, V. and Distaso, W. (2009) "Assessing Market Microstructure Effects via Realized Volatility Measures, with an Application to the Dow Jones Industrial Average Stocks", *Journal of Business and Economic Statistics* 27, No. 2.

[10] Bandi, F. M. and Russell J. R. (2006) "Separating microstructure noise from volatility" *Journal of Financial Economics* 79 (2006) 655–692.

[11] Bandi, F. M. and Russell J. R. (2008) "Microstructure noise, realized variance and optimal sampling", *Review of Economic Studies* 75, 339-369.

[12] Barndorff-Nielsen, O. E., Graversen, S.E., Jacod, J. and Shephard, N. (2006) "Limit theorems for realized power variation in econometrics", *Econometric Theory* 22, 677-719.

[13] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008a) "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise", *Econometrica* 76, 1481-1536.

[14] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008b) "Realised Kernels in Practice: Trades and Quotes", *Econometrics Journal*, forthcoming.

[15] Barndorff-Nielsen, O. E. and Shephard, N. (2002) "Estimating quadratic variation using realized variance". *Journal of Applied Econometrics* 17, 457-477.

[16] Beveridge, S. and Nelson, C. (1981) "A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the business cycle", *Journal of Monetary Economics* 7, 151-174.

[17] Bollen, B. and Inder, B. (2002) "Estimating daily volatility in financial markets utilizing intraday data", *Journal of Empirical Finance* 9, 551-562.

[18] Corradi, V., Distaso, W., and Swanson, N. R. (2008) "Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures", *Journal of Econometrics,* forthcoming.

[19] Eraker, B. (2004) "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices ", *Journal of Finance*, 59, 1367-1403.

[20] French, K. R., Schwert G. W. and Stambaugh, R. F. (1987) "Expected stock returns and volatility", *Journal of Financial Economics 19, 3-29. North-Holland.*

[21] Gatheral, J. and Oomen, R. C. A. (2007) "Zero-intelligence realized variance estimation", *Finance Stochastics*, Forthcoming.

[22] Garcia, R. and Meddahi, N. (2006) "Comment on Hansen and Lunde (2006)", *Journal of Business & Economic Statistics* 24, No 2.

[23] Ghysels, E., Mykland, P and Renault E. (2007) " In-sample Asymptotics and Across-sample Efficiency Gains for High Frequency Data Statistics". *Working paper*, University of North Carolina.

[24] Gonçalves, S. and Meddahi, N. (2009) "Boostrapping realized volatility", *Econometrica,* 77, 283-306.

[25] Hansen, B. E. (2007) "Least Squares Model Averaging" *Econometrica* 75, 1175-1189.

[26] Hansen, P.R and Lunde, A. (2006) "Realized variance and market microstructure noise" *Journal of Business & Economic Statistics*, 24, No. 2.

[27] Hasbrouck, J. (1993) "Assessing the Quality of a Security Market: A New Approach to Transaction Cost Measurement", *The Review of Financial Studies* 6, No. 1, 191-212.

[28] Hasbrouck, J. (1996) "Modelling market microstructure time series", In G.S. Maddala and C.R. Rao (Eds) *Handbook of Statistics*. Elsevier, North-Holland.

[29] Heston, S. (1993) "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options", *The Review of Financial Studies*, 1993, Vol. 6, number 2, 327-343.

[30] Hill, B.M. (1975) "A simple general approach to inference about the tail of a distribud-istribution" *Annals of Statistics*, 3, 1163-1174.

[31] Jacod, J. (1994) "Limit of Random Measures Associated with the Increments of a Brownian Semi-martingale" No 120, *Laboratoire de Probabilities, Université Pierre et Marie Curie, Paris.*

[32] Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. and Vetter, M. (2009) "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach" *Stochastic Processes and their Applications,* 119 2249–2276.

[33] Jacod, J. and Protter, P. (1998) "Asymptotic error distribution for the Euler method for stochastic Differential Equations" *Annals of Probability* 26, 267-307.

[34] Kalnina, I. and Linton, O. (2008) "Estimating quadratic variation consistently in the presence of correlated measurement error", *Journal of Econometrics,* 147, 47-59.

[35] Meddahi, N. (2002) "A Theoretical Comparison Between Integrated and Realized Volatility", *Journal of Applied Econometrics* 17, 475-508.

[36] Newey, W. and West, K. (1987) "A simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica* 55, 703-708.

[37] Podolskij, M. and Vetter, M. (2006) "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps", *Technical Report*, *Ruhr-Universitat Bochum*, 2006.

[38] Roll, R. (1984) "A simple measure of the effective bid-ask spread in an efficient market", *Journal of Finance 39,* 1127-1139.

[39] Stoll, H. R. (1989) "Infering the components of the bid-ask spread: theory and empirical evidence", *Journal of Finance* 44, 115-134.

[40] Stoll, H. R. (2000) "Frictions", *Journal of Finance* 55, 1479-1514.

[41] Ubukata, O. and Oya, K. (2009) "Estimation and Testing for Dependence in Market Microstructure Noise", *Journal of Financial Econometrics*, 2008, 1-46. DOI: 10.1093/jjfinec/nbn021.

[42] Watson, M. (1986) "Univariate detrending methods with stochastic trend", *Journal of Monetary Economics* 18, 49-75.

[43] Zhang, L, Mykland, P.A., and Ait-Sahalia, Y. (2005) "A Tale of Two Time Scale: Determining Integrated Volatility with Noisy High-Frequency Data", *Journal of The American Statistical Association* 100, No 472.

[44] Zhou, B. (1996) "High frequency data and volatility in foreign-exchange rates", *Journal of Business and Economics Statistics* 14, 45-52.

Appendix A: Proofs

The following Lemma will be used in the proof of Theorem 5.

Lemma 8 *Assume that* $r_{t,j} = (1 + a_{t,j})\, r^*_{t,j} - a_{t,j-1} r^*_{t,j-1} + (\varepsilon_{t,j} - \varepsilon_{t,j-1})$ *for some deterministic sequence* $\{a_{t,j}\}$, $j = 1, ..., m$. *Let* $\widetilde{r}_{t,k}$ *be the serie of non-overlapping sums of* q *consecutive observations of* $r_{t,j}$*:*

$$\widetilde{r}_{t,k} = (1 + a_{t,qk})\, r^*_{t,qk} + \sum_{j=qk-q+1}^{qk-1} r^*_{t,j} - a_{t,qk-q} r^*_{t,qk-q} + (\varepsilon_{t,qk} - \varepsilon_{t,qk-q})$$

for $k = 1, ..., m_q$ *and some positive integer* $q \geq 1$ *such that* $m_q = m/q$, *with the convention that* $\sum_{j=qk-q+1}^{qk-1} r^*_{t,j} = 0$ *if* $q = 1$. *Then we have:*

$$\begin{aligned} E\left[RV^{(m_q)}\right] &= IV_t + 2\textstyle\sum_{k=1}^{m_q}\left(a_{t,qk} + a^2_{t,qk}\right)\sigma^{*2}_{t,qk} + a^2_{t,0}\sigma^{*2}_{t,0} - a^2_{t,qm_q}\sigma^{*2}_{t,qm_q} + 2m_q\left(\omega_0 - \omega_{m,q}\right), \\ Var\left[RV^{(m)}\right] &= 2\textstyle\sum_{k=1}^{m_q}\left[(1 + a_{t,qk})^2 + a^2_{t,qk}\right]^2\sigma^{*4}_{t,qk} + 2\sum_{k=1}^{m_q}\left(\sum_{l=qk-q+1}^{qk-1}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,l}\right) \\ &\quad + Var\left[\textstyle\sum_{k=1}^{m_q}(\varepsilon_{t,qk} - \varepsilon_{t,qk-q})^2\right] + 4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}(1 + a_{t,qk})^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk} \\ &\quad + 4\textstyle\sum_{k=1}^{m_q}(1 + a_{t,qk})^2 a^2_{t,qk-q}\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk} + 4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1} a^2_{t,qk-q}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q} \\ &\quad + 8\left(\omega_0 - \omega_{m,q}\right)\textstyle\sum_{k=1}^{m_q}(1 + a_{t,qk})^2\sigma^{*2}_{t,qk} + 8\left(\omega_0 - \omega_{m,q}\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j} \\ &\quad + 8\left(\omega_0 - \omega_{m,q}\right)\textstyle\sum_{k=1}^{m_q} a^2_{t,qk-q}\sigma^{*2}_{t,qk-q} + 2a^4_{t,0}\sigma^{*4}_{t,0} - 2a^4_{t,qm_q}\sigma^{*4}_{t,qm_q} \\ &\quad - 4a^2_{t,qm_q}\left(1 + a_{t,qm_q}\right)^2\sigma^{*4}_{t,qm_q}. \end{aligned}$$

Proof of Lemma 8:

$$RV^{(m_q)} = \sum_{k=1}^{m_q}\widetilde{r}^2_{t,k} = (1) + (2) + (3) + (4) + (5) + (6) + (7) + (8) + (9)$$

where

$(1) = \sum_{k=1}^{m_q}\left[(1 + a_{t,qk})^2 + a^2_{t,qk}\right] r^{*2}_{t,qk} + a^2_{t,0} r^{*2}_{t,0} - a^2_{t,qm_q} r^{*2}_{t,qm_q}$.

$(2) = \sum_{k=1}^{m_q}\left(\sum_{j=qk-q+1}^{qk-1} r^*_{t,j}\right)^2$.

$(3) = \sum_{k=1}^{m_q}(\varepsilon_{t,qk} - \varepsilon_{t,qk-q})^2$.

$(4) = 2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}(1 + a_{t,qk})\, r^*_{t,j} r^*_{t,qk}$.

$(5) = 2\sum_{k=1}^{m_q}(1 + a_{t,qk})\, a_{t,qk-q} r^*_{t,qk-q} r^*_{t,qk}$.

$(6) = 2\sum_{k=1}^{m_q}(1 + a_{t,qk})\,(\varepsilon_{t,qk} - \varepsilon_{t,qk-q})\, r^*_{t,qk}$.

$(7) = -2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1} a_{t,qk-q} r^*_{t,j} r^*_{t,qk-q}$.

$(8) = 2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(\varepsilon_{t,qk}-\varepsilon_{t,qk-q}\right)r^*_{t,j}.$

$(9) = -2\sum_{k=1}^{m_q}a_{t,qk-q}\left(\varepsilon_{t,qk}-\varepsilon_{t,qk-q}\right)r^*_{t,qk-q}.$

Only squared terms have nonzero expectation:

$$\begin{aligned} E\left[RV^{(m_q)}\right] &= m_q E\left[\left(\varepsilon_{t,qk}-\varepsilon_{t,qk-q}\right)^2\right]+\sum_{k=1}^{m_q}\left[\left(1+a_{t,qk}\right)^2+a^2_{t,qk}\right]\sigma^{*2}_{t,qk} \\ &\quad +\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}+a^2_{t,0}\sigma^{*2}_{t,0}-a^2_{t,qm_q}\sigma^{*2}_{t,qm_q} \\ &= 2m_q\left(\omega_0-\omega_{m,q}\right)+IV_t+2\sum_{k=1}^{m_q}\left(a_{t,qk}+a^2_{t,qk}\right)\sigma^{*2}_{t,qk}+a^2_{t,0}\sigma^{*2}_{t,0}-a^2_{t,qm_q}\sigma^{*2}_{t,qm_q}. \end{aligned}$$

where $\omega_{m,q} = E\left[\varepsilon_{t,j}\varepsilon_{t,j-q}\right]$ is independent of t and j. Also, all the terms involved in the expression of $RV^{(m_q)}$ are uncorrelated and thus:

$$\begin{aligned} Var\left[RV^{(m_q)}\right] &= Var((1))+Var((2))+Var((3))+Var((4)) \\ &\quad +Var((5))+Var((6))+Var((7))+Var((8))+Var((9)), \end{aligned}$$

where

$Var((1)) = 2\sum_{k=1}^{m_q}\left[\left(1+a_{t,qk}\right)^2+a^2_{t,qk}\right]^2\sigma^{*4}_{t,qk}+2a^4_{t,0}\sigma^{*4}_{t,0}$
$\qquad -2a^4_{t,qm_q}\sigma^{*4}_{t,qm_q}-4a^2_{t,qm_q}\left(1+a_{t,qm_q}\right)^2\sigma^{*4}_{t,qm_q}.$

$Var((2)) = 2\sum_{k=1}^{m_q}\left(\sum_{l=qk-q+1}^{qk-1}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,l}\right).$

$Var((4)) = 4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(1+a_{t,qk}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk}.$

$Var((5)) = 4\sum_{k=1}^{m_q}\left(1+a_{t,qk}\right)^2a^2_{t,qk-q}\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk}.$

$Var((6)) = 4\sum_{k=1}^{m_q}\left(1+a_{t,qk}\right)^2 Var\left(\varepsilon_{t,qk}-\varepsilon_{t,qk-q}\right)Var\left(r^*_{t,qk}\right)$
$\qquad = 8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m}\left(1+a_{t,qk}\right)^2\sigma^{*2}_{t,qk}.$

$Var((7)) = 4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}a^2_{t,qk-q}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q}.$

$Var((8)) = 8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}.$

$Var((9)) = 8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q}a^2_{t,qk-q}\sigma^{*2}_{t,qk-q}.$

Hence:

$Var\left[RV^{(m_q)}\right] = 2\sum_{k=1}^{m_q}\left[\left(1+a_{t,qk}\right)^2+a^2_{t,qk}\right]^2\sigma^{*4}_{t,qk}$
$\qquad +2\sum_{k=1}^{m_q}\left(\sum_{l=qk-q+1}^{qk-1}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,l}\right)$
$\qquad +Var\left[\sum_{k=1}^{m_q}\left(\varepsilon_{t,qk}-\varepsilon_{t,qk-q}\right)^2\right]+4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(1+a_{t,qk}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk}$

$$
\begin{aligned}
&+4\sum_{k=1}^{m_q}\left(1+a_{t,qk}\right)^2 a_{t,qk-q}^2\sigma_{t,qk-q}^{*2}\sigma_{t,qk}^{*2}+4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1} a_{t,qk-q}^2\sigma_{t,j}^{*2}\sigma_{t,qk}^{*2}\\
&+8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q}\left(1+a_{t,qk}\right)^2\sigma_{t,qk}^{*2}+8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\\
&+8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q} a_{t,qk-q}^2\sigma_{t,qk-q}^{*2}+2a_{t,0}^4\sigma_{t,0}^{*4}-2a_{t,qm_q}^4\sigma_{t,qm_q}^{*4}\\
&-4a_{t,qm_q}^2\left(1+a_{t,qm_q}\right)^2\sigma_{t,qm_q}^{*4}.\blacksquare
\end{aligned}
$$

The following Lemma will be used in the proof of Theorem 6.

Lemma 9 *Under the assumptions of Theorem 6, we have:*

$$\begin{aligned}
E\left[RV_t^{(AC,m,1)}\right] &= IV_t + \left(2a_{t,m} + a_{t,m}^2\right)\sigma_{t,m}^{*2} - \left(2a_{t,0} + a_{t,0}^2\right)\sigma_{t,0}^{*2} \\
Var\left[RV_t^{(AC,m,1)}\right] &= 2\sum_{j=1}^m \sigma_{t,j}^{*4} + 4\sum_{j=1}^m \left(1 + a_{t,j} + a_{t,j}a_{t,j-1}\right)^2 \sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2} \\
&\quad +4\sum_{j=1}^m \left(1 + a_{t,j}\right)^2 a_{t,j-2}^2 \sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2} + 8\omega_0 \sum_{j=1}^m \left(1 + a_{t,j}\right)^2 \sigma_{t,j}^{*2} \\
&\quad +8\omega_0 \sum_{j=1}^m a_{t,j}^2 \sigma_{t,j}^{*2} + 8m\omega_0^2 + 2\left(E\left[\varepsilon_{t,j}^4\right] - \omega_0^2\right) + 2\left(2a_{t,0} + a_{t,0}^2\right)^2 \sigma_{t,0}^{*4} \\
&\quad +2\left(2a_{t,m} + a_{t,m}^2\right)^2 \sigma_{t,m}^{*4} + 2\left(2a_{t,m} + a_{t,m}^2\right)\sigma_{t,m}^{*4} + 4a_{t,-1}^2 a_{t,0}^2 \sigma_{t,-1}^{*2}\sigma_{t,0}^{*2} \\
&\quad -8a_{t,m-1}a_{t,m}\left(1 + a_{t,m} + a_{t,m}a_{t,m-1}\right)\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2} \\
&\quad +4a_{t,m-1}^2 a_{t,m}^2 \sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2} + 8\omega_0\left(\sigma_{t,m-1}^{*2} - \sigma_{t,0}^{*2}\right) \\
&\quad +8\omega_0\left(a_{t,-1}^2\sigma_{t,-1}^{*2} + 2a_{t,0}^2\sigma_{t,0}^{*2} + a_{t,m}\sigma_{t,m}^{*2}\right) \\
&\quad -8\omega_0\left(a_{t,m-1}\sigma_{t,m-1}^{*2} + a_{t,m-1}^2\sigma_{t,m-1}^{*2}\right).
\end{aligned}$$

Proof of Lemma 9: We first note that:

$$\begin{aligned}
RV_t^{(AC,m,1)} &= \sum_{j=1}^m r_{t,j}^2 + 2\sum_{j=1}^m r_{t,j}r_{t,j-1} \\
&= (I) + (II) + (III) + (IV) + (V) + (VI) + (VII) + (VIII) + (IX),
\end{aligned}$$

where

$(I) = \sum_{j=1}^m r_{t,j}^{*2} + \left(2a_{t,m} + a_{t,m}^2\right) r_{t,m}^{*2} - \left(2a_{t,0} + a_{t,0}^2\right) r_{t,0}^{*2}.$

$(II) = 2\sum_{j=1}^m \left(1 + a_{t,j} + a_{t,j}a_{t,j-1}\right) r_{t,j}^* r_{t,j-1}^* + 2a_{t,-1}a_{t,0}r_{t,-1}^* r_{t,0}^* - 2a_{t,m-1}a_{t,m}r_{t,m-1}^* r_{t,m}^*.$

$(III) = -2\sum_{j=1}^m \left(1 + a_{t,j}\right) a_{t,j-2} r_{t,j}^* r_{t,j-2}^*.$

$(IV) = 2\sum_{j=1}^m \left(\varepsilon_{t,j} - \varepsilon_{t,j-1}\right) r_{t,j}^* - 2a_{t,0}\left(\varepsilon_{t,0} - \varepsilon_{t,-1}\right) r_{t,0}^* + 2a_{t,m}\left(\varepsilon_{t,m} - \varepsilon_{t,m-1}\right) r_{t,m}^*.$

$(V) = 2\sum_{j=1}^m \left(1 + a_{t,j}\right)\left(\varepsilon_{t,j-1} - \varepsilon_{t,j-2}\right) r_{t,j}^*.$

$(VI) = 2\sum_{j=1}^m \left(\varepsilon_{t,j} - \varepsilon_{t,j-1}\right) r_{t,j-1}^*.$

$(VII) = -2\sum_{j=1}^m a_{t,j-2}\left(\varepsilon_{t,j} - \varepsilon_{t,j-1}\right) r_{t,j-2}^*.$

$(VIII) = 2\sum_{j=1}^m \left(\varepsilon_{t,j} - \varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-1} - \varepsilon_{t,j-2}\right).$

$(IX) = \sum_{j=1}^m \left(\varepsilon_{t,j} - \varepsilon_{t,j-1}\right)^2.$

Because only squared terms will have nonzero expectation, we have:

$$E\left[RV_t^{(AC,m,1)}\right] = IV_t + \left(2a_{t,m} + a_{t,m}^2\right)\sigma_{t,m}^{*2} - \left(2a_{t,0} + a_{t,0}^2\right)\sigma_{t,0}^{*2}.$$

The calculation of that variance is simplified by noting that only the terms (IV) to (IX) are possibly correlated. Thus we have:

$$
\begin{aligned}
&Var((I)) = 2\sum_{j=1}^{m}\sigma_{t,j}^{*4} + 2\left(2a_{t,0}+a_{t,0}^{2}\right)^{2}\sigma_{t,0}^{*4} + 2\left(2a_{t,m}+a_{t,m}^{2}\right)^{2}\sigma_{t,m}^{*4}\\
&\qquad\qquad +2\left(2a_{t,m}+a_{t,m}^{2}\right)\sigma_{t,m}^{*4}\\
&Var((II)) = 4\sum_{j=1}^{m}\left(1+a_{t,j}+a_{t,j}a_{t,j-1}\right)^{2}\sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2} + 4a_{t,-1}^{2}a_{t,0}^{2}\sigma_{t,-1}^{*2}\sigma_{t,0}^{*2}\\
&\qquad\qquad +4a_{t,m-1}^{2}a_{t,m}^{2}\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}\\
&\qquad\qquad -8a_{t,m-1}a_{t,m}\left(1+a_{t,m}+a_{t,m}a_{t,m-1}\right)\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}\\
&Var((III)) = 4\sum_{j=1}^{m}\left(1+a_{t,j}\right)^{2}a_{t,j-2}^{2}\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}\\
&Var((IV)) = 8\omega_0 IV_t + 8\omega_0\left(a_{t,0}^{2}\sigma_{t,0}^{*2}+a_{t,m}^{2}\sigma_{t,m}^{*2}\right) + 16\omega_0 a_{t,m}\sigma_{t,m}^{*2}\\
&2Cov\left((IV),(V)\right) = 8\sum_{j=1}^{m}\left(1+a_{t,j}\right)E\left[\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right)\right]E\left(r_{t,j}^{*2}\right)\\
&\qquad\qquad = -8\omega_0 IV_t - 8\omega_0\sum_{j=1}^{m}a_{t,j}\sigma_{t,j}^{*2}\\
&2Cov\left((IV),(VI)\right) = 8\sum_{j=1}^{m-1}E\left[\left(\varepsilon_{t,j+1}-\varepsilon_{t,j}\right)\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\right]E\left(r_{t,j}^{*2}\right)\\
&\qquad\qquad\qquad = -8\omega_0 IV_t + 8\omega_0\sigma_{t,m-1}^{*2}\\
&2Cov\left((IV),(VII)\right) = -8\sum_{j=1}^{m-2}a_{t,j}E\left[\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j+2}-\varepsilon_{t,j+1}\right)\right]E\left(r_{t,j}^{*2}\right) = 0\\
&2Cov\left((IV),(VIII)\right) = 2Cov\left((IV),(IX)\right) = 0\\
&Var((V)) = 8\omega_0\sum_{j=1}^{m}\left(1+a_{t,j}\right)^{2}\sigma_{t,j}^{*2}\\
&2Cov\left((V),(VI)\right) = 2Cov\left((V),(VII)\right) =\\
&\qquad\qquad\qquad 2Cov((V),(VIII)) = 2Cov((V),(IX)) = 0\\
&Var((VI)) = 8\omega_0 IV_t - 8\omega_0\sigma_{t,0}^{*2}\\
&2Cov\left((VI),(VII)\right) = -8\sum_{j=1}^{m-2}a_{t,j}E[\left(\varepsilon_{t,j+1}-\varepsilon_{t,j}\right)\left(\varepsilon_{t,j+2}-\varepsilon_{t,j+1}\right)]E\left(r_{t,j}^{*2}\right)\\
&\qquad\qquad\qquad = 8\omega_0\sum_{j=1}^{m}a_{t,j}\sigma_{t,j}^{*2} - 8\omega_0\left(a_{t,m-1}\sigma_{t,m-1}^{*2}+a_{t,m}\sigma_{t,m}^{*2}\right)\\
&2Cov((VI),(VIII)) = 2Cov((VI),(IX)) = 0\\
&Var((VII)) = 8\omega_0\sum_{j=1}^{m}a_{t,j}^{2}\sigma_{t,j}^{*2} + 8\omega_0\left(a_{t,-1}^{2}\sigma_{t,-1}^{*2}+a_{t,0}^{2}\sigma_{t,0}^{*2}-a_{t,m-1}^{2}\sigma_{t,m-1}^{*2}-a_{t,m}^{2}\sigma_{t,m}^{*2}\right)\\
&Cov((VII),(VIII)) = Cov((VII),(IX)) = 0\\
&Var((VIII)) = Var\left[2\sum_{j=1}^{m}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right)\right]\\
&\qquad = 4Var\left[2\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-1}+\varepsilon_{t,0}\varepsilon_{t,-1}+\varepsilon_{t,m}\varepsilon_{t,m-1}+\sum_{j=1}^{m}+\varepsilon_{t,j}\varepsilon_{t,j-2}+\sum_{j=1}^{m}\varepsilon_{t,j-1}^{2}\right]\\
&\qquad = 4mE[\varepsilon_{t,j}^{4}] + 16m\omega_0^{2} - 8\omega_0^{2}\\
&2Cov\left((VIII),(IX)\right) = 4Cov\left[\sum_{j=1}^{m}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right),\ \sum_{k=1}^{m}\left(\varepsilon_{t,k}-\varepsilon_{t,k-1}\right)^{2}\right]\\
&\qquad\qquad\qquad\qquad = -\left(8m-4\right)\left(E[\varepsilon_{t,j}^{4}]+\omega_0^{2}\right)
\end{aligned}
$$

since we have:

$$E[\left(\varepsilon_{t,j+k}-\varepsilon_{t,j+k-1}\right)^{2}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right)] = -2\omega_0^{2}\ \forall\ k\geq 1$$

$E[(\varepsilon_{t,j}-\varepsilon_{t,j-1})^3(\varepsilon_{t,j-1}-\varepsilon_{t,j-2})] = -E[\varepsilon_{t,j}^4]-3\omega_0^2$ (for $k=j$)

$E[(\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-1}-\varepsilon_{t,j-2})^3] = -E[\varepsilon_{t,j}^4]-3\omega_0^2$ (for $k=j-1$)

$E[(\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-1}-\varepsilon_{t,j-2})(\varepsilon_{t,j-k-1}-\varepsilon_{t,j-k-2})^2] = -2\omega_0^2\ \forall\ k\geq 1$

$$\Rightarrow E\left[\left(\sum_{j=1}^m(\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-1}-\varepsilon_{t,j-2})\right)\left(\sum_{j=1}^m(\varepsilon_{t,j}-\varepsilon_{t,j-1})^2\right)\right]$$
$$= (-2m+1)E[\varepsilon_{t,j}^4]+(-2m^2-2m+1)\omega_0^2$$

Also: $E\left(\sum_{j=1}^m(\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-1}-\varepsilon_{t,j-2})\right) = -m\omega_0$

and $E\left(\sum_{j=1}^m(\varepsilon_{t,j}-\varepsilon_{t,j-1})^2\right) = 2m\omega_0$

Thus $Cov\left((VIII),(IX)\right) = (-2m+1)E[\varepsilon_{t,j}^4]+(-2m^2-2m+1)\omega_0^2+2m^2\omega_0^2$
$$= -(2m-1)\left(E[\varepsilon_{t,j}^4]+\omega_0^2\right)$$

$Var((IX)) = 4mE\left[\varepsilon_{t,j}^4\right]+2\left(\omega_0^2-E\left[\varepsilon_{t,j}^4\right]\right)$

The sum of all these terms gives:

$$\begin{aligned}
Var\left[RV_t^{(AC,m,1)}\right] &= 2\sum_{j=1}^m\sigma_{t,j}^{*4}+4\sum_{j=1}^m(1+a_{t,j}+a_{t,j}a_{t,j-1})^2\sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2}\\
&+4\sum_{j=1}^m(1+a_{t,j})^2a_{t,j-2}^2\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}+8\omega_0\sum_{j=1}^m(1+a_{t,j})^2\sigma_{t,j}^{*2}\\
&+8\omega_0\sum_{j=1}^m a_{t,j}^2\sigma_{t,j}^{*2}+8m\omega_0^2+2\left(E\left[\varepsilon_{t,j}^4\right]-\omega_0^2\right)+2\left(2a_{t,0}+a_{t,0}^2\right)^2\sigma_{t,0}^{*4}\\
&+2\left(2a_{t,m}+a_{t,m}^2\right)^2\sigma_{t,m}^{*4}+2\left(2a_{t,m}+a_{t,m}^2\right)\sigma_{t,m}^{*4}+4a_{t,-1}^2a_{t,0}^2\sigma_{t,-1}^{*2}\sigma_{t,0}^{*2}\\
&-8a_{t,m-1}a_{t,m}(1+a_{t,m}+a_{t,m}a_{t,m-1})\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}\\
&+4a_{t,m-1}^2a_{t,m}^2\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}+8\omega_0\left(\sigma_{t,m-1}^{*2}-\sigma_{t,0}^{*2}\right)\\
&+8\omega_0\left(a_{t,-1}^2\sigma_{t,-1}^{*2}+2a_{t,0}^2\sigma_{t,0}^{*2}+a_{t,m}\sigma_{t,m}^{*2}\right)\\
&-8\omega_0\left(a_{t,m-1}\sigma_{t,m-1}^{*2}+a_{t,m-1}^2\sigma_{t,m-1}^{*2}\right).\blacksquare
\end{aligned}$$

Proof of Theorem 5: Substituting for $a_{t,j}=\beta_0+\frac{\beta_1}{\sqrt{m}\sigma_{t,j}^*}$ in Lemma 8, we get for the expectation:

$$\begin{aligned}
E\left[RV_t^{(m_q)}\right] &= IV_t+2m_q(\omega_0-\omega_{m,q})+2\sum_{k=1}^{m_q}\left[\beta_0+\frac{\beta_1}{\sqrt{m}\sigma_{t,qk}^*}+\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma_{t,qk}^*}\right)^2\right]\sigma_{t,qk}^{*2}\\
&+\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma_{t,0}^*}\right)^2\sigma_{t,0}^{*2}-\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma_{t,m}^*}\right)^2\sigma_{t,m}^{*2}
\end{aligned}$$

Hence:

$$\begin{aligned}
E\left[RV_t^{(m_q)}\right] &= IV_t+2m_q(\omega_0-\omega_{m,q})+\frac{2\beta_1^2}{q}+\frac{2(2\beta_0+1)\beta_1}{\sqrt{m}}\sum_{k=1}^{m_q}\sigma_{t,qk}^*\\
&\quad +2\beta_0(\beta_0+1)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*2}+\beta_0^2\left(\sigma_{t,0}^{*2}-\sigma_{t,m}^{*2}\right)+\frac{2\beta_0\beta_1}{\sqrt{m}}\left(\sigma_{t,0}^*-\sigma_{t,m}^*\right).
\end{aligned}$$

For the variance, we have:

$Var\left[RV^{(m_q)}\right] = 2\sum_{k=1}^{m_q}\left[\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2+\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2\right]^2\sigma^{*4}_{t,qk}$
$+2\sum_{k=1}^{m_q}\left(\sum_{l=qk-q+1}^{qk-1}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,l}\right)+Var\left[\sum_{k=1}^{m_q}\left(\varepsilon_{t,kq}-\varepsilon_{t,kq-q}\right)^2\right]$
$+4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk}$
$+4\sum_{k=1}^{m_q}\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk-q}}\right)^2\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk}$
$+4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk-q}}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q}$
$+8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q}\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2\sigma^{*2}_{t,qk}$
$+8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}+8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q}\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk-q}}\right)^2\sigma^{*2}_{t,qk-q}$
$+2\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,0}}\right)^4\sigma^{*4}_{t,0}-2\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,m}}\right)^4\sigma^{*4}_{t,m}-4\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,m}}\right)^2\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,m}}\right)^2\sigma^{*4}_{t,m}.$

In details, we have:

$2\sum_{k=1}^{m_q}\left[\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2+\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2\right]^2\sigma^{*4}_{t,qk}=$
$2\left(1+4\beta_0+8\beta_0^2+8\beta_0^3+4\beta_0^4\right)\sum_{k=1}^{m_q}\sigma^{*4}_{t,qk}+\frac{8\beta_1}{\sqrt{m}}\left(1+4\beta_0+6\beta_0^2+4\beta_0^3\right)\sum_{k=1}^{m_q}\sigma^{*3}_{t,qk}$
$+\frac{16\beta_1^2}{m}\left(1+3\beta_0+3\beta_0^2\right)\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk}+\frac{16\beta_1^3}{m\sqrt{m}}\left(1+2\beta_0\right)\sum_{k=1}^{m_q}\sigma^{*}_{t,qk}+\frac{8\beta_1^4}{qm}.$

$4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk}=4\left(1+2\beta_0+\beta_0^2\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk}$
$+\frac{8\beta_1}{\sqrt{m}}\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*}_{t,qk}+\frac{4\beta_1^2}{m}IV_t.$

$4\sum_{k=1}^{m_q}\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk-q}}\right)^2\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk}=$
$4\beta_0^2\left(1+\beta_0\right)^2\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk}+\frac{8\beta_1}{\sqrt{m}}\beta_0\left(1+\beta_0\right)^2\sum_{k=1}^{m_q}\sigma^{*}_{t,qk-q}\sigma^{*2}_{t,qk}$
$+\frac{8\beta_1}{\sqrt{m}}\beta_0^2\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk-q}\sigma^{*}_{t,qk}+\frac{16\beta_1^2}{m}\beta_0\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sigma^{*}_{t,qk-q}\sigma^{*}_{t,qk}$
$+\frac{4\beta_1^2}{m}\left(1+\beta_0\right)^2\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk}+\frac{4\beta_1^2}{m}\beta_0^2\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk-q}$
$+\frac{8\beta_1^3}{m\sqrt{m}}\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sigma^{*}_{t,qk}+\frac{8\beta_1^3}{m\sqrt{m}}\beta_0\sum_{k=1}^{m_q}\sigma^{*}_{t,qk-q}+\frac{4\beta_1^4}{qm}.$

$4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk-q}}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q}=4\beta_0^2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q}$
$+\frac{8\beta_1}{\sqrt{m}}\beta_0\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*}_{t,qk-q}+\frac{4\beta_1^2}{m}IV_t.$

$8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q}\left(1+\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk}}\right)^2\sigma^{*2}_{t,qk}=8\left(\omega_0-\omega_{m,q}\right)\left(1+\beta_0\right)^2\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk}$
$+\frac{16\beta_1}{m}\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sigma^{*}_{t,qk}+\frac{8\beta_1^2}{q}\left(\omega_0-\omega_{m,q}\right).$

$8\left(\omega_0-\omega_{m,q}\right)\sum_{k=1}^{m_q}\left(\beta_0+\frac{\beta_1}{\sqrt{m}\sigma^*_{t,qk-q}}\right)^2\sigma^{*2}_{t,qk-q}=8\left(\omega_0-\omega_{m,q}\right)\beta_0^2\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk-q}$
$+\frac{16\beta_1}{\sqrt{m}}\left(\omega_0-\omega_{m,q}\right)\beta_0\sum_{k=1}^{m_q}\sigma^{*}_{t,qk-q}+\frac{8\beta_1^2}{q}\left(\omega_0-\omega_{m,q}\right).$

Also, $Var\left[\sum_{k=1}^{m_q}\left(\varepsilon_{t,kq}-\varepsilon_{t,kq-q}\right)^2\right]=O(m_q)$. We thus define:

$$\kappa=\frac{1}{m_q}Var\left[\sum_{k=1}^{m_q}\left(\varepsilon_{t,kq}-\varepsilon_{t,kq-q}\right)^2\right].$$

The sum of all these terms yields:

$$
\begin{aligned}
Var\left[RV^{(m_q)}\right] &= m_q\kappa + \tfrac{16\beta_1^2}{q}\left(\omega_0-\omega_{m,q}\right) + \tfrac{12\beta_1^4}{qm} \\
&+8\left[\tfrac{(3+5\beta_0)\beta_1^3}{m\sqrt{m}} + \tfrac{2(1+\beta_0)\beta_1}{\sqrt{m}} + \tfrac{2\beta_0\beta_1}{\sqrt{m}}\left(\omega_0-\omega_{m,q}\right) + \tfrac{\beta_0\beta_1^3}{m\sqrt{m}}\right]\sum_{k=1}^{m_q}\sigma^*_{t,qk} \\
&+4\left(1+2\beta_0+2\beta_0^2\right)\left[\tfrac{7\left(1+2\beta_0+2\beta_0^2\right)\beta_1^2}{m} + 2\left(\omega_0-\omega_{m,q}\right)\right]\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk} \\
&+8\tfrac{\left(1+4\beta_0+6\beta_0^2+4\beta_0^3\right)\beta_1}{\sqrt{m}}\sum_{k=1}^{m_q}\sigma^3_{t,qk} + 2\sum_{k=1}^{m_q}\left(\sum_{j=qk-q+1}^{qk}\sigma^{*2}_{t,j}\right)^2 \\
&+\tfrac{16\beta_0(1+\beta_0)\beta_1^2}{m}\sum_{k=1}^{m_q}\sigma^*_{t,qk-q}\sigma^*_{t,qk} + \tfrac{8(1+\beta_0)\beta_1}{\sqrt{m}}\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^*_{t,qk} \\
&+\tfrac{8\beta_0^2(1+\beta_0)\beta_1}{\sqrt{m}}\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk-q}\sigma^*_{t,qk} + \tfrac{8\beta_0(1+\beta_0)^2\beta_1}{\sqrt{m}}\sum_{k=1}^{m_q}\sigma^*_{t,qk-q}\sigma^{*2}_{t,qk} \\
&+\tfrac{8\beta_0\beta_1}{\sqrt{m}}\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^*_{t,qk-q} + 2\left(4\beta_0+8\beta_0^2+8\beta_0^3+4\beta_0^4\right)\sum_{k=1}^{m_q}\sigma^4_{t,qk} \\
&+4\left(2\beta_0+\beta_0^2\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk} + 4\beta_0^2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q} \\
&+4\beta_0^2\left(1+\beta_0\right)^2\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk} + 8\left(\omega_0-\omega_{m,q}\right)\left(\beta_0^2+\tfrac{2\beta_0\beta_1}{\sqrt{m}}\right) + Q_m.
\end{aligned}
$$

where

$$
\begin{aligned}
Q_m &= 2\left(\beta_0+\tfrac{\beta_1}{\sqrt{m}\sigma^*_{t,0}}\right)^4\sigma^{*4}_{t,0} - 2\left(\beta_0+\tfrac{\beta_1}{\sqrt{m}\sigma^*_{t,m}}\right)^4\sigma^{*4}_{t,m} - 4\left(\beta_0+\tfrac{\beta_1}{\sqrt{m}\sigma^*_{t,m}}\right)^2\left(1+\beta_0+\tfrac{\beta_1}{\sqrt{m}\sigma^*_{t,m}}\right)^2\sigma^{*4}_{t,m} \\
&+\tfrac{8\beta_1^3\beta_0}{m\sqrt{m}}\left(\sigma^*_{t,0}-\sigma^*_{t,qm_q}\right) + \tfrac{4\beta_1^2\beta_0^2}{m}\left(\sigma^{*2}_{t,0}-\sigma^{*2}_{t,qm_q}\right) = O(m^{-1}).\blacksquare
\end{aligned}
$$

Proof of Theorem 6: Substituting for $a_{t,j} = \beta_0 + \frac{\beta_1}{\sqrt{m}\sigma_{t,j}}$ in Lemma 9, yield:

$$
E\left[RV_t^{(AC,m,1)}\right] = IV_t + \left(\beta_0^2+2\beta_0\right)\left(\sigma^{*2}_{t,m}-\sigma^{*2}_{t,0}\right) - \frac{2\beta_1\left(1+\beta_0\right)}{\sqrt{m}}\left(\sigma^*_{t,m}-\sigma^*_{t,0}\right).
$$

For the variance, we have:

$$
\begin{aligned}
Var\left[RV_t^{(AC,m,1)}\right] &= 2\sum_{j=1}^m\sigma^{*4}_{t,j} + 4\sum_{j=1}^m\left(1+a_{t,j}+a_{t,j}a_{t,j-1}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,j-1} \\
&+4\sum_{j=1}^m\left(1+a_{t,j}\right)^2a^2_{t,j-2}\sigma^{*2}_{t,j}\sigma^{*2}_{t,j-2} + 8\omega_0\sum_{j=1}^m\left(1+a_{t,j}\right)^2\sigma^{*2}_{t,j} \\
&+8\omega_0\sum_{j=1}^m a^2_{t,j}\sigma^{*2}_{t,j} + 8m\omega_0^2 + 2\left(E\left[\varepsilon^4_{t,j}\right]-\omega_0^2\right) + R_m.
\end{aligned}
$$

where

$$
\begin{aligned}
R_m &= 2\left(2a_{t,0}+a^2_{t,0}\right)^2\sigma^{*4}_{t,0} + 2\left(2a_{t,m}+a^2_{t,m}\right)^2\sigma^{*4}_{t,m} + 2\left(2a_{t,m}+a^2_{t,m}\right)\sigma^{*4}_{t,m} \\
&+4a^2_{t,-1}a^2_{t,0}\sigma^{*2}_{t,-1}\sigma^{*2}_{t,0} - 8a_{t,m-1}a_{t,m}\left(1+a_{t,m}+a_{t,m}a_{t,m-1}\right)\sigma^{*2}_{t,m-1}\sigma^{*2}_{t,m} \\
&+4a^2_{t,m-1}a^2_{t,m}\sigma^{*2}_{t,m-1}\sigma^{*2}_{t,m} + 8\omega_0\left(\sigma^{*2}_{t,m-1}-\sigma^{*2}_{t,0}\right) \\
&+8\omega_0\left(a^2_{t,-1}\sigma^{*2}_{t,-1} + 2a^2_{t,0}\sigma^{*2}_{t,0} + a_{t,m}\sigma^{*2}_{t,m} - a_{t,m-1}\sigma^{*2}_{t,m-1} - a^2_{t,m-1}\sigma^{*2}_{t,m-1}\right).
\end{aligned}
$$

$$
R_m = 4\beta_1^4 + 3\beta_1^2\omega_0 + O\left(\beta_0\beta_1 m^{-1/2}\right).
$$

$$
\begin{aligned}
4\sum_{j=1}^m\left(1+a_{t,j}+a_{t,j}a_{t,j-1}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,j-1} &= \tfrac{4\beta_1^4}{m} + \tfrac{8\beta_0\beta_1^3}{m\sqrt{m}}\sum_{j=1}^m\sigma^*_{t,j} \\
&+\tfrac{8\beta_1^2}{m}\left(1+\tfrac{\beta_0\beta_1}{\sqrt{m}}\right)\sum_{j=1}^m\sigma^*_{t,j-1} + \tfrac{8\beta_1^2}{m}\left(1+2\beta_0+2\beta_0^2\right)\sum_{j=1}^m\sigma^*_{t,j}\sigma^*_{t,j-1} \\
&+\tfrac{4\beta_0^2\beta_1^2}{m}\sum_{j=1}^m\sigma^{*2}_{t,j} + \tfrac{4\beta_1^2(1+\beta_0)^2}{m}\sum_{j=1}^m\sigma^{*2}_{t,j-1} + \tfrac{8\beta_0\beta_1}{\sqrt{m}}\left(1+\beta_0+\beta_0^3\right)\sum_{j=1}^m\sigma^{*2}_{t,j}\sigma^*_{t,j-1}
\end{aligned}
$$

$+\frac{8\beta_1}{\sqrt{m}}\left(1+2\beta_0+2\beta_0^2+\beta_0^3\right)\sum_{j=1}^m\sigma_{t,j-1}^{*2}\sigma_{t,j}^{*}+4\left(1+2\beta_0+3\beta_0^2+2\beta_0^3+\beta_0^4\right)\sum_{j=1}^m\sigma_{t,j-1}^{*2}\sigma_{t,j}^{*2}.$

$4\sum_{j=1}^m\left(1+a_{t,j}\right)^2 a_{t,j-2}^2\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}=\frac{4\beta_1^4}{m}+\frac{8\beta_0\beta_1^3}{m\sqrt{m}}\sum_{j=1}^m\sigma_{t,j-2}^{*}$

$+\frac{8(1+\beta_0)\beta_1^3}{m\sqrt{m}}\sum_{j=1}^m\sigma_{t,j}^{*}+\frac{4\beta_0^2\beta_1^2}{m}\sum_{j=1}^m\sigma_{t,j-2}^{*2}+\frac{8\beta_1(1+\beta_0)\beta_0^2}{\sqrt{m}}\sum_{j=1}^m\sigma_{t,j-2}^{*2}\sigma_{t,j}^{*}$

$+\frac{4\beta_1^2(1+\beta_0)^2}{m}\sum_{j=1}^m\sigma_{t,j}^{*2}+\frac{16\beta_0\beta_1^2(1+\beta_0)}{m}\sum_{j=1}^m\sigma_{t,j}^{*}\sigma_{t,j-2}^{*}$

$+\frac{8\beta_1\beta_0(1+\beta_0)^2}{\sqrt{m}}\sum_{j=1}^m\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*}+4\beta_0^2\left(1+\beta_0\right)^2\sum_{j=1}^m\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}.$

$8\omega_0\sum_{j=1}^m\left(1+a_{t,j}\right)^2\sigma_{t,j}^{*2}=8\omega_0\beta_1^2+\frac{16\omega_0\beta_1(1+\beta_0)}{\sqrt{m}}\sum_{j=1}^m\sigma_{t,j}^{*}$

$+8\omega_0\left(1+\beta_0\right)^2\sum_{j=1}^m\sigma_{t,j}^{*2}.$

$8\omega_0\sum_{j=1}^m a_{t,j}^2\sigma_{t,j}^{*2}=8\omega_0\beta_1^2+\frac{16\omega_0\beta_1\beta_0}{\sqrt{m}}\sum_{j=1}^m\sigma_{t,j}^{*}+8\omega_0\beta_0^2\sum_{j=1}^m\sigma_{t,j}^{*2}.$

Hence:

$Var\left[RV_t^{(AC,m,1)}\right]=8m\omega_0^2+2\sum_{j=1}^m\sigma_{t,j}^{*4}+2\left(E\left[\varepsilon_{t,j}^4\right]-\omega_0^2\right)$

$+\frac{\beta_1^4+6\beta_1^2\omega_0}{m}+\frac{8\beta_1^4}{m^2}+\frac{8\beta_1}{\sqrt{m}}\left[\frac{(\beta_0+1)^2\beta_1^2}{m}+\frac{\beta_1}{\sqrt{m}}+2\omega_0\left(1+2\beta_0\right)\right]\sum_{j=1}^m\sigma_{t,j}^{*}$

$+8\left[\frac{\beta_0^2\beta_1^2}{m}+\left(\frac{\beta_1^2}{m}+\omega_0\right)\left(1+\beta_0\right)^2+2\omega_0\beta_0^2\right]\sum_{j=1}^m\sigma_{t,j}^{*2}$

$+\frac{8\beta_1^2}{m}\left(1+2\beta_0+2\beta_0^2\right)\sum_{j=1}^m\sigma_{t,j}^{*}\sigma_{t,j-1}^{*}+\frac{16\beta_0\beta_1^2}{m}\left(1+\beta_0\right)\sum_{j=1}^m\sigma_{t,j}^{*}\sigma_{t,j-2}^{*}$

$+\frac{8\beta_0\beta_1}{\sqrt{m}}\left(1+\beta_0+\beta_0^3\right)\sum_{j=1}^m\sigma_{t,j}^{*2}\sigma_{t,j-1}^{*}+\frac{8\beta_1}{\sqrt{m}}\left(1+2\beta_0+2\beta_0^2+\beta_0^3\right)\sum_{j=1}^m\sigma_{t,j-1}^{*2}\sigma_{t,j}^{*}$

$+\frac{8\beta_1(1+\beta_0)\beta_0^2}{\sqrt{m}}\sum_{j=1}^m\sigma_{t,j-2}^{*2}\sigma_{t,j}^{*}+\frac{8\beta_1\beta_0(1+\beta_0)^2}{\sqrt{m}}\sum_{j=1}^m\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*}$

$+4\left(1+2\beta_0+3\beta_0^2+2\beta_0^3+\beta_0^4\right)\sum_{j=1}^m\sigma_{t,j-1}^{*2}\sigma_{t,j}^{*2}+4\beta_0^2\left(1+\beta_0\right)^2\sum_{j=1}^m\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}+\beta_0 O\left(m^{-1/2}\right).\blacksquare$

Proof of Theorem 7:

We examine the term $K_t^{BNHLS}\left(r^*,\Delta\varepsilon\right)$:

$$K_t^{BNHLS}\left(r^*,\Delta\varepsilon\right)=\gamma_{t,0}\left(r^*,\Delta\varepsilon\right)+2\sum_{h=1}^{H}k\left(\frac{h-1}{H}\right)\gamma_{t,h}\left(r^*,\Delta\varepsilon\right).$$

Let us define $\Phi=\left(1,k\left(\frac{0}{H}\right),k\left(\frac{1}{H}\right),...,k\left(\frac{H-1}{H}\right)\right)'$. Then, we have:

$$K_t^{BNHLS}\left(r^*,\Delta\varepsilon\right)=\Phi'\sum_{j=1}^m r_{t,j}^*\begin{pmatrix}\varepsilon_{t,j}-\varepsilon_{t,j-1}\\ 2\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right)\\ \cdots\\ 2\left(\varepsilon_{t,j-H}-\varepsilon_{t,j-H-1}\right)\end{pmatrix}.$$

Note that:

$$
\begin{aligned}
Var\left[K_t^{BNHLS}(r^*,\Delta\varepsilon)\right] &= Var\left[E\left[K_t^{BNHLS}(r^*,\Delta\varepsilon)\,|\,\{(\varepsilon_{t,j-h}-\varepsilon_{t,j-h-1})\}_{h=0}^{H}\right]\right] \\
&\quad +E\left[Var\left[K_t^{BNHLS}(r^*,\Delta\varepsilon)\,|\,\{(\varepsilon_{t,j-h}-\varepsilon_{t,j-h-1})\}_{h=0}^{H}\right]\right] \\
&= E\left[Var\left[K_t^{BNHLS}(r^*,\Delta\varepsilon)\,|\,\{(\varepsilon_{t,j-h}-\varepsilon_{t,j-h-1})\}_{h=0}^{H}\right]\right] \\
&= IV_t\Phi' Var\left(\Delta\varepsilon^H\right)\Phi,
\end{aligned}
$$

where $\Delta\varepsilon^H = (\varepsilon_{t,j}-\varepsilon_{t,j-1}, 2(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}), ..., 2(\varepsilon_{t,j-H}-\varepsilon_{t,j-H-1}))$.

We now compute explicitely $Var\left(\Delta\varepsilon^H\right)$:

$$
\begin{aligned}
E\left[(\varepsilon_{t,j}-\varepsilon_{t,j-1})^2\right] &= 2(\omega_0-\omega_{m,1}) \\
E\left[(\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-h}-\varepsilon_{t,j-h-1})\right] &= -\omega_{m,h-1}+2\omega_{m,h}-\omega_{m,h+1} \\
E\left[(\varepsilon_{t,j-h}-\varepsilon_{t,j-h-1})(\varepsilon_{t,j-k}-\varepsilon_{t,j-k-1})\right] &= -\omega_{m,h-k-1}+2\omega_{m,h-k}-\omega_{m,h-k+1}, h>k.
\end{aligned}
$$

Let $\Delta\omega_{m,h} = \omega_{m,h}-\omega_{m,h+1}$ for all h. Then:

$$
Var\left(\Delta\varepsilon^H\right) = 2\times
$$

$$
\begin{pmatrix}
\Delta\omega_0 & \bullet & ... & \bullet & \bullet \\
(-\Delta\omega_0+\Delta\omega_{m,1}) & 2\Delta\omega_0 & ... & \bullet & \bullet \\
(-\Delta\omega_{m,1}+\Delta\omega_{m,2}) & 2(-\Delta\omega_0+\Delta\omega_{m,1}) & ... & \bullet & \bullet \\
... & 2(-\Delta\omega_{m,1}+\Delta\omega_{m,2}) & ... & ... & ... \\
... & & ... & 2\Delta\omega_0 & \bullet \\
(-\Delta\omega_{m,H-1}+\Delta\omega_{m,H}) & 2(-\Delta\omega_{m,H-2}+\Delta\omega_{m,H-1}) & ... & 2(-\Delta\omega_0+\Delta\omega_{m,1}) & 2\Delta\omega_0
\end{pmatrix}
$$

To ease the calculations, a simplified representation of $Var\left(\Delta\varepsilon^H\right)$ is needed. To that end, let us define:

$$
\underset{(H+1\times H+1)}{S^0} = \begin{pmatrix}
1 & -1 & \bullet & ... & \bullet \\
-1 & 2 & -1 & ... & ... \\
0 & -1 & 2 & ... & \bullet \\
... & ... & ... & ... & -1 \\
0 & & 0 & -1 & 2
\end{pmatrix}
$$

Also let S^h be the symmetric matrix of size $H+1$ with elements $S^h_{j,k} = 1$ if $j = k+h$ or

$j=k-h$, $S^h_{j,k}=-1$ if $j=k+h+1$ or $j=k-h-1$, and $S^h_{j,k}=0$ otherwise. In fact, S^h is the sparse matrix with ones on the h^{th} diagonals and minus ones on the $h+1^{th}$ diagonals. Finally, let $\widetilde{S}^h$ be the matrix S^h with the nonzero elements of the first row and first column replaced by zero. Then we have:

$$\Phi' Var\left(\Delta\varepsilon^H\right)\Phi = 2\left(\omega_0-\omega_{m,1}\right)\Phi' S^0\Phi + 2\sum_{h=1}^{L}\left(\omega_{m,h}-\omega_{m,h+1}\right)\Phi'\left(S^h+\widetilde{S}^h\right)\Phi.$$

We easily check that:

$$\begin{aligned}
\Phi' S^0\Phi &= \sum_{h=0}^{H}\left(k\left(\frac{h+1}{H}\right)-k\left(\frac{h}{H}\right)\right)^2 \to \frac{1}{H}\int_0^1 k'(x)^2\,dx = \frac{1}{H}.\\
\Phi' S^h\Phi + \Phi'\widetilde{S}^h\Phi &= k\left(\frac{h}{H}\right)-k\left(\frac{h+1}{H}\right)+\frac{4}{H}\sum_{l=0}^{H-h-1}k\left(\frac{l}{H}\right),\\
&= \frac{1}{H}+4\int_0^{1-\frac{h+1}{H}}k(x)\,dx = \frac{1}{H}+2\left[1-\frac{(h+1)^2}{H^2}\right].
\end{aligned}$$

Focusing on the dominant terms, we have:

$$\begin{aligned}
\Phi' Var\left(\Delta\varepsilon^H\right)\Phi &\approx \frac{2}{H}\sum_{h=0}^{L}\left(\omega_{m,h}-\omega_{m,h+1}\right)+4\sum_{h=1}^{L-1}\left(\omega_{m,h}-\omega_{m,h+1}\right)\left[1-\frac{(h+1)^2}{H^2}\right]\\
&\quad +4\omega_{m,L}\left[1-\frac{(L+1)^2}{H^2}\right],\\
&= \frac{2\omega_0}{H}+4\sum_{h=1}^{L-1}\left(\omega_{m,h}-\omega_{m,h+1}\right)\left[1-\frac{(h+1)^2}{H^2}\right]+4\omega_{m,L}\left[1-\frac{(L+1)^2}{H^2}\right].
\end{aligned}$$

This yields the second result. The remaining term to examine is thus $K_t^{BNHLS}(\Delta\varepsilon)$. We have:

$$\begin{aligned}
K_t^{BNHLS}(\Delta\varepsilon) &= \sum_{j=1}^{m}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)^2+2\sum_{h=1}^{H}k\left(\frac{h-1}{H}\right)\sum_{j=1}^{m}\left(\varepsilon_{s,j}-\varepsilon_{s,j-1}\right)\left(\varepsilon_{s,j-h}-\varepsilon_{s,j-h-1}\right),\\
&= V_t^{(AC,m,1)}+2\sum_{h=2}^{H}k\left(\frac{h-1}{H}\right)\sum_{j=1}^{m}\left(\varepsilon_{s,j}-\varepsilon_{s,j-1}\right)\left(\varepsilon_{s,j-h}-\varepsilon_{s,j-h-1}\right).
\end{aligned}$$

$V_t^{(AC,m,1)}=\sum_{j=1}^m\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)^2+2\sum_{j=1}^m\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right)$

$$= -2\sum_{j=1}^{m}\varepsilon_{t,j-2}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)-\varepsilon_{t,0}^{2}+\varepsilon_{t,m}^{2}$$
$$= 2\sum_{j=1}^{m}\varepsilon_{t,j}\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right)-\varepsilon_{t,0}^{2}+\varepsilon_{t,m}^{2}+2\left(\varepsilon_{t,0}\varepsilon_{t,-1}-\varepsilon_{t,m}\varepsilon_{t,m-1}\right).$$

And for $h \geq 2$, we have:

$$\sum_{j=1}^{m}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-h}-\varepsilon_{t,j-h-1}\right)=$$
$$\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-h}-\sum_{j=1}^{m}\varepsilon_{t,j-1}\varepsilon_{t,j-h}-\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-h-1}+\sum_{j=1}^{m}\varepsilon_{t,j-1}\varepsilon_{t,j-h-1}=$$
$$-\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-h+1}+2\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-h}-\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-h-1}$$
$$-\left(\varepsilon_{t,0}\varepsilon_{t,-h+1}-\varepsilon_{t,m}\varepsilon_{t,m-h+1}\right)+\left(\varepsilon_{t,0}\varepsilon_{t,-h}-\varepsilon_{t,m}\varepsilon_{t,m-h}\right).$$

Summing over H yields:

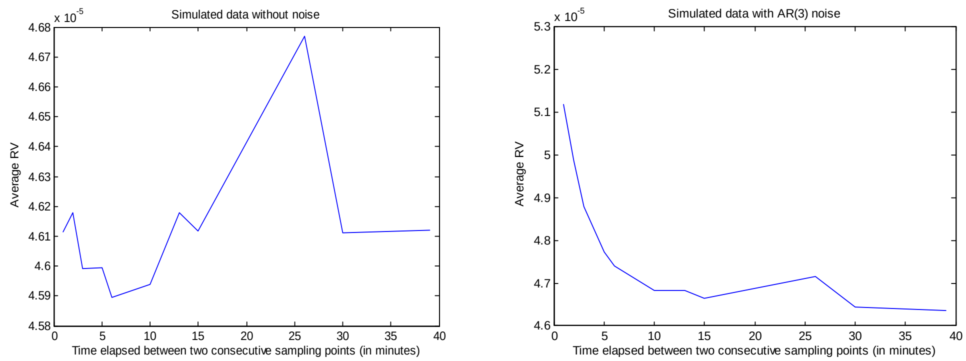
$$2\sum_{h=2}^{H}k\left(\tfrac{h-1}{H}\right)\sum_{j=1}^{m}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-h}-\varepsilon_{t,j-h-1}\right)$$
$$= -2\sum_{j=1}^{m}\varepsilon_{t,j}\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right)-\tfrac{4}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H}$$
$$-\tfrac{2}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H-1}-\tfrac{2}{H}\sum_{h=2}^{H-1}\left(\varepsilon_{t,0}\varepsilon_{t,-h}-\varepsilon_{t,m}\varepsilon_{t,m-h}\right)$$
$$-2\left(\varepsilon_{t,0}\varepsilon_{t,-1}-\varepsilon_{t,m}\varepsilon_{t,m-1}\right)+\tfrac{2}{H}\left(\varepsilon_{t,0}\varepsilon_{t,-H}-\varepsilon_{t,m}\varepsilon_{t,m-H}\right).$$

Finally, we have:

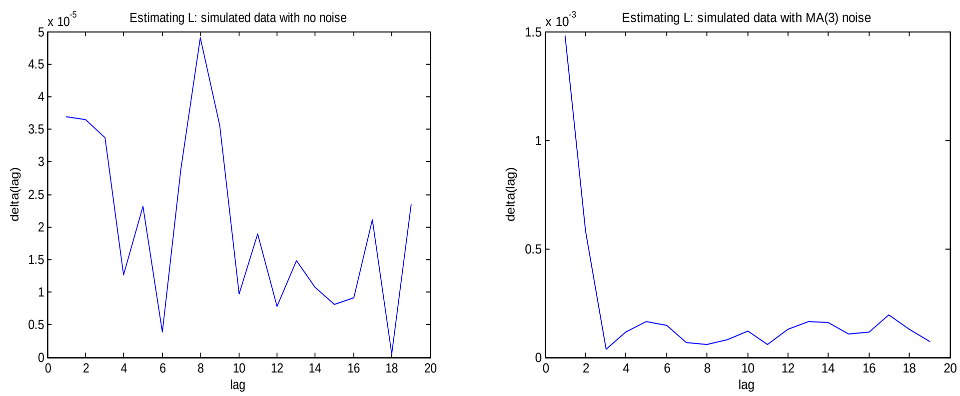
$$\begin{aligned}
K_t^{BNHLS}\left(\Delta\varepsilon\right) &= -\varepsilon_{t,0}^{2}+\varepsilon_{t,m}^{2}-\frac{4}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H}-\frac{2}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H-1}\\
&\quad -\frac{2}{H}\sum_{h=2}^{H-1}\left(\varepsilon_{t,0}\varepsilon_{t,-h}-\varepsilon_{t,m}\varepsilon_{t,m-h}\right)+\frac{2}{H}\left(\varepsilon_{t,0}\varepsilon_{t,-H}-\varepsilon_{t,m}\varepsilon_{t,m-H}\right)\\
&= -\varepsilon_{t,0}^{2}+\varepsilon_{t,m}^{2}+O_p(H^{-1}m^{1/2}).
\end{aligned}$$

■

Appendix B: Simulation Results



Volatility Signature Plots



Plots of $\Delta(l)$ against l.

Figure 2: Detecting Noise in the Data

Left: data with no noise. Right: data with MA(3) noise.

Variance of $\varepsilon_{t,j}$	Frequency	MSE ($\times 10^{-6}$)			Shrinkage weight
ω_0	m	K_t^{BNHLS}	$\theta_{1,t}^{(L)}$	$K_t^{\varpi^*}$	$\widehat{\varpi}^*$
2.25×10^{-8}	780	0.0016	0.0009	0.0009	0.2525
	390	0.0022	0.0018	0.0015	0.3988
	195	0.0029	0.0028	0.0025	0.4721
	78	0.0050	0.0052	0.0047	0.6036
2.5×10^{-7}	780	0.0017	0.0017	0.0012	0.4962
	390	0.0022	0.0022	0.0017	0.5125
	195	0.0030	0.0033	0.0026	0.5719
	78	0.0051	0.0055	0.0049	0.6438
2.25×10^{-6}	780	0.0049	0.0165	0.0048	0.9263
	390	0.0045	0.0113	0.0044	0.8955
	195	0.0050	0.0095	0.0049	0.8387
	78	0.0071	0.0092	0.0070	0.8546
2.5×10^{-5}	780	0.3572	1.6563	0.3571	1.0040
	390	0.2314	0.7405	0.2314	1.0014
	195	0.1597	0.3862	0.1596	0.9845
	78	0.1126	0.1674	0.1126	0.9714

Table 2: Evaluating the performance of the shrinkage estimators of IV_t by Monte Carlo: Case with no Leverage Effect.

Variance of $\varepsilon_{t,j}$	Frequency	MSE ($\times 10^{-6}$)			Shrinkage weight
ω_0	m	K_t^{BNHLS}	$\theta_{1,t}^{(L)}$	$K_t^{\varpi^*}$	$\widehat{\varpi}^*$
2.25×10^{-8}	780	0.0016	0.0008	0.0007	0.2223
	390	0.0021	0.0016	0.0014	0.3172
	195	0.0029	0.0028	0.0025	0.4583
	78	0.0048	0.0055	0.0047	0.7701
2.5×10^{-7}	780	0.0016	0.0014	0.0010	0.4359
	390	0.0022	0.0022	0.0017	0.4867
	195	0.0030	0.0031	0.0026	0.5216
	78	0.0050	0.0057	0.0049	0.7724
2.25×10^{-6}	780	0.0047	0.0173	0.0046	0.9108
	390	0.0044	0.0113	0.0042	0.8696
	195	0.0048	0.0088	0.0047	0.8685
	78	0.0065	0.0084	0.0064	0.8679
2.5×10^{-5}	780	0.3461	1.5821	0.3461	0.9997
	390	0.2217	0.8107	0.2217	1.0065
	195	0.1554	0.3740	0.1554	0.9976
	78	0.1083	0.1529	0.1077	0.9249

Table 3: Evaluating the performance of the shrinkage estimators of IV_t by Monte Carlo: Case with Leverage Effect.

	True (x10^{-7})	Estimate (x10^{-7})	Std. Dev. (x10^{-7})	Student-t
$\widehat{\omega}_0$	2.2500	3.8423	0.4218	9.1095
$\widehat{\omega}_1$	1.0619	1.0722	0.1224	8.7595
$\widehat{\omega}_2$	0.3917	0.3946	0.0512	7.7072
$\widehat{\omega}_3$	0.0870	0.0932	0.0210	4.4365
$\widehat{\omega}_4$	0.0000	0.0082	0.0098	0.8432

m=780

	True (x10^{-7})	Estimate (x10^{-7})	Std. Dev. (x10^{-7})	Student-t
$\widehat{\omega}_0$	2.2500	5.4195	0.5989	9.0489
$\widehat{\omega}_1$	1.0619	1.0885	0.1386	7.8517
$\widehat{\omega}_2$	0.3917	0.4109	0.0689	5.9660
$\widehat{\omega}_3$	0.0870	0.1006	0.0375	2.6806
$\widehat{\omega}_4$	0.0000	0.0021	0.0212	0.0995

m=390

Table 4: Estimated correlogram of the noise (Simulated Data).

Appendix C: Empricial Results

	$\widehat{L}$	$\widehat{\alpha}$ ($\widehat{\sigma}_{\widehat{\alpha}}$)	$\widehat{\delta}$
3M Co (MMM)	12	0.5301 (0.0367)	0.4165
Alcoa Inc (AA)	12	0.4797 (0.0116)	0.4165
American International Group (AIG)	14	0.4715 (0.0123)	0.4423
Americal Express (AXP)	4	0.3151 (0.0087)	0.2324
Dupont and Dupont (DD)	12	0.4805 (0.0104)	0.4165
Walt Disney (DIS)	9	0.4773 (0.0146)	0.3683
General Electric (GE)	14	0.5303 (0.0233)	0.4423
General Motors (GM)	13	0.5117 (0.0658)	0.4299
IBM	12	0.4802 (0.0149)	0.4165
Intel Corp. (INTC)	11	0.5048 (0.0238)	0.4019
Hewlett-Packard (HPQ)	12	0.4942 (0.0115)	0.4165
Microsoft (MSFT)	11	0.4960 (0.0228)	0.4019

Table 5: Estimates of L, α and δ for twelve stocks listed in the DJI.

$\widehat{\sigma}_{\widehat{\alpha}}$ is the estimated standard deviation of $\widehat{\alpha}$

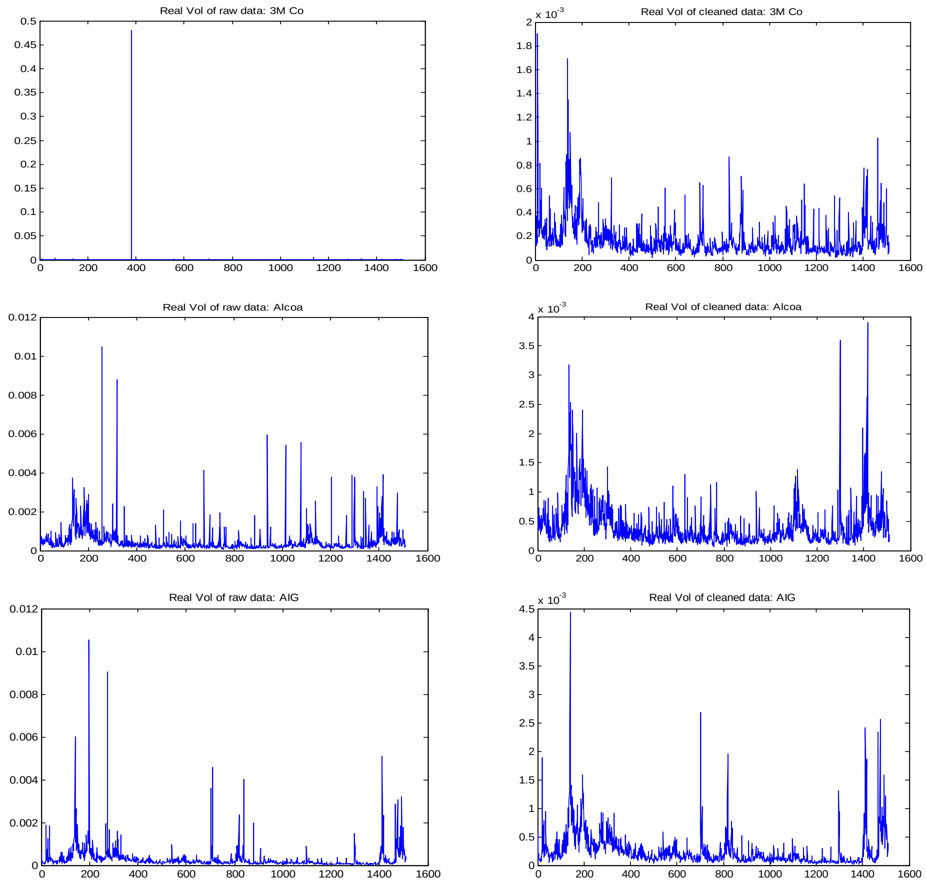
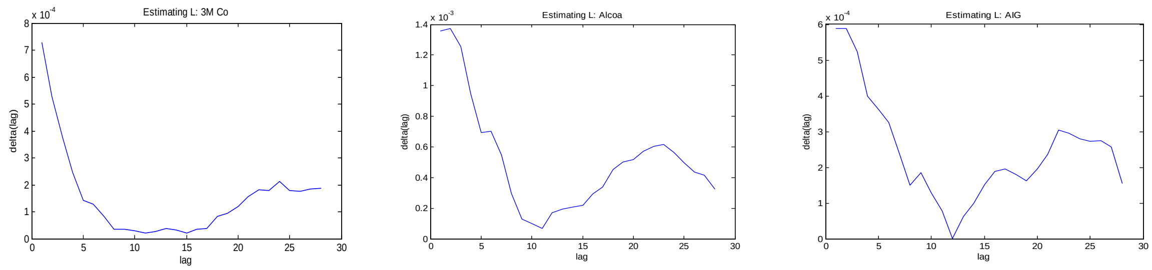
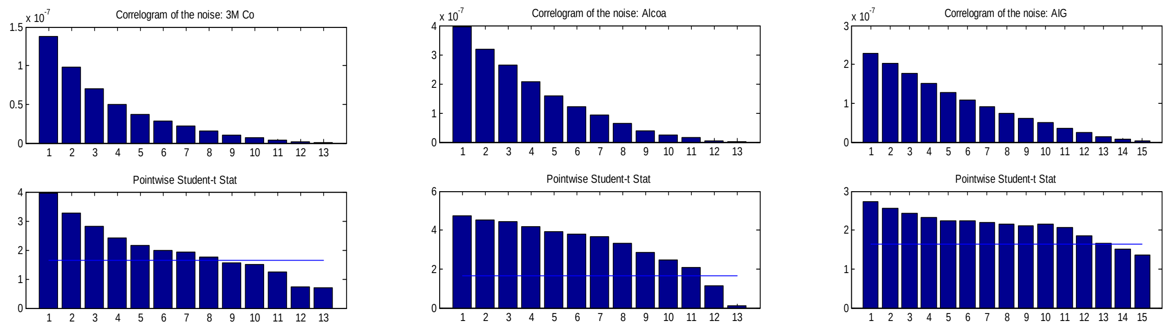


Figure 3: Preprocessing the data.

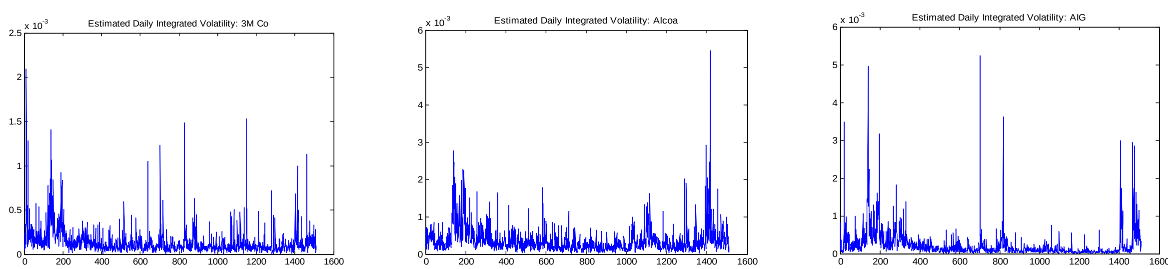
Left: Realized volatility of $r_{t,j}^{OLD}$. Right: Realized volatility of $r_{t,j}^{NEW}$.



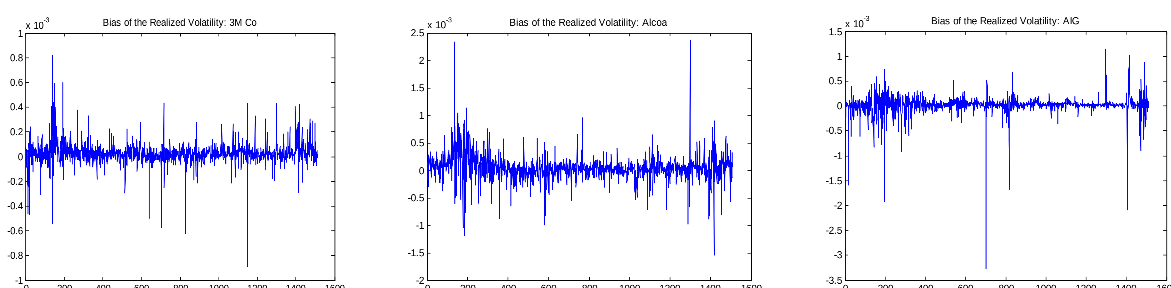
Plot of $\Delta(l)$ against l. The minimum indicates the estimator of L.



The correlogram of the noise $\{\omega_{m,h}\}_{h=0}^{L}$ (top) and the pointwise associated Student statistics (bottom).



Estimated Daily Integrated Volatility $K_t^{\varpi^*}$



Bias of the Realized Volatility $RV^{(m)}$

Figure 4: Estimation Results for 3M Co, Alcoa and AIG.

Chapter 2

Efficient Estimation Using the Characteristic Function

Note: Cet article rédigé en collaboration avec Marine Carrasco est actuellement sous évaluation pour publication dans "*Econometric Theory*"

Mots-Clés: GMM, Continuum of moment conditions, MSE, Stochastic Expansion, Tikhonov regularization.

2.1 Introduction

The empirical characteristic function (henceforth ECF) has played an increasing role in econometrics and finance. Paulson et al. (1975) used a weighted modulus of the difference between the ECF and the theoretical characteristic function (henceforth CF) to estimate the parameters of the stable law. Feuerverger and Mureika (1977) initiated its use for inference. Since then, many interesting applications have been done including Feuerverger and McDunnough (1981*b,c*), Koutrouvelis (1980), Carrasco and Florens (2000), Chacko and Viceira (2003), and Carrasco, Chernov, Florens, and Ghysels (2007)[1]. The CF has the appealing property that, as the Fourier transform of the probability distribution function, it fully characterizes the distribution of the underlying random variable. An estimation procedure based on the CF

[1]See Jun Yu (2004) for a comprehensive review of empirical characteristic function based estimation methods.

is thus expected to have the same level of efficiency as maximum likelihood. Moreover, the likelihood function is often either unavailable in closed form (stable law, discretely sampled continuous time process) or difficult to handle (mixtures of distributions). In the field of finance, it is usually assumed that asset prices follow continuous-time diffusion processes. But, as mentioned by Singleton (2001), the conditional density of discretely sampled returns are not known in closed form except for the very special cases of Gaussian or square-root diffusion. Fortunately, the conditional CF is available in closed form for a wide range of commonly assumed dynamics in finance. In the context of stochastic volatility models, Ait-Sahalia and Kimmel (2006) proposed closed form expansions of the log-likelihood function of various continuous-time processes. But their method cannot be applied to other situations without solving a complicated Kolmogorov forward and backward equation.

For the simplicity of the presentation, let's assume an IID sample $(x_1, \ldots, x_T)$ of a multivariate process $x_t \in \mathbb{R}^p$ and let $\varphi_T(\tau)$ denote its ECF, that is:

$$\varphi_T(\tau) = \frac{1}{T}\sum_{t=1}^{T} e^{i\tau' x_t}, \tau \in \mathbb{R}^p. \tag{2.1}$$

The ECF is related to its theoretical counterpart φ through the following:

$$\varphi(\tau, \theta_0) = E^{\theta_0}\left(\varphi_T(\tau)\right) = E^{\theta_0}\left(e^{i\tau' x_t}\right), \tau \in \mathbb{R}^p \tag{2.2}$$

where $\theta_0 \in \Theta \subset \mathbb{R}^q$ is the true value of the finite dimensional parameter that fully characterizes the true data generating process, and E^{θ_0} is the expectation operator with respect to that data generating process.

Feuerverger and McDunnough (1981b) proposed an estimator that is obtained by minimizing the distance between the ECF and its theoretical counterpart. The objective function they propose involves an optimal weighting function which depends on the true likelihood function that may be unknown. Feuerverger and McDunnough (1981c) proposed to apply the Generalized Method of Moments (GMM) to the discrete set of moment conditions:

$$h_t(\tau_k, \theta) = e^{i\tau_k' x_t} - \varphi(\tau_k, \theta), \tau_k \in (\tau_1, \tau_2, ...\tau_q) \subset \mathbb{R}^p$$

where $h_t(\tau_k, \theta)$ depends on θ_0 through the data. They show that the asymptotic variance of the resulting estimator can be made arbitrarily close to the Cramer-Rao bound by selecting the grid sufficiently fine and extended. But one should note that $q \leq T$ is a necessary (but not sufficient) condition for the covariance matrix of the moment conditions to be invertible. Indeed, singularity will arise whenever the number of points in the grid $(\tau_1, \tau_2, ...\tau_q)$ exceeds the sample size T. Moreover, operator theory is necessary to handle the estimation procedure at the limit because as one refines and extends the grid, the discrete set of moment conditions converges to the moment function:

$$h_t(\tau, \theta) = e^{i\tau' x_t} - \varphi(\tau, \theta), \tau \in \mathbb{R}^p \tag{2.3}$$

and the covariance matrix of the moment condition converges to the covariance operator associated with that moment function. Other discretization approaches can be found in Singleton (2001) and Chacko and Viceira (2003).

To deal with these problems, Carrasco and Florens (2000) (referred to as CaFl subsequently) propose a method that can efficiently use the whole continuum of moment conditions given by

$$E^{\theta_0}\left[h_t(\tau, \theta_0)\right] = 0, \ \forall \ \tau \in \mathbb{R}^p. \tag{2.4}$$

The resulting method has been termed CGMM (Continuum GMM or GMM with a continuum of moment conditions). Their approach is original in that the moment functions are treated as elements of some Hilbert space endowed with a properly defined scalar product denoted $\langle ., . \rangle$. The objective function of the CGMM thus takes the following form:

$$\widehat{\theta} = \arg\min_{\theta} \left\langle B_T \widehat{h}_T(., \theta), \widehat{h}_T(., \theta) \right\rangle. \tag{2.5}$$

where B_T is a sequence of linear operators converging to a limiting operator B as the sample size T goes to infinity, and

$$\widehat{h}_T(\tau, \theta) = \frac{1}{T} \sum_{t=1}^{T} h_t(\tau, \theta) \tag{2.6}$$

In theory, the CGMM estimator is optimal when B_T is the inverse of the empirical covariance operator K_T associated with the moment conditions. But the operator K_T is not necessarily invertible. CaFl thus suggest to use a regularized inverse of the following form:

$$B_{\alpha T} = \left(K_T^2 + \alpha_T I\right)^{-1} K_T \equiv K_{\alpha T}, \tag{2.7}$$

where I denotes the identity operator, K_T is the sample counterpart of K, and α_T is a regularization (or smoothing) parameter that needs to be selected in practice. It is shown in CaFl (2000) that the CGMM estimator $\widehat{\theta}$ is root T consistent, asymptotically normal and asymptotically as efficient as the MLE if α_T converges to zero at a certain rate.

However, the small-sample properties of the estimator will be affected by α_T in a complex way. To gain more insight on the role of α_T, we derive a higher order expansion of the CGMM estimator. Then we address the issue of the optimal selection of α_T. Ideally, we want to select α_T so that it minimizes the MSE of $\widehat{\theta}$ at any given sample size. However, that MSE is not known in closed form. We propose two ways to approximate the MSE. The first method relies on the first terms of the higher-order expansion of the MSE. The second one uses an average over simulated data. We show that our second selection procedure delivers a root T consistent estimator of the optimal α_T. It is an adaptive method in the sense that it does not require the knowledge of the regularity of the moment functions.

The rest of the paper is organized as follow. In Section 2, we briefly review the estimation procedure both in the IID and Markovian cases, and establish the asymptotic properties of the CGMM estimator. In Section 3, we derive the higher-order expansion of the mean square error of the CGMM estimator $\widehat{\theta}_T$ similar to that in Newey and Smith (2004). From that expansion, we gain more insight on the asymptotic behavior of the higher order terms. In Section 4, we study two methods to select the optimal value of the tuning parameter α. Section 5 presents Monte Carlo experiments for a stochastic frontier models involving a convolution of an IID normal with an exponential distribution. Section 6 concludes. The proofs are collected in appendix.

2.2 Overview of the CF-based CGMM

CaFl (2000) focus on the case where the p-dimensional process $\{x_t\}$ is IID and proposed to estimate θ_0 by considering moment restrictions of type (2.3). More recently, Carrasco, Chernov, Florens and Ghysels (2007) (referred to as CCFG) extend the set up to more general situations where the data can be Markovian or weakly dependent. In this paper, we will consider IID and Markov cases only.

The following moment condition has been proposed by CCFG (2007) for the Markov case:

$$h_t(\tau,\theta) = \left[e^{is'x_{t+1}} - \varphi(s,\theta,x_t)\right] e^{ir'x_t}. \tag{2.8}$$

where $\varphi(s,\theta,x_t) = E^\theta(e^{is'x_{t+1}}|x_t)$ is the conditional CF and $\tau = (s,r) \in \mathbb{R}^{2p}$. In equation (2.8), the set of basis functions $\{e^{ir'x_t}\}$ is used as instruments. These instruments are optimal given the Markovian structure of the model (see CCFG (2007) for a discussion). Moment conditions defined by (2.3) are IID whereas equation (2.8) describe a martingale difference sequence. From now on, the generic notation $h_t(\tau,\theta)$, $\tau \in \mathbb{R}^d$ will denote a moment function defined by (2.3) or (2.8) where $d = p$ for (2.3) and $d = 2p$ for (2.8).

Let π be a probability density function on $\mathbb{R}^d$ and $L^2(\pi)$ denote the Hilbert space of complex valued functions that are square integrable with respect to π:

$$\mathbf{L}^2(\pi) = \{f : \mathbb{R}^d \to \mathbf{C} | \int f(\tau)\overline{f(\tau)}\pi(\tau)d\tau < \infty\}. \tag{2.9}$$

If the moment functions are based on the (conditional) CF as it is the case here, $h_t(.,\theta)$ belongs to $L^2(\pi)$ for all $\theta \in \Theta$ and for any probability density function π because $|h_t(.,\theta)|^2 \leq 2$ for all $\theta \in \Theta$. We can thus define the scalar product $\langle .,. \rangle$ on $\mathbf{L}^2(\pi) \times \mathbf{L}^2(\pi)$ in the following way:

$$\langle f, g \rangle = \int f(\tau)\overline{g(\tau)}\pi(\tau)d\tau. \tag{2.10}$$

As shown by CaFl (2000) and CCFG (2007), the optimal CGMM estimator should be defined as:

$$\widehat{\theta} = \arg\min_\theta \left\langle K^{-1}\widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta) \right\rangle. \tag{2.11}$$

where K^{-1} is the inverse of the asymptotic covariance operator K associated with the moment conditions and $\widehat{h}_T(.,\theta)$ is defined in (2.6). In fact, K is the integral operator defined by:

$$Kf\,(\tau_1) = \int_{-\infty}^{\infty} k(\tau_1,\tau) f\,(\tau)\,\pi\,(\tau)\,d\tau \tag{2.12}$$

for any function $f \in L^2(\pi)$, where the kernel $k(\tau_1, \tau)$ is given by:

$$k(\tau_1, \tau_2) = E\left[h_t(\tau_1, \theta)\overline{h_t(\tau_2, \theta)}\right], \tag{2.13}$$

Unfortunately, $K^{-1}g$ does not exist for all g. Moreover, $K^{-1}g$ when it exists is highly unstable to small perturbations in g. To circumvent this difficulty, one may replace K^{-1} by

$$K_\alpha^{-1} = (K^2 + \alpha I)^{-1}K$$

in the objective function. The asymptotic optimality is achieved for the resulting estimator by letting α go to zero as T goes to infinity, provided that there exists $g(\tau, \theta)$ such that $Kg(\tau, \theta) = E\left[\widehat{h}_T(\tau, \theta)\right]$.

With a sample of size T and a consistent first step estimator $\widehat{\theta}^1$ in hand, we can estimate the kernel of K by

$$k_T(\tau_1, \tau_2, \widehat{\theta}^1) = \frac{1}{T}\sum_{t=1}^{T} h_t(\tau_1, \widehat{\theta}^1)\overline{h_t(\tau_2, \widehat{\theta}^1)}. \tag{2.14}$$

In the specific case of IID data, an estimator of the kernel that does not use a first step estimator is available and is defined as

$$k_T(\tau_1, \tau_2) = \frac{1}{T}\sum_{t=1}^{T}\left(e^{i\tau_1' x_t} - \widehat{\varphi}_T(\tau_1)\right)\overline{\left(e^{i\tau_2' x_t} - \widehat{\varphi}_T(\tau_2)\right)}. \tag{2.15}$$

where $\widehat{\varphi}_T(\tau_1) = \frac{1}{T}\sum_{t=1}^{T} e^{i\tau_1' x_t}$. Let K_T be the empirical operator with kernel $k_T(\tau_1, \tau_2, \widehat{\theta}^1)$ or $k_T(\tau_1, \tau_2)$. Then K^{-1} can be estimated by $K_{\alpha T}^{-1} = \left(K_T^2 + \alpha I\right)^{-1} K_T$ and the feasible CGMM estimator is given by:

$$\widehat{\theta}_T(\alpha; \theta_0) = \arg\min_\theta \widehat{Q}_T(\alpha, \theta). \tag{2.16}$$

where $\widehat{Q}_T(\alpha, \theta) = \left\langle K_{\alpha T}^{-1}\widehat{h}_T(., \theta), \widehat{h}_T(., \theta)\right\rangle$.

Note that $K_{\alpha T}^{-1}$ is a regularized inverse of K_T. It has the property that for any function f in the range of K and any $\sqrt{T}$-consistent estimator $\widehat{f}_T$ of f, the function $K_{\alpha T}^{-1}\widehat{f}_T$ converges to $K^{-1}f$ as T goes to infinity and the regularization parameter α goes to zero. Our expression of $K_{\alpha T}^{-1}$ uses Tikhonov regularization, also called ridge regularization. Other forms of regularization could have been used (see Carrasco, Florens, and Renault (2007), Carrasco

(2008)). The advantage of Tikhonov regularization is its simplicity. A simple expression of the objective function $\widehat{Q}_T(\alpha, \theta)$ in matrix form is given in CCFG (Section 3.3.). In the sequel, the following regularity conditions are assumed.

Assumption 1: The probability density function π involved in the scalar product defined by equation (2.10) is strictly positive on $\mathbb{R}^d$ and admits all its moments.

Assumption 2: The equation

$$E^{\theta_0}[h_t(\tau, \theta)] = 0 \text{ for all } \tau \in \mathbb{R}^d, \pi - \text{almost everywhere,}$$

where E^{θ_0} denotes the expectation with respect to the data generating process for $\theta = \theta_0$, has a unique solution θ_0 which is an interior point of a compact set $\boldsymbol{\Theta}$.

Assumption 3: $h_t(\tau, \theta)$ is three time continuously differentiable with respect to θ.

Assumption 4: $E^{\theta_0}[h_T(., \theta)] \in \Phi_\beta$ (for some $\beta \geq 1$) for all $\theta \in \Theta$, and the first two derivatives of $E^{\theta_0}[h_T(., \theta)]$ with respect to θ belong to Φ_β (for the same $\beta \geq 1$) in a neighborhood of θ_0, where

$$\Phi_\beta = \left\{ f \in L^2(\pi) \text{ such that } \left\| K^{-\beta} f \right\| < \infty \right\} \tag{2.17}$$

Assumption 5: The random variable x_t is stationary Markov and satisfies $x_t = r(x_{t-1}, \theta_0, \varepsilon_t)$ where $r(x_{t-1}, \theta_0, \varepsilon_t)$ is three times continuously differentiable with respect to θ_0 and ε_t is a IID white noise whose distribution is known and does not depend on θ_0.

Assumption 1 and 2 are standard and have been used in CaFl (2000). Assumption 3 ensures some smoothness properties for $\widehat{\theta}_T(\alpha; \theta_0)$. The largest real β such that $f \in \Phi_\beta$ in Assumption 4 is the level of regularity of f with respect to K. The larger β is, the better f is approximated by the eigenfunctions of K. Because $Kf(.)$ involve $\dim(\tau)$ integrations, β may be affected by both the dimensionality of the index τ and the smoothness of f. But CCFG show that in the context of CF-based CGMM, we always have $\beta \geq 1$.

Assumption 5 implies that the data can be simulated upon knowing how to draw from the distribution of ε_t. It is satisfied for all random variables that can be written as a location parameter plus a scale parameter time a standardized representative of the family of distribution. Examples include the exponential family and the stable distribution. The IID case is a special case of Assumption 5 where $r(x_{t-1}, \theta_0, \varepsilon_t)$ takes the simpler form $r(\theta_0, \varepsilon_t)$. Further discussions on this type of model can be found in Gourieroux, Monfort, and Renault (1993) in the indirect inference context. Note that the function $r(x_{t-1}, \theta_0, \varepsilon_t)$ may not be available in analytical form. In particular, the relation $x_t = r(x_{t-1}, \theta_0, \varepsilon_t)$ can be the numerical solution of a general equilibrium asset pricing model as in Duffie and Singleton (1993). However, the differentiability of $r(x_{t-1}, \theta_0, \varepsilon_t)$ is crucial for the proof of the optimality of our regularization parameter selection procedure.

Propositions 3.2 and 4.1 of CCFG show that under Assumptions 1 to 5, the CGMM estimator defined by (2.16) is consistent and asymptotically normal and its asymptotic variance reaches the Cramer-Rao bound. Interestingly, the asymptotic distribution of the CGMM estimator does not depend on the weighting function $\pi(.)$ as stated in the following theorem.

Theorem 1 *Under Assumptions 1 to 5, the CGMM estimator defined by (2.11) is consistent and satisfies:*

$$T^{1/2}\left(\widehat{\theta}_T(\alpha;\theta_0) - \theta_0\right) \xrightarrow{L} N(0, I_{\theta_0}^{-1}).$$

as T and $\alpha T^{1/2}$ go to infinity and α goes to zero, where $I_{\theta_0}^{-1}$ denotes the inverse of the Fisher Information Matrix.

Because K is self-adjoint, the ideal objective function of the CGMM can be equivalently defined as

$$\left\langle K^{-1/2}\widehat{h}_T(.,\theta), K^{-1/2}\widehat{h}_T(.,\theta)\right\rangle.$$

In this case, the feasible objective function of the CGMM would be

$$\widetilde{Q}_T(\alpha,\theta) = \left\langle K_{\alpha T}^{-1/2}\widehat{h}_T(.,\theta), K_{\alpha T}^{-1/2}\widehat{h}_T(.,\theta)\right\rangle$$

where $K_{\alpha T}^{-1/2} = (K_T^2 + \alpha I)^{-1/2} K_T^{1/2}$. We have chosen to use (2.11) in this paper because it simplifies some proofs. One advantage of $\widetilde{Q}_T(\alpha,\theta)$ is that it is well defined at the limit for $\beta \geq 1/2$ in Assumption 4, a less restrictive condition than $\beta \geq 1$ (see Appendix A for some basic properties of the operator K). But $\widetilde{Q}_T(\alpha,\theta)$ coincides with $\widehat{Q}_T(\alpha,\theta)$ when $\beta \geq 1$.

In what follows, we investigate how the higher order properties of $\widehat{\theta}_T(\alpha;\theta_0)$ defined in (2.16) are affected by α.

2.3 Stochastic expansion of the CGMM estimator

Many studies have investigated the higher order properties of various types of estimators. For the GMM type estimators, example of such studies include Rothenberg (1983, 1984), Koenker et al. (1994), Rilstone et al. (1996) and Newey and Smith (2004). Other examples in the linear simultaneous equation framework can be found in Nagar (1959), or more recently in Buse (1992) and Donald and Newey (2001). The approach we follow is similar to Nagar (1959) and Newey and Smith (2004) in that the expression we derive for the MSE is based on the leading terms in an expansion of the estimator.

The terms of the expansion depend on a combination of α and T. Two difficulties arise when analyzing the terms of the expansion. First, when the rate of decrease of α as a function of T is unknown, it is not always possible to order the terms of the expansion in monotonically decreasing order. A second difficulty lies in a result that dramatically differs from the case with a finite number of moment conditions. Indeed, when the number of moment conditions is finite, the quadratic form

$$T\widehat{h}_T(\theta_0)' K^{-1}\widehat{h}_T(\theta_0),$$

follows asymptotically a chi-square distribution with degrees of freedom given by the number of moment conditions, hence it is $O_p(1)$. However, the quadratic form $\left\|K^{-1/2}\sqrt{T}\widehat{h}_T(\theta_0)\right\|^2$ is not well defined in presence of a continuum of moment conditions. Also, $\left\|K_\alpha^{-1/2}\sqrt{T}\widehat{h}_T(\theta_0)\right\|^2$ is well defined but diverges as T goes to infinity and α goes to zero. We are able to state the following rate for the norm of $K_\alpha^{-1/2}\sqrt{T}\widehat{h}_T(\theta_0)$:

$$\left\|K_\alpha^{-1/2}\sqrt{T}\widehat{h}_T(\theta_0)\right\| = O_p\left(\alpha^{-1/4}\right).$$

Indeed, we have

$$\begin{aligned} & \left\|K_\alpha^{-1/2}\sqrt{T}\widehat{h}_T(\theta_0)\right\| \\ \leq\ & \underbrace{\left\|\left(K^2+\alpha I\right)^{-1/4}\right\|}_{\leq\alpha^{-1/4}}\underbrace{\left\|\left(K^2+\alpha I\right)^{-1/4}K^{1/2}\right\|}_{\leq 1}\underbrace{\left\|\sqrt{T}\widehat{h}_T(\theta_0)\right\|}_{=O_p(1)} \qquad (2.18) \\ =\ & O_p\left(\alpha^{-1/4}\right). \end{aligned}$$

The expansion we derived for $\widehat{\theta}_T(\alpha;\theta_0)-\theta_0$ is of the same form for both the IID and Markov case. Namely, we have:

$$\widehat{\theta}_T(\alpha;\theta_0)-\theta_0=\Delta_1+\Delta_2+\Delta_3+o_p\left(\alpha^{-1}T^{-1}\right)+o_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right) \tag{2.19}$$

where

$$\begin{aligned}
\Delta_1 &= O_p\left(T^{-1/2}\right);\\
\Delta_2 &= O_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right);\\
\Delta_3 &= O_p\left(\alpha^{-1}T^{-1}\right).
\end{aligned}$$

Appendix B provides details about the above expansion whose validity is ensured by the consistency result stated by Theorem 1. The following result gives an insight on the behavior of the higher order MSE of the CGMM estimator.

Theorem 2 *Assume that Assumptions* 1 *to* 5 *hold. Then we have the followings:*

(i) The MSE matrix of $\widehat{\theta}_T(\alpha;\theta_0)$ is decomposed as the sum of the squared bias and variance:

$$TE\left[\left(\widehat{\theta}_T(\alpha;\theta_0)-\theta_0\right)\left(\widehat{\theta}_T(\alpha;\theta_0)-\theta_0\right)'\right]=TBias*Bias'+TVar$$

where

$$\begin{aligned}
TBias*Bias' &= O\left(\alpha^{-2}T^{-1}\right),\\
TVar &= I_{\theta_0}^{-1}+O\left(\alpha^{\min\left(2,\frac{2\beta-1}{2}\right)}\right)+O\left(\alpha^{-1}T^{-1/2}\right).
\end{aligned}$$

(ii) The α that minimizes the MSE of $\widehat{\theta}_T(\alpha;\theta_0)$, noted $\alpha_T(\theta_0)$, satisfies $\alpha_T^2(\theta_0)T\to\infty$ and

$$\alpha_T(\theta_0)=O\left(T^{-\max\left(\frac{1}{6},\frac{1}{2\beta+1}\right)}\right).$$

Remarks.

1. We have the usual trade-off between a term that is decreasing in α and another that is increasing in α. Interestingly here, the bias term is dominated by two variance terms whose rates are equated to obtain the optimal regularization parameter. The same happens for the

Limited Information Maximum Likelihood estimator for which the bias is also dominated by variance terms (see Donald and Newey, 2001).

2. The rate for the $O\left(\alpha^{\min\left(2,\frac{2\beta-1}{2}\right)}\right)$ variance term does not improve for $\beta > 3$. This is due to a property of Tikhonov regularization which is well documented in the statistics literature on inverse problems, see Carrasco, Florens and Renault (2007). The use of another regularization such as spectral cut-off or Landweber Fridman would permit to improve the rate of convergence for large values of β, however it would come at the cost of a greater complexity in the proofs (in the spectral cut-off, we would lose the differentiability of the estimator with respect to α).

3. Our expansion is consistent with the condition of Theorem 1, since the optimal regularization parameter α_T satisfies $\alpha_T^2(\theta_0)\,T \to \infty$.

4. The rate for α_T does not coincide exactly with $T^{-\max\left(\frac{1}{6},\frac{1}{2\beta+1}\right)}$, unless the derived big O rates in the expression of the MSE are based on equivalent expressions rather than on upper bounds.

5. It follows from Theorem 2 that the optimal regularization parameter α_T, that is, the α_T that minimizes the MSE of the CGMM estimator is of the form

$$\alpha_T(\theta_0) = c(\theta_0)\,T^{-g(\beta)}, \tag{2.20}$$

for some positive function $c(\theta_0)$ that does not depend on T and a positive function $g(\beta)$ that satisfies $\max\left(\frac{1}{6},\frac{1}{2\beta+1}\right) \leq g(\beta) < 1/2$.

An expression of the form (2.20) is often used as starting point for optimal bandwidth selection in nonparametric density estimation. Examples in the semiparametric context include Linton (2002) or Jacho-Chavez (2007). In what follows, we approximate the MSE of $\widehat{\theta}_T(\alpha_T;\theta_0)$ in order to provide a data-driven method for selecting $\alpha_T(\theta_0)$.

2.4 Optimal selection of the regularization parameter

In Section 3, we showed that the optimal rate regularization parameter satisfies $\alpha_T(\theta_0) = c(\theta_0)\,T^{-g(\beta)}$, where $c > 0$ and $g(\beta)$ is some positive and bounded function of β. But the constants $c(\theta_0)$ and β and the function $g(\beta)$ are unknown in general so that α_T needs to be estimated in practice. The question of the selection of the regularization parameter α_T in our CGMM context is analogous to that of selecting the optimal bandwidth in the kernel density estimation. Ideally, it should be selected so as to minimize the mean square error of $\widehat{\theta}_T(\alpha;\theta_0)$ as a function of α for a given sample of size T:

$$\alpha_T(\theta_0) = \arg \min_{\alpha\in[0,1]} \Sigma_T(\alpha;\theta_0). \tag{2.21}$$

where $\Sigma_T(\alpha;\theta_0)$ is defined as the trace of the MSE matrix, that is:

$$\Sigma_T(\alpha;\theta_0) = TE\left[\left(\widehat{\theta}_T(\alpha;\theta_0) - \theta_0\right)'\left(\widehat{\theta}_T(\alpha;\theta_0) - \theta_0\right)\right]. \tag{2.22}$$

From now on, we shall use the term MSE to refer to $\Sigma_T(\alpha;\theta_0)$.

Note that $\Sigma_T(\theta_0)$ is a function of θ_0 according to (2.22). Finding $\alpha_T(\theta_0)$ gives rise to at least two problems: (i) θ_0 is unknown, and (ii) even if θ_0 were given, the exact finite sample distribution of $\widehat{\theta}_T(\alpha;\theta_0) - \theta_0$ is not known so that $\Sigma_T(\alpha;\theta_0)$ can only be approximated. It should be stressed that in this paper, we adopted a fully parametric approach from the beginning where the model is completely specified. Indeed, it would not be possible to obtain MLE efficiency otherwise. Hence, the model can be simulated. We are going to exploit this feature to approximate the unknown MSE by simulations. The estimation of $\Sigma_T(\alpha;\theta_0)$ will rely on a first step consistent estimator $\widehat{\theta}^1_T$ of θ_0 and a large number of independently simulated samples, say $X_T^{(j)}(\widehat{\theta}^1)$ for $j = 1, 2, ..., M$.

The first step estimator $\widehat{\theta}^1_T$ can be either a classical GMM estimator that uses a finite number of moment conditions built from a discretization scheme, or a CGMM estimator obtained by replacing the covariance operator with the identity operator. The samples $X_T^{(j)}(\widehat{\theta}^1_T)$ can be generated by path simulation (see Gouriéroux and Monfort (1996)). From assumption 5, we have $x_t = r\left(x_{t-1}, \theta, \varepsilon_t\right)$. One generates MT IID random realizations drawn from the known distribution of the errors, denoted $\varepsilon_t^{(j)}$. For arbitrary starting values $x_0^{(j)}$, one can generates M time-series $x_t^{(j)} = r\left(x_{t-1}^{(j)}, \widehat{\theta}^1_T, \varepsilon_t^{(j)}\right)$ of size T for $j = 1, 2, ..., M$. To avoid transient effects, it is usual to simulate more data than needed and discard the first observations, or alternatively, to condition the simulation on the first observations of the actual sample.

We consider two different ways to compute the approximate MSE, leading to two different ways to estimate $\alpha_T(\theta_0)$. The first method uses the analytical approximation of $\widehat{\theta}-\theta_0$ given in the preceding section, and the second method uses a large number of IID copies of a consistent approximation of $\widehat{\theta}_T(\alpha;\theta_0)-\theta_0$. We discuss each of these approaches below.

2.4.1 Estimating the MSE using its higher order expansion

This first approach consists in approximating the unknown $\Sigma_T(\alpha;\theta_0)$ by $\widehat{\Sigma}^A_{TM}(\alpha,\widehat{\theta}^1)$, a Monte Carlo estimation of the MSE based on the higher order expansion of $\widehat{\theta}_T(\alpha;\theta_0)-\theta_0$ given in (2.19):

$$\Delta_T = \Delta_1+\Delta_2+\Delta_3$$

The expressions of $\Delta_j, j=1,2,3$ depend on both deterministic and random quantities. The deterministic quantities are the true parameter θ_0, the covariance operator K, the probability limit of the gradient of the moment function $G(\tau,\theta_0)$ and the regularization parameter α. The random quantities are the moment function $\widehat{h}_T(\tau,\theta_0)$ and the empirical covariance operator K_T. We can write

$$\Delta_T = \Delta\left(K(\theta_0), G(.,\theta_0), X_T(\theta_0)\right)$$

where Δ depends on the sample $X_T(\theta_0)$ through $\widehat{h}_T(\tau,\theta_0)$ and K_T.

To implement the current approach, the first step consists in replacing θ_0 by a first step consistent estimator $\widehat{\theta}^1$ computed from the actual data. Secondly, one uses $\widehat{\theta}^1$ to simulate a single very large sample (e.g. 100 times larger than T whenever possible). That large sample is used to compute highly accurate estimations of K and $G(\tau,\theta_0)$ denoted $\widetilde{K}(\widehat{\theta}^1)$ and $\widetilde{G}(.,\widehat{\theta}^1)$ respectively[2]. If closed form expressions were available for all the terms involved in $E(\Delta'\Delta)$, it would be possible to estimate Δ in one shot using $\widetilde{K}(\widehat{\theta}^1)$, $\widetilde{G}(.,\widehat{\theta}^1)$ and the actual data $X_T(\theta_0)$. Unfortunately, $Cov(\Delta_1,\Delta_3)$ cannot be computed explicitly even though Δ_1 and Δ_3 are given in closed form. We shall thus resort to simulations.

Let $X_T^{(j)}(\theta_0), j=1,...,M$ be M samples of size T simulated using $\widehat{\theta}^1$ and define

$$\Delta_T^{(j)} = \Delta\left(\widetilde{K}(\widehat{\theta}^1), \widetilde{G}(.,\widehat{\theta}^1), X_T^{(j)}(\theta_0)\right).$$

[2]We choose the size of the simulated sample large enough to ensure that $\widetilde{K}^{-1}(\widehat{\theta}^1)$ can be computed without regularizing the numerical approximation of $\widetilde{K}(\widehat{\theta}^1)$. See Appendix D for the numerical algorithm.

Finally, $\widehat{\Sigma}^{A}_{TM}(\alpha,\widehat{\theta}^{1})$ is given by

$$\widehat{\Sigma}^{A}_{TM}(\alpha,\widehat{\theta}^{1}) = \frac{T}{M}\sum_{j=1}^{M}\Delta_T^{(j)\prime}\Delta_T^{(j)}, \tag{2.23}$$

where the superscript A stands for "analytical approximation", the subscripts T and M indicates respectively the sample size and the number of Monte Carlo replications. Our estimation of the optimal α is thus:

$$\widehat{\alpha}_{TM}\left(\widehat{\theta}^{1}\right) = \underset{\alpha\in[0,1]}{\arg\min}\widehat{\Sigma}^{A}_{TM}(\alpha,\widehat{\theta}^{1}). \tag{2.24}$$

The Law of Large Numbers ensures that for sufficiently large M, $\widehat{\Sigma}^{A}_{TM}(\alpha,\widehat{\theta}^{1}) \approx \Sigma_T(\alpha,\widehat{\theta}^{1})$ so that if $\widehat{\theta}^{1}$ is close enough to θ_0, we also obtain $\Sigma_T(\alpha,\widehat{\theta}^{1}) \approx \Sigma_T(\alpha,\theta_0)$. This gives an intuition of why $\widehat{\alpha}_{TM}\left(\widehat{\theta}^{1}\right)$ can be a good estimator of $\alpha_T(\theta_0)$. This procedure can be fast as it does not require a parameter estimation at each Monte Carlo replication. On the other hand, it may need very large sample sizes to obtain accurate estimates of $\widehat{\theta}^{1}$, K, and G.

2.4.2 Estimating the MSE by standard Monte Carlo

From a large number of independently simulated samples of size T, $X_T^j(\theta_0), j = 1, 2, ..., M$, we can compute M IID copies of the CGMM estimator $\widehat{\theta}_T^j(\alpha;\theta_0)$ for any fixed α. A natural estimator of the MSE is then:

$$\widehat{\Sigma}^{MC}_{TM}(\alpha;\theta_0) = \frac{T}{M}\sum_{j=1}^{M}\left(\widehat{\theta}_T^j(\alpha;\theta_0)-\theta_0\right)'\left(\widehat{\theta}_T^j(\alpha;\theta_0)-\theta_0\right). \tag{2.25}$$

The above estimator of the MSE would thus naturally yield an estimator of the optimal α of the form:

$$\widehat{\alpha}_{TM}(\theta_0) = \underset{\alpha\in[0,1]}{\arg\min}\widehat{\Sigma}^{MC}_{TM}(\alpha,\theta_0), \tag{2.26}$$

where the superscript MC stands for "Monte Carlo approximation" and the subscript M indicates the number of Monte Carlo replications.

For a sufficiently large value of M, the Law of Large Numbers ensures that $\widehat{\Sigma}^{MC}_{TM}(\alpha;\theta_0)$ converges to its expectation $\Sigma_T(\alpha;\theta_0)$. But as θ_0 is not known, a feasible Monte Carlo approach simply consists in replacing θ_0 with a consistent first step estimator $\widehat{\theta}^1$ in the simulation scheme, that is, choosing the optimal regularization parameter according to

$$\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right) = \underset{\alpha\in[0,1]}{\arg\min}\widehat{\Sigma}_{TM}^{MC}\left(\alpha,\widehat{\theta}^1\right). \tag{2.27}$$

It is important to note that $\widehat{\Sigma}_{TM}^{MC}\left(\alpha,\widehat{\theta}^1\right)$ is simulated conditional on $\widehat{\theta}^1$. As a consequence, $\lim_{M\to\infty}\widehat{\Sigma}_{TM}^{MC}\left(\alpha,\widehat{\theta}^1\right) = \Sigma_T\left(\alpha,\widehat{\theta}^1\right)$. Minimizing this limiting MSE would yield the true value of the optimal α if the true value of the parameter of interest was the point estimate $\widehat{\theta}^1$:

$$\alpha_T\left(\widehat{\theta}^1\right) = \underset{\alpha\in[0,1]}{\arg\min}\Sigma_T\left(\alpha,\widehat{\theta}^1\right). \tag{2.28}$$

As such, $\alpha_T\left(\widehat{\theta}^1\right)$ is a deterministic and continuous function of a stochastic argument, while $\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right)$ is doubly random, being a stochastic function of a stochastic argument.

Given this construction and $\alpha_T\left(\widehat{\theta}^1\right) = c\left(\widehat{\theta}^1\right)T^{-g(\beta)}$, it is easy to show that $c(.)$ is a continuous function using the Maximum Theorem (see Lemma 17 in Appendix). Although $\alpha_T\left(\widehat{\theta}^1\right)$ is not feasible, its properties will be a key ingredient in establishing the consistency of its feasible counterpart $\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right)$.

This second approach is computationally demanding because it requires computing a CGMM estimation at each of the M Monte Carlo iterations. But it has the main advantage of avoiding the cumbersome calculations involved in the higher order expansion of the MSE. In the next subsection, we prove the optimality of $\widehat{\alpha}_{TM}\left(\widehat{\theta}_T^1\right)$.

2.4.3 Optimality of the data-driven selection of the regularization parameter

We focus below on the optimality of the second approach. We are able to prove the following result.

Theorem 3 *Let $\widehat{\theta}^1$ be $\sqrt{T}-$consistent estimator of θ_0. Then under Assumptions 1 to 5, we have*

$\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1$ converges in probability to zero as T goes to infinity.

In Theorem 3, the function $\alpha_T(.)$ is deterministic and continuous but the argument $\widehat{\theta}^1$ is stochastic. Note that as T goes to infinity, $\widehat{\theta}^1$ gets closer and closer to θ_0, but at the same time $\alpha_T(\theta_0)$ converges to zero at some rate that depends on T. This prevents us from claiming without caution that $\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1 = o_p(1)$ since the denominator is not bounded away from zero. To be able to write this, one needs to ensure that the residual $\alpha_T(\theta_0) - \alpha_T(\widehat{\theta}^1)$ converges to zero faster than $\alpha_T(\theta_0)$ itself.

The next theorem gives the rate of convergence of $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)}$.

Theorem 4 *Under assumptions 1 to 5, the term* $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)} - 1$ *converges in probability to zero at rate* $M^{-1/2}$ *as* M *goes to infinity and* T *is fixed.*

In Theorem 4, $\widehat{\alpha}_{TM}(\theta_0)$ is the minimum of the empirical MSE simulated with the true θ_0. In the proof, we first show that the conditions of the uniform convergence in probability of the empirical MSE are satisfied. Next, we use the Theorem 2.1 of Newey and McFadden (1994) and the fact that $\alpha_T(\theta_0)$ is bounded away from zero for finite T to establish the consistency of $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)}$. In the next theorem, we revisit the results of Theorem 4 when θ_0 is replaced by a $\sqrt{T}-$consistent estimator $\widehat{\theta}^1$.

Theorem 5 *Let* $\widehat{\theta}^1$ *be a* $\sqrt{T}-$*consistent estimator of* θ_0. *Then under assumptions 1 to 5, we have:* $\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1 = O_p(T^{-1/2}) + O_p(M^{-1/2})$ *as* M *goes to infinity first and* T *goes to infinity second.*

The result of Theorem 5 is obtained using a sequential limit in M and T. Such sequential approach has been used in other fields of econometrics, such as panel data. It is also used implicitly in the theoretical analysis of bootstrap. Theorem 5 implies that $\widehat{\alpha}_{TM}(\widehat{\theta}^1)$ benefits from an increase in both M and T. In practice, one clearly has more freedom in setting M than in setting T. For a given sample size T, setting $M \approx T$ permits to achieve the best rate for $\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\theta_0)}$ at the lowest computing time. In fact, increasing M significantly increases the precision of the estimator $\widehat{\alpha}_{TM}(\widehat{\theta}^1)$ only in the region $M \leq T$. When M becomes larger than T, the $O_p(M^{-1/2})$ term in the expression of $\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\theta_0)}$ becomes a higher order term so that the rate $\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1 = O_p(T^{-1/2})$ does not improve.

Overall, our selection procedure of α_T is optimal and adaptive as the a priori knowledge of the regularity of the moment function and its derivatives (the value of β) is not needed.

Using these results, we obtain a fully feasible CGMM estimator of the form $\widehat{\theta}_T(\widehat{\alpha};\theta_0)$. In what follows, we provide some Monte Carlo simulations that illustrate the relevance of tuning the regularization parameter in the CGMM estimation.

2.5 Monte Carlo Simulations

Our focus in this Monte Carlo experiment is to illustrate the properties of the estimator $\widehat{\alpha}_{TM}(\widehat{\theta}^1)$ proposed in the previous section. For this purpose, we select a model that is at the same time simple and widely used in the empirical literature. We present below the model and the design of the experiment.

2.5.1 The simulation design

For this study, we consider a model of convolution of a normal distribution with an exponential distribution. This type of mixture is commonly encountered in the stochastic cost frontier literature. A cost frontier typically describes the minimum level of cost involved in the production of a given level of output at a given level of prices. The following specification has been used in Meeusen and van den Broeck (1977):

$$y_t = X_t\beta_0 + u_t + v_t, \qquad t = 1, 2, ..., T. \tag{2.29}$$

where y_t is the log cost for plant t, X_t is a vector of independent variables, u_t and v_t are independent, $u_t \overset{IID}{\sim} N(0,\sigma_0^2)$ is a term accounting for measurement error, $v_t \overset{IID}{\sim} Exp(\lambda_0)$ is a non negative error term measuring the plant inefficiency, and $Exp(\lambda_0)$ denotes the exponential random variable with mean $1/\lambda_0$. Many authors including van den Broeck, Koop, Osiewalski and Steel (1994) provide a Bayesian treatment to this model. We illustrate below an alternative CGMM approach.

In our simulations, we shall focus on the following simplified version of the model:

$$y_t = u_t + v_t,\ t = 1, 2, ..., T. \tag{2.30}$$

The moment function for this model is:

$$h_t\left(\tau, \sigma^2, \lambda\right) = \exp(i\tau y_t) - \frac{\lambda \exp\left(-\tau^2\sigma^2/2\right)}{\lambda - i\tau}. \tag{2.31}$$

In the sequel, we simulate M independent samples of size T and estimate $\alpha_T(\theta_0)$ by standard Monte Carlo by varying α on the finite grid:

$$\begin{aligned}&[5\times 10^{-8}; 10^{-7}; 5\times 10^{-7}; 10^{-6}; 5\times 10^{-6}; 10^{-5}; 5\times 10^{-5}; 7\times 10^{-5}; 10^{-4}; \\ &3\times 10^{-4}; 5\times 10^{-4}; 7\times 10^{-4}; 10^{-3}; 3\times 10^{-3}; 5\times 10^{-3}; 7\times 10^{-3}; 10^{-2}].\end{aligned} \tag{2.32}$$

The bounds of the grid must be adapted to the particular application under consideration. We choose a Gaussian form for the weighting function $\pi()$ embedded in the objective function of the CGMM, which allows us to use Gauss-Hermite quadrature points to compute the integrals (see Appendix D). We use a common random numbers simulation scheme in order to facilitate the comparision of the MSE accross the values of α. Finally, we consider different sample sizes T and numbers of replications M ($T, M =$ 100, 250, 500 and 1000).

2.5.2 The simulation results

Table 1 shows the values of $\widehat{\alpha}_{TM}$ obtained by using the procedure described in the previous subsection. It is seen that for each sample size T, the estimated values differ as M varies. Because the empirical criterion $\widehat{\Sigma}_{TM}^{MC}(\alpha;\theta_0)$ gets closer and closer to the targeted criterion $\Sigma_T(\alpha;\theta_0)$ as M increases, we can claim that our best approximation of $\alpha_T(\theta_0)$ in this table is given by $\widehat{\alpha}_{T,1000}$ for each T. Hence using $\widehat{\alpha}_{T,1000}$ as a benchmark, we see that the precision of $\widehat{\alpha}_{T,M}$ increases as M increases.

M	T=100	T=250	T=500	T=1000
100	0.0070 (9.59)	0.0070 (7.35)	0.0003 (7.72)	0.0005 (8.03)
250	0.0070 (9.67)	0.0050 (8.15)	0.0003 (8.05)	0.0003 (8.18)
500	0.0070 (10.55)	0.0030 (9.08)	0.0003 (8.38)	0.0007 (8.90)
1000	0.0050 (10.62)	0.0030 (8.94)	0.0005 (8.58)	0.0007 (9.02)

Table 2.1: Estimation of the optimal regularisation parameter for different T and M.

The numbers in brackets are T times the MSE of $\widehat{\theta}_T\left(\widehat{\alpha}_{TM}\right)$.

Also, it is seen that $\widehat{\alpha}_{T,M}$ decreases as T increases from 100 to 250 and from 250 to 500 while the difference between $\widehat{\alpha}_{500,M}$ and $\widehat{\alpha}_{1000,M}$ is negligible for all M. The fact that $\widehat{\alpha}_{500,1000}$ is a bit lower than $\widehat{\alpha}_{1000,1000}$ should not be controversial because the MSE of $\widehat{\theta}_T(\alpha;\theta_0)$ is very flat around the optimal α, and as we are approximating $\alpha_T(\theta_0)$ from a finite grid, it is easy to slightly overestimate or underestimate the target. This is well illustrated by the following plots of $T \times MSE\left(\widehat{\theta}_T(\widehat{\alpha};\theta_0)\right)$ of the CGMM estimator in the scenarios ($T = 500$, $M = 1000$) and ($T = 1000$, $M = 1000$).

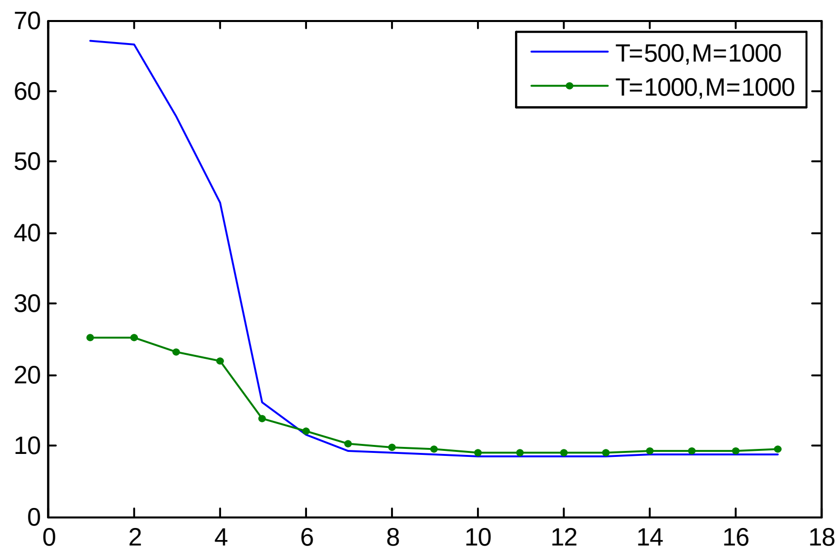


Figure 2.1: MSE of the CGMM estimator

Y-axis: T times the MSE of $\widehat{\theta}_T(\widehat{\alpha};\theta_0)$ for T=500 and T=1000.

X-axis: Rank of α in the grid given by (2.32).

2.6 Conclusion

The goal of this paper is the provide a way to optimally choose the regularization parameter in the CGMM estimation proposed by Carrasco and Florens (2000). To this end, we derive a higher order expansion of the CGMM estimator that sheds light on how the finite sample MSE depends on the regularization parameter α. We propose two selection procedures based on two methods for approximating the MSE. In the first procedure, the approximation relies on its higher-order expansion while in the second procedure, the approximation is based on simulations. The second method is shown to be optimal. The optimal selection of the regularization parameter in the CGMM estimation procedure permits to devise a fully feasible estimator and to establish CGMM as a real competitor to maximum likelihood estimation.

Our simulation-based selection procedure has the advantage to be easily applicable to other estimators, for instance it could be used to select the number of polynomial terms in the efficient method of moments procedure of Gallant and Tauchen (1996).

Bibliography

[1] Ait-Sahalia, Y., Kimmel, R. (2007), "Maximum likelihood estimation of stochastic volatility models", *Journal of Financial Economics* 83 (2007) 413–452.

[2] Buse, A. (1992) "The Bias of Instrumental Variable Estimators", *Econometrica* 60:1, (1992) 173–180.

[3] Carrasco M. (2008) "A regularization approach to the many instruments problem", mimeo, University of Montreal.

[4] Carrasco M., Chernov M., Florens J. P. and Ghysels E. (2007) "Efficient estimation of general dynamic models with a continuum of moment conditions", *Journal of Econometrics,* 140 (2007) 529–573.

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

[6] Carrasco, M. and J. P. Florens (2002) "Efficient GMM Estimation Using the Empirical Characteristic Function", mimeo, CREST, Paris.

[7] Carrasco, M. and J. P. Florens (2004) "On the Asymptotic Efficiency of GMM", mimeo, University of Montréal.

[8] Carrasco, M., J. P. Florens, and E. Renault (2007) "Linear Inverse Problems in Structural Econometrics: Estimation based on spectral decomposition and regularization", in the *Handbook of Econometrics*, Vol. 6, edited by J.J. Heckman and E.E. Leamer.

[9] Chacko, G. and L. Viceira (2003) "Spectral GMM estimation of continuous-time processes", *Journal of Econometrics*, 116, 259-292.

[10] Donald, S. and W. Newey (2001) "Choosing the number of instruments", *Econometrica*, 69, 1161-1191.

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

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

[13] Feuerverger, A. and P. McDunnough (1981a) "On Efficient Inference in Symmetry Stable Laws and Processes", *Csorgo, M., ed. Statistics and Related Topics.* New York: North Holland, pp. 109-122.

[14] Feuerverger, A. and P. McDunnough (1981b) "On some Fourier Methods for Inference", *J. R. Statist. Assoc.* 76:379-387.

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

[16] Feuerverger, A. and Mureika, R. (1977) "The empirical characteristic function and its applications", *The annals of statistics*, Vol. 5, No 1, 88-97.

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

[18] Gouriéroux, C., A. Monfort (1996) "Simulation Based Econometric Methods", CORE Lectures, *Oxford University Press*, New York.

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

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

[21] Jacho-Chavez, D.T (2007) "Optimal Bandwidth Choice for Estimation of Inverse Conditional–Density–Weighted Expectations". mimeo, Indiana University.

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

[23] Koenker, R., Machado, J. A. F., Skeels, C. L. and Welsh, A. H. I. (1994): "Momentary Lapses: Moment Expansions and the Robustness of Minimum Distance Estimators." *Econometric Theory*, 10, 172-197.

[24] Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of stable laws. J. Am. Statist. Assoc. 75:918–928.

[25] Linton, O. (2002) "Edgeworth Approximation for Semiparametric Instrumental Variable and Test Statistics", *Journal of Econometrics*, 106 (2002) 325-368.

[26] Liu, Q. and Pierce, D. A. (1994) "A Note on Gauss-Hermite Quadrature", *Biometrika*, Vol. 81, No. 3, pp. 624-629.

[27] Meeusen, W. and J. van den Broeck (1977) "Efficiency Estimation from Cobb-Douglas Production Function with Composed Error", *International Economic Review*, 18, 435-444.

[28] Nagar, A. L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations" *Econometrica*, 27. 573-595.

[29] Newey, W K and D. McFadden (1994) "Large Sample Estimation and Hypotheses Testing", *Handbook of Econometrics*, Vol IV, ed. by R.F. Engle and D.L. McFadden.

[30] Newey, W K. and Smith, R. J. (2004) "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", *Econometrica* 72 (1), 219–255.

[31] Paulson, A. S., Holcomb W. E. and Leitch, R. A. (1975) "The estimation of the parameters of the stable laws", *Biometrika* (1975) 62(1):163-170.

[32] Rilstone, P., Srivastava, V. K., and Ullah, A. (1996): "The Second-Order Bias and Mean-Squared Error of Nonlinear Estimators," *Journal of Econometrics*, 75,369-395.

[33] Rothenberg, T. J. (1983): "Asymptotic properties of Some Estimators in Structural Models", *Studies in Econometrics, Time Series and Multivariate Statistics*, edited by S. Karlin, T. Amemiya and L. A. Goodman, New York: Academic Press.

[34] Rothenberg, T. J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. D. Intriligator. New York: North-Holland.

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

[36] van den Broeck, J. & Koop, G. & Osiewalski, J. and Steel, M. F. J. (1994) "Stochastic frontier models : A Bayesian perspective," *Journal of Econometrics*, 61(2), 273-303.

[37] Yu, J. (2004) "Empirical Characteristic Function Estimation and Its Applications", *Econometric Reviews*, Vol. 23, No 2, pp. 93-123.

Some basic properties of the covariance operator

For more formal proofs of the results mentioned in this appendix, see Carrasco, Florens and Renault (2007). Let K be the covariance operator defined in (2.12) and (2.13), and $\widehat{h}_t(\tau,\theta)$ the moment function defined in (2.3) and (2.8). Finally, let Φ_β be the subset of $L^2(\pi)$ defined in Assumption 4.

Definition 6 *The range of K denoted $R(K)$ is the set of functions g such that $Kf = g$ for some f in $L^2(\pi)$.*

Proposition 7 *$R(K)$ is a subspace of $L^2(\pi)$.*

Note that the kernel functions $k(s,.)$ and $k(.,r)$ are elements of $L^2(\pi)$ because

$$|k(s,r)|^2 = \left|E\left[h_t(\theta,s)\overline{h_t(\theta,r)}\right]\right|^2 \leq 4,\ \forall (s,r) \in \mathbb{R}^{2p}$$

Thus for any $f \in L^2(\pi)$, we have

$$\begin{aligned} |Kf(s)|^2 &= \left|\int k(s,r)f(r)\pi(r)\,dr\right|^2 \leq \int |k(s,r)f(r)|^2\pi(r)\,dr \\ &\leq 4\int |f(r)|^2\pi(r)\,dr < \infty, \end{aligned}$$

implying

$$\|Kf\|^2 = \int |Kf(s)|^2\pi(s)\,ds < \infty \Rightarrow Kf \in L^2(\pi).$$

Definition 8 *The null space of K denoted $N(K)$ is the set of functions f in $L^2(\pi)$ such that $Kf = 0$.*

The covariance operator K associated with a moment function based on the CF is such that $N(K) = \{0\}$ (See CCFG, 2007, for a proof).

Definition 9 *ϕ is an eigenfunction of K associated with eigenvalue μ if and only if $K\phi = \mu\phi$.*

Proposition 10 *Suppose $\mu_1 \geq \mu_2 \geq .... \geq \mu_j \geq ...$ are the eigenvalues of K. Then the sequence $\{\mu_j\}$ satisfies: (i) $\mu_j > 0$ for all j, (ii) $\mu_1 < \infty$ and $\lim\limits_{j\to\infty}\mu_j = 0$.*

Remark. The covariance operator associated with the CF-based moment function is necessarily compact.

Proposition 11 *Every $f \in L^2(\pi)$ can be decomposed as: $f = \sum_{j=1}^{\infty} \langle f, \phi_j \rangle \phi_j$.*

As a consequence, $Kf = \sum_{j=1}^{\infty} \langle f, \phi_j \rangle K\phi_j = \sum_{j=1}^{\infty} \langle f, \phi_j \rangle \mu_j \phi_j$.

Proposition 12 *If $0 < \beta_1 \leq \beta_2$, then $\Phi_{\beta_2} \subset \Phi_{\beta_1}$.*

We recall that Φ_β is the set of functions such that $\left\| K^{-\beta} f \right\| < \infty$. In fact, $f \in R(K^{\beta_2}) \Rightarrow K^{-\beta_2} f$ exist and $\left\| K^{-\beta_2} f \right\|^2 = \sum_{j=1}^{\infty} \mu_j^{-2\beta_2} \left| \langle f, \phi_j \rangle \right|^2 < \infty$. Thus if $f \in R(K^{\beta_2})$, we have:

$$\begin{aligned} \left\| K^{-\beta_1} f \right\|^2 &= \sum_{j=1}^{\infty} \mu_j^{2(\beta_2 - \beta_1)} \mu_j^{-2\beta_2} \left| \langle f, \phi_j \rangle \right|^2 \\ &\leq \mu_1^{2(\beta_2 - \beta_1)} \sum_{j=1}^{\infty} \mu_j^{-2\beta_2} \left| \langle f, \phi_j \rangle \right|^2 < \infty \end{aligned}$$

$\Rightarrow K^{-\beta_1} f$ exist $\Rightarrow f \in R(K^{\beta_1})$. This means $R(K) \subset R(K^{1/2})$ so that the function $K^{-1/2} f$ is defined on a wider subset of $L^2(\pi)$ compared to $K^{-1} f$. When $f \in \Phi_1$, $\langle K^{-1/2} f, K^{-1/2} f \rangle = \langle K^{-1} f, f \rangle$. But when $f \in \Phi_\beta$ for $1/2 \leq \beta < 1$, the quadratic form $\langle K^{-1/2} f, K^{-1/2} f \rangle$ is well defined while $\langle K^{-1} f, f \rangle$ is not.

Expansion of the MSE and proof of Theorems 1 and 2

Preliminary results and proof of Theorem 1

Lemma 13 *Let $K_\alpha^{-1} = (K^2 + \alpha I)^{-1} K$ and assume that $f \in \Phi_\beta$ for some $\beta > 1$. Then as α goes to zero and n goes to infinity, we have:*

$$\begin{aligned} \left\| K_{\alpha T}^{-1} - K_\alpha^{-1} \right\| &= O_p\left(\alpha^{-3/2} T^{-1/2} \right), & (2.33) \\ \left\| \left(K_{\alpha T}^{-1} - K_\alpha^{-1} \right) f \right\| &= O_p\left(\alpha^{-1} T^{-1/2} \right), & (2.34) \\ \left\| \left(K_\alpha^{-1} - K^{-1} \right) f \right\| &= O\left(\alpha^{\min\left(1, \frac{\beta-1}{2}\right)} \right), & (2.35) \\ \left\langle \left(K^{-1} - K_\alpha^{-1} \right) f, f \right\rangle &= O\left(\alpha^{\min\left(1, \frac{2\beta-1}{2}\right)} \right). & (2.36) \end{aligned}$$

Proof of Lemma 13. In the sequel, $\phi_j, j = 1, 2..., \infty$ denote the eigenfunctions of the covariance operator K associated respectively with the eigenvalues $\mu_j, j = 1, 2..., \infty$. We first consider (2.33):

$$\begin{aligned}
\left\|(K_T^2+\alpha I)^{-1}K_T-(K^2+\alpha I)^{-1}K\right\| &\leq \\
\left\|(K_T^2+\alpha I)^{-1}(K_T-K)\right\|+\left\|(K_T^2+\alpha I)^{-1}K-(K^2+\alpha I)^{-1}K\right\| &\leq \\
\underbrace{\left\|(K_T^2+\alpha I)^{-1}\right\|}_{\leq\alpha^{-1}}\underbrace{\|K_T-K)\|}_{=O_p\left(T^{-1/2}\right)}+\left\|\left[(K_T^2+\alpha I)^{-1}-(K^2+\alpha I)^{-1}\right]K\right\| &\leq
\end{aligned}$$

We have:

$$\begin{aligned}
& \left\|\left[(K_T^2+\alpha I)^{-1}-(K^2+\alpha I)^{-1}\right]K\right\| \\
= & \left\|(K_T^2+\alpha I)^{-1}\left(K^2-K_T^2\right)(K^2+\alpha I)^{-1}K\right\| \\
\leq & \underbrace{\left\|(K_T^2+\alpha I)^{-1}\right\|}_{\leq\alpha^{-1}}\underbrace{\left\|\left(K^2-K_T^2\right)\right\|}_{=O_p\left(T^{-1/2}\right)}\underbrace{\left\|(K^2+\alpha I)^{-1/2}\right\|}_{\leq\alpha^{-1/2}}\underbrace{\left\|(K^2+\alpha I)^{-1/2}K\right\|}_{\to 1}
\end{aligned}$$

This proves (2.33).

The difference between (2.33) and (2.34) is that in (2.34) we exploit the fact that $f\in\Phi_\beta$ with $\beta>1$, hence $\|K^{-1}f\|<\infty$. We can rewrite (2.34) as

$$\begin{aligned}
\left\|\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)f\right\| &= \left\|\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)KK^{-1}f\right\| \\
&\leq \left\|\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)K\right\|\left\|K^{-1}f\right\|.
\end{aligned}$$

We have

$$\begin{aligned}
& \left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)K \\
= & (K_T^2+\alpha I)^{-1}K_TK-(K^2+\alpha I)^{-1}K^2 \\
= & (K_T^2+\alpha I)^{-1}(K_T-K)K \qquad (2.37)\\
& +\left[(K_T^2+\alpha I)^{-1}-(K^2+\alpha I)^{-1}\right]K^2. \qquad (2.38)
\end{aligned}$$

The term (2.37) can be bounded in the following manner

$$\begin{aligned}
\left\|(K_T^2+\alpha I)^{-1}(K_T-K)K\right\| &\leq \left\|(K_T^2+\alpha I)^{-1}\right\|\|K_T-K\|\|K\| \\
&= O_p\left(\alpha^{-1}T^{-1/2}\right).
\end{aligned}$$

For the term (2.38), 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

$$\begin{aligned}
& \left\|\left[(K_T^2 + \alpha I)^{-1} - (K^2 + \alpha I)^{-1}\right] K^2\right\| \\
= & \left\|(K_T^2 + \alpha I)^{-1}\left(K^2 - K_T^2\right)(K^2 + \alpha I)^{-1} K^2\right\| \\
\leq & \underbrace{\left\|(K_T^2 + \alpha I)^{-1}\right\|}_{\leq \alpha^{-1}} \underbrace{\left\|K^2 - K_T^2\right\|}_{= O_p\left(T^{-1/2}\right)} \underbrace{\left\|(K^2 + \alpha I)^{-1} K^2\right\|}_{\to 1} \\
= & O_p\left(\alpha^{-1} T^{-1/2}\right).
\end{aligned}$$

This proves (2.34).

Now we turn our attention toward equation (2.35). We can write

$$\begin{aligned}
(K^2 + \alpha I)^{-1} K f - K^{-1} f &= \sum_{j=1}^{\infty}\left[\frac{\mu_j}{\alpha + \mu_j^2} - \frac{1}{\mu_j}\right]\langle f, \phi_j\rangle \phi_j \\
&= \sum_{j=1}^{\infty}\left(\frac{\mu_j^2}{\alpha + \mu_j^2} - 1\right)\frac{\langle f, \phi_j\rangle}{\mu_j}\phi_j.
\end{aligned}$$

We now take the norm:

$$\begin{aligned}
(2.35) &= \left\|(K^2 + \alpha I)^{-1} K f - K^{-1} f\right\| \\
&= \left(\sum_{j=1}^{\infty}\left(\frac{\mu_j^2}{\alpha + \mu_j^2} - 1\right)^2 \frac{\left|\langle f, \phi_j\rangle\right|^2}{\mu_j^2}\right)^{1/2} \\
&= \left(\sum_{j=1}^{\infty}\mu_j^{2\beta-2}\left(\frac{\mu_j^2}{\alpha + \mu_j^2} - 1\right)^2 \frac{\left|\langle f, \phi_j\rangle\right|^2}{\mu_j^{2\beta}}\right)^{1/2} \\
&\leq \left(\sum_{j=1}^{\infty}\frac{\left|\langle f, \phi_j\rangle\right|^2}{\mu_j^{2\beta}}\right)^{1/2} \sup_{1\leq j\leq\infty}\mu_j^{\beta-1}\frac{\alpha}{\alpha + \mu_j^2}.
\end{aligned}$$

Recall that as K is a compact operator, its largest eigenvalue μ_1 is bounded. We need to find an equivalent to

$$\sup_{0\leq\mu\leq\mu_1}\mu^{\beta-1}\frac{\alpha}{\alpha + \mu_j^2} = \sup_{0\leq\lambda\leq\mu_1^2}\lambda^{\frac{\beta-1}{2}}\left(1 - \frac{1}{\alpha/\lambda + 1}\right) \tag{2.39}$$

Case where $1 \leq \beta \leq 3$: We apply another change of variables $x = \alpha/\lambda$:

$$\sup_{x \geq 0} \frac{\alpha^{\beta/2-1/2}}{x^{\beta/2-1/2}} \left(\frac{x}{1+x}\right).$$

We see that an equivalent to (2.39) is $\alpha^{\beta/2-1/2}$ provided that $\frac{1}{x^{\beta/2-1/2}}\left(\frac{x}{1+x}\right)$is bounded on $\mathbb{R}^{+}$. Note that $g(x) \equiv \frac{x^{(3-\beta)/2}}{1+x}$ is continuous and therefore bounded on any interval of $(0, +\infty)$. It goes to 0 at $+\infty$ and its limit at 0 also equals 0 for $1 \leq \beta < 3$. For $\beta = 3$, we have:

$$g(x) \equiv \frac{1}{1+x}.$$

Then $g(x)$ goes to 1 at $x = 0$ and to 0 at $+\infty$.

Case where $\beta > 3$: We rewrite the left hand side of (2.39) as

$$\mu_j^{\beta-1} \frac{\alpha}{\alpha + \mu_j^2} = \alpha \mu_j^{\beta-3} \underbrace{\frac{\mu_j^2}{\alpha + \mu_j^2}}_{\in(0,1)} \leq \alpha \mu_1^{\beta-3} = O(\alpha).$$

To summarize, we have for $f \in \Phi_\beta$:

$$(2.35) = O\left(\alpha^{\min(1, \frac{\beta-1}{2})}\right).$$

Finally, we consider (2.36). We have:

$$\begin{aligned}
(2.36) &= \sum_j \left(\frac{1}{\mu_j} - \frac{\mu_j}{\mu_j^2 + \alpha}\right) \langle f, \phi_j \rangle^2 \\
&= \sum_j \left(1 - \frac{\mu_j^2}{\mu_j^2 + \alpha}\right) \frac{\langle f, \phi_j \rangle^2}{\mu_j} \\
&= \sum_j \mu_j^{2\beta-1} \left(1 - \frac{\mu_j^2}{\mu_j^2 + \alpha}\right) \frac{\langle f, \phi_j \rangle^2}{\mu_j^{2\beta}} \\
&\leq \sum_j \frac{\langle f, \phi_j \rangle^2}{\mu_j^{2\beta}} \sup_{\mu \leq \mu_1} \mu^{2\beta-1} \frac{\alpha}{\mu^2 + \alpha}.
\end{aligned}$$

For $\beta \geq 3/2$, we have:

$$\sup_{\mu \leq \mu_1} \mu^{2\beta-1} \frac{\alpha}{\mu^2 + \alpha} \leq \alpha \mu_1^{2\beta-3} = O(\alpha).$$

For $\beta < 3/2$, we apply the change of variables $x = \alpha/\mu^2$ and obtain

$$\sup_{x \geq 0} \frac{x}{1+x} \left(\frac{\alpha}{x}\right)^{\frac{2\beta-1}{2}} = O\left(\alpha^{\frac{2\beta-1}{2}}\right).$$

due to the fact that $f(x) = \frac{x}{1+x} x^{-\frac{2\beta-1}{2}}$ is bounded on $\mathbb{R}^+$. Finally,

$$(2.36) = O\left(\alpha^{\min(1, \frac{2\beta-1}{2})}\right).$$

■

Lemma 14 *Suppose we have a particular function $f(\theta) \in \Phi_\beta$ for some $\beta > 1$, and a sequence of functions $f_T(\theta) \in \Phi_\beta$ such that $\sup_{\theta \in \Theta} \|f_T(\theta) - f(\theta)\| = O_p(T^{-1/2})$. Then as α goes to zero, we have*

$$\sup_{\theta \in \Theta} \left\| K_{\alpha T}^{-1/2} f_T(\theta) - K^{-1/2} f(\theta) \right\| = O_p(\alpha^{-1} T^{-1/2}) + O\left(\alpha^{\min(1, \frac{\beta-1}{2})}\right).$$

Proof of Lemma 14.

$$\sup_{\theta \in \Theta} \left\| K_{\alpha T}^{-1} f_T(\theta) - K^{-1} f(\theta) \right\| \leq B_1 + B_2,$$

with

$$\begin{aligned} B_1 &= \sup_{\theta \in \Theta} \left\| K_{\alpha T}^{-1} f_T(\theta) - K_{\alpha T}^{-1} f(\theta) \right\|, \\ B_2 &= \sup_{\theta \in \Theta} \left\| \left(K_{\alpha T}^{-1} - K^{-1}\right) f(\theta) \right\|. \end{aligned}$$

We have

$$\begin{aligned} B_1 &\leq \left\| K_{\alpha T}^{-1} \right\| \sup_{\theta \in \Theta} \|f_T(\theta) - f(\theta)\| \\ &\leq \underbrace{\left\| \left(\alpha_T + K_T^2\right)^{-1/2} \right\|}_{\leq \alpha_T^{-1/2}} \underbrace{\left\| \left(\alpha_T + K_T^2\right)^{-1/2} K_T \right\|}_{\to 1} \underbrace{\sup_{\theta \in \Theta} \|f_T(\theta) - f(\theta)\|}_{= O_p(T^{-1/2})} \\ &= O_p(\alpha_T^{-1/2} T^{-1/2}). \end{aligned}$$

On the other hand, Lemma 13 implies that:

$$\begin{aligned}
B_2 &= \left\| \left(K_{\alpha T}^{-1} - K^{-1} \right) f(\theta) \right\| \\
&\leq \left\| \left(K_{\alpha T}^{-1} - K_{\alpha}^{-1} \right) f(\theta) \right\| + \left\| \left(K_{\alpha}^{-1} - K^{-1} \right) f(\theta) \right\| \\
&= O_p\left(\alpha^{-1} T^{-1/2} \right) + O\left(\alpha^{\min\left(1, \frac{\beta-1}{2}\right)} \right).
\end{aligned}$$

Hence, B_1 is negligible with respect to B_2 and the result follows.■

Lemma 15 *For all nonrandom functions* (u, v), *we have:*

$$E\left[\left\langle u, \widehat{h}_T(., \theta) \right\rangle \overline{\left\langle v, \widehat{h}_T(., \theta) \right\rangle} \right] = \frac{1}{T} \langle u, Kv \rangle$$

Proof of Lemma 15. We have:

$$\begin{aligned}
& E\left[\left\langle u, \widehat{h}_T(., \theta) \right\rangle \overline{\left\langle v, \widehat{h}_T(., \theta) \right\rangle} \right] \\
= \; & E\left[\left(\int u(\tau) \overline{\widehat{h}_T(\tau, \theta)} \pi(\tau) \, d\tau \right) \left(\int \overline{v(\tau)} \widehat{h}_T(\tau, \theta) \pi(\tau) \, d\tau \right) \right] \\
= \; & E\left[\int \int \overline{\widehat{h}_T(\tau_1, \theta)} \widehat{h}_T(\tau_2, \theta) u(\tau_1) \overline{v(\tau_2)} \pi(\tau_1) \pi(\tau_2) \, d\tau_1 d\tau_2 \right] \\
= \; & \int \int E\left[\overline{\widehat{h}_T(\tau_1, \theta)} \widehat{h}_T(\tau_2, \theta) \right] u(\tau_1) \overline{v(\tau_2)} \pi(\tau_1) \pi(\tau_2) \, d\tau_1 d\tau_2.
\end{aligned}$$

By noting that

$$E\left[\widehat{h}_T(\tau_1, \theta) \overline{\widehat{h}_T(\tau_2, \theta)} \right] = \frac{1}{T} E\left[h_t(\tau_1, \theta) \overline{h_t(\tau_2, \theta)} \right] = \frac{1}{T} k(\tau_1, \tau_2),$$

we have

$$\begin{aligned}
& E\left[\left\langle u, \widehat{h}_T(., \theta) \right\rangle \overline{\left\langle v, \widehat{h}_T(., \theta) \right\rangle} \right] \\
= \; & \frac{1}{T} \int \underbrace{\left(\int \overline{k(\tau_1, \tau_2) v(\tau_2)} \pi(\tau_2) \, d\tau_2 \right)}_{\overline{Kv(\tau_1)}} u(\tau_1) \pi(\tau_1) \, d\tau_1 \\
\equiv \; & \frac{1}{T} \langle u, Kv \rangle.
\end{aligned}$$

■

Lemma 16 *For all nonrandom points $\widetilde{\theta}$ in a neighborhood of θ_0 and for sufficiently large T, we have:*

$$\operatorname{Im}\left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\widetilde{\theta}),\widehat{h}_T(.,\widetilde{\theta})\right\rangle = O_p\left[\left(\widetilde{\theta}-\theta_0\right)'\left(\widetilde{\theta}-\theta_0\right)\right]$$

where $\widehat{G}_T(.,\theta)=\frac{\partial \widehat{h}_T(.,\theta)}{\partial\theta}$.

Proof of Lemma 16 We have:

$$\begin{aligned}
\widehat{h}_T(.,\widetilde{\theta}) &\approx \widehat{h}_T(.,\theta_0)+\widehat{G}_T(.,\theta_0)\left(\widetilde{\theta}-\theta_0\right)\\
\widehat{G}_T(.,\widetilde{\theta}) &\approx \widehat{G}_T(.,\theta_0)+\sum_{j=1}^{q}\widehat{H}_{j,T}(.,\theta_0)\left(\widetilde{\theta}_j-\theta_{j,0}\right)
\end{aligned}$$

Hence

$$\begin{aligned}
\left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\widetilde{\theta}),\widehat{h}_T(.,\widetilde{\theta})\right\rangle &\approx \left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\theta_0),\widehat{h}_T(.,\theta_0)\right\rangle\\
&\quad+\left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\theta_0),\widehat{G}_T(.,\theta_0)\right\rangle\left(\widetilde{\theta}-\theta_0\right)\\
&\quad+\sum_{j=1}^{q}\left\langle K_{\alpha T}^{-1}\widehat{H}_{j,T}(.,\theta_0),\widehat{G}_T(.,\theta_0)\right\rangle\left(\widetilde{\theta}_j-\theta_{j,0}\right)\left(\widetilde{\theta}-\theta_0\right)
\end{aligned}$$

where $\widehat{H}_{j,T}(.,\theta)=\frac{\partial \widehat{G}_T(.,\theta)}{\partial\theta_j}$ and a higher order term is omitted. Note that $\widetilde{\theta}$ is deterministic and does not depend of T. For sufficiently large T, we have:

$$\begin{aligned}
\left\langle K_{\alpha T}^{-1}G(.,\widetilde{\theta}),E\left[\widehat{h}_t(.,\widetilde{\theta})\right]\right\rangle &\approx \left\langle K_{\alpha T}^{-1}G(.,\theta_0),G(.,\theta_0)\right\rangle\left(\widetilde{\theta}-\theta_0\right)\\
&\quad+\sum_{j=1}^{q}\left\langle K_{\alpha T}^{-1}H_{j,T}(.,\theta_0),G(.,\theta_0)\right\rangle\left(\widetilde{\theta}_j-\theta_{j,0}\right)\left(\widetilde{\theta}-\theta_0\right)
\end{aligned}$$

where $G(.,\theta)=P\lim\widehat{G}_T(.,\theta)$ and $H_{j,T}(.,\theta_0)=P\lim\widehat{H}_{j,T}(.,\theta)$. It is seen that the first term of the right hand side is real. Consequently, the imaginary part of $\left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\widetilde{\theta}),\widehat{h}_T(.,\widetilde{\theta})\right\rangle$ can only come from the second term which is proportional to $\left(\widetilde{\theta}_j-\theta_{j,0}\right)\left(\widetilde{\theta}-\theta_0\right)$ ■

Proof of Theorem 1. The proof follows the same step as that of Proposition 3.2 in CCFG. However, we now exploit the fact $E\nabla_\theta h_t(\theta)\in\Phi_\beta$ with $\beta\geq 1$. The consistency follows from Lemma 14 provided $\alpha T^{1/2}\to\infty$ and $\alpha\to 0$. For the asymptotic normality to hold, we need to find a bound for the term B.10 of CCFG. We have:

$$
\begin{aligned}
|B.10| &= \left| \left\langle K_{\alpha T}^{-1} \nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right) - K^{-1} E\left(\nabla_\theta \hat{h}_T\left(\theta_0\right)\right), \sqrt{T}\hat{h}_T\left(\theta_0\right) \right\rangle \right| \\
&\leq \left\| K_{\alpha T}^{-1/2} \nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right) - K^{-1/2} E\left(\nabla_\theta \hat{h}_T\left(\theta_0\right)\right) \right\| \underbrace{\left\| \sqrt{T}\hat{h}_T\left(\theta_0\right) \right\|}_{=O_p(1)} \\
&= O_p\left(\alpha^{-1/2} T^{-1/2}\right) + O\left(\alpha^{\min\left(1, \frac{\beta-1}{2}\right)}\right)
\end{aligned}
$$

Hence the asymptotic normality requires the same conditions as the consistency, that is, $\alpha T^{1/2} \to \infty$ and $\alpha \to 0$.■

Stochastic expansion of the CGMM estimator: IID case

The objective function is

$$
\widehat{\theta} = \arg\min_{\theta} \left\{ Q_{\alpha T}(\theta) = \left\langle K_{\alpha T}^{-1} \widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta) \right\rangle \right\}.
$$

where $\widehat{h}_T(\tau,\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(e^{i\tau' x_t} - \varphi(\tau,\theta)\right)$. The optimal $\widehat{\theta}$ solves:

$$
\frac{\partial Q_{\alpha T}\left(\widehat{\theta}\right)}{\partial \theta} = 2\,\mathrm{Re}\left\langle K_{\alpha T}^{-1} G(.,\widehat{\theta}), \widehat{h}_T(.,\widehat{\theta}) \right\rangle = 0 \tag{2.40}
$$

where $G(.,\theta) = -\frac{\partial \varphi(\tau,\theta)}{\partial \theta}$.

A third order expansion gives

$$
0 = \frac{\partial Q_{\alpha T}\left(\theta_0\right)}{\partial \theta} + \frac{\partial^2 Q_{\alpha T}\left(\theta_0\right)}{\partial \theta \partial \theta'}\left(\widehat{\theta} - \theta_0\right) + \sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right) \frac{\partial^3 Q_{\alpha T}\left(\bar{\theta}\right)}{\partial \theta_j \partial \theta \partial \theta'}\left(\widehat{\theta} - \theta_0\right),
$$

where $\bar{\theta}$ lies between $\widehat{\theta}$ and θ_0. The dependence of $\widehat{\theta}$ on (α_T, θ_0) is hidden for convenience. Let us define

$$
\begin{aligned}
G_j(.,\theta) &= -\frac{\partial \varphi(\tau,\theta)}{\partial \theta_j}, H(.,\theta) = -\frac{\partial^2 \varphi(\tau,\theta)}{\partial \theta \partial \theta'}, \\
H_j(.,\theta) &= -\frac{\partial^2 \varphi(\tau,\theta)}{\partial \theta \partial \theta_j}, L_j = -\frac{\partial^3 \varphi(\tau,\theta)}{\partial \theta_j \partial \theta \partial \theta'}.
\end{aligned}
$$

and

$$\begin{aligned}
\Psi_T(\theta_0) &= \operatorname{Re}\left\langle K_{\alpha T}^{-1}G(.,\theta_0),\widehat{h}_T(.,\theta_0)\right\rangle, \\
W_T(\theta_0) &= \left\langle K_{\alpha T}^{-1}G(.,\theta_0),G(.,\theta_0)\right\rangle+\operatorname{Re}\left\langle K_{\alpha T}^{-1}H(.,\theta_0),\widehat{h}_T(.,\theta_0)\right\rangle, \\
B_{j,T}(\bar{\theta}) &= 2\operatorname{Re}\left\langle K_{\alpha T}^{-1}G(.,\bar{\theta}),H_j(.,\bar{\theta})\right\rangle+\operatorname{Re}\left\langle K_{\alpha T}^{-1}L_j(.,\bar{\theta}),\widehat{h}_T(.,\bar{\theta})\right\rangle \\
&\quad +\operatorname{Re}\left\langle K_{\alpha T}^{-1}H(.,\bar{\theta}),G_j(.,\bar{\theta})\right\rangle.
\end{aligned}$$

Then we can write:

$$0=\Psi_T(\theta_0)+W_T(\theta_0)\left(\widehat{\theta}-\theta_0\right)+\sum_{j=1}^{q}\left(\widehat{\theta}_j-\theta_{j,0}\right)B_{j,T}(\bar{\theta})\left(\widehat{\theta}-\theta_0\right).$$

Note that the derivatives of the moment functions are deterministic in the IID case. We decompose $\Psi_T(\theta_0)$, $W_T(\theta_0)$ and $B_{j,T}(\bar{\theta})$ as follows:

$$\Psi_T(\theta_0)=\Psi_{T,0}(\theta_0)+\Psi_{T,\alpha}(\theta_0)+\widetilde{\Psi}_{T,\alpha}(\theta_0),$$

where

$$\begin{aligned}
\Psi_{T,0}(\theta_0) &= \operatorname{Re}\left\langle K^{-1}G,\widehat{h}_T\right\rangle=O_p\left(T^{-1/2}\right) \\
\Psi_{T,\alpha}(\theta_0) &= \operatorname{Re}\left\langle \left(K_\alpha^{-1}-K^{-1}\right)G,\widehat{h}_T\right\rangle=O_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right) \\
\widetilde{\Psi}_{T,\alpha}(\theta_0) &= \operatorname{Re}\left\langle \left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G,\widehat{h}_T\right\rangle=O_p\left(\alpha^{-1}T^{-1}\right)
\end{aligned}$$

where the rates of convergences are obtained using the fact that $|\langle f,g\rangle|\leq\|f\|\,\|g\|$ and the results of Lemma 13. Similarly, we decompose $W_T(\theta_0)$ into various terms with distinct rates of convergence:

$$W_T(\theta_0)=W_0(\theta_0)+W_\alpha(\theta_0)+\widetilde{W}_\alpha(\theta_0)+W_{T,0}(\theta_0)+\widetilde{W}_{T,\alpha}(\theta_0),$$

where

$$
\begin{aligned}
W_0(\theta_0) &= \left\langle K^{-1}G, G\right\rangle = O(1), \\
W_\alpha(\theta_0) &= \left\langle \left(K_\alpha^{-1} - K^{-1}\right) G, G\right\rangle = O\left(\alpha^{\min(1, \frac{2\beta-1}{2})}\right), \\
\widetilde{W}_\alpha(\theta_0) &= \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G, G\right\rangle = O_p\left(\alpha^{-1}T^{-1/2}\right), \\
W_{T,0}(\theta_0) &= \operatorname{Re}\left\langle K^{-1}H(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle = O_p\left(T^{-1/2}\right), \\
\widetilde{W}_{T,\alpha}(\theta_0) &= \operatorname{Re}\left\langle \left(K_{\alpha T}^{-1} - K^{-1}\right) H(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle = O_p\left(\alpha^{-1}T^{-1}\right).
\end{aligned}
$$

We consider a simpler decomposition for $B_{j,T}(\bar{\theta})$:

$$
B_{j,T}(\bar{\theta}) = B_j(\bar{\theta}) + \left(B_{j,T}(\bar{\theta}) - B_j(\bar{\theta})\right)
$$

where

$$
\begin{aligned}
B_j(\bar{\theta}) &= 2\operatorname{Re}\left\langle K^{-1}G(.,\bar{\theta}), H_j(.,\bar{\theta})\right\rangle + \operatorname{Re}\left\langle K^{-1}H(.,\bar{\theta}), G_j(.,\bar{\theta})\right\rangle = O(1), \\
B_{j,T}(\bar{\theta}) &= B_j(\bar{\theta}) + O\left(\alpha^{\min(1, \frac{\beta-1}{2})}\right) + O_p\left(\alpha^{-1}T^{-1/2}\right).
\end{aligned}
$$

By replacing these decompositions into the expansion of the FOC, we can solve for $\widehat{\theta} - \theta_0$ to obtain:

$$
\begin{aligned}
\widehat{\theta} - \theta_0 &= -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\
&\quad -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0) + W_\alpha(\theta_0)\left(\widehat{\theta} - \theta_0\right)\right] \\
&\quad -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0) + \widetilde{W}_\alpha(\theta_0)\left(\widehat{\theta} - \theta_0\right)\right] \\
&\quad -W_0^{-1}(\theta_0)W_{T,0}(\theta_0)\left(\widehat{\theta} - \theta_0\right) \\
&\quad -\sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right)W_0^{-1}(\theta_0)B_j(\bar{\theta})\left(\widehat{\theta} - \theta_0\right) \\
&\quad -\sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right)W_0^{-1}(\theta_0)(B_{j,T}(\bar{\theta}) - B_j(\bar{\theta}))\left(\widehat{\theta} - \theta_0\right)
\end{aligned}
$$

To complete the expansion, we replace $\widehat{\theta}-\theta_0$ by $-W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)$ in the higher order terms:

$$\widehat{\theta}-\theta_0=\Delta_1+\Delta_2+\Delta_3+\Delta_4+\Delta_5+\widehat{R}$$

where $\widehat{R}$ is a remainder that goes to zero faster than the following terms:

$$\begin{aligned}
\Delta_1 &= -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0),\\
\Delta_2 &= -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0)-W_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right],\\
\Delta_3 &= -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0)-\widetilde{W}_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right],\\
\Delta_4 &= W_0^{-1}(\theta_0)W_{T,0}(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\\
&\quad -\sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)B_j(\bar{\theta})W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0),\\
\Delta_5 &= -\sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)(B_{j,T}(\bar{\theta})-B_j(\bar{\theta}))W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0).
\end{aligned}$$

To obtain the rates of these terms, we use the fact that $|Af|\le\|A\|\,|f|$. This yields immediately:

$$\begin{aligned}
\Delta_1 &= O_p\left(T^{-1/2}\right);\ \Delta_2=O_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right),\\
\Delta_3 &= O_p\left(\alpha^{-1}T^{-1}\right);\ \Delta_4=O_p\left(T^{-1}\right),\\
\Delta_5 &= O\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1}\right)+O_p\left(\alpha^{-1}T^{-3/2}\right).
\end{aligned}$$

To summarize, we have:

$$\widehat{\theta}-\theta_0=\Delta_1+\Delta_2+\Delta_3+o_p\left(\alpha^{-1}T^{-1}\right)+o_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right) \tag{2.41}$$

Stochastic expansion of the CGMM estimator: Markov case

The objective function here is given by:

$$\widehat{\theta}=\arg\min_{\theta}\left\{Q_{\alpha T}(\theta)=\left\langle K_{\alpha T}^{-1}\widehat{h}_T(.,\theta),\widehat{h}_T(.,\theta)\right\rangle\right\}.$$

where $\widehat{h}_T(\tau,\theta)=\frac{1}{T}\sum_{t=1}^{T}\left(e^{is'x_{t+1}}-\varphi(s,\theta,x_t)\right)e^{ir'x_t}$and $\tau=(s,r)\in\mathbb{R}^{2p}$. The optimal $\widehat{\theta}$

solves

$$\frac{\partial Q_{\alpha T}\left(\widehat{\theta}\right)}{\partial \theta} = 2\operatorname{Re}\left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\widehat{\theta}), \widehat{h}_T(.,\widehat{\theta})\right\rangle = 0 \tag{2.42}$$

where $\widehat{G}_T(\tau,\theta) = -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\varphi(s,\theta,x_t)}{\partial\theta}e^{ir'x_t}$.

The third order Taylor expansion of (2.42) around θ_0 yields:

$$0 = \frac{\partial Q_{\alpha T}\left(\theta_0\right)}{\partial\theta} + \frac{\partial^2 Q_{\alpha T}\left(\theta_0\right)}{\partial\theta\partial\theta'}\left(\widehat{\theta}-\theta_0\right) + \sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right)\frac{\partial^3 Q_{\alpha T}\left(\bar{\theta}\right)}{\partial\theta_j\partial\theta\partial\theta'}\left(\widehat{\theta}-\theta_0\right),$$

where $\bar{\theta}$ lies between $\widehat{\theta}$ and θ_0.

Let us define:

$$\begin{aligned}
\widehat{H}_T(\tau,\theta) &= -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\varphi(s,\theta,x_t)}{\partial\theta\partial\theta'}e^{ir'x_t},\\
\widehat{G}_{j,T}(\tau,\theta) &= -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\varphi(s,\theta,x_t)}{\partial\theta_j}e^{ir'x_t},\ \ \widehat{H}_{j,T}(\tau,\theta) = -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\varphi(s,\theta,x_t)}{\partial\theta_j\partial\theta}e^{ir'x_t},\\
\widehat{L}_{j,T}(\tau,\theta) &= -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3\varphi(s,\theta,x_t)}{\partial\theta_j\partial\theta\partial\theta'}e^{ir'x_t},
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\Psi}_T(\theta_0) &= \operatorname{Re}\left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle,\\
\widehat{W_T}(\theta_0) &= \left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\theta_0), \widehat{G}_T(.,_0)\right\rangle + \operatorname{Re}\left\langle K_{\alpha T}^{-1}\widehat{H}_T(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle,\\
\widehat{B}_{j,T}(\bar{\theta}) &= 2\operatorname{Re}\left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\bar{\theta}), \widehat{H}_{j,T}(.,\bar{\theta})\right\rangle + \operatorname{Re}\left\langle K_{\alpha T}^{-1}\widehat{H}_T(.,\bar{\theta}), \widehat{G}_{j,T}(.,\bar{\theta})\right\rangle\\
&\quad + \operatorname{Re}\left\langle K_{\alpha T}^{-1}\widehat{L}_{j,T}(.,\bar{\theta}), \widehat{h}_T(.,\bar{\theta})\right\rangle.
\end{aligned}$$

Then the expansion of the FOC becomes:

$$0 = \widehat{\Psi}_T(\theta_0) + \widehat{W_T}(\theta_0)\left(\widehat{\theta}-\theta_0\right) + \sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right)\widehat{B}_{j,T}(\bar{\theta})\left(\widehat{\theta}-\theta_0\right),$$

Unlike in the IID case, the derivatives of the moment function are not deterministic. We thus define:

$$\begin{aligned}
G(\tau,\theta) &= \underset{T\to\infty}{P\lim}\widehat{G}_T(\tau,\theta) \equiv E\left[\frac{\partial\varphi(s,\theta,x_t)}{\partial\theta}e^{ir'x_t}\right], \\
H(\tau,\theta) &= \underset{T\to\infty}{P\lim}\widehat{H}_T(\tau,\theta) \equiv E\left[\frac{\partial^2\varphi(s,\theta,x_t)}{\partial\theta\partial\theta'}e^{ir'x_t}\right], \\
G_j(\tau,\theta) &= \underset{T\to\infty}{P\lim}\widehat{G}_{j,T}(\tau,\theta) \equiv E\left[\frac{\partial\varphi(s,\theta,x_t)}{\partial\theta_j}e^{ir'x_t}\right], \\
H_j(\tau,\theta) &= \underset{T\to\infty}{P\lim}\widehat{H}_{j,T}(\tau,\theta) \equiv E\left[\frac{\partial^2\varphi(s,\theta,x_t)}{\partial\theta\partial\theta_j}e^{ir'x_t}\right].
\end{aligned}$$

Because x_t is Markov, we have:

$$\begin{aligned}
G(\tau,\theta)-\widehat{G}_T(\tau,\theta) &= O_p\left(T^{-1/2}\right), \\
H(\tau,\theta)-\widehat{H}_T(\tau,\theta) &= O_p\left(T^{-1/2}\right), \\
G_j(\tau,\theta)-\widehat{G}_{j,T}(\tau,\theta) &= O_p\left(T^{-1/2}\right), \\
H_j(\tau,\theta)-\widehat{H}_{j,T}(\tau,\theta) &= O_p\left(T^{-1/2}\right).
\end{aligned}$$

We have the following decomposition for $\widehat{\Psi}_T(\theta_0)$:

$$\widehat{\Psi}_T(\theta_0) = \Psi_{T,0}(\theta_0) + \Psi_{T,\alpha}(\theta_0) + \widetilde{\Psi}_{T,\alpha}(\theta_0) + \widehat{\Psi}_{T,\alpha}(\theta_0) + \widehat{\widetilde{\Psi}}_{T,\alpha}(\theta_0).$$

Using the fact that $|Af| \leq \|A\|\,|f|$, the rates of convergence are:

$$\begin{aligned}
\Psi_{T,0}(\theta_0) &= \mathrm{Re}\left\langle K^{-1}G,\widehat{h}_T\right\rangle = O_p\left(T^{-1/2}\right), \\
\Psi_{T,\alpha}(\theta_0) &= \mathrm{Re}\left\langle \left(K_\alpha^{-1}-K^{-1}\right)G,\widehat{h}_T\right\rangle = O_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right), \\
\widetilde{\Psi}_{T,\alpha}(\theta_0) &= \mathrm{Re}\left\langle \left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G,\widehat{h}_T\right\rangle = O_p\left(\alpha^{-1}T^{-1}\right), \\
\widehat{\Psi}_{T,\alpha}(\theta_0) &= \mathrm{Re}\left\langle K_\alpha^{-1}\left(\widehat{G}_T-G\right),\widehat{h}_T\right\rangle = O_p\left(\alpha^{-1/2}T^{-1}\right), \\
\widehat{\widetilde{\Psi}}_{T,\alpha}(\theta_0) &= \mathrm{Re}\left\langle \left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)\left(\widehat{G}_T-G\right),\widehat{h}_T\right\rangle = O_p\left(\alpha^{-3/2}T^{-3/2}\right).
\end{aligned}$$

The difference between the above decomposition of $\widehat{\Psi}_T(\theta_0)$ and the one in the IID case only comes from the additional higher order terms $\widehat{\Psi}_{T,\alpha}(\theta_0)$ and $\widehat{\widetilde{\Psi}}_{T,\alpha}(\theta_0)$. Hence we can write

$\widehat{\Psi}_T(\theta_0)$ as:

$$\widehat{\Psi}_T(\theta_0) = \Psi_{T,0}(\theta_0) + \Psi_{T,\alpha}(\theta_0) + \widetilde{\Psi}_{T,\alpha}(\theta_0) + R_\Psi$$

where $R_\Psi = o_p\left(\alpha^{-1}T^{-1}\right) + o_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right)$.

We have a similar decomposition for $\widehat{W}_T(\theta_0)$:

$$\begin{aligned}
\widehat{W}_T(\theta_0) &= W_0(\theta_0) + W_\alpha(\theta_0) + \widetilde{W}_\alpha(\theta_0) + \widehat{W}_\alpha(\theta_0) + \widehat{\widetilde{W}}_\alpha(\theta_0) \\
&\quad + W_1(\theta_0) + W_{1,\alpha}(\theta_0) + \widetilde{W}_{1,\alpha}(\theta_0) + \widehat{W}_{1,\alpha}(\theta_0) + \widehat{\widetilde{W}}_{1,\alpha}(\theta_0)
\end{aligned}$$

where

$$\begin{aligned}
W_0(\theta_0) &= \left\langle K^{-1}G, G\right\rangle = O(1), \\
W_\alpha(\theta_0) &= \left\langle \left(K_\alpha^{-1} - K^{-1}\right)G, G\right\rangle = O\left(\alpha^{\min(1,\frac{2\beta-1}{2})}\right), \\
\widetilde{W}_\alpha(\theta_0) &= \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)G, G\right\rangle = O_p\left(\alpha^{-1}T^{-1/2}\right), \\
\widehat{W}_\alpha(\theta_0) &= \left\langle K_\alpha^{-1}\left(\widehat{G}_T - G\right), G\right\rangle = O_p\left(\alpha^{-1/2}T^{-1/2}\right), \\
W_1(\theta_0) &= \operatorname{Re}\left\langle K^{-1}H, \widehat{h}_T\right\rangle + \left\langle K^{-1}G, \widehat{G}_T - G\right\rangle = O_p\left(T^{-1/2}\right), \\
W_{1,\alpha}(\theta_0) &= \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)\left(\widehat{G}_T - G\right), G\right\rangle = O_p\left(\alpha^{-3/2}T^{-1}\right).
\end{aligned}$$

$$\begin{aligned}
\widehat{\widetilde{W}}_{1,\alpha}(\theta_0) &= \operatorname{Re}\left\langle \left(K_\alpha^{-1} - K^{-1}\right)H, \widehat{h}_T\right\rangle + \left\langle \left(K_\alpha^{-1} - K^{-1}\right)G, \widehat{G}_T - G\right\rangle \\
&= O\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right), \\
\widetilde{W}_{1,\alpha}(\theta_0) &= \operatorname{Re}\left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)H, \widehat{h}_T\right\rangle + \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)G, \widehat{G}_T - G\right\rangle \\
&= O_p\left(\alpha^{-1}T^{-1}\right), \\
\widehat{W}_{1,\alpha}(\theta_0) &= \operatorname{Re}\left\langle K_\alpha^{-1}\left(\widehat{H}_T - H\right), \widehat{h}_T\right\rangle + \left\langle K_\alpha^{-1}\left(\widehat{G}_T - G\right), \widehat{G}_T - G\right\rangle \\
&= O_p\left(\alpha^{-1/2}T^{-1}\right), \\
R_{W,1} &= \operatorname{Re}\left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)\left(\widehat{H}_T - H\right), \widehat{h}_T\right\rangle \\
&\quad + \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)\left(\widehat{G}_T - G\right), \widehat{G}_T - G\right\rangle = O_p\left(\alpha^{-3/2}T^{-3/2}\right).
\end{aligned}$$

For the purpose of finding the optimal α, it is enough to consider the shorter decomposition:

$$\widehat{W}_T(\theta_0) = W_0(\theta_0) + W_\alpha(\theta_0) + \widetilde{W}_\alpha(\theta_0) + \widehat{W}_\alpha(\theta_0) + W_1(\theta_0) + W_{1,\alpha}(\theta_0) + R_W$$

with

$$R_W \equiv \widehat{\widetilde{W}}_{1,\alpha}(\theta_0) + \widetilde{W}_{1,\alpha}(\theta_0) + \widehat{W}_{1,\alpha}(\theta_0) + R_{W,1} = O_p\left(\alpha^{-1}T^{-1}\right) + O\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right).$$

Finally, we consider again a simpler decomposition for $B_{j,T}(\bar{\theta})$:

$$B_{j,T}(\bar{\theta}) = B_j(\bar{\theta}) + \left(B_{j,T}(\bar{\theta}) - B_j(\bar{\theta})\right)$$

where

$$\begin{aligned} B_j(\bar{\theta}) &= 2\,\mathrm{Re}\left\langle K^{-1}G(.,\bar{\theta}), H_j(.,\bar{\theta})\right\rangle + \mathrm{Re}\left\langle K^{-1}H(.,\bar{\theta}), G_j(.,\bar{\theta})\right\rangle = O(1) \\ B_{j,T}(\bar{\theta}) &= B_j(\bar{\theta}) + O\left(\alpha^{\min(1,\frac{\beta-1}{2})}\right) + O_p\left(\alpha^{-1}T^{-1/2}\right). \end{aligned}$$

We replace these decompositions into the expansion of the FOC and solve for $\widehat{\theta} - \theta_0$ to obtain:

$$\begin{aligned} \widehat{\theta} - \theta_0 &= -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\ &\quad -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0) + W_\alpha(\theta_0)\left(\widehat{\theta} - \theta_0\right)\right] \\ &\quad -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0) + \widetilde{W}_\alpha(\theta_0)\left(\widehat{\theta} - \theta_0\right)\right] - W_0^{-1}(\theta_0)\widehat{W}_\alpha(\theta_0)\left(\widehat{\theta} - \theta_0\right) \\ &\quad -W_0^{-1}(\theta_0)W_1(\theta_0)\left(\widehat{\theta} - \theta_0\right) - \sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right)W_0^{-1}(\theta_0)B_j(\bar{\theta})\left(\widehat{\theta} - \theta_0\right) \\ &\quad -W_0^{-1}(\theta_0)W_{1,\alpha}(\theta_0)\left(\widehat{\theta} - \theta_0\right) \\ &\quad -\sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right)W_0^{-1}(\theta_0)(B_{j,T}(\bar{\theta}) - B_j(\bar{\theta}))\left(\widehat{\theta} - \theta_0\right) \\ &\quad -W_0^{-1}(\theta_0)R_W\left(\widehat{\theta} - \theta_0\right) - W_0^{-1}(\theta_0)R_\Psi \end{aligned}$$

Next, we replace $\widehat{\theta} - \theta_0$ by $-W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(T^{-1/2}\right)$ in the higher order terms. This yields:

$$\widehat{\theta} - \theta_0 = \Delta_1 + \Delta_2 + \Delta_3 + \widehat{R}_1 + \widehat{R}_2 + \widehat{R}_3 + \widehat{R}_4$$

where:

$$\begin{aligned}
\Delta_1 &= -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(T^{-1/2}\right), \\
\Delta_2 &= -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0) - W_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right] = O_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right), \\
\Delta_3 &= -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0) - \widetilde{W}_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right] = O_p\left(\alpha^{-1}T^{-1}\right), \\
\widehat{R}_1 &= W_0^{-1}(\theta_0)\widehat{W}_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(\alpha^{-1/2}T^{-1}\right), \\
\widehat{R}_2 &= W_0^{-1}(\theta_0)W_1(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\
&\quad - \sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)B_j(\bar{\theta})W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(T^{-1}\right), \\
\widehat{R}_3 &= W_0^{-1}(\theta_0)W_{1,\alpha}(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(\alpha^{-3/2}T^{-3/2}\right),
\end{aligned}$$

and

$$\begin{aligned}
\widehat{R}_4 &= -W_0^{-1}(\theta_0)R_\Psi + W_0^{-1}(\theta_0)R_W W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\
&\quad - \sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)(B_{j,T}(\bar{\theta}) - B_j(\bar{\theta}))W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\
&= o_p\left(\alpha^{-1}T^{-1}\right) + o_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right).
\end{aligned}$$

In summary, we have:

$$\widehat{\theta} - \theta_0 = \Delta_1 + \Delta_2 + \Delta_3 + o_p\left(\alpha^{-1}T^{-1}\right) + o_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right) \tag{2.43}$$

which is of the same form as in the IID case.

Proof of Theorem 2.

Using the expansions given in (2.41) and (2.43), we obtain:

$$\widehat{\theta} - \theta_0 = \Delta_1 + \Delta_2 + \Delta_3 + O_p\left(T^{-1}\right)$$

As T increases, the consistent estimator $\widehat{\theta}$ eventually falls (with probability approching one) in a neighborhood of θ_0 on which Lemma 16 holds. This ensures that all the terms that are slower than $O_p\left(T^{-1}\right)$ in the expansion of the FOC are real so that the Re symbol may be removed from the expression of Δ_1, Δ_2 and Δ_3.

Asymptotic Variance

The asymptotic variance of $\widehat{\theta}$ is given by

$$\begin{aligned}
TVar\left(\Delta_1\right) &= TW_0^{-1}E\left[\Psi_{T,0}(\theta_0)\Psi_{T,0}(\theta_0)'\right]W_0^{-1} \\
&= TW_0^{-1}E\left[\left\langle K^{-1}G,\widehat{h}_T\right\rangle \overline{\left\langle K^{-1}G,\widehat{h}_T\right\rangle}'\right]W_0^{-1} \\
&= W_0^{-1}\left\langle K^{-1}G,G\right\rangle W_0^{-1}
\end{aligned}$$

where the last equality follows from Lemma 15. Hence,

$$TVar\left(\Delta_1\right) = W_0^{-1}\left\langle K^{-1}G,G\right\rangle W_0^{-1} = W_0^{-1}.$$

Higher Order Bias

The terms Δ_1 and Δ_2 have zero expectations. Hence, the bias can only come from Δ_3:

$$\begin{aligned}
Bias &= E\left[\widehat{\theta}-\theta_0\right] = E\left[\Delta_3\right] \\
&= -W_0^{-1}E\left[\widetilde{\Psi}_{T,\alpha}\right] + W_0^{-1}E\left[\widetilde{W}_\alpha W_0^{-1}\Psi_{T,0}\right] \\
&= O\left(\alpha^{-1}T^{-1}\right)
\end{aligned}$$

where we used the rule $E\left[O_p\left(\alpha^{-1}T^{-1}\right)\right] = O\left(\alpha^{-1}T^{-1}\right)$. This rule gives the worst case rate because the true bias may converge to zero faster. The squared bias satisfies:

$$TBias * Bias' = O\left(\alpha^{-2}T^{-1}\right).$$

Higher Order Variance

The dominant terms in the higher order variance are

$$Cov\left(\Delta_1,\Delta_2\right) + Var\left(\Delta_2\right) + Cov\left(\Delta_1,\Delta_3\right).$$

We first consider $Cov\left(\Delta_1, \Delta_2\right)$:

$$Cov\left(\Delta_1, \Delta_2\right) = W_0^{-1} E\left[\Psi_{T,0} \Psi_{T,\alpha}(\theta_0)'\right] W_0^{-1} - W_0^{-1} E\left[\Psi_{T,0} \Psi_{T,0}'\right] W_0^{-1} W_\alpha W_0^{-1}.$$

From Lemma 15, we have:

$$E\left[\Psi_{T,0} \Psi_{T,\alpha}'\right] = \frac{1}{T} \left\langle \left(K_\alpha^{-1} - K^{-1}\right) G, G \right\rangle = W_\alpha.$$

and $E\left[\Psi_{T,0} \Psi_{T,0}'\right] = W_0$. Hence,

$$\begin{aligned} Cov\left(\Delta_1, \Delta_2\right) &= \frac{1}{T} W_0^{-1} W_\alpha W_0^{-1} - \frac{1}{T} W_0^{-1} W_0 W_0^{-1} W_\alpha W_0^{-1} \\ &= 0. \end{aligned}$$

Now we consider the term $Cov\left(\Delta_1, \Delta_3\right)$:

$$Cov\left(\Delta_1, \Delta_3\right) = W_0^{-1} E\left[\Psi_{T,0} \widetilde{\Psi}_{T,\alpha}'\right] W_0^{-1} - W_0^{-1} E\left[\Psi_{T,0} \Psi_{T,0}' W_0^{-1} \widetilde{W}_\alpha\right] W_0^{-1}.$$

We first consider:

$$E\left[\Psi_{T,0} \widetilde{\Psi}_{T,\alpha}'\right] = E\left[\left\langle K^{-1} G, \widehat{h}_T \right\rangle \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G, \widehat{h}_T \right\rangle\right].$$

Using Equation (2.34) in Lemma 13, we have

$$\begin{aligned} \left\langle K^{-1} G, \widehat{h}_T \right\rangle \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G, \widehat{h}_T \right\rangle &= O_p\left(T^{-1/2}\right) \times O_p\left(\alpha^{-1} T^{-1/2}\right) \times O_p\left(T^{-1/2}\right) \\ &= O_p\left(\alpha^{-1} T^{-3/2}\right). \end{aligned}$$

Hence we conclude for the first term of $Cov\left(\Delta_1, \Delta_3\right)$ that

$$W_0^{-1} E\left[\Psi_{T,0} \widetilde{\Psi}_{T,\alpha}'\right] W_0^{-1} = O\left(\alpha^{-1} T^{-3/2}\right).$$

We now turn to examine the second term of $Cov\left(\Delta_1,\Delta_3\right)$.

$$\begin{aligned}
& E\left[\Psi_{T,0}\Psi_{T,0}^{\prime}W_0^{-1}\widetilde{W}_\alpha\right] \\
= \; & E\left[\left\langle K^{-1}G,\widehat{h}_T\right\rangle\left\langle K^{-1}G,\widehat{h}_T\right\rangle^{\prime}W_0^{-1}\left\langle\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G,G\right\rangle\right].
\end{aligned}$$

Using again Lemma 13, we have

$$\begin{aligned}
& \left\langle K^{-1}G,\widehat{h}_T\right\rangle\left\langle K^{-1}G,\widehat{h}_T\right\rangle^{\prime}W_0^{-1}\left\langle\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G,G\right\rangle \\
= \; & O_p\left(T^{-1/2}\right)\times O_p\left(T^{-1/2}\right)\times O\left(1\right)\times O_p\left(\alpha^{-1}T^{-1/2}\right)=O_p\left(\alpha^{-1}T^{-3/2}\right).
\end{aligned}$$

so that we can conclude:

$$E\left[\left\langle K^{-1}G,\widehat{h}_T\right\rangle\left\langle K^{-1}G,\widehat{h}_T\right\rangle^{\prime}W_0^{-1}\left\langle\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G,G\right\rangle\right]=O\left(\alpha^{-1}T^{-3/2}\right).$$

Overall, the rate of $Cov\left(\Delta_1,\Delta_3\right)$ is thus given by

$$T\times Cov\left(\Delta_1,\Delta_3\right)=O\left(\alpha^{-1}T^{-1/2}\right).$$

We are left with the rate of $Var\left(\Delta_2\right)$. We recall that

$$\Delta_2=-W_0^{-1}\Psi_{T,\alpha}+W_0^{-1}W_\alpha W_0^{-1}\Psi_{T,0}.$$

We have

$$\begin{aligned}
Var\left(\Delta_2\right)=W_0^{-1}E\left[\Psi_{T,\alpha}\Psi_{T,\alpha}^{\prime}\right]W_0^{-1} \quad & -W_0^{-1}E\left[\Psi_{T,\alpha}\Psi_{T,0}^{\prime}\right]W_0^{-1}W_\alpha W_0^{-1} \\
-W_0^{-1}W_\alpha W_0^{-1}E\left[\Psi_{T,0}\Psi_{T,\alpha}^{\prime}\right]W_0^{-1} \quad & +W_0^{-1}W_\alpha W_0^{-1}E\left[\Psi_{T,0}\Psi_{T,0}^{\prime}\right]W_0^{-1}W_\alpha W_0^{-1}.
\end{aligned}$$

Replacing $E\left[\Psi_{T,0}\Psi_{T,\alpha}^{\prime}\right]=\frac{1}{T}W_\alpha$ and $E\left[\Psi_{T,0}\Psi_{T,0}^{\prime}\right]=\frac{1}{T}W_0$, we see immediately that the last two terms cancel out so that

$$Var\left(\Delta_2\right)=W_0^{-1}E\left[\Psi_{T,\alpha}\Psi_{T,\alpha}^{\prime}\right]W_0^{-1}-W_0^{-1}W_\alpha W_0^{-1}W_\alpha W_0^{-1}.$$

For the first term of $Var\left(\Delta_2\right)$, we use Lemma 15 to obtain:

$$
\begin{aligned}
E\left[\Psi_{T,\alpha}\Psi'_{T,\alpha}\right] &= E\left[\left\langle \left(K_\alpha^{-1}-K^{-1}\right)G,\widehat{h}_T\right\rangle \left\langle \left(K_\alpha^{-1}-K^{-1}\right)G,\widehat{h}_T\right\rangle\right] \\
&= \frac{1}{T}\left\langle \left(K_\alpha^{-1}-K^{-1}\right)G,\left(K_\alpha^{-1}-K^{-1}\right)KG\right\rangle \\
&= \sum_j \left(\frac{\mu_j}{\mu_j^2+\alpha}-\frac{1}{\mu_j}\right)^2 \mu_j \left\langle G,\phi_j\right\rangle^2 \\
&= \sum_j \left(\frac{\mu_j}{\mu_j^2+\alpha}-\frac{1}{\mu_j}\right)^2 \mu_j^{2\beta+1}\frac{\left\langle G,\phi_j\right\rangle^2}{\lambda_j^{2\beta}} \\
&\leq \sum_j \frac{\left\langle G,\phi_j\right\rangle^2}{\mu_j^{2\beta}} \sup_{\mu\leq\mu_1}\left(\frac{\mu}{\mu^2+\alpha}-\frac{1}{\mu}\right)^2 \mu^{2\beta+1}.
\end{aligned}
$$

We focus on the square-root of $\left(\frac{\mu}{\mu^2+\alpha}-\frac{1}{\mu}\right)^2\mu^{2\beta+1}$, namely:

$$
\sup_{\mu\leq\mu_1}\left(\frac{1}{\mu}-\frac{\mu}{\mu^2+\alpha}\right)\mu^{(2\beta+1)/2} = \sup_{\mu\leq\mu_1}\left(1-\frac{\mu^2}{\mu^2+\alpha}\right)\mu^{\beta-1/2}.
$$

Case where $\beta \geq 5/2$

$$
\begin{aligned}
\sup_{\mu\leq\mu_1}\left(1-\frac{\mu^2}{\mu^2+\alpha}\right)\mu^{\beta-1/2} &= \alpha\sup_{\mu\leq\mu_1}\frac{\mu^{\beta-1/2}}{\mu^2+\alpha} \\
&\leq \alpha\sup_{\mu\leq\mu_1}\mu^{\beta-5/2} \leq \alpha\mu_1^{\beta-5/2}.
\end{aligned}
$$

Case where $\beta < 5/2$

We apply the change of variable $x = \alpha/\mu^2$ and obtain

$$
\begin{aligned}
\sup_{\mu\leq\mu_1}\left(1-\frac{\mu^2}{\mu^2+\alpha}\right)\mu^{\beta-1/2} &= \sup_{x\geq 0}\left(1-\frac{1}{1+x}\right)\left(\frac{\alpha}{x}\right)^{\frac{\beta-1/2}{2}} \\
&= \alpha^{\frac{2\beta-1}{4}}\sup_{x\geq 0}\frac{x}{1+x}x^{-\frac{2\beta-1}{4}}.
\end{aligned}
$$

The function $f\left(x\right) = \frac{x}{1+x}x^{-\frac{2\beta-1}{4}}$ is continuous and hence bounded for x away from 0 and infinity. When x goes to infinity, $f\left(x\right)$ goes to zero because $2\beta-1>0$. When x goes to zero,

$$
f\left(x\right) = \frac{x^{\frac{5-2\beta}{4}}}{1+x}
$$

goes to zero because $5-2\beta>0$. Hence, $f(x)$ is bounded on $\mathbb{R}^{+}$. In conclusion, the rate of convergence of $E\left(\Psi_{T,\alpha}\Psi_{T,\alpha}'\right)$ is given by:

$$\alpha^{\min\left(2,\frac{2\beta-1}{2}\right)}T^{-1}.$$

Note that this rate is an equivalent, not a O.

For the second term of $Var\left(\Delta_2\right)$, we use the fact that $W_\alpha = O\left(\alpha^{\min\left(1,\frac{2\beta-1}{2}\right)}\right)$ according to Equation (2.36) in Lemma 13:

$$\begin{aligned}
& \frac{1}{T}W_0^{-1}W_\alpha W_0^{-1}W_\alpha W_0^{-1} \\
= & \frac{1}{T}\times O(1)\times O\left(\alpha^{\min\left(1,\frac{2\beta-1}{2}\right)}\right)\times O(1)\times O\left(\alpha^{\min\left(1,\frac{2\beta-1}{2}\right)}\right)\times O(1) \\
= & O\left(\alpha^{\min(2,2\beta-1)}T^{-1}\right).
\end{aligned}$$

Optimal Rate for α

Note that the bias term $TBias*Bias' = O\left(\alpha^{-2}T^{-1}\right)$ goes to zero faster than the covariance term $TCov\left(\Delta_1,\Delta_3\right) = O\left(\alpha^{-1}T^{-1/2}\right)$. Hence the optimal α is the one that achieves the best trade-off between $TVar\left(\Delta_2\right)\sim\alpha^{\min\left(2,\beta-\frac{1}{2}\right)}$ which is increasing in α and $TCov\left(\Delta_1,\Delta_3\right)$ which is decreasing in α. We have

$$\alpha^{\min\left(2,\beta-\frac{1}{2}\right)} = \alpha^{-1}T^{-1/2} \Rightarrow \alpha = T^{-\max\left(\frac{1}{6},\frac{1}{2\beta+1}\right)}.$$

Consistency of $\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right)$

First we recall the notations. Let $\widehat{\theta}_T(\alpha;\theta)$ be a CGMM estimator of θ built from $X_T(\theta)$, a sample of size T and the value α for the regularization parameter. The sample $X_T(\theta)$ is supposed to be drawn from the DGP indexed by θ. One can construct M IID copies of the random variable $\widehat{\theta}_T(\alpha;\theta)$ by drawing M independent samples from the DGP to obtain $\widehat{\theta}_T^j(\alpha;\theta)$, $j=1,2,...,M$ as described in Section 4.

Let $\Sigma_T(\alpha,\theta)$ denote the MSE of $\widehat{\theta}_T(\alpha;\theta)$, that is:

$$\Sigma_T(\alpha,\theta) = TE\left[\left(\widehat{\theta}_T(\alpha;\theta)-\theta\right)'\left(\widehat{\theta}_T(\alpha;\theta)-\theta\right)\right].$$

We estimate $\Sigma_T(\alpha;\theta_0)$ in a natural way as:

$$\widehat{\Sigma}_{TM}^{MC}(\alpha,\theta)=\frac{T}{M}\sum_{j=1}^{M}\left(\widehat{\theta}_T^j(\alpha;\theta)-\theta\right)'\left(\widehat{\theta}_T^j(\alpha;\theta)-\theta\right).$$

Note that $\Sigma_T(\alpha,\theta)=E\left[\widehat{\Sigma}_{TM}^{MC}(\alpha,\theta)\right]$. We further define:

$$\begin{aligned}\alpha_T\left(\theta\right) &= \arg\min_{\alpha\in[0,1]}\Sigma_T(\alpha;\theta)\\ \widehat{\alpha}_{TM}\left(\theta\right) &= \arg\min_{\alpha\in[0,1]}\widehat{\Sigma}_{TM}^{MC}(\alpha;\theta)\end{aligned}$$

The estimator of $\alpha_T\left(\theta_0\right)$ we are considering is $\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right)$ where $\widehat{\theta}^1$ is a $\sqrt{T}$-consistent first step estimator of θ_0. The following lemma is useful for the subsequent derivations.

Lemma 17 *Under Assumptions 1 to 5, $\widehat{\theta}_T(\alpha;\theta_0)$ is once continuously differentiable with respect to α and twice continuously differentiable with respect to θ_0 and $\alpha_T(\theta_0)$ is a continuous function of θ_0.*

Proof of Lemma 17 Note that $\widehat{Q}_T\left(\alpha,\theta\right)$ involves

$$K_{\alpha T}^{-1}\widehat{h}_T(.,\theta)=\sum_{j=1}^{T}\frac{\widehat{\mu}_j}{\alpha+\widehat{\mu}_j^2}\left\langle\widehat{h}_T(.,\theta),\widehat{\phi}_j\right\rangle\widehat{\phi}_j$$

where $\widehat{\phi}_j$ is the eigenfunction of K_T associated with the eigenvalue $\widehat{\mu}_j$. By assumption 3, the moment function $\widehat{h}_T(.,\theta)$ is three times continuously differentiable with respect to θ, the argument with respect to which we minimize the objective function of the CGMM. By assumption 5, $x_t = x\left(x_{t-1},\theta_0,\varepsilon_t\right)$ where r is three times continuously differentiable with respect to θ_0 (the true unknown parameter) and ε_t is an IID white noise whose distribution does not depend on θ_0. Thus as an exponential function of x_t, the moment function is also three times continuously differentiable with respect to θ_0. Thus Assumptions 3 and 5 imply that the objective function of the CGMM is three times continuously differentiable with respect to θ and θ_0. Now we turn our attention toward the differentiability with respect to α. It is easy to check that

$$\frac{\partial^3 K_{\alpha T}^{-1}\widehat{h}_T(.,\theta)}{\partial\alpha^3}=\widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta)$$

where $\widetilde{K}_{\alpha T} \equiv -\left(K_T^2 + \alpha_T I\right)^{-2} K_T$ which is well defined on $L^2(\pi)$ for α_T fixed. When α_T goes to zero, we have to be more careful. We check that

$$\left|\left\langle \widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta)\right\rangle\right|$$

is bounded. We have

$$\begin{aligned}
& \left|\left\langle \widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta)\right\rangle\right| \\
\leq & \left\|\widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta)\right\| \left\|\widehat{h}_T(.,\theta)\right\| \\
\leq & \left\|\left(K_T^2 + \alpha_T I\right)^{-2} K_T\right\| \left\|\widehat{h}_T(.,\theta)\right\|^2 \\
= & \underbrace{\left\|\left(K_T^2 + \alpha_T I\right)^{-3/2}\right\|}_{\leq \alpha_T^{-3/2}} \underbrace{\left\|\left(K_T^2 + \alpha_T I\right)^{-1/2} K_T\right\|}_{\leq 1} \underbrace{\left\|\widehat{h}_T(.,\theta)\right\|^2}_{=O_p(T^{-1})} \\
= & O_p\left(\alpha_T^{-3/2} T^{-1}\right) \\
= & o_p(1),
\end{aligned}$$

where the last equality follows from Theorem 2(ii). This shows that $\widehat{Q}_T(\alpha,\theta)$ is once continuously differentiable with respect to α and three times continuously differentiable with respect to (θ,θ_0). By the implicit function theorem,

$$\widehat{\theta}_T(\alpha;\theta_0) = \arg\min_{\theta} \widehat{Q}_T(\alpha,\theta)$$

is once continuously differentiable with respect to α and twice continuously differentiable w.r.t. θ_0. The MSE $\widehat{\theta}_T(\alpha;\theta_0)$ is an expectation of a quadratic function in $\widehat{\theta}_T(\alpha;\theta_0)$:

$$\Sigma_T(\alpha) = TE\left[\left(\widehat{\theta}_T(\alpha;\theta_0) - \theta_0\right)'\left(\widehat{\theta}_T(\alpha;\theta_0) - \theta_0\right)\right],$$

hence $\Sigma_T(\alpha;\theta_0)$ is also once continuously differentiable w.r.t. α and twice continuously differentiable w.r.t. θ_0. Now, the Maximum theorem implies that

$$\alpha_T(\theta_0) = \arg\min_{\alpha\in[0,1]} \Sigma_T(\alpha;\theta_0)$$

is continuous w.r.t. θ_0.■

Proof of Theorem 3

Using (**??**), we see that

$$\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} = \frac{c(\widehat{\theta}^1)}{c(\theta_0)}.$$

Moreover by Lemma 17, $\alpha_T(\theta)$ and hence c are continuous functions of θ. Since $\widehat{\theta}^1$ is a consistent estimator of θ_0, the continuous mapping theorem implies that

$$\frac{c(\widehat{\theta}^1)}{c(\theta_0)} \xrightarrow{P} 1.$$

■

Proof of Theorem 4

We can write $\widehat{\Sigma}_{TM}^{MC}(\alpha;\theta_0) = \frac{1}{M}\sum_{j=1}^{M} m_j(\alpha)$ where

$$m_j(\alpha) = \left(\widehat{\theta}_T^j(\alpha;\theta_0) - \theta_0\right)' \left(\widehat{\theta}_T^j(\alpha;\theta_0) - \theta_0\right)$$

are IID (across j) and continuous in α. We have

$$\widehat{\Sigma}_{TM}^{MC}(\alpha;\theta_0) - \Sigma_T(\alpha;\theta_0) = \frac{1}{M}\sum_{j=1}^{M}\left(m_j(\alpha) - E(m_j(\alpha))\right).$$

If we can show that there exists a function $b_T > 0$ independent of α such that

$$\left\|\frac{\partial m_j(\alpha)}{\partial\alpha}\right\| < b_T, \tag{2.44}$$

and $E(b_T) < \infty$, then, by Lemma 2.4 of Newey and McFadden (1994), we would have

$$\sup_{\alpha\in[0,1]}\left|\widehat{\Sigma}_{TM}^{MC}(\alpha;\theta_0) - \Sigma_T(\alpha;\theta_0)\right| = O_p\left(M^{-1/2}\right).$$

Hence, it would follow from Theorem 2.1 of Newey and McFadden (1994) that

$$\widehat{\alpha}_{TM}(\theta_0) - \alpha_T(\theta_0) = O_p\left(M^{-1/2}\right).$$

As T is fixed, $\alpha_T(\theta_0)$ is bounded away from zero and we would obtain

$$\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)} - 1 = O_p\left(M^{-1/2}\right).$$

To prove inequality (2.44), we first compute:

$$\frac{\partial m(\alpha)}{\partial \alpha} = 2\frac{\partial \widehat{\theta}_T(\alpha;\theta_0)}{\partial \alpha}' \left(\widehat{\theta}_T(\alpha;\theta_0) - \theta_0\right)$$

where by the implicit function theorem:

$$\frac{\partial \widehat{\theta}_T(\alpha;\theta_0)}{\partial \alpha} = -\left[\frac{\partial^2 \widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\theta'}\right]^{-1} \frac{\partial^2 \widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\alpha}.$$

The expressions involved are:

$$\begin{aligned}
\frac{\partial^2 \widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\theta'} &= \left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\theta_0), \widehat{G}_T(.,\theta_0)\right\rangle + \left\langle K_{\alpha T}^{-1}\widehat{H}_T(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle, \\
\frac{\partial^2 \widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\alpha} &= \left\langle K_{\alpha T}^{*}\widehat{G}_T(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle + \left\langle K_{\alpha T}^{*}\widehat{h}_T(.,\theta_0), \widehat{G}_T(.,\theta_0)\right\rangle
\end{aligned}$$

and $K_{\alpha T}^{*} \equiv -\left(K_T^2 + \alpha I\right)^{-2} K_T$.

Next recall that for fixed T, $\alpha_T(\theta_0)$ is bounded away from zero because the objective function of the CGMM as well as the MSE of the CGMM estimator diverge as α approach zero and T is fixed. Thus there exists at least one sequence $\underline{\alpha}_T$ such that $\alpha_T(\theta_0) \geq \underline{\alpha}_T$ for all T so that the problem of choosing α may be written as:

$$\alpha(\theta_0) = \arg\min_{\alpha \in [\underline{\alpha}_T, 1]} \Sigma_T(\alpha;\theta_0)$$

For one such sequence,

$$\left\|b_T(\overline{\alpha}_T)\right\| = \left\|\frac{\partial m(\overline{\alpha}_T)}{\partial\alpha}\right\| < \infty$$

where $\overline{\alpha}_T = \arg\sup_{\alpha\in[\underline{\alpha}_T,1]} \left\|\frac{\partial m(\alpha)}{\partial\alpha}\right\|$ because $\frac{\partial^2 \widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\theta'}$ and $\frac{\partial^2 \widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\alpha}$ are continuous with respect to α. Hence the result.■

Proof of Theorem 5

We first make the following decomposition

$$\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1 = \left(\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1\right) + \left(\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1\right)\left(\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\widehat{\theta}^1)} - 1\right) + \left(\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\widehat{\theta}^1)} - 1\right).$$

By Theorem 3, $\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1 = O_p(T^{-1/2})$. According to Theorem 4, $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)} - 1 = O_p(M^{-1/2})$ for T fixed. We can keep this rate of convergence if one takes a sequential limit, first in M and second in T. Noting that the product of $\left(\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1\right)$ and $\left(\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\widehat{\theta}^1)} - 1\right)$ is negligible with respect to either of the other two terms, the result follows, that is:

$$\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1 = O_p(T^{-1/2}) + O_p(M^{-1/2}).$$

■

Numerical algorithms: Computing the objective function of the CGMM

Let $L^2(\pi)$ be defined by (2.9). It turns out that $h_t(\theta,\tau) \in L^2(\pi)$ for any finite measure π. Thus without loss of generality, we can take $\pi(\tau)$ to be the standard normal density up to a multiplicative constant: $\pi(\tau) = \exp\{-\tau'\tau\}$. Recall that:

$$K_T \hat{h}_T(\theta,\tau) = \int_{R^d} \widehat{k}_T(s,\tau)\,\widehat{h}_T(\theta,s)\exp\{-s's\}\,ds.$$

This kind of integral can easily be well approximated numerically using the Gauss-Hermite quadrature. This amounts to find m points $s_1, s_2, ...s_m$ and weighs $\omega_1, \omega_2, ...\omega_m$ such that:

$$\int_{R^d} p(s)\exp\{-s's\}dx = \sum_{k=1}^{m} \omega_k p(s_k)$$

for any polynomial function $p(.)$ of order smaller than or equal to $2m-1$. See for example Liu and Pierce (1994).

If f is differentiable at any order (for example an analytic function), it can be shown that for any positive ε arbitrarily small, there exist m such that:

$$\left|\int_{R^d} f(s)\exp\{-s's\}dx - \sum_{k=1}^{m} \omega_k f(s_k)\right| < \varepsilon.$$

More importantly, the choice of the quadrature point does not depend on the function f. The quadrature points and weights are determined by solving:

$$\int s^l \exp\{-s^2\} ds = \sum_{k=1}^{n} \omega_k s_k^l \text{ for all } l = 1, ..., 2n-1$$

Applying that method to evaluate the above integral, we get

$$K_T \widehat{h}_T(\theta, \tau) \approx \sum_{k=1}^{m} \omega_k \widehat{k}_T(s_k, \tau) \widehat{h}_T(\theta, s_k).$$

Let $\widehat{h}_T(\theta)$ denote the vector $\left(\widehat{h}_T(\theta, s_k), \widehat{h}_T(\theta, s_k), ..., \widehat{h}_T(\theta, s_k)\right)'$ and $\widehat{W}_T$ denote the matrix with elements: $W_{jk} = \omega_k \widehat{k}_T(s_k, s_j)$. Thus we can simply write:

$$K_T \widehat{h}_T(\theta) \approx \widehat{W}_T \widehat{h}_T(\theta).$$

For any given level of precision, the matrix $\widehat{W}_T$ can be looked at as the best finite dimensional reduction of the operator K_T. From the spectral decomposition of $K_{\alpha T}^{-1}$, it is easy to deduce the approximation:

$$K_{\alpha T}^{-1} \widehat{h}_T(\theta) \approx \left(\widehat{W}_T^2 + \alpha I\right)^{-1} \widehat{W}_T \widehat{h}_T(\theta) \equiv \widetilde{h}_T(\theta).$$

Finally, the objective function of the CGMM is computed as:

$$\begin{aligned} \left\langle K_{\alpha T}^{-1} \widehat{h}_T(\theta, .), \widehat{h}_T(\theta, .)\right\rangle &= \int \left| K_{\alpha T}^{-1/2} \widehat{h}_T(\theta, \tau)\right|^2 \exp\{-\tau' \tau\} d\tau \\ &\approx \sum_{k=1}^{m} \omega_k \left|\widetilde{h}_T(\theta, s_k)\right|^2 \end{aligned}$$

where $\widetilde{h}_T(\theta, s_k)$ is the k^{th} component of $\widetilde{h}_T(\theta)$.

Chapter 3

Applications of the Characteristic Function Based Continuum GMM in Finance

Note: Cet article dont je suis l'unique auteur est actuellement sous évaluation pour publication dans "*Computational Statistics & Data Analysis*". Nous remercions Marine Carrasco pour ses commentaires utiles.

Mots-Clés: Continuum of Moments Conditions, Simulation, Stable Distribution, Autoregressive variance Gamma model

3.1 Introduction

For many interesting financial econometric models, the characteristic functions (CF) is available in closed form while the likelihood functions is not. This is for example the case of the stable distribution, or a discretely sampled continuous time process. Exceptionally, a discrete sample from the square root diffusion model has a closed form conditional likelihood, but its expression is an infinite sum that must be truncated in practice. In the discrete time literature also, many models (e.g the variance gamma model) has known closed form likelihood functions that are not convenient for numerical optimization. In these situations, the use of the CF for inference is attractive. In fact, two random variables have the same distribution if and only if their CF coincide on the whole real line. This suggests that an inference method that adequately exploits the information content of the CF have the potential to achieve the same level of efficiency as a likelihood-based approach. One such inference method proposed by Carrasco and Florens (2000) for IID models exploits the whole continuum of moment conditions based on the difference between the empirical and theoretical characteristic functions to estimate a model. The method has been extended by Carrasco, Chernov, Florens and Ghysels (2007) to deal with Markov and dependent models. Other leading works in this area include Singleton (2001), Knight and Yu (2002), Knight, Satchell and Yu (2002) and Chacko and Viceira (2003). A good review of this literature is provided by Yu (2004).

The goal of this paper is to make the CF based GMM with a continuum of moment conditions (henceforth CGMM) accessible to applied researchers. Our focus will be on the approach proposed by Carrasco and Florens (2000) and its extension by Carrasco and al (2007). First of all, we review the theory underlying the CGMM. We recall the main assumptions that are useful for the consistency and asymptotic normality of the CGMM estimator. Next, we discuss in details the important steps of the implementation of the CGMM in practice. Finally, we provide a simulation study with the stable distribution and an empirical application with the autoregressive variance gamma model.

The Stable Distribution have been introduced in finance to fit the asymmetry and fat tails observed empirically in the distributions of assets returns (Mandelbrot (1963) or McCulloch (1986)). Different parametrizations coexist in the literature for this distribution, and this has sometimes been a source of confusion. In the common parametrization adopted in this paper, the stable distribution has a stability parameter $\alpha \in]0, 2]$, a skewness parameter $\beta \in [-1, 1]$, a scale parameter $\sigma > 0$ and a location parameter $\mu \in \mathbb{R}$. The moments of order larger than α do not exist for the stable distribution when $\alpha < 2$. When $\alpha = 2$, all the moments exist but the asymmetry parameter β is no longer identifiable. We present an overview of the Stable Distribution and discuss its different parametrizations and simulation strategies. Our results show that the CGMM produces reliable estimators for the parameters of the model. However, the variance of the estimators cannot be computed analytically when the vector of parameters is close to the non identification region (that is, when α is close to 2). One then has to rely on Monte Carlo simulations to build confidence intervals.

The fact that the asymmetry and fat -tailedness of the stable distribution vanish when its variance exist is a limit to its use for the purpose of modeling assets returns. A simple solution to this limitation consists in modeling the variance of the returns as a Gamma variable. This yields the variance gamma model. The symmetric variance gamma model has been propose by Madan and Senata (1990). Madan, Carr and Chang (1998) extend the basic model to include asymmetry. These two models unfortunately assume that the variance is IID. We relax this assumption by assuming that the variance follows the autoregressive gamma process studied in Gourieroux and Jasiak (2005). The resulting model for assets returns is termed the *"autoregressive variance gamma model"*. We propose an estimation strategy in two steps. In the first step, we fit the autoregressive gamma model to a consistent estimator of the integrated variance (used as a proxy for the true variance). And secondly, we estimate a relation between the returns and the volatility that allows to disentangle the risk premium from the leverage effect. An empirical application with the Alcoa index (listed in the Dow Jones Industrial Average) shows that investors require a positive premium for bearing the expected risk while a possibly time varying negative premium is attached to the unexpected risk.

We organize the rest of the paper as follows. The next section reviews the main theoretical results on the CGMM. In section 3, we discuss the numerical aspects of its implementation. In section 4, we present a simulation study of the performance of the CGMM to estimate the stable distribution. In section 5, we present and estimate the autoregressive variance gamma model with real data. Section 6 concludes. A few graphs and mathematical formulas are left in appendix.

3.2 The CGMM: a Brief Theoretical Review

In this section, we present the theoretical framework underlying the CGMM estimation. The first subsection reviews the IID framework while the second subsection deals with the dependent case. In the third subsection, we discuss the assumptions needed in order for the CGMM estimator to have good asymptotic properties .

3.2.1 The CGMM in the IID Case

Let $(x_1, \ldots, x_T)$ be an IID sample of an $m-$dimensional vector process whose CF is given by $E^{\theta_0}\left(e^{i\tau x_t}\right) = \varphi(\tau, \theta_0)$, where θ_0 is a finite dimensional parameter that fully characterizes the distribution of $\{x_t\}$. By the definition of $\varphi(\tau, \theta_0)$, the following set of moment functions can be considered for the purpose of estimating the parameter θ_0:

$$h_t(\tau, \theta_0) = e^{i\tau x_t} - \varphi(\tau, \theta_0), \text{ for all } \tau \in \mathbb{R}^m. \tag{3.1}$$

Note that these moment functions are indexed by $\tau \in \mathbb{R}^m$, and hence we have a continuum of moment conditions. Because the CF contains the same information as the likelihood function, an efficient use of the whole continuum of moment conditions can permit to achieve the maximum likelihood efficiency.

As in Feuerverger and McDunnough (1981b), Singleton (2001) or Chacko and Viceira (2003), one may choose to use the GMM based on a discrete subset of the continuum (3.1). The implementation of such a GMM is done in the standard way, as proposed by Hansen (1982). More precisely, let $\{h_t(\tau_k, \theta_0)\}_{k=1}^{q}$ be a discrete subset of $\{h_t(\tau, \theta_0), \tau \in \mathbb{R}^m\}$, and define the vector $g_t(\theta_0)$ by:

$$g_t(\theta_0) = (\operatorname{Re} h_t(\tau_1, \theta_0), ..., \operatorname{Re} h_t(\tau_q, \theta_0), \operatorname{Im} h_t(\tau_1, \theta_0), ..., \operatorname{Im} h_t(\tau_q, \theta_0))'$$

The objective function of the GMM may be defined as:

$$\min \widehat{g}(\theta_0)' \widehat{W}^{-1} \widehat{g}(\theta_0)$$

where $\widehat{g}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta_0)$ and $\widehat{W} = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta_0) g_t(\theta_0)'$.

Feuerverger and McDunnough (1981b) claim that the asymptotic variance of the resulting estimator can be made arbitrarily close to the Cramer-Rao bound by selecting the grid $(\tau_1, ..., \tau_q)$ sufficiently refined and extended. This confirms in fact that the maximum likelihood efficiency can be achieved only by using the whole continuum of moment function $\{h_t(\tau, \theta_0), \tau \in \mathbb{R}^m\}$. However, as one refines and extends the grid the discrete set of moment conditions converge to the continuous moment function $h_t(\tau, \theta) = e^{i\tau x_t} - \varphi(\tau, \theta)$, $\tau \in \mathbb{R}^m$, while the covariance matrix $\widehat{W}$ converges to the covariance operator associated with that moment function. Moreover, one should note that $2q \leq T$ is a necessary condition for the covariance matrix $\widehat{W}$ to be invertible.

Different methods that match continuously the empirical CF to its theoretical counterpart has been proposed as far back as in Press (1972) and Paulson and al. (1975), but the ideal objective function has been introduced more recently by Carrasco and Florens (2000). That objective function is given by:

$$Q = \left\langle K^{-1/2}\widehat{h}_T(.,\theta), K^{-1/2}\widehat{h}_T(.,\theta)\right\rangle, \tag{3.2}$$

where $\langle\ .\ ,\ .\ \rangle$ is a scalar product on the Hilbert space of square integrable functions, and K is a linear operator.

To be more precise, let π be a probability density function on $\mathbb{R}^m$ and $L^2(\pi)$ denote the Hilbert space of complex valued functions that are square integrable with respect to π, that is:

$$\mathbf{L}^2(\pi) = \{f : \mathbb{R}^m \to \mathbf{C} \text{ such that } \int f(\tau)\overline{f(\tau)}\pi(\tau)d\tau < \infty\}.$$

Interestingly, $h_t(\tau,\theta)$ is bounded in modulus and consequently belongs to $L^2(\pi)$ for all $\theta \in \Theta$ and any choice of π. The scalar product $\langle .,. \rangle$ on $\mathbf{L}^2(\pi) \times \mathbf{L}^2(\pi)$ is defined by:

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

Carrasco and Florens (2000) show that the maximum likelihood efficiency is achieved when K is the asymptotic covariance operator associated with the moment function $h_t(\tau,\theta_0)$. The kernel of K is given by:

$$k(s,\tau) = E\left[h_t(s,\theta)\overline{h_t(\tau,\theta)}\right] \tag{3.3}$$

and for any function $f \in \mathbf{L}^2(\pi)$, $Kf = \int k(s,\tau)f(s)\pi(s)\,ds$. It can be shown that $Kf \in \mathbf{L}^2(\pi)$ for all $f \in \mathbf{L}^2(\pi)$ so that $K^{-1}f$ exists for all $f \in \mathbf{L}^2(\pi)$.

In practice, one has to use the empirical counterpart K_T of K. The operator K_T is the one got by replacing the kernel $k(s,\tau)$ by a consistent estimator. A natural estimator of $k(s,\tau)$ is given by:

$$k_T(s,\tau,\widehat{\theta}^1) = \frac{1}{T}\sum_{t=1}^{T} h_t(s,\widehat{\theta}^1)\overline{h_t(\tau,\widehat{\theta}^1)}, \tag{3.4}$$

where $\widehat{\theta}^1$ is a consistent first step estimator of θ_0. In the specific case of IID data, an estimator of the kernel that does not use a first step estimator is given by:

$$k_T(s,\tau) = \frac{1}{T}\sum_{t=1}^{T}\left(e^{isx_t} - \widehat{\varphi}_T(s)\right)\overline{\left(e^{isx_t} - \widehat{\varphi}_T(s)\right)}. \tag{3.5}$$

where $\widehat{\varphi}_T(s) = \frac{1}{T}\sum_{t=1}^{T} e^{isx_t}$.

It turns out that K_T is not invertible on the whole $\mathbf{L}^2(\pi)$ space. Carrasco and Florens (2000) then proposed to work with a generalized inverse of type[1] $K_{T,\lambda}^{-1} = \left(K_T^2 + \lambda_T I_T\right)^{-1} K_T$, where I_T is the identity operator and λ_T is a regularization parameter that is function of the sample size T. The feasible CGMM estimator is thus given by:

$$\begin{aligned} \widehat{\theta}_{T,\lambda} &= \arg\min \widehat{Q}_{T,\lambda}, \text{ where} \\ \widehat{Q}_{T,\lambda} &= \left\langle K_{T,\lambda}^{-1/2}\widehat{h}_T(.,\theta), K_{T,\lambda}^{-1/2}\widehat{h}_T(.,\theta)\right\rangle \end{aligned} \tag{3.6}$$

and $K_{T,\lambda}^{-1/2} = \left(K_T^2 + \lambda_T I_T\right)^{-1/2} K_T^{1/2}$. It is shown in Carrasco and Florens (2000) that the ML efficiency is achieved when λ_T converges to zero at a certain rate as the sample size diverges to infinity. We discuss the assumption underlying these results in Section 2.3.

3.2.2 The CGMM with Dependent Data

When $\{x_t\}$ is Markov instead of being IID, it may not be possible to identify θ_0 from the marginal CF. In this case, Carrasco, Chernov, Florens and Ghysels (2007) proposed to use moment functions based on the conditional CF:

$$h_t(\tau,\theta) = \left[e^{isx_{t+1}} - \varphi(\theta,s,x_t)\right] e^{irx_t}, \tag{3.7}$$

where $\varphi(s,\theta,x_t) = E^{\theta}(e^{isx_{t+1}}|x_t)$, $\tau = (s,r) \in \mathbb{R}^{2m}$. In the above expression, the manifold $\{e^{irx_t}, r \in \mathbb{R}^m\}$ is used as instruments. Carrasco, Chernov, Florens and Ghysels (2007) show that these instruments are optimal given the Markov assumption.

There also exist many interesting situations where the process $\{x_t\}$ is mixing instead of being Markov or IID. In a typical stochastic volatility models for instance, the joint process

[1] This is the Tikhonov regularization. Other types of regularized inverse can also be used (e.g spectral cutt-off).

of the observed return and the latent volatility is Markov but the return process alone is not. In that case, the idea is to use the moment conditions built from the joint CF:

$$h_t(\tau,\theta) = e^{i\tau Y_t} - E^{\theta}(e^{i\tau Y_t}), \tau \in \mathbb{R}^{mp} \tag{3.8}$$

where $Y_t = (x_t, x_{t-1}, ..., x_{t-p+1})$. In theory, the larger p the more efficient the estimator. But in practice, the quest for efficiency must be balanced with the computing cost. For more discussions on this point, see Feuerverger (1990), Carrasco and Florens (2002), Jiang and Knight (2002), Yu (2004) and Carrasco, Chernov, Florens and Ghysels (2007).

The objective function of the CGMM for Markov and dependent models has the same expression as in (3.2), except that the kernel of the asymptotic covariance operator K associated with the moments conditions is now given by (see Carrasco, Chernov, Florens and Ghysels (2007)):

$$\begin{aligned} k(s,\tau) &= E\left[h_t(s,\theta)\overline{h_t(\tau,\theta)}\right] \\ &\quad + \sum_{j=1}^{\infty} E\left[h_t(s,\theta)\left(\overline{h_{t-j}(\tau,\theta)} + \overline{h_{t+j}(\tau,\theta)}\right)\right] \end{aligned} \tag{3.9}$$

Note that moments conditions of type (3.1) are IID while those of type (3.7) are martingale difference sequence. Hence in the Markov case, $k(s,\tau)$ reduces to (3.3) and can thus be estimated by (3.4). On the other hand, the moment conditions described by (3.8) are autocorrelated even if the process $\{x_t\}$ is Markov. In the latter case, $k(s,\tau)$ can be estimated as in Newey and West (1987) or Andrews and Monahan (1992) using the Bartlett kernel:

$$\begin{aligned} k_T(s,\tau,\widehat{\theta}^1) &= \frac{1}{T}\sum_{t=1}^{T} h_t(s,\widehat{\theta}^1)\overline{h_t(\tau,\widehat{\theta}^1)} \\ &\quad + \sum_{j=1}^{J_T}\left(1-\frac{j-1}{J_T}\right)\sum_{t=1}^{T} h_t(s,\widehat{\theta}^1)\left(\overline{h_{t-j}(\tau,\widehat{\theta}^1)} + \overline{h_{t+j}(\tau,\widehat{\theta}^1)}\right), \end{aligned} \tag{3.10}$$

where $\widehat{\theta}^1$ is a consistent first step estimator of θ_0 and J_T is a bandwidth that is increasing in T. Again, the operator K_T with kernel $k_T(s,\tau,\widehat{\theta}^1)$ is not invertible on the whole reference space, and the feasible CGMM estimator is defined in the same fashion as in (3.6).

In the sequel, we shall focus on the IID and Markov case and use the generic notation

$h_t(\tau,\theta)$, $\tau \in \mathbb{R}^d$, where $d = m$ for moments conditions of type (3.1) and $d = 2m$ for moments conditions of type (3.7).

3.2.3 Basic Assumptions of the CGMM

To derive the theoretical properties of the CGMM estimator, the following regularity conditions are assumed.

Assumption 1: The p.d.f π involved in the definition of scalar product $\langle\ .\ ,\ .\ \rangle$ is strictly positive on $\mathbf{R}^d$ and admits all its moments.

Assumption 2: The equation

$$E^{\theta_0}\left[h_t(\tau,\theta)\right] = 0 \text{ for all } \tau \in \mathbf{R}^d, \pi - \text{almost everywhere},$$

where E^{θ_0} denotes the expectation with respect to the distribution of x_t for $\theta = \theta_0$, has a unique solution θ_0 which is an interior point of a compact set $\mathbf{\Theta}$.

Assumption 3: $h_t(\tau,\theta)$ is three time continuously differentiable with respect to θ.

Assumption 4: For all θ, $E^{\theta_0}\left[h_T(.,\theta)\right]$ and its first three derivatives with respect to θ belong to the range of K^β for some $\beta \geq 1/2$.

Assumption 5: The random variable x_t is stationary and satisfies $x_t = x\left(\theta_0, \varepsilon_t, Z_{t-1}\right)$ where $x\left(., \varepsilon_t, Z_{t-1}\right)$ is three times continuously differentiable with respect to θ_0, ε_t is a IID white noise whose distribution is known and does not depend on θ_0, and Z_{t-1} can only contain lagged values of x_t.

Assumption 1 ensures that the norm associated with the scalar product $\langle .,. \rangle$ is well defined while Assumption 2 is a global identification requirement. The CGMM estimator is still well defined if Assumption 3 is weaker, for example if $h_t(\tau,\theta)$ is only once continuously differentiable, but the derivation of some of the asymptotic properties of the estimator become difficult. Assumption 4 ensures that the limit of the objective function as T goes to infinity is well defined. The real number β in this assumption is the level of regularity of $E^{\theta_0}\left[h_T(.,\theta)\right]$ with respect to the operator K, that is, the largest real number such that $\left\|K^{-\beta}E^{\theta_0}\left[h_T(.,\theta)\right]\right\| < \infty$.

Under assumptions 1 and 2, the estimator of the covariance operator satisfies in the IID and Markov case:

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

The regularized inverse $K_{T,\lambda}^{-1/2}$ has the property that for any function f in the range of $K^{1/2}$, $K_{T,\lambda}^{-1/2}f$ converges to $K^{-1/2}f$ as T goes to infinity and λ_T goes to zero at some rate.

In the IID and Markov case, assumptions 1 to 4 ensure that the CGMM estimator satisfies:

$$T^{1/2}\left(\widehat{\theta}_T(\lambda_T) - \theta_0\right) \xrightarrow{L} N(0, I_{\theta_0}^{-1}). \tag{3.11}$$

as T and $\lambda_T^2 T$ go to infinity and λ_T goes to zero, where $I_{\theta_0}^{-1}$ denote the inverse of the Fisher Information Matrix. Carrasco and Kotchoni (2008) show that this result still holds even when $\lambda_T^{3/2}T$ diverges as T goes to infinity.

Assumption 5 is not crucial for the good properties of the CGMM. It has been used in Carrasco and Kotchoni (2008) to derive the properties of the optimal sequence of regularization parameters λ_T. A similar assumption is also used in Gourieroux, Monfort and Renault (1993) to derive the properties of indirect inference estimators.

3.3 The CGMM in Practice

In this section, we discuss two numerical methods to evaluate the objective function of the CGMM. The first method is based on Gauss-Hermite quadratures while the second uses Monte Carlo integration. We show how to compute the variance of the CGMM estimator and review the simulation based selection of the regularization parameter λ_T.

3.3.1 Computing the Objective Function by Quadrature Method

The challenge in implementing the CGMM is the accurate computation of the multiple integrals embedded in its objective function:

$$\widehat{Q}_T = \int_{R^d} \left| K_{T,\lambda}^{-1/2}\widehat{h}_T(\tau,\theta)\right|^2 \pi(\tau)d\tau$$

To start with, let us consider the univariate case $d = 1$, and assume that a function $f(\tau,\theta)$ is continuously differentiable up to order $2n$. Then $f(\tau,\theta)$ can be well approximated by a polynomial function of τ, that is:

$$f(\tau,\theta)=\sum_{k=0}^{2n-1}a_k(\theta)\,\tau^k+\varepsilon(\tau,\theta) \tag{3.12}$$

where the residual of the polynomial approximation $\varepsilon(\tau,\theta)$ is negligible for large n. In that case, the weighting function $\pi(s)=\exp(-s^2)$ is quite convenient to work with. We have:

$$K_T\hat{h}_T(\tau,\theta)=\int k_T(\tau,s)\,\hat{h}_T(s,\theta)\exp\left(-s^2\right)ds. \tag{3.13}$$

Interestingly, $K_T\hat{h}_T(\tau,\theta)$ can be well approximated by the Gauss-Hermite quadrature. This amounts to find n points $(s_1,...,s_n)$ and weighs $(\omega_1,...,\omega_n)$ such that:

$$\int p(s)\exp\{-s^2\}ds=\sum_{k=1}^{n}\omega_k p(s_k) \tag{3.14}$$

for any polynomial function $p(.)$ of order smaller or equal to $2n-1$. We thus have:

$$\left|\int f(s)\exp\{-s^2\}ds-\sum_{k=1}^{n}\omega_k f(s_k)\right|=\left|\int\varepsilon(\tau,\theta)\exp\left(-s^2\right)ds\right|$$

If $K_T\hat{h}_T(\tau,\theta)$ is analytic as a function of τ, the residual $\left|\int\varepsilon(\tau,\theta)\exp(-s^2)\,ds\right|$ can be made arbitrarily small by increasing n. Note that the choice of the quadrature points and weights does not depend on the particular function $f(\tau,\theta)$. The quadrature points and weights are determined by solving:

$$\int s^l\exp\{-s^2\}ds=\sum_{k=1}^{n}\omega_k s_k^l \text{ for all } l=1,...,2n-1$$

Applying this quadrature method to (3.13) yields:

$$K_T\widehat{h}_T(\theta)\approx\widehat{W}_T\widehat{h}_T(\theta), \tag{3.15}$$

where $\widehat{W}_T$ is the matrix with (j,k) elements $W_{jk}=\omega_k k_T(s_j,s_k)$, and:

$$\widehat{h}_T(\theta)=\left(\widehat{h}_T(s_1,\theta),...,\widehat{h}_T(s_n,\theta)\right)'.$$

For any given level of precision, the matrix $\widehat{W}_T$ can be looked at as the best finite dimensional reduction of the operator K_T. The resulting approximation of $K_{T,\lambda}^{-1/2}$ is:

$$K_{T,\lambda}^{-1/2} = \left(\widehat{W}_T^2 + \lambda I\right)^{-1/2} \widehat{W}_T^{1/2},$$

that is:

$$K_{T,\lambda}^{-1/2}\widehat{h}_T(\theta) \approx \left(\widehat{W}_T^2 + \lambda I\right)^{-1/2} \widehat{W}_T^{1/2}\widehat{h}_T(\theta). \tag{3.16}$$

Substituting for $K_{T,\lambda}^{-1/2}\widehat{h}_T(\theta)$ in the objective function of the CGMM yields:

$$\begin{aligned} \left\langle K_{T,\lambda}^{-1/2}\widehat{h}_T(.,\theta), K_{T,\lambda}^{-1/2}\widehat{h}_T(.,\theta)\right\rangle &= \int \left|K_{T,\lambda}^{-1/2}\widehat{h}_T(\tau,\theta)\right|^2 \exp\left\{-\tau^2\right\} d\tau \\ &\approx \sum_{k=1}^{n} \omega_k \left|K_{T,\lambda}^{-1/2}\widehat{h}_T(s_k,\theta)\right|^2 \end{aligned} \tag{3.17}$$

where $K_{T,\lambda}^{-1/2}\widehat{h}_T(s_k,\theta)$ is the k^{th} element of the vector $K_{T,\lambda}^{-1/2}\widehat{h}_T(\theta)$ given in (3.16).

In theory, the extension of the above quadrature method to the multivariate case is straightforward. When $\tau \in \mathbf{R}^d$, the $d-$dimensional set of multivariate quadrature points is given by the Cartesian product:

$$D = \left\{\tau = \left(\tau_{(1)}, ..., \tau_{(d)}\right) : \tau_{(i)} \in \{s_1, ..., s_n\} \text{ for all } i = 1 \text{ to } d\right\}, \tag{3.18}$$

where $\{s_1, ..., s_n\}$ is the set of n univariate quadrature points with weights $\{\omega_1, ..., \omega_n\}$, and $\tau_{(i)}$ is the i^{th} coordinate of τ. Associated with each $\tau \in D$ is the weight:

$$\overline{\omega}(\tau) = \omega(\tau_1)\,\omega(\tau_2)\,...\omega(\tau_d) \tag{3.19}$$

where $\omega(\tau_i) = \omega_k$ if $\tau_{(i)} = s_k$, $i = 1, ..., d$.

The multivariate Gauss-Hermite quadrature has the undesirable feature that $Card(D) = n^d$. This raises a "curse of dimensionality" because the size of the matrix $\widehat{W}_T$ is precisely n^d while we need to take n quite large ($n \approx 10$) to accurately evaluate the objective function of the CGMM. Because $\widehat{W}_T$ must be inverted at each iteration of the optimization algorithm, the CGMM becomes virtually unfeasible by quadrature method when $d \geq 3$. We shall thus limit ourselves to the case $d \leq 2$ in the sequel, leaving the discussion on the large d case for further research.

3.3.2 Computing the Objective Function by Monte Carlo Integration

This approach does not require the differentiability of $f(\tau,\theta) = k_T(\tau,s)\,\widehat{h}_T(\tau,\theta)$ and relies on the alternative formula of the objective function of the CGMM provided in Carrasco, Chernov, Florens and Ghysels (2007):

$$Q_T = v(\theta)' \left[\lambda I_T + \widehat{C}_T^2\right]^{-1} \overline{v(\theta)}, \tag{3.20}$$

where $\widehat{C}_T \equiv \widehat{C}_T\left(\widehat{\theta}_1\right)$ is the square matrix of size T with (t,l) element $c_{t,l}/(T-q)$, I_T is the identity matrix of size T, and $v(\theta) = (v_1, ..., v_T)'$ with:

$$v_t = \int \overline{h_t(\tau, \widehat{\theta}^1)} h_T(\tau,\theta)\pi(\tau)\,d\tau, \tag{3.21}$$

$$c_{t,l} = \int \overline{h_t(\tau, \widehat{\theta}^1)} h_l(\tau,\widehat{\theta}^1)\pi(\tau)\,d\tau, \tag{3.22}$$

The main drawback of the above expressions lies in that it involves the inverse of the matrix $\widehat{C}_T$ which has size T. But this should be balanced with at least one computational advantage: the integrals embedded in v_t and $c_{t,l}$ can be approximated by Monte Carlo. If we set $\pi(\tau)$ to be the multivariate standard normal density and $\left(\tau^{(1)}, ..., \tau^{(M)}\right)$ be M values of τ simulated according to $\pi(\tau)$, the Monte Carlo approximations of v_t and $c_{t,l}$ are:

$$v_t \approx \frac{1}{M}\sum_{k=1}^{M} U h_t(\tau^{(k)}, \widehat{\theta}^1) h_T(\tau^{(k)}, \theta), \tag{3.23}$$

$$c_{t,l} \approx \frac{1}{M}\sum_{k=1}^{M} U h_t(\tau^{(k)}, \widehat{\theta}^1) h_l(\tau^{(k)}, \widehat{\theta}^1). \tag{3.24}$$

For the optimization algorithm to converge, it is crucial to simulate the set $\left(\tau^{(1)}, ..., \tau^{(M)}\right)$ only once at the beginning of the estimation process and supply this as a fixed array to the code that evaluates the objective function of the CGMM.

3.3.3 Computing the Variance of the CGMM Estimator

The asymptotic variance of the optimal CGMM estimator is derived in Carrasco and Florens (2000):

$$
\begin{aligned}
AVar\left(\widehat{\theta}\right) &= Var\left[\sqrt{T}\left(\widehat{\theta}-\theta_0\right)\right] \\
&= \left\langle K^{-1/2}E\left(\widehat{G}_t(.,\theta)\right), K^{-1/2}E\left(\widehat{G}_t(.,\theta)\right)\right\rangle^{-1} \qquad (3.25)
\end{aligned}
$$

where $\widehat{G}_t(\tau,\theta)=\frac{\partial \widehat{h}_t(\tau,\theta)}{\partial\theta}$ is a column vector of length q whose i^{th} element is $\widehat{G}_{t,i}(\tau,\theta)=\frac{\partial \widehat{h}_t(\tau,\theta)}{\partial\theta_i}$, and for every two vectors functions f and g, we have: $\langle f,g\rangle_{i,j}=\langle f_i,g_j\rangle$. This asymptotic variance can be consistently estimated by:

$$
\widehat{AVar}\left(\widehat{\theta}\right)=\left\langle K_{T,\lambda}^{-1/2}\widehat{G}_T(\tau,\widehat{\theta}), K_{T,\lambda}^{-1/2}\widehat{G}_T(\tau,\widehat{\theta})\right\rangle^{-1} \qquad (3.26)
$$

where $\widehat{G}_T(\tau,\widehat{\theta})=\frac{1}{T}\sum_{t=1}^{T}\widehat{G}_t(\tau,\widehat{\theta})$. The above formula is convenient to work with when the scalar products are evaluated by quadrature methods. Define:

$$
\begin{aligned}
\widehat{G}_{T,i}(\theta) &= \left(\widehat{G}_{T,i}(\tau_1,\theta),...,\widehat{G}_{T,i}(\tau_N,\theta)\right)' \\
K_{\lambda T}^{-1/2}\widehat{G}_{T,i}(\theta) &= \left(K_{T,\lambda}^{-1/2}\widehat{G}_{T,i}(\tau_1,\theta),...,K_{T,\lambda}^{-1/2}\widehat{G}_{T,i}(\tau_N,\theta)\right)'
\end{aligned}
$$

where $N=n^d$ and $\widehat{G}_{T,i}(\tau,\widehat{\theta})=\frac{1}{T}\sum_{t=1}^{T}\widehat{G}_{t,i}(\tau,\widehat{\theta})$. Then we have:

$$
K_{T,\lambda}^{-1/2}\widehat{G}_{T,i}(\theta)=\left(\widehat{W}_T^2+\lambda_T I\right)^{-1/2}\widehat{W}_T^{1/2}\widehat{G}_{T,i}(\theta),
$$

where $\widehat{W}$ is defined in (3.15). The (i,j) element of $\widehat{AVar}\left(\widehat{\theta}\right)^{-1}$ can then be computed as:

$$
\left(\widehat{AVar}\left(\widehat{\theta}\right)^{-1}\right)_{i,j}=\sum_{k=1}^{N}\omega_k\left(K_{T,\lambda}^{-1/2}\widehat{G}_{T,i}(\theta)\right)_k\overline{\left(K_{T,\lambda}^{-1/2}\widehat{G}_{T,j}(\theta)\right)_k} \qquad (3.27)
$$

where $\left(K_{T,\lambda}^{-1/2}\widehat{G}_{T,i}(\theta)\right)_k$ is the k^{th} coordinate of $K_{T,\lambda}^{-1/2}\widehat{G}_{T,i}(\theta)$.

Carrasco, Chernov, Florens and Ghysels (2007) established the following alternative expression for $\widehat{AVar}\left(\widehat{\theta}\right)$:

$$\widehat{AVar}\left(\widehat{\theta}\right)=\left(\frac{1}{T-q}V(\widehat{\theta})'\left[\lambda_T I_T+\widehat{C}_T^2\right]^{-1}\overline{V(\widehat{\theta})}\right)^{-1} \tag{3.28}$$

where $\widehat{C}$ is the same square matrix as in (3.20), $V(\widehat{\theta})$ is the (T,q) matrices with (t,i) element:

$$V_{t,i}=\int\overline{h_t(\tau,\widehat{\theta})}\widehat{G}_{T,i}(\tau,\widehat{\theta})\pi\left(\tau\right)d\tau,$$

The formula (3.28) best suites when the Monte Carlo integration is used to evaluate the scalar products. In this case, $V_{t,i}$ is approximated by:

$$V_{t,i}\approx\frac{1}{M}\sum_{k=1}^{M}\overline{h_t(\tau^{(k)},\widehat{\theta})}\widehat{G}_{T,i}(\tau^{(k)},\widehat{\theta}),$$

where $\left(\tau^{(1)},...,\tau^{(M)}\right)$ are M values of τ simulated according to the multivariate normal density $\pi\left(\tau\right)$.

3.3.4 Data-driven Selection of the Regularization Parameter

The CGMM estimator is consistent for any reasonable choice of the regularization parameter λ_T. In most applications, an arbitrary choice of λ_T between 10^{-6} and 10^{-2} works quite well. However, if the spectrum of the empirical covariance operator is severely discontinuous, such an arbitrary choice is not advised. To get close to the optimal CGMM in the mean square error (MSE) sense, Carrasco and Kotchoni (2008) proposed two simulation based methods to select the λ_T. The first method uses the higher-order closed form approximation of the MSE whereas the second method relies on the Monte Carlo simulations of the MSE. We briefly review the second method here.

Let $\lambda_T\left(\theta_0\right)$ be the optimal value of the regularization parameter when θ_0 is the true parameter of interest and T is the sample size:

$$\lambda_T\left(\theta_0\right)=\underset{\lambda\in[0,1]}{\arg\min}\Sigma_T\left(\lambda,\theta_0\right). \tag{3.29}$$

where $\Sigma_T(\lambda_T, \theta_0) = E\left[\left(\widehat{\theta}_T(\lambda_T;\theta_0) - \theta_0\right)'\left(\widehat{\theta}_T(\lambda_T;\theta_0) - \theta_0\right)\right]$ is the trace of the MSE matrix and $\widehat{\theta}_T(\lambda_T;\theta_0)$ is the CGMM estimator computed from an arbitrary sample of size T generated from the true distribution, and using λ as the regularization parameter. To approximate the MSE $\Sigma_T(\lambda_T, \theta_0)$, assume that we can draw samples of size T from the true data generating process of $\{x_t\}$, and let $\widehat{\theta}_T^j(\lambda_T;\theta_0)$ denote the CGMM estimator of θ_0 computed using the j^{th} independently simulated sample. A natural estimator of $\Sigma_T(\lambda_T, \theta_0)$ is given by:

$$\widehat{\Sigma}_{TM}(\lambda_T, \theta_0) = \frac{1}{M}\sum_{j=1}^{M}\left(\widehat{\theta}_T^j(\lambda_T;\theta_0) - \theta_0\right)'\left(\widehat{\theta}_T^j(\lambda_T;\theta_0) - \theta_0\right). \tag{3.30}$$

where the subscript TM indicate that T is the sample size and M is the number of Monte Carlo replications. If feasible, the above estimator of the MSE would naturally yield an estimator of the optimal λ of the form:

$$\widehat{\lambda}_{TM}(\theta_0) = \arg\min_{\lambda\in[0,1]}\widehat{\Sigma}_{TM}(\lambda, \theta_0), \tag{3.31}$$

For a sufficiently large value of M, the Law of Large Numbers ensures that $\widehat{\Sigma}_{TM}(\lambda, \theta_0)$ converges to its expectation $\Sigma_T(\lambda, \theta_0)$. But as θ_0 is not known, a feasible Monte Carlo approach simply consists in replacing θ_0 with a consistent first step estimator $\widehat{\theta}^1$ in the simulation scheme, that is, choosing the optimal regularization parameter according to:

$$\widehat{\lambda}_{TM}\left(\widehat{\theta}^1\right) = \arg\min_{\lambda\in[0,1]}\widehat{\Sigma}_{TM}^{MC}\left(\lambda, \widehat{\theta}^1\right). \tag{3.32}$$

It is important to note that $\widehat{\Sigma}_{TM}^{MC}\left(\lambda, \widehat{\theta}^1\right)$ is simulated conditional on the first step estimator $\widehat{\theta}^1$, and that $\widehat{\Sigma}_{TM}^{MC}\left(\lambda, \widehat{\theta}^1\right)$ converges to $\Sigma_T\left(\lambda, \widehat{\theta}^1\right)$ as M goes to infinity. The minimizer of this limiting MSE is $\lambda_T\left(\widehat{\theta}^1\right)$, the theoretically optimal λ if the true θ_0 was the point estimate $\widehat{\theta}^1$ and the sample size is T:

$$\lambda_T\left(\widehat{\theta}^1\right) = \arg\min_{\lambda\in[0,1]}\Sigma_T\left(\lambda, \widehat{\theta}^1\right). \tag{3.33}$$

Under Assumption 1 to 5, Carrasco and Kotchoni (2008) established that as M and T go to infinity, we have:

$$\frac{\widehat{\lambda}_{TM}(\widehat{\theta}^1)}{\lambda_T(\theta_0)} - 1 = O_p(T^{-1/2}) + O_p(M^{-1/2}) \tag{3.34}$$

Finally, the optimal feasible CGMM estimator is $\widehat{\theta}_T^{(2)} = \widehat{\theta}_T(\widehat{\lambda}_{TM}^*; \theta_0)$, that is, the second step estimator of θ_0 computed with the actual data by using the point estimate of the optimal regularization parameter $\widehat{\lambda}_{TM}^* = \widehat{\lambda}_{TM}\left(\widehat{\theta}^1\right)$. In practice, $\widehat{\Sigma}_{TM}(\lambda_T, \theta_0)$ should be simulated using common random numbers accross the different values of λ_T.

In the sequel, we propose two illustrative applications of the CGMM.

3.4 Estimating the Stable Distribution by CGMM: a Simulation Study

In modeling time series, the stable distributions is a way to depart from the usual normality assumption in case the latter seems too restrictive. This family is rich enough to capture heavy tails as well as asymmetry, as pointed out by Mandelbrot (1963), Fama (1965) or McCulloch (1986). However, the stable distribution does not admit a closed form likelihood function. This has led researcher to investigate alternative inference methods. CF based inference has been used in Paulson, Holcomb and Leitch (1975) and Feuerverger and McDunnough (1981a), while a regression-based approach is presented in Koutrouvelis (1980). Garcia, Renault and Veredas (2006) have resorted to indirect inference. Cornea and Davidson (2009) proposed a refined bootstrap method for testing an hypothesis about the mean of the stable distribution.

The stable distribution has been represented under different parametrizations in the literature. Some of these parametrizations are more or less tied to particular simulation strategies. Because our inferences are based on Monte Carlo simulation, we will carefully review below the most used parametrizations and simulation methods.

3.4.1 Parametrizations of the Stable Distribution

The standard Stable Distribution has two parameters: the stability parameter $\alpha \in]0, 2]$, and the skewness parameter $\beta \in [-1, 1]$. A random variable Z is said to follow the standard stable distribution if and only if its CF is given by:

$$E\left[\exp\left(i\tau Z\right)\right] = \exp\left\{-\left|\tau\right|^{\alpha}\left[1 + i\beta_0 sign\left(\tau\right) g\left(\tau,\alpha\right)\right]\right\} \tag{3.35}$$

where $g(\tau,\alpha) = -\tan\frac{\alpha\pi}{2}$ if $\alpha \neq 1$ and $g\left(\tau,\alpha\right) = \frac{2}{\pi}\ln\left|\tau\right|$ if $\alpha = 1$. A random variable X follows a α-Stable Distribution if and only if it is linked to the standard variable Z by:

$$X = \begin{cases} \sigma Z + \mu_0, & \alpha \neq 1 \\ \sigma Z + \mu_0 + \frac{2}{\pi}\beta_0\sigma\ln\sigma, & \alpha = 1 \end{cases}$$

The CF of X is given by:

$$E\left[\exp\left(i\tau X\right)\right] = \exp\left\{i\mu_0\tau - \sigma_0^{\alpha}\left|\tau\right|^{\alpha}\left[1 + i\beta_0 sign\left(\tau\right) g\left(\tau,\alpha\right)\right]\right\} \tag{3.36}$$

where μ_0 is a location parameter and σ_0 is the scale parameter. The notation $X \sim S_{\alpha}\left(\beta_0,\sigma_0,\mu_0\right)$ is often used to mean that the random variable X has a $\alpha-$Stable Distribution with CF (3.36).

Unfortunately, this CF is discontinuous around $\alpha = 1$ whenever $\beta_0 \neq 0$. To circumvent this, Zorotalev (1986) proposed to parametrize:

$$\mu_1 = \begin{cases} \mu_0 + \beta_0\sigma_0^{\alpha}\tan\frac{\alpha\pi}{2}, & \alpha \neq 1 \\ \mu_0, & \alpha = 1 \end{cases}$$

This results in the following expression for the CF which is continuous with respect to all the parameters:

$$E\left[\exp\left(i\tau X\right)\right] = \begin{cases} \exp\left\{i\mu_1\tau - \sigma_0^{\alpha}\left[\left|\tau\right|^{\alpha} - i\tau\beta_0\left(\left|\tau\right|^{\alpha-1} - 1\right)\tan\frac{\alpha\pi}{2}\right]\right\}, & \alpha \neq 1 \\ \exp\left\{i\mu_1\tau - \sigma_0\left|\tau\right|\left[1 + i\frac{2}{\pi}\beta_0 sign\left(\tau\right)\ln\left|\tau\right|\right]\right\}, & \alpha = 1 \end{cases} \tag{3.37}$$

We will refer to the parametrization (3.36) as $S_{\alpha}^{0}\left(\beta_0,\sigma_0,\mu_0\right)$ and to (3.37) as $S_{\alpha}^{1}\left(\beta_0,\sigma_0,\mu_1\right)$.

Using (3.36) as starting point, Nolan (1997) proposed:

$$\mu_2 = \begin{cases} \mu_0 + \beta_0\sigma_0\tan\frac{\alpha\pi}{2}, & \alpha \neq 1 \\ \mu_0 + \frac{2}{\pi}\beta_0\sigma_0\ln\sigma_0, & \alpha = 1 \end{cases}$$

This yields another continuous representation of the CF:

$$E\left[\exp\left(i\tau X\right)\right] \tag{3.38}$$
$$= \begin{cases} \exp\left\{i\mu_2\tau - \sigma_0^\alpha \left|\tau\right|^\alpha \left[1 + i\beta_0 sign\left(\tau\right)\left(\left|\sigma_0\tau\right|^{1-\alpha} - 1\right)\tan\frac{\alpha\pi}{2}\right]\right\}, & \alpha \neq 1 \\ \exp\left\{i\mu_2\tau - \sigma_0\left|\tau\right|\left[1 + i\frac{2}{\pi}\beta_0 sign\left(\tau\right)\ln\left|\sigma_0\tau\right|\right]\right\}, & \alpha = 1 \end{cases}$$

The parametrization (3.38) will be referred to as $S_\alpha^2\left(\beta_0, \sigma_0, \mu_2\right)$. An important feature of this parametrization is that $\frac{X-\mu_2}{\sigma_0} \sim S_\alpha^2\left(\beta_0, 1, 0\right)$, no matter the value of α. This is true for the two other parametrizations only when $\alpha \neq 1$.

An alternative parametrization $S_\alpha^3\left(\beta_0, \sigma_0, \mu_3\right)$ tied to the data simulation method of Chambers, Mallows and Stuck (1976) is got by setting:

$$\mu_3 = \begin{cases} \mu_0 + \beta_0\sigma_0\tan\frac{\alpha\pi}{2}, & \alpha \neq 1 \\ \mu_0, & \alpha = 1 \end{cases}$$

This is identical to $S_\alpha^2\left(\beta_0, \sigma_0, \mu_2\right)$ for the case $\alpha \neq 1$. As pointed out by Nolan (2008), these small changes in parameterization have caused many confusions in the literature. For instance, some papers build their theoretical framework on the parametrization $S_\alpha^0\left(\beta_0, \sigma_0, \mu_0\right)$ but simulate the data under the parametrization $S_\alpha^3\left(\beta_0, \sigma_0, \mu_3\right)$.

Another important parametrization proposed in Zorotalev (1986) allows to derive an integral representation of the probability distribution function of α-stable random variables[2]. In a few cases, the density of the stable distribution is available in a tractable closed form. The case $\alpha = 2$ for example reduces to a normal distribution $N\left(\alpha, 2\sigma_0^2\right)$. When $\alpha = 1$ and $\beta_0 = 0$, we get the Cauchy distribution. The case $\alpha = 1/2$ and $\beta_0 = 1$ results in to the so called Levy distribution. Finally, an identity established by Zorotalev (1986) and commented in Weron (1996) allows to get the density for the case $\alpha = 1/2$ and $\beta_0 = -1$ from the previous one. For all other values of the parameter, numerical approximations of the likelihood function must be used. Important progress have been made in that direction by Nolan (1997, 1999) and McCulloch (1998).

The parametrization $S_\alpha^0\left(\beta_0, \sigma_0, \mu_0\right)$ will used in the sequel.

[2] See Zolotarev (1986) Remark 1, page 78 or Weron (1995).

3.4.2 Simulating from the Stable Distribution

A method to simulate from the parameterization $S_\alpha^0(\beta_0, \sigma_0, \mu_0)$ is presented in Weron (1996). To start with, one draws two independent uniforms v and w in [0,1] and calculate: $V = \pi(u - 1/2)$ and $W = -\ln w$. Then $Z \sim S_\alpha^0(\beta_0, 1, 0)$ and $X \sim S_\alpha^0(\beta_0, \sigma_0, \mu_0)$ are obtained as follows:

- If $\alpha \neq 1$, one computes:

$$Z = S_{\alpha,\beta_0} \frac{\sin(\alpha V + \alpha B_{\alpha,\beta_0})}{(\cos V)^{1/\alpha}} \left(\frac{\cos((1-\alpha)V - \alpha B_{\alpha,\beta_0})}{W} \right)^{-1+1/\alpha} \tag{3.39}$$

where $B_{\alpha,\beta_0} = \frac{\arctan\left(\beta_0 \tan \frac{\alpha\pi}{2}\right)}{\alpha}$ and $S_{\alpha,\beta_0} = \left(1 + \beta_0^2 \tan^2 \frac{\alpha\pi}{2}\right)^{\frac{1}{2\alpha}}$. Then we have

$$\begin{aligned} X &= \sigma_0 Z + \mu_0 \sim S_\alpha^0(\beta_0, \sigma_0, \mu_0) \\ X &= \sigma_0 Z + \mu_1 - \beta_0 \sigma_0^\alpha \tan \frac{\alpha\pi}{2} \sim S_\alpha^1(\beta_0, \sigma_0, \mu_1) \\ X &= \sigma_0 \left(Z - \beta_0 \tan \frac{\alpha\pi}{2}\right) + \mu_2 \sim S_\alpha^2(\beta_0, \sigma_0, \mu_2) \end{aligned}$$

- If $\alpha = 1$, one computes instead:

$$Z = \frac{2}{\pi} \left[\left(\frac{\pi}{2} + \beta_0 V \right) \tan V - \beta \log \left(\frac{W \cos V}{\frac{\pi}{2} + \beta_0 V} \right) \right] \tag{3.40}$$

We then have:

$$\begin{aligned} X &= \sigma_0 Z + \mu_0 + \frac{2}{\pi} \beta_0 \sigma_0 \ln \sigma_0 \sim S_1^0(\beta_0, \sigma_0, \mu_0) \\ X &= \sigma_0 Z + \mu_1 + \frac{2}{\pi} \beta_0 \sigma_0 \ln \sigma_0 \sim S_1^1(\beta_0, \sigma_0, \mu_1) \\ X &= \sigma_0 Z + \mu_2 \sim S_1^2(\beta_0, \sigma_0, \mu_2) \end{aligned}$$

The simulation strategy of Z for the case $\alpha = 1$ is quite standard in the literature. However, other methods (than the one above) have been used in the literature for the case

$\alpha \neq 1$. We show the link between (3.39) and two of them below. To start with, note that:

$$\begin{aligned}
\sin\left(\alpha V + \alpha B_{\alpha,\beta_0}\right) &= \sin\alpha V \sin\alpha B_{\alpha,\beta_0} + \cos\alpha V \cos\alpha B_{\alpha,\beta_0} \\
\cos\left((1-\alpha) V - \alpha B_{\alpha,\beta_0}\right) &= \cos(1-\alpha) V \cos\alpha B_{\alpha,\beta_0} + \sin(1-\alpha) V \sin\alpha B_{\alpha,\beta_0}
\end{aligned}$$

Also, due to $\sin a = \cos a \tan a$ for all a, we have: $\sin\alpha B_{\alpha,\beta_0} = \beta_0 \tan\frac{\alpha\pi}{2}\cos\alpha B_{\alpha,\beta_0}$ so that:

$$\cos^2\alpha B_{\alpha,\beta_0} = 1 - \sin^2\alpha B_{\alpha,\beta_0} = 1 - \beta_0^2\tan^2\frac{\alpha\pi}{2}\cos^2\alpha B_{\alpha,\beta_0},$$

The last equation implies:

$$\cos^2\alpha B_{\alpha,\beta_0} = \frac{1}{1+\beta_0^2\tan^2\frac{\alpha\pi}{2}}$$

Replacing this in the expressions of $\sin\left(\alpha V + \alpha B_{\alpha,\beta_0}\right)$ and $\cos\left((1-\alpha) V - \alpha B_{\alpha,\beta_0}\right)$ yields:

$$\begin{aligned}
S_{\alpha,\beta_0} &= \left(\cos\alpha B_{\alpha,\beta_0}\right)^{-1/\alpha} \\
\sin\left(\alpha V + \alpha B_{\alpha,\beta_0}\right) &= \frac{\beta_0\tan\frac{\alpha\pi}{2}\sin\alpha V + \cos\alpha V}{\sqrt{1+\beta_0^2\tan^2\frac{\alpha\pi}{2}}} \\
\cos\left((1-\alpha) V - \alpha B_{\alpha,\beta_0}\right) &= \left(\frac{\cos(1-\alpha) V + \beta_0\tan\frac{\alpha\pi}{2}\sin(1-\alpha) V}{\sqrt{1+\beta_0^2\tan^2\frac{\alpha\pi}{2}}}\right)^{-1+1/\alpha}
\end{aligned}$$

Putting these expression together in Equation (3.39) yields:

$$Z = \frac{\beta_0\tan\frac{\alpha\pi}{2}\sin\alpha V + \cos\alpha V}{(\cos V)^{1/\alpha}}\left(\frac{\cos(1-\alpha) V + \beta_0\tan\frac{\alpha\pi}{2}\sin(1-\alpha) V}{W}\right)^{-1+1/\alpha} \tag{3.41}$$

Adding $-\beta\tan\frac{\alpha\pi}{2}$ to the above expression yields the formula of Chambers, Mallows and Stuck (1976). To get the alternative expression of Nolan (2008), Theorem 1.19, it suffices to substitute for $S_{\alpha,\beta_0} = \left(\cos\alpha B_{\alpha,\beta_0}\right)^{-1/\alpha}$ in Equation (3.39). This yields:

$$Z = \frac{\sin\left(\alpha V + \alpha B_{\alpha,\beta_0}\right)}{\left(\cos\alpha B_{\alpha,\beta_0}\cos V\right)^{1/\alpha}}\left(\frac{\cos\left((1-\alpha) V - \alpha B_{\alpha,\beta_0}\right)}{W}\right)^{-1+1/\alpha} \tag{3.42}$$

As pointed out by Nolan (2008), the evaluation of (3.42) for values very close to $\alpha = 1$ raises some numerical problems. By avoiding the division by $\cos\alpha B_{\alpha,\beta_0}$, the expressions

(3.39) and (3.41) are more numerically stable and accurate.

3.4.3 Monte Carlo Experiments

With in hand a method to simulate data from the stable distribution, we can now evaluate by Monte Carlo the ability of the CGMM to identify the true parameters from finite sample. To this end, we consider a stable AR(1) model specified as:

$$y_t = \rho_0 + \rho_1 y_{t-1} + \varepsilon_t \tag{3.43}$$

where $\varepsilon_t \sim S_\alpha^0(\beta_0, \sigma_0, 0)$ is IID. Note that this amount to say that $y_t \sim S_\alpha^0(\beta_0, \sigma_0, \mu_t)$ with $\mu_t = \rho_0 + \rho_1 y_{t-1}$. The parameter of the model are gathered in $\theta = (\rho_0, \rho_1, \alpha, \beta_0, \sigma_0)'$.

To estimate θ, the following continuum of moment conditions is considered:

$$h_t(\tau, \theta) = \left[e^{i\tau_1 y_t} - \varphi_t(\tau_1, \theta)\right] e^{i\tau_2 y_{t-1}} \tag{3.44}$$

where $\tau = (\tau_1, \tau_2) \in \mathbf{R}^2$ and:

$$\varphi_t(\tau_1, \theta) = \exp\left\{i(\rho_0 + \rho_1 y_{t-1})\tau_1 - \sigma^\alpha |\tau_1|^\alpha \left[1 - i\beta sign(\tau_1) \tan\frac{\alpha\pi}{2}\right]\right\} \tag{3.45}$$

The following gradients are useful for the analytical computation of the variance of the CGMM estimator:

$$\begin{aligned}
\frac{\partial h_t(\tau,\theta)}{\partial \rho_0} &= -i\tau_1 \varphi_t(\tau_1,\theta) e^{i\tau_2 r_{t-1}} \\
\frac{\partial h_t(\tau,\theta)}{\partial \rho_1} &= -i\tau_1 y_{t-1} \varphi_t(\tau_1,\theta) e^{i\tau_2 r_{t-1}} \\
\frac{\partial h_t(\tau,\theta)}{\partial \alpha} &= \left\{\log(\sigma|\tau_1|)\left[1 - i\beta sign(\tau_1)\tan\frac{\alpha\pi}{2}\right] - \frac{i\pi\beta sign(\tau_1)}{2\cos^2\frac{\alpha\pi}{2}}\right\} \\
&\quad \times \sigma^\alpha |\tau_1|^\alpha \varphi_t(\tau_1,\theta) e^{i\tau_2 r_{t-1}} \\
\frac{\partial h_t(\tau,\theta)}{\partial \beta} &= -\sigma^\alpha |\tau_1|^\alpha i sign(\tau_1) \tan\frac{\alpha\pi}{2} \varphi_t(\tau_1,\theta) e^{i\tau_2 r_{t-1}} \\
\frac{\partial h_t(\tau,\theta)}{\partial \sigma} &= \alpha\sigma^{\alpha-1}|\tau_1|^\alpha \left[1 - i\beta sign(\tau_1)\tan\frac{\alpha\pi}{2}\right] \varphi_t(\tau_1,\theta) e^{i\tau_2 r_{t-1}}
\end{aligned}$$

We consider the following two vectors of true parameters in our simulations: $\theta_{01} =$

$(0, 0.1, 1.5, 0, 0.5)$ and $\theta_{02} = (0, 0.1, 1.95, 0, 0.5)$, the specificity of θ_{02} being that it is quite close to the non-identification region of β_0. To ease the numerical optimizations, the following transformations are imposed on the parameter space:

$$\begin{aligned}
\alpha_0 &= 1 + \frac{\exp\left(\widetilde{\alpha}_0\right)}{1+\exp\left(\widetilde{\alpha}_0\right)} \in]1,2] \text{ for all } \widetilde{\alpha}_0 \in \mathbb{R} \\
\beta_0 &= \frac{2\exp\left(\widetilde{\beta}_0\right)}{1+\exp\left(\widetilde{\beta}_0\right)} - 1 \in [-1,1] \text{ for all } \widetilde{\beta}_0 \in \mathbb{R} \\
\sigma_0 &= \exp(\widetilde{\sigma}_0) > 0 \text{ for all } \widetilde{\sigma}_0 \in \mathbb{R}
\end{aligned}$$

After these transformations, the new objective function of the CGMM is written in terms of the unconstrained parameters $\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{\sigma}_0$ and μ_0.

The Monte Carlo experiments are conducted in two steps. First, we run a small scale simulation (100 replications) for the purpose of selecting an approximately optimal λ. In this small scale simulation, we compute the objective function with $N = 64$ Hermitian quadrature points in $\mathbb{R}^2$. The first simulated sample is used to compute the following first step estimator:

$$\widehat{\theta}_T^1 = \arg\min_\theta \left\| \widehat{h}_T(.,\theta) \right\|^2 \tag{3.46}$$

For each λ_k and each simulated sample, we compute the second step estimator as:

$$\widehat{\theta}_T^{(j)}(\lambda_k) = \arg\min_\theta \left\| K_{T,\lambda_k}^{-1/2} \widehat{h}_T(.,\theta) \right\|^2$$

where:

$$K_{T,\lambda_k}^{-1/2}\left(\widehat{\theta}_T^1\right) = \left(K_T^2\left(\widehat{\theta}_T^1\right) + \lambda_k I\right)^{-1/2} K_T^{1/2}\left(\widehat{\theta}_T^1\right)$$

for $\lambda_k \in \{10^{-7}, 5\times 10^{-7}, ..., 5\times 10^{-4}\}$. The selection of λ_k is based on the criterion:

$$\lambda^* = \arg\min_{\lambda_k} \widehat{\Sigma}_{TM}(\lambda_k)$$

where $\widehat{\Sigma}_{TM}(\lambda_k) = \frac{1}{M}\sum\left(\widehat{\theta}_T^{(j)}(\lambda_k) - \widehat{\theta}_T^1\right)'\left(\widehat{\theta}_T^{(j)}(\lambda_k) - \widehat{\theta}_T^1\right)$. The following figure shows the plot of $\widehat{\Sigma}_{TM}(\lambda_k)$ as a function of λ_k. On the considered grid, $\widehat{\Sigma}_{TM}(\lambda_k)$ is minimized at $\lambda^* = 5\times 10^{-7}$.

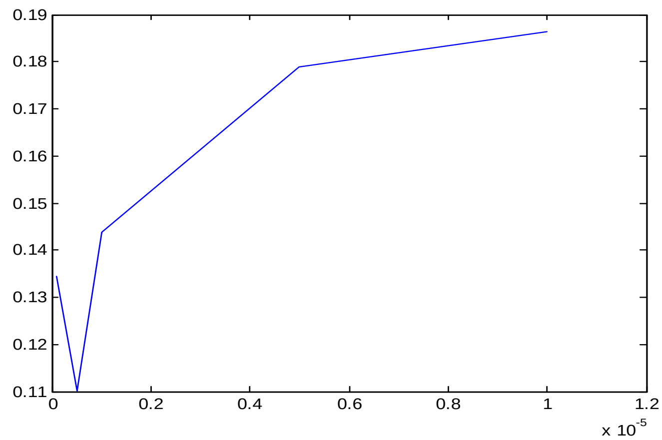


Figure 3.1: Choosing the optimal regularization parameter.

y-axis: MSE $\widehat{\theta}_T(\lambda_k)$of as function of λ_k.

x-axis: values of λ_k

In a second step, the selected λ^* is used in a large scale simulation to assess the performance of the CGMM estimator. We draw $M = 1001$ samples of size $T = 501$ and estimate $\theta_{0i}, i = 1, 2$. To speed up the simulation, we reduce the number of quadrature points to $N = 36$. The following table shows some statistical properties of $\widehat{\theta}_T(\lambda^*)$ when the true parameter is $\theta_{01} = (0, 0.1, 1.5, 0, 0.5)$.

	ρ_0	ρ_1	α	β_0	σ_0
True Values	0	0.1	1.5	0	0.5
Mean Bias	0.0044	-0.0032	0.0072	0.0075	-0.0033
Median Bias	0.0007	-0.0026	0.0066	0.0078	-0.0038
Emp. Std. Dev.	0.0869	0.0482	0.1033	0.2133	0.0305
Ana. Std. Dev.	0.1137	0.0496	0.1027	0.2597	0.0243
IC1(95%)	-0.0032	0.0925	1.4982	-0.0112	0.4940
IC2(95%)	0.0120	0.1010	1.5163	0.0262	0.4993

Table 3.1: Simulation Results for the Stable Distribution with $\alpha = 1.5$ (far from 2)

(1000 Monte Carlo replications)

In this table, 'Emp. Std. Dev' is the standard deviation of the simulated empirical distribution of $\widehat{\theta}_T(\lambda^*)$, while 'Ana. Std. Dev.' is the average standard deviation computed according to the analytical formula (3.28); Interestingly, the standard deviations computed in these two ways are quite close for all the parameters. IC1(95%) and IC2(95%) are respectively the lower and upper bound of the 95% confidence interval for the true mean of the empirical distribution, assuming normality for the empirical mean of the estimates:

$$\begin{aligned} IC1(95\%) &= \overline{\overline{\theta}}_i - 1.96 * \widehat{s}_{\widehat{\theta}_i}/\sqrt{M} \\ IC2(95\%) &= \overline{\overline{\theta}}_i + 1.96 * \widehat{s}_{\widehat{\theta}_i}/\sqrt{M} \end{aligned}$$

where $\widehat{\theta}_i$ is the i^{th} component of $\widehat{\theta}$, $\overline{\overline{\theta}}_i$ and $\widehat{s}_{\widehat{\theta}_i}$ are respectively the empirical mean and standard deviation of $\widehat{\theta}_i$ and M is the number of independently simulated copies of $\widehat{\theta}_i$. The confidence intervals reveal that the estimator of σ_0 is slightly biased downward. This problem can be fixed by increasing the number of quadrature points used to approximate the objective function of the CGMM. All the other estimators display quite good statistical properties. The following table displays the same Monte Carlo statistics when the true parameter is $\theta_{02} = (0, 0.1, 1.95, 0, 0.5)$.

	ρ_0	ρ_1	α	β_0	σ_0
True Values	0	0.1	1.95	0	0.5
Mean Bias	0.0020	-0.0043	0.0081	-0.0516	-0.0011
Median Bias	0.0023	-0.0020	0.0366	-0.0146	-0.0019
Emp. Std. Dev.	0.0383	0.0583	0.0536	0.5752	0.0211
IC1(95%)	-0.0014	0.0906	1.9534	-0.1019	0.4971
IC2(95%)	0.0054	0.1008	1.9628	-0.0012	0.5008

Table 3.2: Simulation Results for the Stable Distribution with $\alpha = 1.95$ (close to 2)

(1000 Monte Carlo replications)

In this case, the estimator of β_0 is highly volatile due the fact that the objective function is unable to identify β_0 when α is close to 2. The empirical gradient turns out to be badly conditioned so that it is not possible to compute the variance analytically. Also, the distributions of the $\widehat{\alpha}$ and $\widehat{\beta}_0$ are far from normality in the region around θ_{02} (see figures at the end of this section). The Monte Carlo statistics of the estimators other than $\widehat{\beta}_0$ are quite good.

3.5 Fitting the Autoregressive Variance Gamma Model to Assets Returns

The basic Variance Gamma model has been proposed by Madan and Seneta (1990). A random variable r_t is said to follow a symmetric Variance Gamma distribution if:

$$r_t|V_t \sim N(\delta, \sigma^2 V_t), \text{ with } V_t \overset{IID}{\sim} Gamma\,(1/\gamma, 1/\gamma) \tag{3.47}$$

The density of V_t is given by:

$$f_V\,(v) = \frac{v^{1/\gamma - 1}}{\gamma^{1/\gamma}\Gamma\,(1/\gamma)} \exp\,(-v/\gamma)$$

where $\Gamma\,(1/\gamma) = \int_0^\infty u^{1/\gamma-1} e^{-u} du$. It can be easily checked that $E\,(V_t) = 1$.

Unlike the stable distribution, all the conditional and unconditional moments of r_t exist. It can be checked that $E\,[r_t] = \delta$ and $E\left[(r_t - \delta)^2\right] = \sigma^2$. The kurtosis of r_t is given by:

$$\frac{E\left[(r_t - \delta)^4\right]}{E\left[(r_t - \delta)^2\right]^2} = 3\,(1+\gamma)$$

which shows that the distribution of r_t is more fat tailed than the normal whenever $\gamma > 0$. To introduce skewness into this basic set up, Madan, Carr and Chang (1998) expressed the mean of r_t as a linear function of V_t:

$$r_t|V_t \sim N(\delta_0 + \delta_1 V_t, \sigma^2 V_t), \text{ with } V \sim Gamma\,(1/\gamma, 1/\gamma) \tag{3.48}$$

where $\gamma > 0$. If r_t is a series of returns, the parameter δ_1 captures the so-called leverage effect while δ_0 measures the risk premium. Note that when $\delta_1 = -\delta_0$, the leverage effect offsets the risk premium so that the conditional mean of r_t is zero, but the skewness is nonzero unless $\delta_1 = 0$.

Many studies have diagnosed patterns like persistence and clustering in the time series properties of the volatility assets returns. Unfortunately, the basic Variance Gamma models assumes that V_t follows an IID process. In an effort to correctly measure the volatility, Engle (1982) and Bollerslev (1986) introduced respectively the ARCH and GARCH models that usually have good filtering properties. In the stochastic volatility literature, the volatility is often specified as a latent state variable that determines the distribution of the returns. For example, Jacquier, Polson and Rossi (1994) postulated:

$$\begin{aligned} r_t &= \sqrt{V_t}\varepsilon_t \\ \log V_t &= a + b \log V_{t-1} + u_t \end{aligned}$$

where ε_t and u_t are uncorrelated and $r_t|V_t \sim N(0, V_t)$. This model may be viewed as a discrete time version of Hull and White (1987). It has been extended in Jacquier, Polson and Rossi (2004) to allow for correlation between ε_t and u_t. Other famous examples in continuous time include Stein and Stein (1991) and Heston (1993).

In the next subsection, we extend of the basic variance Gamma model.

3.5.1 The Autoregressive Variance Gamma Model

The Autoregressive Variance Gamma Model (henceforth ARVG) is a stochastic volatility model in which the return process r_t is a function of the expected variance $E[V_t|V_{t-1}]$ and the innovation $V_t - E[V_t|V_{t-1}]$:

$$r_t = \mu_0 + \mu_1 \sqrt{E[V_t|V_{t-1}]} + \delta (V_t - E[V_t|V_{t-1}]) + \sqrt{V_t}\varepsilon_t \tag{3.49}$$

where $\varepsilon_t \overset{IID}{\sim} N(0,1)$ is uncorrelated with past, current and future realizations of V_t, $\mu \geq 0$ and $\delta \leq 0$. In turn, V_t follows an Autoregressive Gamma process with conditional density:

$$\begin{aligned} f(V_t| \{V_{t-k}\}_{k=1}^{\infty}) &= f(V_t|V_{t-1}) \\ &= \sum_{j=0}^{\infty} \frac{V_t^{j+q-1} c^{j+q}}{\Gamma(j+q)} \exp(-cV_t)\, p_j(V_{t-1}) \end{aligned} \tag{3.50}$$

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

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

The term $\mu_1\sqrt{E\left[V_{t-1}|V_{t-1}\right]}$ in the expression of the return aims to capture the premium investors require for bearing the expected risk while $\delta\left(V_t - E\left[V_t|V_{t-1}\right]\right)$ is a penalty attached to the unexpected risk.

The postulated distribution for V_t is also known as the non-centered Chi-square[3]. It has been proposed in Gourieroux and Jasiak (2005) as a model for intertrade durations. The conditional CF of V_t is given by:

$$E\left[e^{i\tau V_t}|V_{t-1}\right] = \left(1 - \frac{i\tau}{c}\right)^{-q} \exp\left(\frac{i\tau e^{-\kappa}V_{t-1}}{1-\frac{i\tau}{c}}\right) \tag{3.51}$$

By looking at the expression above, we see that the autoregressive Gamma family nests the univariate Wishart autoregressive process of Gourieroux, Jasiak and Sufana (2005). The expressions of the conditional expectation and variance of V_t are the following:

$$E\left[V_t|V_{t-1}\right] = \beta\left(1-e^{-\kappa}\right) + e^{-\kappa}V_{t-1} \tag{3.52}$$

$$Var\left[V_t|V_{t-1}\right] = \frac{1}{c}\left[\beta\left(1-e^{-\kappa}\right) + 2e^{-\kappa}V_{t-1}\right] \tag{3.53}$$

To assess the potential of the ARVG model to capture asymmetry and fat tails in the distribution of stock returns, we examine below the third and fourth conditional moments of r_t. We have:

$$E\left[\left(r_t - E\left[r_t|V_{t-1}\right]\right)^3|V_{t-1}\right] = \delta^3 E\left[\left(V_t - E\left[V_t|V_{t-1}\right]\right)^3|V_{t-1}\right] \tag{3.54}$$

Because V_t is positively skewed like any Gamma distribution, r_t has a time varying negative skewness whenever $\delta < 0$. It is difficult to say wether r_t is fat-tailed in general. However, it can be shown that when $\delta = 0$ the conditional kurtosis of r_t is:

$$\frac{E\left[\left(r_t - E\left[r_t|V_{t-1}\right]\right)^4|V_{t-1}\right]}{Var\left[r_t|V_{t-1}\right]^2} = 3 + \frac{3Var\left[V_t|V_{t-1}\right]}{Var\left[r_t|V_{t-1}\right]^2} \tag{3.55}$$

We present a method to simulate the ARVG model below.

[3] When $2\kappa\beta > \sigma^2$, the process V_t can be viewed as discrete observations of the following Square-Root diffusion (see Feller, 1951): $dV_s = \kappa\left(\beta - V_s\right)ds + \sigma\sqrt{V_s}dW_s$.

3.5.2 Simulating the ARVG model

A method to simulate V_t can be inferred from the Poisson-Mixing-Gamma representation of its density given in (3.50) (Devroye (1986)). The simulation algorithm may be initialized to the unconditional mean $V_0 = \beta$ or by drawing V_0 from the stationary distribution[4] of V_t. At $t = 1$, one draws an integer j_0 from the Poisson distribution with parameter $ce^{-\kappa}V_0$. The current realization V_1 of the autoregressive Gamma process is then drawn from the Gamma distribution with density $f_{j_0}(v)$ given by:

$$f_{j_0}(v) = \frac{v^{j+q-1}c^{j+q}}{\Gamma(j_0+q)} \exp(-cv)$$

A $t = 2$, one draws again an integer j_1 from the Poisson distribution with parameter $ce^{-\kappa}V_1$. The new realization V_2 of the autoregressive Gamma process 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 $f_{j_{t-1}}(v)$, where j_{t-1} is a draw from the Poisson distribution with parameter $ce^{-\kappa}V_{t-1}$. To minimize transient effects, a good idea is to simulate $T + T_0 + 1$ observations and keep only the last $T + 1$ ones.

Let $(V_0, V_1, ..., V_T)$ be the simulated path for the volatility process. Because r_t depends on two consecutive realizations of V_t, its simulation starts at $t = 1$. We generate a sample of size T of the return process using the equation:

$$r_t = \mu_0 + \mu_1\sqrt{\beta(1-e^{-\kappa}) + e^{-\kappa}V_{t-1}} + \delta\left[V_t - \beta\left(1-e^{-\kappa}\right) - e^{-\kappa}V_{t-1}\right] + \sqrt{V_t}\varepsilon_t$$

for $t = 1, ..., T$, where ε_t is an IID draw from the standard normal distribution.

In what follows, we present an estimation strategy for the ARVG model.

3.5.3 Estimating the ARVG Model from High Frequency Data

This section explains why and how one can construct a proxy for V_t using high frequency data. Let us consider an arbitrary asset whose instantaneous log-price p_s follows a Brownian

[4]The stationary distribution of V_t is a Gamma with density: $f(V_t) = \frac{V_t^{q-1}}{\Gamma(q)}\left(\frac{2\kappa}{\sigma^2}\right)^q \exp\left(\frac{-2\kappa}{\sigma^2}V_t\right)$.

diffusion with drift:

$$dp_s = m(s, \sigma_s)\, ds + \sigma_s dW_s \tag{3.56}$$

where W_s is a standard Brownian motion uncorrelated with σ_s. It is further assumed that σ_s itself follows a positive diffusion. If we normalize a trading day to be one period, then the daily returns satisfies:

$$r_t \equiv p_t - p_{t-1} = \int_{t-1}^{t} m(s, \sigma_s)\, ds + \int_{t-1}^{t} \sigma_s dW_s \tag{3.57}$$

and $r_t | \{\sigma_s\}_{s=0}^{T} \sim N\left(\int_{t-1}^{t} m(s, \sigma_s)\, ds, IV_t\right)$, where $IV_t = \int_{t-1}^{t} \sigma_s^2 ds$ is the integrated volatility.

A strategy to estimate the ARVG model from high frequency data consists in assuming the following intuitive matching:

$$\begin{aligned} V_t &\equiv IV_t & (3.58) \\ \int_{t-1}^{t} m(s, \sigma_s)\, ds &\equiv \mu_0 + \mu_1 \sqrt{E[V_t|V_{t-1}]} + \delta\left(V_t - E[V_t|V_{t-1}]\right) & (3.59) \end{aligned}$$

Equation (3.58) suggests that the integrated volatility can be used as proxy for V_t, while the second equation can be interpreted as the linear projection of the integrated drift onto $\sqrt{E[V_t|V_{t-1}]}$.

If IV_t is observed, the ARVG model can be estimated in two steps. In the first step, we estimate an Autoregressive Gamma model for V_t by CGMM, using the moment function:

$$h_t(\tau, \theta_1) = \left(\exp(i\tau_1 V_t) - E[\exp(i\tau_1 V_t)|V_{t-1}]\right) \exp(i\tau_2 V_{t-1}), \tag{3.60}$$

where $E[\exp(i\tau_1 V_t)|V_{t-1}]$ is given by (3.51), $\tau = (\tau_1, \tau_2)$ and $\theta_1 = (\kappa, \beta, \sigma^2)$. In the second step estimation, $\theta_2 = (\mu_0, \mu_1, \delta)$ is estimated by Gaussian maximum likelihood based on the distribution of ε_t conditional on V_{t-1}, as postulated in (3.49). We have:

$$\widehat{\varepsilon}_t = \frac{r_t - \mu_0 - \mu_1\sqrt{\widehat{V}_t} - \delta\left(V_t - \widehat{V}_t\right)}{\sqrt{V_t}} \sim N(0,1) \tag{3.61}$$

where $\widehat{V}_t = \widehat{\beta}\left(1 - e^{-\widehat{\kappa}}\right) + e^{-\widehat{\kappa}} V_{t-1}$.

In reality, IV_t is not observed. To construct a proxy, let us assume that in each day trading we observe $m+1$ equidistant prices. These prices can be used to compute exactly m high frequency returns $r_{t,1}, r_{t,2}, ..., r_{t,m}$, that is:

$$r_{t,j} = p_{t-1+j/m} - p_{t-1+(j-1)/m}$$

Jacod (1994), Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002) show that for large m, the realized volatility $RV_t^{(m)} = \sum_{j=1}^{m} r_{t,j}^2$ is a fairly good proxy for IV_t. In practice however, the observed prices are contaminated with the market microstructure noise which causes the naive realized volatility to be a biased estimator of IV_t. The following estimator proposed by Barndoff-Nielsen, Hansen, Lunde and Shephard (2008) is known to be consistent for IV_t even in the presence of microstructure noise:

$$K_{H,t}^{BNHLS} = \gamma_{t,0} + \sum_{h=1}^{H} \left(1 - \frac{h-1}{H}\right) \left(\gamma_{t,h} + \gamma_{t,-h}\right) \tag{3.62}$$

where $\gamma_{t,h} = \sum_{j=1}^{m} r_{t,j} r_{t,j-h}$. To further reduce the variance of $K_{H,t}^{BNHLS}$, we will use the following shrinkage estimator proposed in Carrasco and Kotchoni (2009):

$$K_{H,t}^{\varpi} = \varpi K_{H,t}^{BNHLS} + (1-\varpi)\widehat{IV}_t \tag{3.63}$$

where

$$\widehat{IV}_t = \gamma_{t,0} + \gamma_{t,1} + \gamma_{t,-1} + \frac{1}{T} \sum_{s=1}^{T} \sum_{l=2}^{L+1} \left(\gamma_{s,l} + \gamma_{s,-l}\right) \tag{3.64}$$

The shrinkage weight ϖ is chosen so as to minimize the marginal variance of $K_{H,t}^{\varpi}$:

$$\varpi_t^* = \arg\min_{\varpi} E\left[\left(K_{H,t}^{\varpi} - IV_t\right)^2\right]$$

It is easy to show that $\varpi_t^* = \frac{Cov\left[\widehat{IV}_t, \widehat{IV}_t - K_{H,t}^{BNHLS}\right]}{Var\left[\widehat{IV}_t - K_{H,t}^{BNHLS}\right]}$, which we estimate in the simplest way from the data by:

$$\widehat{\varpi}^* = \frac{\sum_{t=1}^{T} \left(\widehat{IV}_t - K_{H,t}^{BNHLS}\right) \widehat{IV}_t}{\sum_{t=1}^{T} \left(\widehat{IV}_t - K_{H,t}^{BNHLS}\right)^2}.$$

An empirical application is presented in the next section.

3.5.4 Empirical Application

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 one minute from January 1^{st}, 2002 to December 31^{th}, 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.

The estimation takes place in several steps. First of all, we compute the first step CGMM estimator of $\theta_1 = (\kappa, \beta, \sigma^2)$ based on the moment function (3.60):

$$\widehat{\theta}_1^1 = \arg\min_{\theta_1} \left\| \widehat{h}_T(., \theta_1) \right\|^2$$

Secondly, we use $\widehat{\theta}_1^1$ to estimate the covariance operator $K_T\left(\widehat{\theta}_1^1\right)$ associated with the moment function. We shall use the ad-hoc value $\lambda = 10^{-6}$ for the regularization of the inverse of $K_T\left(\widehat{\theta}_{1,T}^1\right)$, that is:

$$K_{T,10^{-6}}^{-1/2}\left(\widehat{\theta}_1^1\right) = \left(K_T^2\left(\widehat{\theta}_1^1\right) + 10^{-6} \times I\right)^{-1/2} K_T^{1/2}\left(\widehat{\theta}_1^1\right)$$

The second step estimator of θ_1 is thus:

$$\widehat{\theta}_1^2 = \arg\min_{\theta_1} \left\| K_{T,10^{-6}}^{-1/2} \widehat{h}_T(., \theta_1) \right\|^2$$

where $K_{T,10^{-6}}^{-1/2} \equiv K_{T,10^{-6}}^{-1/2}\left(\widehat{\theta}_1^1\right)$.

Thirdly, we estimate the variance of $\widehat{\theta}_1^2$. Unfortunately, the analytical expression (3.26) is unusable because the gradient of $\widehat{h}_T(\tau, \theta_1)$ is extremely badly scaled[5]. This is due to the fact that the likelihood function of Gamma distributions (like some Student distributions) are very flat around the true value of the degree of freedom parameter. As a result, the derivative of the objective function with respect to the degree of freedom parameter is very small relatively to the derivatives with respect to the remaining parameters. In other words, the matrix of gradient is so badly scaled that it is numerically singular. The problem is even

[5] The gradients of the moments conditions (3.60) are given in appendix.

more severe in the Autoregressive Gamma model because the degree of freedom is $q = \frac{2\kappa\beta}{\sigma^2}$, that is, a function of all three parameters of interest. However, this numerical singularity does not imply that the model is not identified[6]. We can thus resort to bootstrap to evaluate the variance of $\widehat{\theta}_1^2$. The experiment is conducted as follows.

We use the initial sample of size $T = 1510$ to compute 1509 moment functions: $\{h_t(\tau, \theta_1)\}_{t=2}^{T}$. Note that $h_t(\tau, \theta_1)$ is a function of two consecutive observations (V_t, V_{t-1}). Next, we draw 500 moment functions with equal probability and replacement from the above set to get $\left\{\widetilde{h}_j^{(b)}(\tau, \theta_1)\right\}_{j=1}^{500}$, for $b = 1, 2, ..., B = 1000$. Each sample $\left\{\widetilde{h}_j^{(b)}(\tau, \theta_1)\right\}_{j=1}^{500}$ is then used to compute an estimator $\widehat{\theta}_{1,b}$ for $\theta_1 = (\kappa, \beta, \sigma^2)$. Finally, $\widehat{\theta}_{1,b}$ is used together with the realizations of V_t on which $\left\{\widetilde{h}_j^{(b)}(\tau, \theta_1)\right\}_{j=1}^{500}$ depend to compute an estimator $\widehat{\theta}_{2,b}$ for $\theta_2 = (\mu_0, \mu_1, \delta)$. In computing $\widehat{\theta}_{2,b}$, the constraints $\mu \geq 0$ and $\delta \leq 0$ are explicitly imposed. Likewise, $(\kappa, \beta, \sigma^2) > 0$ is imposed in the estimation of θ_1. The following table summarizes the empirical distributions of $\widehat{\theta}_{1,b}$ and $\widehat{\theta}_{2,b}$.

	$\widehat{\theta}_{1,b}$			$\widehat{\theta}_{2,b}$		
	$\widehat{\kappa}_b$	$\widehat{\beta}_b$	$\widehat{\sigma}_b^2$	$\widehat{\mu}_{0,b}$	$\widehat{\mu}_{1,b}$	$\widehat{\delta}_b$
Mean	0.8825	0.0004	0.0006	-0.0032	0.1665	-3.2039
Median	0.8859	0.0004	0.0006	-0.0010	0.0023	-0.0320
Std. Dev.	0.0281	2.3×10^{-5}	7.8×10^{-5}	0.0046	0.2422	5.3613
IC1(95)	0.8135	0.0003	0.0005	-0.0154	0.0000	-19.807
IC2(95)	0.8917	0.0004	0.0008	0.0008	0.8009	0.0000

Table 3.3: Bootstrap statistics ($B = 1000$ samples of size 500).

The estimator $\widehat{\theta}_{1,b}$ is less volatile than $\widehat{\theta}_{2,b}$. This shows up as relatively large confidence regions for the components of $\widehat{\theta}_{2,b}$. Part of the variability of the latter can be explained by the fact it has been computed conditional on the point estimated $\widehat{\theta}_{1,b}$. However, it may be the case that the true parameter θ_2 is not constant across time. If θ_2 happens to be constant across time, the large standard deviations of $\widehat{\theta}_{2,b}$ primarily mean that the true θ_2 is not significantly different from zero. However, if θ_2 is time varying, the large standard deviations of $\widehat{\theta}_{2,b}$ take a quite different meaning. Namely, they may be interpreted as the level of heterogeneity of the possible realizations of the true parameter θ_2.

[6]Even in the discrete GMM, it is possible to construct an overidentified set of restrictions that is first order underidentified. See Dovonon and Renault (2009).

In this application, the large standard deviations of $\widehat{\theta}_{2,b}$ are most likely due to time variation in θ_2. In fact, the empirical distributions of $\widehat{\mu}_{1,b}$ and $\widehat{\delta}_b$ are very skewed: there are large differences between the means and the medians. Also, 50% of the realization of $\widehat{\mu}_{1,b}$ are above 0.0023 while 50% of the realizations of $\widehat{\delta}_b$ fall below -0.0320. Hence a significant proportion of samples lead to the conclusion that the returns are positively correlated with the expected risk and negatively correlated with the unpredictable risk.

3.6 Conclusions

The goal of this paper was to illustrate how to implement the CGMM. To start with, we briefly reviewed the useful theoretical properties of the CGMM estimator. Next, we exposed in details some helpful numerical recipes for the implementation of the CGMM. Finally, we applied the estimation method to the stable distribution and the autoregressive variance Gamma model.

When the parameter α of the stable distribution is close to 2, the asymmetry parameter β becomes hard to identify. As a result, the gradient of the moment function is numerically singular and one has to rely on Monte Carlo simulations for inference on the identifiable parameters. When α is far from 2, the gradient of the moment function is of full rank and the standard errors of the estimators can be computed using the standard analytical formulas. Overall, the parameters of the stable distribution can be reliably estimated by CGMM.

In the autoregressive Gamma model, the variances of the estimators cannot be computed analytically because the gradient of the moment is numerically singular. This problem is due to the fact that the objective function is extremely flat around the true values of the parameters, and can be linked to the difficulties inherent to the estimation of the degree of freedom parameter in Gamma distributions or Student distributions. We elude this problem by generating the empirical distributions of the estimates by resampling from the original sample. The empirical application with the Alcoa index suggest that the returns are positively correlated with the expected risk and negatively correlated with the unpredictable risk.

Bibliography

[1] Andrew, D.W.K. and J.C. Monahan (1992) "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator", *Econometrica,* vol. 60, 953-966.

[2] Barndorff-Nielsen, O. E. and Shephard, N. (2002) "Estimating quadratic variation using realized variance". *Journal of Applied Econometrics,* vol. 17, 457-477.

[3] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008) "Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise", *Econometrica,* vol. 76, 1481-1536.

[4] Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, vol. 31, 307-327.

[5] Carrasco M., Chernov M., Florens J. P. and E. Ghysels (2007) "Efficient estimation of general dynamic models with a continuum of moment conditions" *Journal of Econometrics,* vol. 140, 529–573.

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

[7] Carrasco, M. and J. P. Florens (2002) "Efficient GMM Estimation Using the Empirical Characteristic Function", *Working Paper,* Université de Montréal.

[8] Carrasco, M., J. P. Florens, and E. Renault (2007) "Linear Inverse Problems in Structural Econometrics: Estimation based on spectral decomposition and regularization", in the *Handbook of Econometrics*, vol. 6, edited by J.J. Heckman and E.E. Leamer.

[9] Carrasco, M. and R. Kotchoni (2008) "Efficient Estimation Using the Characteristic Function". *Working paper*, Université de Montréal.

[10] Carrasco, M. and R. Kotchoni (2009) "Assessing the Nature of Pricing Inefficiencies via Realized Measures". *Working paper*, Université de Montréal.

[11] Chacko, G. and L. Viceira (2003) "Spectral GMM estimation of continuous-time processes", *Journal of Econometrics*, vol. 116, 259-292.

[12] Chambers, J. M., Mallows, C. L. and Stuck, B.W. (1976), "A Method for Simulating Stable Random Variables", *Journal of the American Statistical Association,* vol. 71, 340-344. Corrections 82 (1987):704, 83 (1988): 581.

[13] Cornea, A. and Davidson, R. (2008) "A Refined Bootstrap for Heavy Tailed Distributions", *Working paper*, McGill University.

[14] Devroye, L. (1986) *"Non-Uniform Random Variate Generation"*, Spinger-Verlag.

[15] Dovonon, P. and E. Renault (2009) "GMM Overidentification Test with First Order Underidentification". *Working Paper*, University of North Carolina at Chapel Hill.

[16] Engle, R. F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflations", *Econometrica*, vol. 50, 987-1008.

[17] Fama, E. (1965) "The behavior of stock prices", *Journal of Business,* vol 47:244–280.

[18] Feller, W. (1951) "Two Singular Diffusion Problems", *Annals of Mathematics*, vol. 54, 173-182.

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

[20] Feuerverger, A. and P. McDunnough (1981a) "On Efficient Inference in Symmetry Stable Laws and Processes", *Csorgo, M., ed. Statistics and Related Topics.* New York: North Holland, 109-122.

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

[22] Garcia, R., Renault, E. and D. Veredas (2006) "Estimation of Stable Distributions by Indirect Inference", *CORE discussion paper* 2006/112.

[23] Gourieroux, C., Jasiak, J., (2005) "Autoregressive Gamma Processes". *Journal of Forecasting,* vol. 25, 129–152. Published online in Wiley InterScience (*www.interscience.wiley.com*) DOI: 10.1002/for.978.

[24] Gourieroux, C., Jasiak, J. and Sufana, R. (2005) "The Wishart Autoregressive Process". *Journal of Econometrics,* vol. 150 (2009) 167-181.

[25] Gourieroux, C., Monfort, A. and Renault, E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, vol. 8, s85-s118.

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

[27] Hansen, P.R and Lunde, A. (2006) "Realized variance and market microstructure noise" *Journal of Business & Economic Statistics*, vol. 24, No. 2.

[28] Heston, S. (1993) "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options", *The Review of Financial Studies*, 1993, vol. 6, number 2, 327-343.

[29] Hull, J., White, A., (1987) "The pricing of options on assets with stochastic volatility", *Journal of Finance,* vol. 3, 281–300.

[30] Jacod, J. (1994) "Limit of Random Measures Associated with the Increments of a Brownian Semi-martingale" No 120, *Laboratoire de Probabilities, Université Pierre et Marie Curie, Paris.*

[31] Jacod, J. & Protter, P. (1998) "Asymptotic error distribution for the Euler method for stochastic Differential Equations" *Annals of Probability,* vol. 26, 267-307.

[32] Jacquier, E., Polson, N., Rossi, P., 1994. "Bayesian analysis of stochastic volatility models (with discussion)". *Journal of Business and Economic Statistics,* vol. 12 (4), 371–417.

[33] Jacquier, E., Polson, N. G., and Rossi, P. E., (2004) "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors", *Journal of Econometrics,* vol. 122, 185 – 212.

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

[35] Knight, J. L. and J. Yu (2002) "The empirical characteristic function in time series estimation", *Econometric Theory,* vol. 18:691–721.

[36] Knight, J. L., Satchell, S. E. and J. Yu (2002) "Estimation of the stochastic volatility model by the empirical characteristic function method", *Australian New Zealand J. Statist.,* vol. 44:319–335.

[37] Koutrouvelis, I. A. (1980). "Regression-type Estimation of the Parameters of Stable Laws", *J. Am. Statist. Assoc.,* vol. 75:918–928.

[38] Madan, D. B., Carr, P., and Chang, E. C. (1998), "The variance gamma process and option pricing", *European Finance Review*, vol. 2, 79-105.

[39] Madan, D. B. and Senata, E. (1990)."The Variance Gamma Model for Share Market Returns", *Journal of Business*, 1990, vol. 63, no. 4

[40] Mandelbrot, B. (1963) "The variations of certain speculative prices" *Journal of Business* vol. 36:394–419.

[41] McCulloch, J.H. (1986), "Simple Consistent Estimators of Stable Distribution Parameters", *Communications in Statistics - Simulation and Computation*, vol. 15, 1109-1136.

[42] McCulloch, J. H. (1998) "Numerical approximation of the symmetric stable distribution and density" In *R. J. Adler, R. E. Feldman, & M. S. Taqqu (eds.), A Practical Guide to Heavy Tails*, 489–500. Boston: Birkhäuser.

[43] Newey, W. and West, K. (1987) "A simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica, vol.* 55, 703-708.

[44] Nolan, J. (1997) "Numerical calculation of stable densities and distribution functions", *Communication Statistics: Stochastic Models,* vol. 13, 759–774.

[45] Nolan, J. (1999) "Maximum Likelihood Estimation and Diagnostics for Stable Distributions", Manuscript, Math/Stat Department, American University.

[46] Nolan, J. (2008) "Stable Distributions Models for Heavy Tailed Data", Manuscript, Math/Stat Department, American University.

[47] Paulson, A. S., Holcomb W. E. and Leitch, R. A. (1975) "The estimation of the parameters of the stable laws", *Biometrika*, vol. 62(1):163-170.

[48] Press, S. J. (1972). "Estimation of univariate and multivariate stable distributions" *J. Am. Statist. Assoc.,* vol. 67:842–846.

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

[50] Stein, E. M. and Stein, J. C., (1991) "Stock Price Distribution with Stochastic Volatility: an Analytic Approach" *The Review of Financial Studies*, vol. 4 number 4, 727-752.

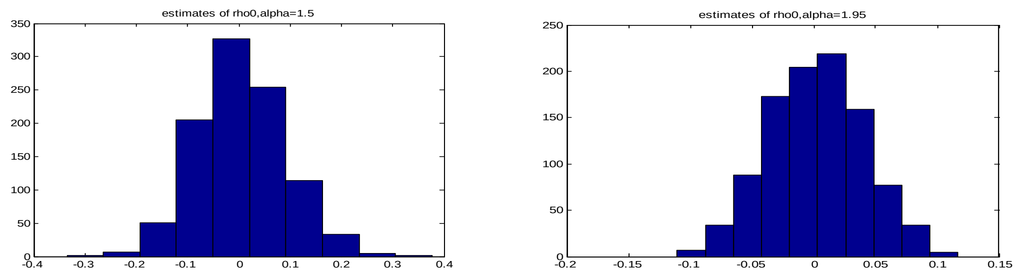
[51] Weron, R. (1996) "On the Chambers-Mallows-Stuck method for simulating skewed stable random variables" *Statistics & Probability Letters,* vol. 28, 165-171.

[52] Yu, J. (2004) "Empirical Characteristic Function Estimation and Its Applications", *Econometric Reviews*, vol. 23, No 2, 93-123.

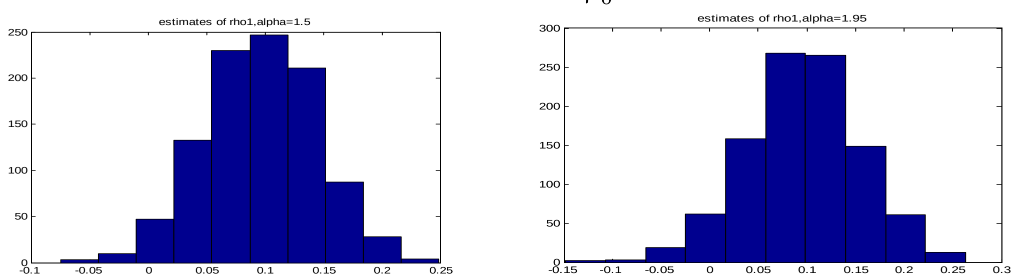
[53] Zolotarev, V. (1986), "One-Dimensional Stable Distributions" *American Mathematical Society, Providence, RI.* Russian original, 1983.

A. Estimation of Stable Distribution. Empirical distributions of the estimators for 1000 Monte Carlo replications.

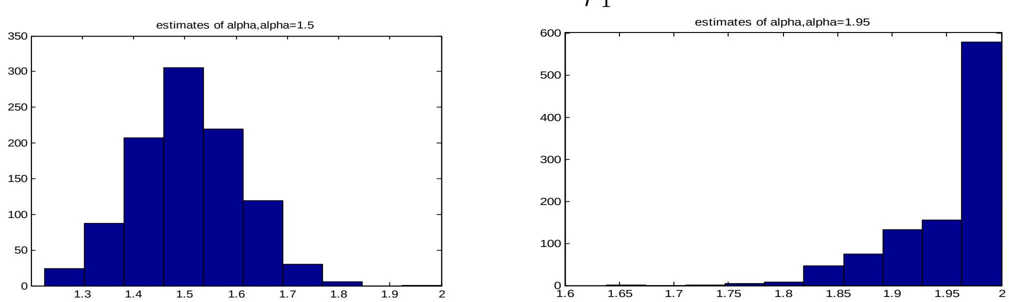
Left: $\theta_{01} = (0, 0.1, 1.5, 0, 0.5)$ Right: $\theta_{02} = (0, 0.1, 1.95, 0, 0.5)$



Estimates of ρ_0

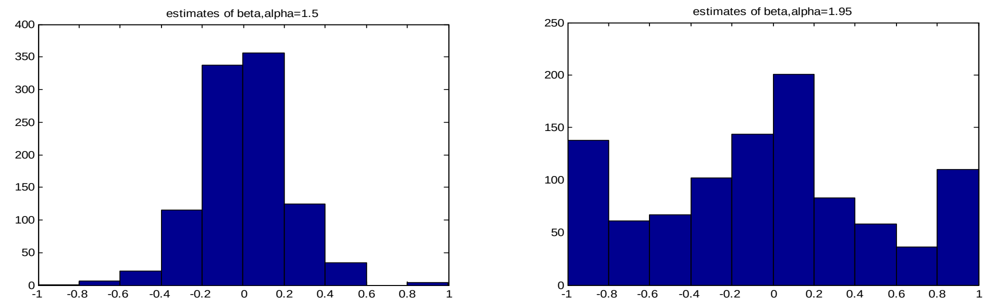


Estimates of ρ_1

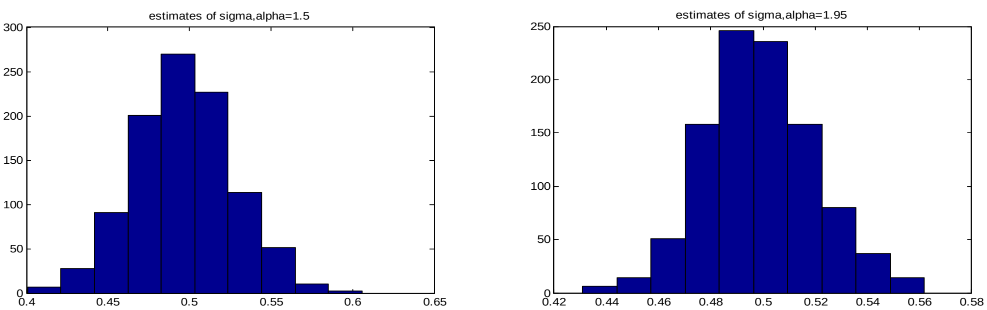


Estimates of α

Left: $\theta_{01} = (0, 0.1, 1.5, 0, 0.5)$ Right: $\theta_{02} = (0, 0.1, 1.95, 0, 0.5)$



Estimates of β



Estimates of σ

B. Derivatives of the Conditional CF of V_t in the ARVG model

The CF of V_t conditional on V_{t-1} is:

$$\varphi\left(\theta, V_{t-1}\right)=\left(1-\frac{i\tau\sigma^2\left(1-e^{-\kappa}\right)}{2\kappa}\right)^{-\frac{2\kappa\beta}{\sigma^2}} \exp\left(\frac{2i\kappa\tau e^{-\kappa}V_{t-1}}{2\kappa-i\tau\sigma^2\left(1-e^{-\kappa}\right)}\right)$$

Derivative with respect to β:

$$\begin{aligned}\frac{\partial\varphi\left(\theta, V_{t-1}\right)}{\partial\beta} &= \frac{-2\kappa}{\sigma^2}\ln\left(1-\frac{i\tau\sigma^2\left(1-e^{-\kappa}\right)}{2\kappa}\right)\exp\left(\frac{2\kappa i\tau e^{-\kappa}V_{t-1}}{2\kappa-i\tau\sigma^2\left(1-e^{-\kappa}\right)}\right)\\ &\times\left(1-\frac{i\tau\sigma^2\left(1-e^{-\kappa}\right)}{2\kappa}\right)^{-\frac{2\kappa\beta}{\sigma^2}}\end{aligned}$$

Derivative with respect to κ:

$$\begin{aligned}\frac{\partial\varphi\left(\theta, V_{t-1}\right)}{\partial\kappa} &= \exp\left(\frac{2\kappa i\tau e^{-\kappa}V_{t-1}}{2\kappa-i\tau\sigma^2\left(1-e^{-\kappa}\right)}\right)\left(1-\frac{i\tau\sigma^2\left(1-e^{-\kappa}\right)}{2\kappa}\right)^{-\frac{2\kappa\beta}{\sigma^2}}\\ &\times\left[\frac{-2\beta}{\sigma^2}\left(\ln\left(1-\frac{i\tau\sigma^2\left(1-e^{-\kappa}\right)}{2\kappa}\right)+\frac{i\tau\sigma^2\left(1-e^{-\kappa}-\kappa e^{-\kappa}\right)}{2\kappa-i\tau\sigma^2\left(1-e^{-\kappa}\right)}\right)\right.\\ &\left.-\frac{2i\tau e^{-\kappa}V_{t-1}}{2\kappa-i\sigma^2\tau\left(1-e^{-\kappa}\right)}\left(\frac{i\sigma^2\tau\left(\kappa-1+e^{-\kappa}\right)-2\kappa^2}{2\kappa-i\sigma^2\tau\left(1-e^{-\kappa}\right)}\right)\right]\end{aligned}$$

Derivative with respect to σ^2:

$$\begin{aligned}\frac{\partial\varphi\left(\theta, V_{t-1}\right)}{\partial\sigma^2} &= \exp\left(\frac{2\kappa i\tau e^{-\kappa}V_{t-1}}{2\kappa-i\tau\sigma^2\left(1-e^{-\kappa}\right)}\right)\left(1-\frac{i\tau\sigma^2\left(1-e^{-\kappa}\right)}{2\kappa}\right)^{-\frac{2\kappa\beta}{\sigma^2}}\\ &\times\left[\frac{2\kappa\beta}{\sigma^4}\left(\frac{i\tau\sigma^2\left(1-e^{-\kappa}\right)}{2\kappa-i\tau\sigma^2\left(1-e^{-\kappa}\right)}+\ln\left(1-\frac{i\sigma^2\tau\left(1-e^{-\kappa}\right)}{2\kappa}\right)\right)\right.\\ &\left.+\frac{i\tau\left(1-e^{-\kappa}\right)}{2\kappa-i\tau\sigma^2\left(1-e^{-\kappa}\right)}\frac{2i\kappa\tau e^{-\kappa}V_{t-1}}{2\kappa-i\tau\sigma^2\left(1-e^{-\kappa}\right)}\right]\end{aligned}$$

Chapter 4

A Solution to the Curse of Dimensionality in the Continuum GMM

Note: Cet article dont je suis l'unique auteur sera bientôt soumis pour publication dans un journal d'économétrie appliquée. Nous remercions Marine Carrasco et Pierre Evariste Nguimkeu pour leurs commentaires utiles.

Mots-Clés: Autoregressive Gamma, Bootstrap, Continuum of Moments Conditions, Realized Volatility

4.1 Introduction

The generalized method of moment (henceforth GMM) has been extended to handle a continuum of moments conditions by Carrasco and Florens (2000), and the resulting estimation procedure has been termed continuum-GMM (henceforth CGMM). A continuum of moment conditions arises for instance when one tries to estimate a parameter using moment conditions based on the characteristic function (henceforth CF). More precisely, let $x_t \in \mathbb{R}^d$ be a series of 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 $h_t(\tau, \theta_0)$ given by:

$$h_t(\tau, \theta_0) = \exp(i\tau' x_t) - \varphi(\tau, \theta_0), \tau \in \mathbb{R}^d, \tag{4.1}$$

where $\varphi(\tau,\theta_0) = E^{\theta_0}[\exp(i\tau' x_t)]$ and E^{θ_0} is the expectation operator with respect to the true data generating process. Because $E^{\theta_0}[h_t(\tau,\theta_0)] = 0$ for all $\tau \in \mathbb{R}^d$, we have a continuum of valid moments conditions that can be used to estimate θ_0 from observed sample. Moments conditions based on the CF are useful when the likelihood is either unavailable in closed form or non convenient to work with. For instance, the stable distribution and certain discretely sampled diffusion processes have known CFs but unknown likelihood functions.

Basically, 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 while the scalar product of the GMM is defined on a finite dimensional vector space, that of the CGMM is defined on an infinite dimensional Hilbert space. To fix ideas, let $\pi(\tau)$ be a probability density function on $\mathbb{R}^d$ and $L^2(\pi)$ denote the Hilbert space of complex valued functions that are square integrable with respect to π, that is:

$$\mathbf{L}^2(\pi) = \{f : \mathbb{R}^d \to \mathbb{C} \text{ such that } \int f(\tau)\overline{f(\tau)}\pi(\tau)d\tau < \infty\}. \tag{4.2}$$

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

$$\langle f, g \rangle = \int f(\tau)\overline{g(\tau)}\pi(\tau)d\tau = \underset{\pi(\tau)}{E}\left[f(\tau)\overline{g(\tau)}\right], \tag{4.3}$$

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

$$\begin{aligned} Q_T(\theta) &= \left\langle K^{-1/2}\widehat{h}_T(.,\theta), K^{-1/2}\widehat{h}_T(.,\theta)\right\rangle \\ &\equiv \underset{\pi(\tau)}{E}\left[K^{-1/2}\widehat{h}_T(\tau,\theta)\overline{K^{-1/2}\widehat{h}_T(\tau,\theta)}\right], \end{aligned} \tag{4.4}$$

where $\widehat{h}_T(\tau,\theta) = \frac{1}{T}\sum_{t=1}^{T} h_t(\tau,\theta)$ and K is the covariance operator K associated with the moment function. Finally, the CGMM estimator is defined as the particular value of θ that minimizes $Q_T(\theta)$.

In implementing the CGMM when x_t is multivariate ($d > 1$), a major difficulty lies in the evaluation of the multiple integrals embedded in the objective function $Q_T(\theta)$. Typically, one would choose the weighting function $\pi(\tau) = \exp(-\tau'\tau)$ in order to be able to use Gauss-Hermite quadrature methods. Quadrature methods are fast and accurate when $d \leq 2$. However, the complexity of the numerical integration grows exponentially as the dimension of τ increases. More precisely, if 10 quadrature points are needed to achieve a certain level of precision for a one-dimensional integration, about 10^d quadrature points are required to obtain the same level of precision in evaluating $Q_T(\theta)$. This is a well known "curse of dimensionality" in computational fields.

The situation gets even worse when x_t is not IID but Markov of order one. In this case, the moment function would be defined as:

$$h_{t+1}(\tau, \theta_0) = [\exp(i\tau_1' x_{t+1}) - \varphi(\tau_1, \theta_0, x_t)] \exp(i\tau_2' x_t), \tag{4.5}$$

where $\tau = (\tau_1, \tau_2) \in \mathbb{R}^{2d}$ and $\varphi(\tau_1, \theta_0, x_t) = E^{\theta_0}[\exp(i\tau_1' x_{t+1}) | x_t]$ is the CF of x_{t+1} given x_t. Accordingly, we would define the scalar product $\langle ., . \rangle$ on $\mathbf{L}^2(\pi) \times \mathbf{L}^2(\pi)$ in the same way as above but now using a probability measure $\pi(\tau_1, \tau_2)$ on $\mathbb{R}^{2d}$. If 10 quadrature points are required to get a desired level of precision for a one dimensional integration, about 10^{2d} quadrature points are required to obtain the same precision in evaluating $Q_T(\theta)$. This implies 10000 quadrature points when x_t is bivariate and 1000000 quadrature points when x_t is trivariate. Hence for values of d as low as 3, the implementation of the CGMM procedure becomes quickly an unfeasible task[1]. To circumvent this problem, a solution may consists in *(i)* discarding quadrature points that have very low weights, or *(ii)* reducing the number of quadrature points. Unfortunately, none of these solutions provide a substantial numerical efficiency gain without jeopardizing the accuracy of the overall estimation procedure.

The goal of this paper is to propose a solution to the type of curse of dimensionality just described. Our approach consists in turning an unfeasible optimization problem involving d-dimensional integrals into several feasible small scale optimizations problem involving only one-dimensional integrals. More precisely, instead of using the CF of $x_t \in \mathbb{R}^d$ in the CGMM

[1] This is particularly true when a large number of iterative evaluations of the objective function is required for the convergence of the minimization algorithm

procedure, we use the CF of $y_{\tau,t} = \tau' x_t \in \mathbb{R}$, for some fixed τ in the normalized set $\mathbf{S}^d$ given by:

$$\mathbf{S}^d = \left\{\tau \in \mathbb{R}^d : \|\tau\|_E = 1\right\}, \tag{4.6}$$

and $\|\|_E$ is the Euclidian norm. In the IID case for example, we would use the moments conditions:

$$h_{\tau,t}(u,\theta) = \exp\left(iuy_{\tau,t}\right) - E\left[\exp\left(iuy_{\tau,t}\right)\right], \;\; u \in \mathbb{R}, \tag{4.7}$$

where τ is fixed and u is the Fourier transformation variable. A suboptimal CGMM estimator is given by:

$$\widehat{\theta}^{(1)}(\tau) = \arg\min_{\theta} Q_{\tau,T}(\theta), \tag{4.8}$$

where

$$Q_{\tau,T}(\theta) = \underset{\omega(u)}{E}\left[\widehat{h}_{\tau,T}(u,\theta)\overline{\widehat{h}_{\tau,T}(u,\theta)}\right],$$

$\omega(u)$ is a univariate density on $\mathbb{R}$ and $\widehat{h}_{\tau,T}(u,\theta) = \frac{1}{T}\sum_{t=1}^{T} h_{\tau,t}(u,\theta)$. To make the overall estimation procedure independent of τ, we define the final estimator as the average:

$$\widehat{\theta}_\pi = \underset{\pi(\tau)}{E}\left[\widehat{\theta}^{(1)}(\tau)\right], \tag{4.9}$$

where $\pi(\tau)$ is a density on $\mathbf{S}^d$.

Because each draw from the distribution $\pi(\tau)$ gives rise to a new sample $y_{\tau,t} = \tau' x_t$, the proposed solution is basically a resampling technique. The sample $y_{\tau,t}$ contains the signature of the original multivariate distribution of x_t at the particular frequency τ. We shall thus refer to $\widehat{\theta}_\pi$ as the frequency domain resampling CGMM estimator (henceforth FCGMM). Two major theoretical issues are discussed in the sequel: the design of the optimal aggregating weight $\pi^*(\tau)$ and the efficiency of $\widehat{\theta}_{\pi^*}$ relatively to the maximum likelihood estimator. We found that the optimal weighting scheme is closely related to the inverse of the covariance operator associated with $\widehat{\theta}^{(1)}(\tau)$ viewed as a function of τ. We review the conditions under which $\widehat{\theta}_{\pi^*}$ is as optimal as the unfeasible maximum likelihood estimator.

A similar approach has been advocated by Chen, Jacho-Chavez and Linton (2009). These authors face a set of conditional moment restriction of type $E\left[\rho\left(Z_t,\theta_0\right)|X_t\right] = 0$, for some scalar function $\rho\left(Z_t,\theta_0\right)$. The standard approach in this literature consists in turning these conditional moment restrictions into unconditional moment restrictions by using $E\left[\rho\left(Z_t,\theta_0\right)A\left(X_t\right)\right]=0$, for any vector function $A\left(X_t\right)$. One then estimates the optimal instrument function $A_{oiv}\left(X_t\right)$, and the GMM estimator $\widehat{\theta}_{oiv}$ based on the unconditional moment restrictions $E\left[\rho\left(Z_t,\theta_0\right)A_{oiv}\left(X_t\right)\right]=0$ is called the optimal instrumental variable estimator. Chen, Jacho-Chavez and Linton (2009) proposed the alternative estimator $\widehat{\theta}_w=\sum_{j=1}^{N}w_j\widehat{\theta}_j$, where N is allowed to increase with the sample size, $\widehat{\theta}_j$ is the GMM estimator based on the moment restrictions $E\left[\rho\left(Z_t,\theta_0\right)A_j\left(X_t\right)\right]=0$, and $\left\{A_j\left(X_t\right)\right\}_{j=1}^{\infty}$ are basis functions chosen by the econometrician. It is shown in the paper that $\widehat{\theta}_w$ is as efficient as $\widehat{\theta}_{oiv}$ 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. In Section 3 we discuss the properties of the CGMM estimators $\widehat{\theta}^{(1)}(\tau)$. In Section 4 we derive the theoretically optimal aggregating weight $\pi^*(\tau)$ for the FCGMM estimator $\widehat{\theta}_\pi$. In particular, we compare the best FCGMM estimator $\widehat{\theta}_{\pi^*}$ to the maximum likelihood estimator. In Section 5 we present the feasible FCGMM estimator and show its asymptotic equivalence with $\widehat{\theta}_{\pi^*}$. Section 6 presents a Monte Carlo study based on the Autoregressive Factor Gamma Model. In Section 7 we used the FCGMM to fit an Autoregressive Variance Gamma of order p to the joint dynamic of the daily return on Alcoa and its realized variance. Finally, Section 8 concludes the paper. The proofs are gathered in appendix.

4.2 The General Framework

4.2.1 The Objective functions

Depending on the model under consideration, three types of moment functions can be used to implement the CF based CGMM. When x_t is IID, the moment function to use is given by (4.1). In the Markov case, the appropriate moment function is given by (4.5). 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:

$$h_t(\tau,\theta)=e^{i\tau' Y_t}-E^{\theta}(e^{i\tau' Y_t}),\tau\in\mathbb{R}^{dp}, \tag{4.10}$$

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 will show up quickly as p increases. For more discussions on the use of the moment function (4.10), see Jiang and Knight (2002), Yu (2004) and Carrasco, Chernov, Florens and Ghysels (2007).

From now on, let us use the generic notation $h_t(\tau, \theta)$, $\tau \in \mathbf{S}^d$ for any of the moments functions described above (that is, d denotes the dimension of τ in either model), and define:

$$\widehat{h}_{\tau,t}(u, \theta) = h_t(u\tau, \theta),\ u \in \mathbb{R},$$

where we recall that $\mathbf{S}^d = \left\{\tau \in \mathbb{R}^d : \|\tau\|_E = 1\right\}$. Note that in the IID case, the moment function $\widehat{h}_{\tau,t}(u, \theta)$ reduces to (4.7) while in the Markov case, $\widehat{h}_{\tau,t}(u, \theta)$ can be written as:

$$\widehat{h}_{\tau,t}(u, \theta) = \left[\exp\left(iuy_{\tau_1,t}\right) - E\left(\exp\left(iuy_{\tau_1,t}\right)\right)\right]\exp\left(iuy_{\tau_2,t-1}\right), \tag{4.11}$$

where $y_{\tau_1,t} = \tau_1' x_t$ and $\tau = (\tau_1, \tau_2) \in \mathbf{S}^d$. Note that $y_{\tau_2,t-1} = \tau_2' x_t$ is being used as instrument in (4.11). Here if τ_1 was not kept fixed, the FCGMM estimator would be the continuum version of the alternative optimal instrumental variable of Chen, Jacho-Chavez and Linton (2009). Finally in the dependent case, the moment function may be designed as follows:

$$\widehat{h}_{\tau,t}(u, \theta) = e^{iuY_{\tau,t}} - E^{\theta}(e^{iuY_{\tau,t}}),$$

where $\tau = (\tau_1, ..., \tau_p) \in \mathbf{S}^d$, $Y_{\tau,t} = \sum_{k=1}^{p} y_{\tau_k,t}$ and $y_{\tau_k,t} = \tau_k' x_{t-k+1}$.

Having defined the appropriate moment function for the model under consideration, the suboptimal CGMM estimator $\widehat{\theta}^{(1)}(\tau)$ indexed by τ is defined as in (4.8). In practice, $\widehat{\theta}^{(1)}(\tau)$ may be use to estimate the covariance operator associated with the moments function K_τ that enters in the definition of the optimal second step CGMM estimator indexed by τ:

$$\widehat{\theta}^{(2)}(\tau) = \arg\min_{\theta} Q_{\tau,T}^{(2)}(\theta), \tag{4.12}$$

where

$$Q_{\tau,T}^{(2)}(\theta) = \underset{\omega(u)}{E}\left[K_\tau^{-1/2}\widehat{h}_{\tau,T}(u, \theta) \overline{K_\tau^{-1/2}\widehat{h}_{\tau,T}(u, \theta)}\right],$$

In IID and Markov models, K_τ is the linear operator with kernel

$$k_\tau\left(u_1,u_2\right)=E^{\theta_0}\left[h_{\tau,t}(u_1,\theta)\overline{h_{\tau,t}(u_2,\theta)}\right], \tag{4.13}$$

and $K_\tau f\left(u_1\right)=\int k_\tau\left(u_1,u_2\right)f\left(u_2\right)du_2$ for all $f:\mathbb{R}\to\mathbb{R}$. In dependent models, we have:

$$\begin{aligned} k_\tau(u_1,u_2) &= E^{\theta_0}\left[h_{\tau,t}(u_1,\theta)\overline{h_{\tau,t}(u_2,\theta)}\right] \\ &\quad+\sum_{j=1}^{\infty}E^{\theta_0}\left[h_{\tau,t}(u_1,\theta)\left(\overline{h_{\tau,t-j}(u_2,\theta)}+\overline{h_{\tau,t+j}(u_2,\theta)}\right)\right]. \end{aligned} \tag{4.14}$$

We now discuss the useful assumptions in this framework.

4.2.2 The Assumptions

The following assumptions will be used to study the properties of the estimators.

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

Assumption 2: For all $\tau\in\mathbf{S}^d\backslash\aleph$, the equation

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

has a unique solution θ_0 which is an interior point of a compact set $\mathbf{\Theta}$, where $\aleph$ is a null set with respect to π, E^{θ_0} denotes the expectation with respect to the distribution of the data at $\theta=\theta_0$.

Assumption 3: For all $\tau\in\mathbf{S}^d\backslash\aleph$, $h_{\tau,t}(u,\theta)$ is three time continuously differentiable with respect to θ.

Assumption 4: For all θ and $\tau\in\mathbf{S}^d\backslash\aleph$, $E^{\theta_0}\left[h_{\tau,T}(.,\theta)\right]$ and its first three derivatives with respect to θ belong to the range of K_τ^{β} for $\beta\geq 1/2$, but not to the range of $K_\tau^{\beta+\varepsilon}$ for all $\varepsilon>0$, where K_τ is the asymptotic covariance operator associated with the moment function $h_{\tau,t}(.,\theta)$.

Assumption 5: $h_{\tau,t}(u,\theta)$ is at least twice continuously differentiable with respect to τ in $\mathbf{S}^d\backslash\aleph$.

Assumption 6: (i) $\frac{\partial^2 Q_{\tau,T}}{\partial\theta\partial\theta'}$ is positive definite and (ii) $\frac{\partial^2 Q_{\tau,T}}{\partial\theta\partial\tau}$ is of full rank in $\mathbf{S}^d\backslash\aleph$.

Assumption 7: The measure $\pi(\tau)$ on $\mathbf{S}^d$ 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 $x(., \varepsilon_t, Z_{t-1})$ is three times continuously differentiable with respect to θ, ε_t is a IID white noise whose distribution does not depend on θ_0, and Z_{t-1} can only contain lagged values of x_t.

The first assumption ensures that $0 < \underset{\omega(u)}{E}\left[f(u)\overline{f(u)}\right] < \infty$ for all $f \neq 0$. Assumption 2 is an identification assumption that may not hold in all situations. This assumption will hold in a model of a univariate process x_t conditional on past realizations $(x_{t-1}, ..., x_{t-d+1})$, where the curse of dimensionality comes from conditioning on long lagged. This assumption may not hold in the joint model of $(x_{1,t}, x_{2,t}, ..., x_{d,t})$ if each $x_{i,t}$ is draw from a given stable distribution, since in this case $y_{\tau,t} = \sum_{i=1}^{d} \tau_i x_{i,t}$ also belongs to a stable distribution with same stability index. In this case, the mean parameters of the individual $x_{i,t}$s may not be identifiable from the marginal distribution of $y_{\tau,t}$. This problem can be circumvented by adopting a copula-type approach, where the specification of the margins of the $x_{i,t}$s are made independent from the form of the co-dependence between the $x_{i,t}$s.

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. Assumption 4 ensures that the limit of the objective function as T goes to infinity is well defined. Assumptions 5 and 6 ensures that $\widehat{\theta}^{(1)}(\tau)$ is unique and is a smooth function of τ. Assumption 2 already ensures the positive definiteness of $\frac{\partial^2 Q_{\tau,T}}{\partial\theta\partial\theta'}$ as T goes to infinity, but we request this in finite sample for simplicity. The measure $\pi(\tau)$ in Assumption 7 need not be positive for all τ. 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 $\widehat{\theta}^{(2)}(\tau)$.

In the next section, we recall the properties of the CGMM estimators $\widehat{\theta}^{(1)}(\tau)$ and $\widehat{\theta}^{(2)}(\tau)$.

4.3 Properties of the CGMM Estimators

Under assumptions 1 to 4, $\widehat{\theta}^{(1)}(\tau)$ is consistent for θ_0 (for almost all τ) 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 $\widehat{\theta}^{(1)}(\tau)$.

Proposition 1 *Under Assumptions 1 to 6, $\widehat{\theta}^{(1)}(\tau)$ is unique for each τ in $\mathbf{S}^d \backslash \aleph$. Moreover, $\widehat{\theta}^{(1)}(\tau)$ is continuously differentiable with respect to τ.*

This result is useful for the derivation of the minimum variance FCGMM estimator and for the comparison of the latter with the maximum likelihood estimator. A key ingredient for the derivation of this result is the positive definiteness of the matrix $\frac{\partial^2 Q_{\tau,T}}{\partial\theta\partial\theta'}$ which is guaranteed to hold according to Assumption 6.

The estimator $\widehat{\theta}^{(1)}(\tau)$ is not optimal in the sense that its variance does not reach the Cramer-Rao bound associated with the likelihood of $y_{\tau,t}$. However, it can be used to consistently estimate the covariance operator K_τ. In IID and Markov models, a natural estimator of K_τ is given by the linear empirical operator $K_{\tau,T}$ with kernel:

$$\widehat{k}_\tau(u_1, u_2) = \frac{1}{T}\sum_{t=1}^{T} h_{\tau,t}(u_1, \widehat{\theta}^{(1)})\overline{h_{\tau,t}(u_2, \widehat{\theta}^{(1)})}, \tag{4.15}$$

where $\widehat{\theta}^{(1)} \equiv \widehat{\theta}^{(1)}(\tau)$ is defined in (4.8). In IID models specifically, the first step estimator $\widehat{\theta}^{(1)}(\tau)$ may be bypassed by using:

$$\widehat{k}_\tau(u_1, u_2) = \frac{1}{T}\sum_{t=1}^{T} \left(e^{iu_1 y_{\tau,t}} - \widehat{\varphi}_{\tau,T}\right)\overline{\left(e^{iu_1 y_{\tau,t}} - \widehat{\varphi}_{\tau,T}\right)}, \tag{4.16}$$

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

$$\begin{aligned} \widehat{k}_\tau(u_1, u_2) &= \frac{1}{T}\sum_{t=1}^{T} h_{\tau,t}(u_1, \widehat{\theta}^{(1)})\overline{h_{\tau,t}(u_2, \widehat{\theta}^{(1)})} \qquad (4.17) \\ &+ \sum_{j=1}^{J_T}\left(1 - \frac{j-1}{J_T}\right)\sum_{t=1}^{T} h_{\tau,t}(u_1, \widehat{\theta}^{(1)})\left(\overline{h_{\tau,t-j}(u_2, \widehat{\theta}^{(1)})} + \overline{h_{\tau,t+j}(u_2, \widehat{\theta}^{(1)})}\right), \end{aligned}$$

and J_T is a bandwidth that is increasing in T.

The operator K_τ 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 $\widehat{l}_{\tau,i}$ be an eigenvalue of $K_{\tau,T}$ associated with the eigenfunction $\widehat{\psi}_{\tau,i}$, then we can label $\widehat{l}_{\tau,i}$ and $\widehat{\psi}_{\tau,i}$ so that: (i) $\widehat{l}_{\tau,1} < \infty$, (ii) $\widehat{l}_{\tau,i} > \widehat{l}_{\tau,i+1} \geq 0$

for all i, and (iii) $\widehat{l}_{\tau,i} = 0$ for all $i > T$, where T is the sample size[2]. As a result, $K_{\tau,T}$ is not invertible on $L^2(\omega)$. To estimate K_τ^{-1}, the following generalized inverse is used:

$$K_{\tau,T,\alpha_T}^{-1} = \left(K_{\tau,T}^2 + \alpha I\right)^{-1} K_{\tau,T}.$$

With the same notations as above, it can be checked that $\widehat{\psi}_{\tau,i}$ is an eigenfunction of K_{τ,T,α_T}^{-1} associated with the eigenvalue $\frac{\widehat{l}_{\tau,i}}{\widehat{l}_{\tau,i}^2+\alpha_T}$.

In IID and Markov models, Under Assumptions 1 and 2 $K_{\tau,T}$ satisfies:

$$\|K_{\tau,T} - K\| = O_p\left(T^{-1/2}\right).$$

where K is the covariance operator defined in equation (4.4). The regularized inverse K_{τ,T,α_T}^{-1} has the property that for any function f in the range of $K_{\tau,T}^{1/2}$, the function $K_{\tau,T,\alpha_T}^{-1/2} f$ converges to $K^{-1/2} f$ as T goes to infinity and α_T goes to zero. Assumptions 1 to 4 then ensure that replacing $K_\tau^{-1/2}$ by $K_{\tau,T,\alpha_T}^{-1/2}$ in (4.12) yields[3]:

$$T^{1/2}\left(\widehat{\theta}^{(2)}(\tau) - \theta_0\right) \xrightarrow{L} N(0, I_{\tau,\theta_0}^{-1}), \tag{4.18}$$

as T and $\alpha_T^{3/2} T$ go to infinity and α_T goes to zero, where I_{τ,θ_0}^{-1} denote the inverse of the Fisher Information Matrix associated with the likelihood of $y_{\tau,t}$.

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 $\widehat{\theta}^{(2)}(\tau)$ is optimal among the following class indexed by a linear operator B:

$$\arg\min_{\theta} \underset{\omega(u)}{E}\left[B\widehat{h}_{\tau,T}(u,\theta)\overline{B\widehat{h}_{\tau,T}(u,\theta)}\right].$$

[2]See Carrasco and Florens (2000) and Carrasco, Florens and Renault (2007) for more details on covariance operators.

[3]The consistency and optimality is guaranted for $\beta \geq 1/2$. However, 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.

In order for $\widehat{\theta}^{(2)}(\tau)$ to be truly optimal in the sense of (4.18), the regularization parameter α_T needs to be calibrated in practice. Let $\alpha_T(\theta_0)$ be the value of α_T that optimal in the mean square error sense, that is:

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

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

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

and $\widehat{\theta}_T^{(2,k)}(\tau)$ is the second step CGMM estimator of θ_0 computed using a sample simulated from the data generating process indexed by the point estimate $\widehat{\theta}^{(1)}(\tau)$, and M is the total number of simulated samples.

In the next section, we discuss the properties of the FCGMM estimator.

4.4 The Ideal FCGMM Estimator

In Equation (4.9), we have defined the FCGMM estimator as the weighted sum of an infinite number of $\sqrt{T}$-consistent estimators indexed by τ, that is:

$$\widehat{\theta}_{\pi} = \int \widehat{\theta}^{(1)}(\tau)\, \pi(\tau)\, d\tau$$

where $\pi(\tau)$ is a pdf on $\mathbf{S}^d$. The continuity of $\widehat{\theta}^{(1)}(\tau)$ as a function of τ allows to consider the use of continuous pdfs $\pi(\tau)$ for the weighting function. Below we discuss the consistency of $\widehat{\theta}_{\pi}$ and derive the weighting function π^* that minimizes the variance. The ideal FCGMM estimator $\widehat{\theta}_{\pi^*}$ is then compared to the maximum likelihood estimator.

4.4.1 Consistency and Optimal Aggregating Measure

For any $\lambda \in \mathbb{R}^q$, the variance of $\lambda'\widehat{\theta}_\pi$ is given by:

$$Var\left(\sqrt{T}\lambda'\widehat{\theta}_\pi\right) = \int\int g_{T,\lambda}\left(\tau_1, \tau_2\right)\pi\left(\tau_1\right)\pi\left(\tau_2\right)d\tau_1 d\tau_2, \tag{4.19}$$

where

$$g_{T,\lambda}\left(\tau_1, \tau_2\right) = \lambda' Cov\left(\sqrt{T}\widehat{\theta}^{(1)}\left(\tau_1\right), \sqrt{T}\widehat{\theta}^{(1)}\left(\tau_2\right)\right)\lambda. \tag{4.20}$$

A first order Taylor expansion of $\widehat{\theta}^{(1)}\left(\tau\right)$ deduced from the first order condition it solves can be used to show that:

$$\begin{aligned} g_\lambda\left(\tau_1, \tau_2\right) &\equiv P\lim g_{T,\lambda}\left(\tau_1, \tau_2\right) \\ &= \lambda' W_{\tau_1}^{-1}\left\langle G_{\tau_1}(., \theta_0), K_{\tau_1,\tau_2} G_{\tau_2}(., \theta_0)\right\rangle W_{\tau_2}^{-1}\lambda, \end{aligned} \tag{4.21}$$

where K_{τ_1,τ_2} is the operator with kernel:

$$k_{\tau_1,\tau_2}\left(u, v\right) = Cov\left(\sqrt{T}\widehat{h}_{\tau_1,T}(u, \theta_0), \sqrt{T}\widehat{h}_{\tau_2,T}(v, \theta_0)\right),$$

and

$$W_\tau = \left\langle G_\tau(., \theta_0), G_\tau(., \theta_0)\right\rangle; \;\; G_\tau(u, \theta_0) = P\lim \frac{\partial \widehat{h}_{\tau,T}(u, \widehat{\theta}^{(1)}\left(\tau\right))}{\partial\theta}$$

We have the following consistency result for $\widehat{\theta}_\pi$.

Proposition 2 *Under Assumptions 1 to 4 and Assumptions 7, the FCGMM estimator satisfies:* $\widehat{\theta}_\pi - \theta_0 = O_p\left(T^{-1/2}\right)$.

Clearly, the optimal aggregating measure depends on the choice of λ. In practice, λ may be set according to some particular hypothesis one which to test on $\widehat{\theta}_\pi$. The ideal measure $\pi_\lambda^*\left(\tau\right)$ solves:

$$\pi_\lambda^* = \arg\min_\pi \int\int g_\lambda\left(\tau_1, \tau_2\right)\pi\left(\tau_1\right)\pi\left(\tau_2\right)d\tau_1 d\tau_2, \tag{4.22}$$

subject to $\int \pi\left(\tau_1\right)d\tau_1 = 1$.

Let V_λ be the linear operator with kernel $g_\lambda\left(\tau_1, \tau_2\right)$, that is, the asymptotic covariance operator associated with $\lambda'\widehat{\theta}^{(1)}(\tau)$. The operator V_λ is compact if we have:

$$\int_{\mathbf{S}^d}\int_{\mathbf{S}^d}\left[g_\lambda(\tau_1,\tau_2)\right]^2 d\tau_1 d\tau_2<\infty$$

This condition is satisfied because $\mathbf{S}^d$ is a bounded set while $g_\lambda(\tau_1,\tau_2)$ is finite and continuous at all (τ_1,τ_2). These properties of $g_\lambda(\tau_1,\tau_2)$ follow from the consistency of $\widehat{\theta}^{(1)}(\tau)$ and its continuity as a function of τ. The compactness of the covariance operator V_λ ensures that it has a discrete spectrum. If we let $\phi_{\lambda,j}(\tau_1)$ denote the eigenfunction of V_λ 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 (4.22) $\pi^*_\lambda(\tau)$ with minimal norm is given by:*

$$\pi^*_\lambda(\tau)=\left[\sum_{j=1}^{\infty}\frac{1}{\nu_{\lambda,j}}\left(\int\phi_{\lambda,j}(\tau_1)\,d\tau_1\right)^2\right]^{-1}\sum_{j=1}^{\infty}\frac{1}{\nu_{\lambda,j}}\left(\int\phi_{\lambda,j}(\tau_1)\,d\tau_1\right)\phi_{\lambda,j}(\tau), \tag{4.23}$$

At the optimum, the variance of $\lambda'\widehat{\theta}_\pi$ is:

$$Var\left(\lambda'\widehat{\theta}_{\pi^*_\lambda}\right)=\left[\sum_{j=1}^{\infty}\frac{1}{\nu_{\lambda,j}}\left(\int\phi_{\lambda,j}(\tau)\,d\tau\right)^2\right]^{-1}. \tag{4.24}$$

Note that $\pi^*_\lambda(\tau)+\widetilde{f}(\tau)$ is also a solution of (4.22) for any function $\widetilde{f}(\tau_1)$ in the null set of V_λ, and $\pi^*_\lambda(\tau)$ is the unique solution if the null set of V_λ reduces to the null function. Below, we compare the efficiency of $\widehat{\theta}_{\pi^*_\lambda}$ to that of the maximum likelihood estimator.

4.4.2 Comparison with the Maximum Likelihood

The CGMM estimator may be used when the computation of the maximum likelihood is more costly. In a situation where even the CGMM estimator itself is unfeasible due to the curse of dimensionality, the FCGMM becomes an excellent alternative that delivers at least a $\sqrt{T}$-consistent estimator of the parameter of interest. In this section, we discuss the conditions under which $\widehat{\theta}_{\pi^*_\lambda}$ is as efficient as the unfeasible maximum likelihood estimator.

Let $\widehat{\theta}_{MLE}$ be the unknown maximum likelihood estimator of θ_0, and define the linear manifold $\widehat{D}_T(\theta_0)$ by:

$$\widehat{D}_T(\theta_0) = \left\{\theta \in \mathbb{R}^q \text{ s.t } \theta = \int \pi(\tau)\,\widehat{\theta}^{(1)}(\tau)\,d\tau \text{ and } \int \pi(\tau)\,d\tau = 1\right\}. \tag{4.25}$$

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

$$\int \pi^*(\tau)\,\widehat{\theta}^{(1)}(\tau)\,d\tau = \widehat{\theta}_{MLE}, \tag{4.26}$$

In this case, $\widehat{\theta}_{MLE} \in \widehat{D}_T(\theta_0)$ for each sample and we have:

$$Var\left(\lambda'\widehat{\theta}_{\pi^*_\lambda}\right) = \underset{\widehat{\theta}_\pi \in \widehat{D}_T(\theta_0)}{\arg\min}\, Var\left(\lambda'\widehat{\theta}_\pi\right) \leq Var\left(\lambda'\widehat{\theta}_{MLE}\right).$$

When $d = 1$, the normalized set $\mathbf{S}^d$ reduces to the singleton $\{\tau = 1\}$ and $\widehat{D}_T(\theta_0) = \left\{\widehat{\theta}^{(1)}(1)\right\}$. In this case, it is clear that $\lambda'\widehat{\theta}_{\pi^*_\lambda}$ is never as efficient as the $\lambda'\widehat{\theta}_{MLE}$. But when $d \geq 2$, $\mathbf{S}^d$ contains a continuum of normalized vectors τ and $\left\{\widehat{\theta}^{(1)}(\tau), \tau \in \mathbf{S}^d\right\}$ is a continuum of estimators. The following proposition then gives a condition under which Equation (1.31) is satisfied.

Proposition 4 *Under Assumptions 1 to 7 and $d \geq \max\{q, 2\}$, the optimal FCGMM estimator $\lambda'\widehat{\theta}_{\pi^*_\lambda}$ is as efficient as the unfeasible maximum likelihood estimator $\lambda'\widehat{\theta}_{MLE}$.*

The intuition behind this result is the following. Around a particular τ, we have:

$$\widehat{\theta}^{(1)}(\tau + \tau_0) = \widehat{\theta}^{(1)}(\tau) + \frac{\partial\widehat{\theta}^{(1)}(\tau)}{\partial\tau}\tau_0. \tag{4.27}$$

When the rank of $\frac{\partial\widehat{\theta}_T(\tau)}{\partial\tau}$ is equal to the dimensionality of θ_0, the linear manifold $\widehat{D}_T(\theta_0)$ replicates the entire parameter space Θ. In this case, $\lambda'\widehat{\theta}_{\pi^*_\lambda}$ is as efficient as $\lambda'\widehat{\theta}_{MLE}$ because (4.26) then holds. Note that a necessary condition for this rank condition to be satisfies is $q \leq d$, that is, we must have less coordinates in θ_0 than there are dimensions in τ.

If $\widehat{D}_T(\theta_0)$ has less than q dimensions, it can still encompass $\widehat{\theta}_{MLE}$ but there is no simple way to verify this. Intuitively, $\widehat{D}_T(\theta_0)$ is more likely to encompass $\widehat{\theta}_{MLE}$ is there is enough variability in the set $\widehat{\theta}^{(1)}(\tau)$ across τ. In this regard, using the suboptimal CGMM estimator

$\widehat{\theta}^{(1)}(\tau)$ in the definition of $\widehat{\theta}_{\pi}$ has two advantages. First of all, $\widehat{\theta}^{(1)}(\tau)$ is less efficient than $\widehat{\theta}^{(2)}(\tau)$ and thus has more variability than the latter, thus allowing the manifold $\widehat{D}_T(\theta_0)$ to have more probability to encompass the maximum likelihood estimator. And secondly, the use of $\widehat{\theta}^{(1)}(\tau)$ makes the computation of the FCGMM estimator easier because this allows to bypass the estimation of $K_{\tau}^{-1/2}$.

In the next section, we discuss the feasible FCGMM.

4.5 The Feasible Optimal FCGMM

If we knew how to draw from the optimal aggregating measure $\pi_{\lambda}^{*}()$, a natural way to implement the CGMM would be to draw $\tau_1, ..., \tau_S$ from this distribution, compute $\widehat{\theta}(\tau_s)$ for each τ_s and take the average. We would have:

$$\widehat{\theta}_{\pi_{\lambda}^{*}} \simeq \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}(\tau_s).$$

where τ_s is a draw from $\pi_{\lambda}^{*}()$. This would be a Monte Carlo approximation of the integral $\widehat{\theta}_{\pi_{\lambda}^{*}} = \int \widehat{\theta}(\tau)\pi_{\lambda}^{*}(\tau)d\tau$. Unfortunately, $\pi_{\lambda}^{*}(\tau)$ has an intractable form and we do not know how to draw directly from this distribution. Interestingly, another Monte Carlo approximation of this integral is given by:

$$\widehat{\theta}_{\pi_{\lambda}^{*}} \simeq \sum_{s=1}^{S}\pi_{\lambda}^{*}(\tau_s)\widehat{\theta}(\tau_s), \tag{4.28}$$

where τ_s is a draw from the multivariate uniform distribution on $\mathbf{S}^d$. When implementing (4.28), the main challenge is to estimate the optimal aggregating weight $\pi_{\lambda}^{*}()$. This requires the estimation of the asymptotic covariance operator V_{λ} whose kernel $g_{\lambda}(\tau_1, \tau_2)$ is given by (4.21). The expression of $g_{\lambda}(\tau_1, \tau_2)$ involves the gradient $G_{\tau}(., \theta_0)$ which is not always easy to compute by hand. To avoid this difficulty, we suggest to use a simulation approach presented below.

To start with, we note that the kernel function $g_{\lambda}(\tau_1, \tau_2)$ can be simulated using the formula:

$$\widehat{g}_{\lambda}(\tau_1, \tau_2) = \frac{1}{L}\sum_{l=1}^{L}\lambda'\left(\widehat{\theta}(\tau_1, l) - \overline{\widehat{\theta}}(\tau_1, l)\right)\left(\widehat{\theta}(\tau_2, l) - \overline{\widehat{\theta}}(\tau_2, l)\right)'\lambda. \tag{4.29}$$

where $\overline{\overline{\widehat{\theta}}}(\tau_1, l) = \frac{1}{L}\sum_{l=1}^{L}\widehat{\theta}(\tau_1, l)$ and $\left\{\widehat{\theta}(\tau_i, l)\right\}_{l=1}^{L}$ are l independent copies of $\widehat{\theta}(\tau_i, l), i = 1, 2$.

Let $\tau_1, ..., \tau_S$ be S draws from the multivariate uniform distribution on $\mathbf{S}^d$, and assume that we can simulate from the true data generating process. Further let $\left\{x_t^{(l)}\right\}_{t=1}^{T}, l = 1, ..., L$ be L independent samples of size T simulated from the distribution of x_t. For each sample indexed by l and each τ_s, we compute the univariate samples:

$$\left\{y_{\tau_s,t}^{(l)}\right\} = \left\{\tau_s' x_t^{(l)}\right\}, \ s = 1, ..., S \text{ and } l = 1, ..., L. \tag{4.30}$$

Finally, let $\widehat{\theta}(\tau_s, l)$ be the first step CGMM estimator based on the sample $\left\{y_{\tau_s,t}^{(l)}\right\}$, and $\widehat{\Theta}_\lambda$ be the $L \times S$ matrix with (l, s) element given by $\lambda'\widehat{\theta}(\tau_s, l)$. Note that $\widehat{\theta}(\tau_s, l), l = 1, ..., L$ are IID copies of the CGMM estimator $\widehat{\theta}^{(1)}(\tau_s)$. Rigorously, it is not possible to draw from the true data generating process because θ_0 is unknown. However, one can proxy θ_0 by the consistent estimator $\widehat{\theta}_S = \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}^{(1)}(\tau_s)$ computed from the actual data.

A degenerate estimator of the covariance operator associated with $\widehat{\theta}^{(1)}(\tau)$ is given by the $(S \times S)$ empirical covariance matrix of $\widehat{\Theta}_\lambda$:

$$\widehat{V}_\lambda = \frac{1}{L}\left(\widehat{\Theta}_\lambda - \overline{\overline{\widehat{\Theta}}}_\lambda\right)'\left(\widehat{\Theta}_\lambda - \overline{\overline{\widehat{\Theta}}}_\lambda\right), \tag{4.31}$$

where $\overline{\overline{\widehat{\Theta}}}_\lambda$ is the matrix with (l, s) element given by $\lambda'\overline{\overline{\widehat{\theta}}}(\tau_s, l)$. The following proposition shows that $\widehat{V}_\lambda$ is consistent for V_λ in the following sense:

Proposition 5 *Let us define $\underline{f} = (f(\tau_1), ..., f(\tau_S))'$ where $\tau_1, ..., \tau_S$ are S draws from the multivariate uniform distribution on $\mathbf{S}^d$. Then as L and S go to infinity, we have:*

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

for all τ_i■

We may thus use $\widehat{V}_\lambda$ to estimate the optimal aggregating weight π_λ^* by:

$$\widehat{\pi}_{\lambda,\alpha}^* = \left(\widehat{\pi}_{\lambda,\alpha}^*(\tau_1), ..., \widehat{\pi}_{\lambda,\alpha}^*(\tau_S)\right)' = \left(\iota'\widehat{V}_{\lambda,\alpha}^{-1}\iota\right)^{-1}\iota'\widehat{V}_{\lambda,\alpha}^{-1}, \tag{4.32}$$

where ι is a vector of ones and $\widehat{V}_{\lambda,\alpha}^{-1}$ is the regularized inverse of $\widehat{V}_{\lambda}$ defined as:

$$\widehat{V}_{\lambda,\alpha}^{-1} = \left(\widehat{V}_{\lambda}^{2} + \alpha I\right)^{-1} \widehat{V}_{\lambda},\ \alpha \in (0,1). \tag{4.33}$$

This regularization is necessary because $\widehat{V}_{\lambda}$ is nearly singular or singular for sufficiently large S due to the fact that any two elements of the set $\tau_1, ..., \tau_S$ can eventually be arbitrarily close. The following result are the main ingredient for the proof of the consistency of the feasible optimal FCGMM estimator given by:

$$\widehat{\theta}_{\widehat{\pi}_{\lambda,\alpha}^{*}} = \sum_{s=1}^{S} \widehat{\pi}_{\lambda,\alpha}^{*}(\tau_s)\, \widehat{\theta}^{(1)}(\tau_s). \tag{4.34}$$

where $\widehat{\theta}^{(1)}(\tau_s)$ is computed from the actual data.

Proposition 6 *Let f be a function such that ϵ is the largest real number for which $\left\|V_{\lambda}^{-\epsilon} f\right\| < \infty$, and $\epsilon \geq 1$. Then under Assumptions 1 to 7, we have:*

$$\begin{aligned} \left\|\widehat{V}_{\lambda,\alpha}^{-1} - V_{\lambda,\alpha}^{-1}\right\| &= O_p\left(\alpha^{-3/2} L^{-1/2}\right) + O_p\left(\alpha^{-3/2} S^{-1/2}\right), && (4.35)\\ \left\|\left(\widehat{V}_{\lambda,\alpha}^{-1} - V_{\lambda,\alpha}^{-1}\right) f\right\| &= O_p\left(\alpha^{-1} L^{-1/2}\right) + O_p\left(\alpha^{-1} S^{-1/2}\right), && (4.36)\\ \left\|\left(V_{\lambda,\alpha}^{-1} - V_{\lambda}^{-1}\right) f\right\| &= O\left(\alpha^{\min\left(1, \frac{\epsilon-1}{2}\right)}\right). && (4.37) \end{aligned}$$

The next step is to show that the estimated optimal weighting function converges to the theoretical one. We have the following proposition:

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

$$\widehat{\pi}_{\lambda,\alpha}^{*} - \pi_{\lambda}^{*} = O\left(\alpha^{\min\left(1, \frac{\epsilon-1}{2}\right)}\right) + O_p\left(\alpha^{-1} L^{-1/2}\right) + O_p\left(\alpha^{-1} S^{-1/2}\right)$$

as L and S go to infinity and α, $\alpha^{-1} L^{-1/2}$ and $\alpha^{-1} S^{-1/2}$ goes to zero. Moreover, the asymptotic variances of $\widehat{\theta}_{\widehat{\pi}_{\lambda,\alpha}^{*}}$ and $\widehat{\theta}_{\pi_{\lambda}^{*}}$ are the same.

We can also define the simulated estimators:

$$\widehat{\theta}_{\widehat{\pi}_{\lambda,\alpha}^{*}}^{(l)} = \sum_{s=1}^{S} \widehat{\pi}_{\lambda,\alpha}^{*}(\tau_s)\, \widehat{\theta}(\tau_s, l),\ l = 1, ..., L. \tag{4.38}$$

By noting that $\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}}, l = 1, ..., L$ are IID copies of $\widehat{\theta}_{\widehat{\pi}^*_{\lambda,\alpha}}$, the covariance matrix of $\widehat{\theta}_{\widehat{\pi}^*_{\lambda,\alpha}}$ can be estimated by:

$$\widehat{Var}\left(\widehat{\theta}_{\widehat{\pi}^*_{\lambda,\alpha}}\right) = \frac{1}{L}\sum_{l=1}^{L}\left(\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}} - \overline{\widehat{\theta}}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}}\right)\left(\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}} - \overline{\widehat{\theta}}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}}\right)' \tag{4.39}$$

where $\overline{\widehat{\theta}}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}} = \frac{1}{L}\sum_{l=1}^{L}\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}}$.

Clearly, the higher order asymptotic of $\widehat{\theta}_{\widehat{\pi}^*_{\lambda,\alpha}}$ will depend on the regularization parameter α. The optimal α can be estimated by minimizing the following approximate mean square error:

$$\widehat{\alpha}^* = \underset{\alpha\in(0,1)}{\arg\min}\frac{1}{L}\sum_{k=1}^{L}\left(\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}} - \widehat{\theta}_S\right)'\left(\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}} - \widehat{\theta}_S\right), \tag{4.40}$$

where $\widehat{\theta}_S = \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}^{(1)}(\tau_s)$ is the proxy used for θ_0.

Instead of using Monte Carlo simulations to obtain the samples, one can also resort to bootstrap. To this end, we let $\left\{\widehat{h}_{\tau_s,t}(u,\theta), u\in\mathbb{R}\right\}_{t=1}^{T}$ be the set of moments functions based on the observations $\{y_{\tau_s,t}\}_{t=1}^{T}$, and $\left(t_1^{(l)}, ..., t_B^{(l)}\right)$ be B independent uniform draws with replacement from the discrete set $\{1, 2, ..., T\}$, for $l = 1, ..., L$. Define the sets of moment functions:

$$\left\{\widehat{h}_{\tau_s,t_b^{(l)}}(u,\theta)\right\}_{b=1}^{B}, l = 1, ..., L \text{ and } s = 1, ..., S. \tag{4.41}$$

The remaining steps of the estimation are the same as in the Monte Carlo simulation case once we define $\widehat{\theta}(\tau_s, l)$ as the first step CGMM estimator based on the set of moment functions $\left\{\widehat{h}_{\tau_s,t_b^{(l)}}(u,\theta)\right\}_{b=1}^{B}$.

Two illustrations of the use of the FCGMM procedure are presented in the sequel.

4.6 A Simulation Study

In a universe where agents are rational and risk averse, the expected return should be positively correlated with the expected 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 the expected and unexpected volatility. 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.

4.6.1 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 $V_t - E\left[V_t|V_{t-1}\right]$, that is:

$$r_t = \mu_0 + \mu_1 V_{t-1} + \delta\left(V_t - E\left[V_t|V_{t-1}\right]\right) + \sigma_\varepsilon \varepsilon_t, \tag{4.42}$$

where $\varepsilon_t \overset{IID}{\sim} N(0,1)$ is uncorrelated with V_{t-1} and $V_t - E\left[V_t|V_{t-1}\right]$. The risk premium is modeled as a positive relationship between the return and the expected risk ($\mu_1 \geq 0$) while the leverage effect is modeled as a negative relation between the return and the unexpected risk ($\delta \leq 0$). The latent variable V_t is assumed to follow an Autoregressive Gamma process of order one:

$$f\left(V_t|V_{t-1}\right) = \sum_{j=0}^{\infty} \frac{V_t^{j+q-1} c^{j+q}}{\Gamma\left(j+q\right)} \exp\left(-cV_t\right) p_j\left(V_{t-1}\right), \tag{4.43}$$

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

$$p_j\left(V_{t-1}\right) = \frac{\left(ce^{-\kappa}V_{t-1}\right)^j}{j!} \exp\left(-ce^{-\kappa}V_{t-1}\right). \tag{4.44}$$

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

$$f\left(V_t\right) = \frac{V_t^{q-1}}{\Gamma\left(q\right)} \left(\frac{2\kappa}{\sigma^2}\right)^q \exp\left(\frac{-2\kappa}{\sigma^2} V_t\right). \tag{4.45}$$

Its conditional and unconditional CF are:

$$E\left[e^{i\tau V_t}|V_{t-1}\right] = \left(1-\frac{i\tau}{c}\right)^{-q}\exp\left(\frac{i\tau e^{-\kappa}V_{t-1}}{1-\frac{i\tau}{c}}\right) \tag{4.46}$$

$$E\left[e^{i\tau V_t}\right] = \left(1-\frac{i\sigma^2\tau}{2\kappa}\right)^{-q}.$$

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[4]:

$$E\left[V_t|V_{t-1}\right] = \beta\left(1-e^{-\kappa}\right)+e^{-\kappa}V_{t-1}, \tag{4.47}$$

$$Var\left[V_t|V_{t-1}\right] = \frac{\beta\sigma^2}{2\kappa}\left(1-e^{-\kappa}\right)^2+\frac{\sigma^2}{\kappa}e^{-\kappa}\left(1-e^{-\kappa}\right)V_{t-1}. \tag{4.48}$$

This implies that the conditional mean and variance of r_t are linear in the lagged realization of the risk factor:

$$E\left[r_t|V_{t-1}\right] = \mu_0+\mu_1 V_{t-1}, \tag{4.49}$$

$$\begin{aligned} Var\left[r_t|V_{t-1}\right] &= \delta^2 Var\left[V_t|V_{t-1}\right]+\sigma_\varepsilon^2 \\ &= \frac{\delta^2\beta\sigma^2}{2\kappa}\left(1-e^{-\kappa}\right)^2+\sigma_\varepsilon^2+\frac{\delta^2\sigma^2}{\kappa}e^{-\kappa}\left(1-e^{-\kappa}\right)V_{t-1}. \end{aligned} \tag{4.50}$$

In particular, $E\left[r_t|V_{t-1}\right]$ is a linear function of $Var\left[r_t|V_{t-1}\right]$.

We derive similarly the third and fourth conditional moments of V_t:

$$E\left[\left(V_t-E\left[V_t|V_{t-1}\right]\right)^3|V_{t-1}\right] = \frac{\beta\sigma^4}{2\kappa^2}\left(1-e^{-\kappa}\right)^3+\frac{3\sigma^4e^{-\kappa}}{2\kappa^2}\left(1-e^{-\kappa}\right)^2V_{t-1}, \tag{4.51}$$

$$\begin{aligned} E\left[\left(V_t-E\left[V_t|V_{t-1}\right]\right)^4|V_{t-1}\right] &= 3Var\left[V_t|V_{t-1}\right]^2 \\ &+\frac{3\sigma^6}{4\kappa^3}\left(\beta\left(1-e^{-\kappa}\right)^4+4e^{-\kappa}\left(1-e^{-\kappa}\right)^3V_{t-1}\right). \end{aligned} \tag{4.52}$$

Equation (4.52) shows that V_t has a positive excess kurtosis. The third and fourth conditional

[4]The noncentered conditional moments of are given by $E\left[(V_t)^n|V_{t-1}\right] = \frac{1}{i^n}\frac{\partial^n E\left[e^{i\tau V_t}|V_{t-1}\right]}{\partial\tau^n}|_{\tau=0}$. These derivatives may be computed using a mathematical software.

moments of r_t are linked to those of V_t by:

$$E\left[(r_t - E[r_t|V_{t-1}])^3 |V_{t-1}\right] = \delta^3 E\left[(V_t - E[V_t|V_{t-1}])^3 |V_{t-1}\right], \tag{4.53}$$

$$E\left[(r_t - E[r_t|V_{t-1}])^4 |V_{t-1}\right] = \delta^4 E\left[(V_t - E[V_t|V_{t-1}])^4 |V_{t-1}\right] + 3\sigma_\varepsilon^4. \tag{4.54}$$

Hence r_t has a time varying negative skewness whenever $\delta < 0$. To see the implications of the model in term of kurtosis, we note that the last equality implies:

$$\begin{aligned} E\left[(r_t - E[r_t|V_{t-1}])^4 |V_{t-1}\right] &= 3Var[r_t|V_{t-1}]^2 + \frac{3\delta^4\sigma^6 e^{-\kappa}}{2\kappa^3}\left(1-e^{-\kappa}\right)^3 V_{t-1}, \\ &+ \frac{3\delta^2\sigma^2}{\kappa}\left(\frac{\delta^2\sigma^4}{4\kappa^2}\left(1-e^{-\kappa}\right)^2 - \sigma_\varepsilon^2\right)\beta\left(1-e^{-\kappa}\right)^2 \\ &+ \frac{6\delta^2\sigma^2}{\kappa}\left(\frac{\delta^2\sigma^4}{4\kappa^2}\left(1-e^{-\kappa}\right)^2 - \sigma_\varepsilon^2\right)e^{-\kappa}\left(1-e^{-\kappa}\right)V_{t-1}. \end{aligned} \tag{4.55}$$

It is seen that this model can reproduce fat tailed distributions. In particular, the distribution of r_t given V_t is fat tailed when $\frac{\delta^2\sigma^4}{4\kappa^2}\left(1-e^{-\kappa}\right)^2 - \sigma_\varepsilon^2$ is positive[5].

In what follows, we present an estimation strategy for the ARFG model.

4.6.2 Estimating 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* $(r_t, ..., r_{t+1-d})$ *is given by:*

$$\begin{aligned} &E\left[\exp\left(\sum_{k=1}^{d} i\tau_k r_{t+1-k}\right)\right] \\ =\ &\exp\left(\left[\mu_0 - \delta\beta\left(1-e^{-\kappa}\right)\right]\sum_{k=1}^{d} i\tau_k - \frac{\sigma_\varepsilon^2}{2}\sum_{k=1}^{d}\tau_k^2\right) \\ &\times\left(1 - \frac{iu_{L+1}\sigma^2}{2\kappa}\right)^{-q}\prod_{k=1}^{d}\left(1-\frac{iu_k}{c}\right)^{-q}. \end{aligned} \tag{4.56}$$

[5]This is only a sufficient condition, not necessary.

where:

$$\begin{aligned}
u_1 &= \tau_1\delta; \\
u_k &= \frac{u_{k-1}e^{-\kappa}}{1-\frac{iu_{k-1}}{c}} + \tau_{k-1}\left(\mu_1 - \delta e^{-\kappa}\right) + \tau_k\delta,\ k = 2, ..., d; \\
u_{d+1} &= \frac{u_d e^{-\kappa}}{1-\frac{iu_d}{c}} + \tau_d\left(\mu_1 - \delta e^{-\kappa}\right).
\end{aligned}$$

The details of the derivation of this CF are left in Appendix. The moment function we will use in the FCGMM procedure is:

$$h_{\tau,t}(u,\theta) = \exp\left(iu\sum_{k=1}^{d}\tau_k r_{t+1-k}\right) - E\left[\exp\left(iu\sum_{k=1}^{d}\tau_k r_{t+1-k}\right)\right], \tag{4.57}$$

where $\tau = (\tau_1, ..., \tau_d) \in \mathbf{S}^d$, $u \in \mathbb{R}$, $\theta = \left(\mu_0, \mu_1, \delta, \beta, \kappa, \sigma, \sigma_\varepsilon^2\right)'$. Note that Equation (4.57) is a moment function of type (4.10).

In the next section, we evaluate by Monte Carlo the performance of the FCGMM in estimating this model.

4.6.3 Monte Carlo Experiments

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 (4.43) as suggested by Devroye (1986). At time $t = 0$, one draws an initial value V_0 from the stationary Gamma distribution (4.45). At $t = 1$, one draws an integer j_0 from the Poisson distribution with parameter $ce^{-\kappa}V_0$. The current realization V_1 of the state variable is then drawn from the Gamma distribution with density $f_{j_0}(v)$, where:

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

At $t = 2$, one draws again an integer j_1 from the Poisson distribution with parameter $ce^{-\kappa}V_1$. 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 $f_{j_{t-1}}(v)$, where j_{t-1} is a draw from the Poisson distribution with parameter $ce^{-\kappa}V_{t-1}$. 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 (4.42).

In this Monte Carlo experiment, we set $d = 10$ so that the estimation of θ_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 $\widehat{\pi}^*_{\lambda,\alpha}$ and regularization parameter α^* are estimated by generating $S = 100$ draws of τ for each Monte Carlo replication. We arbitrarily fixed $\lambda = (1, ..., 1)'$ and compute $\widehat{\pi}^*_{\lambda,\alpha}$ for α on the grid:

$$\begin{aligned} \alpha \in & \left[7{\times}10^{-4},\ 5{\times}10^{-4},\ 3{\times}10^{-4},\ 1{\times}10^{-4},\ 7{\times}10^{-5},\right. \\ & \left. 5{\times}10^{-5},\ 1{\times}10^{-5},\ 5{\times}10^{-6},\ 1{\times}10^{-6},\ 1{\times}10^{-7}\right] \end{aligned}$$

For each $\widehat{\pi}^*_{\lambda,\alpha}$, we compute the mean square error of the FCGMM estimator using the formula:

$$MSE\left(\alpha\right) = \frac{1}{L}\sum_{k=1}^{L}\left(\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}} - \theta_0\right)'\left(\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}} - \theta_0\right) \tag{4.58}$$

where $\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}}$ is defined in (4.38) and:

$$\begin{aligned} \theta_0 &= \left(\mu_0, \mu_1, \delta, \beta, \kappa, \sigma, \sigma^2_\varepsilon\right)' \\ & \left(0, 10^{-2}, -5\times10^{-2}, 10^{-4}, 2\times10^{-2}, 5\times10^{-2}, 2\times10^{-4}\right)' \end{aligned}$$

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

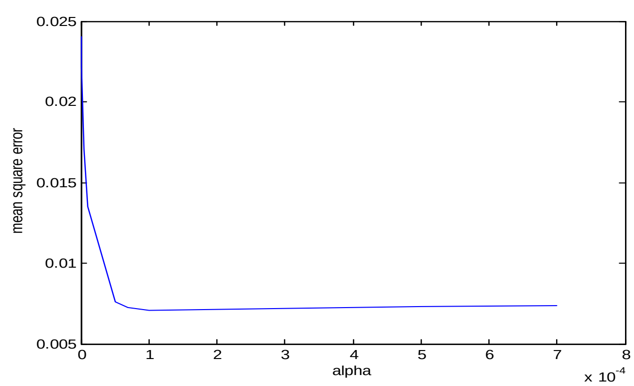


Figure 4.1: The mean square error of the FCGMM estimator as a function of the regularization parameter α.

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 $\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}}$. IC1 and IC2 are respectively the lower and upper bound of the 90% confidence interval.

Parameters	Mean	Median	Std. Dev.	IC1	IC2
μ_0	-4.0×10^{-5}	-3.5×10^{-5}	3.9×10^{-4}	-6.9×10^{-4}	5.8×10^{-4}
μ_1	1.0×10^{-2}	1.0×10^{-2}	1.6×10^{-4}	9.8×10^{-3}	1.0×10^{-2}
δ	-4.8×10^{-2}	-4.8×10^{-2}	2.6×10^{-3}	-5.4×10^{-2}	-4.5×10^{-2}
β	1.0×10^{-4}	1.0×10^{-4}	5.9×10^{-6}	9.0×10^{-5}	1.0×10^{-4}
κ	1.9×10^{-2}	1.9×10^{-2}	8.9×10^{-4}	1.7×10^{-2}	2.0×10^{-2}
σ	5.0×10^{-2}	4.8×10^{-2}	7.7×10^{-3}	4.1×10^{-2}	6.3×10^{-2}
σ^2_ε	1.9×10^{-4}	2.0×10^{-4}	4.5×10^{-5}	9.3×10^{-5}	2.3×10^{-4}

Table 4.1: Monte Carlo simulations results for the ARFG Model estimated by FCGMM. We draw $L = 100$ independent samples, $l = 1, .., L = 100$. For each sample l, the FCGMM $\widehat{\theta}^{(l)}_{\widehat{\pi}^*_{\lambda,\alpha}}$ is computed using $S = 100$ draws of τ. The true vector of parameters is

$$\theta_0 = \left(0, 10^{-2}, -5 \times 10^{-2}, 10^{-4}, 2 \times 10^{-2}, 5 \times 10^{-2}, 2 \times 10^{-4}\right)'$$

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 draws S of τ are quite moderate, the results of this experiment suggest that the suggested FCGMM is a reliable inference method.

In what follows, we present an empirical applications.

4.7 An Empirical Application

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

4.7.1 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 $E\left[V_t \middle| \{V_{t-k}\}_{k=1}^{p}\right]$ and the innovation $V_t - E\left[V_t \middle| \{V_{t-k}\}_{k=1}^{p}\right]$:

$$r_t = \mu_0 + \mu_1 \sqrt{E\left[V_t \middle| \{V_{t-k}\}_{k=1}^{p}\right]} + \delta\left(V_t - E\left[V_t \middle| \{V_{t-k}\}_{k=1}^{p}\right]\right) + \sqrt{V_t}\varepsilon_t \qquad (4.59)$$

where $\varepsilon_t \overset{IID}{\sim} N(0,1)$ 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 μ_1 captures the premium for bearing the expected risk while δ 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:

$$\begin{aligned} f\left(V_t \middle| \{V_{t-k}\}_{k=1}^{\infty}\right) &= f\left(V_t \middle| \{V_{t-k}\}_{k=1}^{p}\right) \\ &= \sum_{j=0}^{\infty} \frac{V_t^{j+q-1} c^{j+q}}{\Gamma(j+q)} \exp(-cV_t)\, p_j\left(\{V_{t-k}\}_{k=1}^{p}\right) \end{aligned}$$

where $p_j\left(\{V_{t-k}\}_{k=1}^{p}\right)$ are Poisson weights given by:

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

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

$$\begin{aligned} (\kappa, \beta, \sigma) &> 0, \quad \{\rho_1\}_{k=0}^{p} \geq 0, \quad \sum_{k=0}^{p} \rho_k = 1 \\ c &= \frac{2\kappa}{\sigma^2 \rho_0} \text{ and } q = \frac{2\beta\kappa}{\sigma^2}. \end{aligned}$$

The specified dynamic for V_t extends the model of Gourieroux and Jasiak (2005) which is an autoregressive Gamma or order one. Likewise, the model studied in Kotchoni (2009) is the Autoregressive Variance Gamma model of order one. The conditional CF of V_t is an exponential affine form given by:

$$E\left[e^{i\tau V_t} |\left\{V_{t-k}\right\}_{k=1}^{p}\right] = \left(1-\frac{i\tau}{c}\right)^{-q}\exp\left(\frac{i\tau}{1-\frac{i\tau}{c}}\sum_{k=1}^{p}\rho_k V_{t-k}\right) \tag{4.60}$$

This CF shows that the Autoregressive Gamma model of order p is identical (up to a reparametrization) to a univariate Wishart autoregressive process of order p discussed in Gourieroux, Jasiak and Sufana (2005)[6]. The following moments can be computed by using the two first derivatives of the above conditional CF evaluated at zero[7]:

$$E\left[V_t |\left\{V_{t-k}\right\}_{k=1}^{p}\right] = \beta\left(1-\sum_{k=1}^{p}\rho_k\right)+\sum_{k=1}^{p}\rho_k V_{t-k} \tag{4.61}$$

$$Var\left[V_t |\left\{V_{t-k}\right\}_{k=1}^{p}\right] = \frac{1}{c}\left[\beta\left(1-\sum_{k=1}^{p}\rho_k\right)+2\sum_{k=1}^{p}\rho_k V_{t-k}\right] \tag{4.62}$$

It is seen that the conditional mean and variance of V_t are linear in its lagged realizations. Moreover, the ARVG model has the potential to generate asymmetry and fat tails. In fact we have:

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

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:

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

Equation (4.61) provides a good forecasting formula for the volatility. If the lagged variables $(V_{t-1},...V_{t-p})$ are such that $E\left[V_{t-k}\right] = \beta$ for all $k = 1,...,p$, then we also have $E\left[V_t\right] = \beta$. This indicates that the stationary autoregressive Gamma process of order p has to satisfy $E\left[V_t\right] = \beta$ for all t. In the next subsection, we present an estimation strategy for the ARVG(p).

---

[6] See Gourieroux, Jasiak and Sufana (2005), Section 2.3, Definition 2.

[7] The noncentrered conditional moments of V_t are given by: $E\left[(V_t)^n |\left\{V_{t-k}\right\}_{k=1}^{p}\right] = \frac{1}{i^n}\frac{\partial^n E\left[e^{i\tau V_t}|\left\{V_{t-k}\right\}_{k=1}^{p}\right]}{\partial\tau^n}|_{\tau=0}$

4.7.2 Estimation of the ARVG(p) Using High Frequency Data

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

$$Var\left[r_t|\left\{V_{t-k}\right\}_{k=1}^{p}\right] = V_t \tag{4.63}$$

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:

$$Var\left[r_t|\left\{\sigma_s\right\}_{s=-\infty}^{\infty}\right] = \int_{t-1}^{t} \sigma_s^2 ds. \tag{4.64}$$

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 (2009).

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:

$$h_t\left(\tau, \theta_1\right) = \left(\exp\left(i\underline{\tau}_1 V_t\right) - E\left[\exp\left(i\underline{\tau}_1 V_t\right) | \left\{V_{t-k}\right\}_{k=1}^{p}\right]\right) \exp\left(\sum_{k=1}^{p} i\underline{\tau}_{k+1} V_{t-k}\right),$$

where $E\left[\exp\left(i\underline{\tau}_1 V_t\right) | \left\{V_{t-k}\right\}_{k=1}^{p}\right]$ is given by (4.60), $\tau = \left(\underline{\tau}_1, ..., \underline{\tau}_{p+1}\right)$ and $\theta_1 = \left(\rho_0, ..., \rho_p, \kappa, \beta, \sigma^2\right)$. 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 (2009).

Having computed $\widehat{\theta}_1$, the expected variance $\widehat{V}_t$ is estimated by:

$$\widehat{V}_t = \widehat{\beta}\left(1 - \sum_{k=1}^{p} \widehat{\rho}_k\right) + \sum_{k=1}^{p} \widehat{\rho}_k V_{t-k}$$

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 ε_t, where the following proxy is used for ε_t:

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

We implement the ARVG(p) with real data in the next subsection.

4.7.3 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 one minute from January 1^{st}, 2002 to December 31^{th}, 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. As in Kotchoni (2009), we construct the proxy of V_t using the shrinkage realized kernels of Carrasco and Kotchoni (2009).

The implementation of the FCGMM 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 (4.41). We resample $L = 100$ times in the time domain and use $S = 50$ draws of τ. Finally, we set $p = 30$ (six weeks) in order to assess the level of persistence of the volatility process.

To select the regularization parameter α, we choose to minimize the tracking error:

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

The following graph suggests that the optimal regularization parameter is around $\alpha^* = 10^{-4}$ for these data.

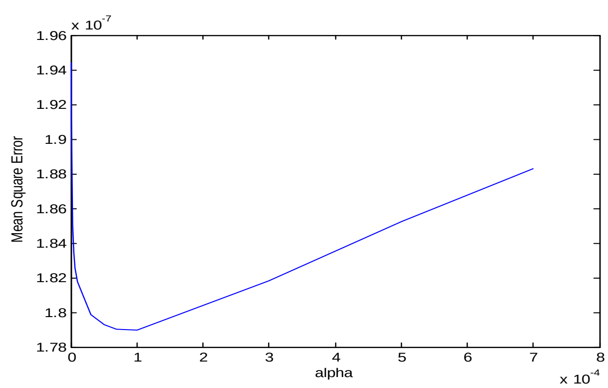


Figure 4.2: Selecting the regularization parameter

We compute the optimal weighting function using that value of α^*. The following Table shows the summary of the results for the parameters $(\kappa, \beta, \sigma^2)$ and (μ_0, μ_1, δ). One important difference between the current results and those of the case $p = 1$ presented in Kotchoni

(2009) is that the estimates of κ are considerably lower here. The variances of the estimators are relatively high due the the fact that they are estimated using only $L = 100$ samples. Accordingly, the confidence intervals are also large.

	$\widehat{\theta}_1^2$			$\widehat{\theta}_2$		
	$\widehat{\kappa}$	$\widehat{\beta}$	$\widehat{\sigma}^2$	$\widehat{\mu}_0$	$\widehat{\mu}_1$	$\widehat{\delta}$
Mean	0.0528	0.0017	0.0082	−0.0007	0.0262	−0.1886
Median	0.1358	0.0003	0.0077	−0.0005	0.0000	0.0000
Std. Dev.	0.3656	0.0023	0.0045	0.0014	0.0689	0.9688
IC1(95)	−0.8730	−0.0000	0.0007	−0.0043	0.0000	−0.7724
IC2(95)	0.4516	0.0060	0.0145	0.0009	0.1618	−0.0000

Table 4.2: Summary of the Estimation Results

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

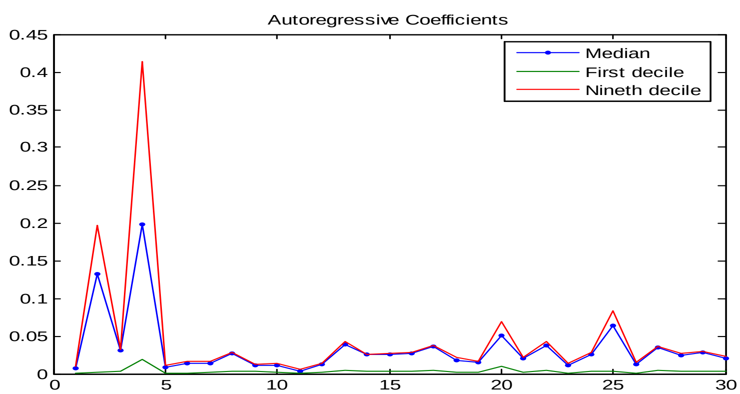


Figure 4.3: Estimated Autoregressive Roots for the Volatility

It is seen that the volatility strongly responds to its own lags lying within one week. This finding is consistent with the volatility clustering. There also seems to be some responses of smaller magnitude to lags lying between 20 and 25 days.

Finally, the following graph shows the volatility and its estimated expectation conditional on past realizations.

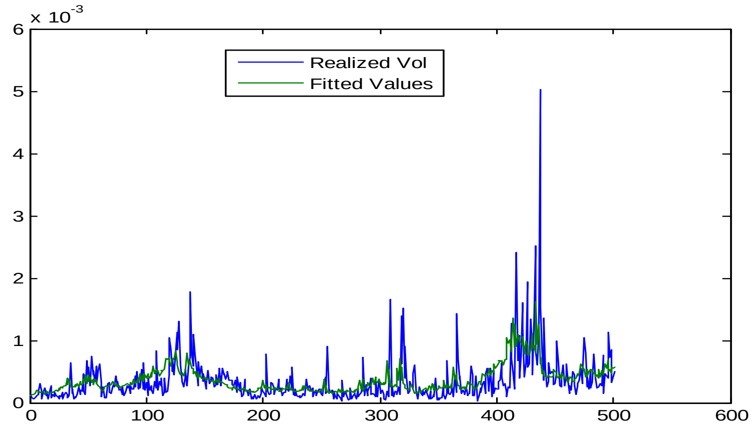


Figure 4.4: Fitted Values for the Volatility

It is seen that the conditional expectation of the volatility is very smooth compared to actual series. In fact, the Autoregressive Gamma model ignores the erratic fluctuations and jumps of the volatility and focus on 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. More precisely, one can test the presence of jumps in the volatility by using the residuals of the Autoregressive Gamma model.

4.8 Conclusion

The Generalized method of moments with a continuum of moment conditions (CGMM) introduced by Carrasco and Florens (2000) aims to deliver estimators that are as efficient as the maximum likelihood. The objective function of the CGMM is a quadratic form associated with a scalar products on Hilbert spaces. When the characteristic function is used to build the moment conditions, that objective function involves as many integrals as there are dimensions in the data. Unfortunately, the complexity of the numerical integration grows as an exponential function of the dimensionality of the vector of observation. This makes the use of the CGMM unattractive in multivariate Markov model and non-Markov models. To circumvent this "curse of dimensionality", we propose to work with univariate samples obtained by taking linear combinations of the initial vector of observations. Each sample obtained in this way is called a frequency domain sample and can be used to estimate the parameter of interest by CGMM. Finally, all the possible estimators obtained in this fashion are aggregated to obtain what we called a frequency domain resampling estimator.

We derived the optimal aggregation rule for the CGMM frequency domain resampling estimator and propose two illustrations. The first one 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. The second illustration is an empirical application based on the Autoregressive Variance Gamma model of order p. Although the estimated variances of the estimators are relatively high due to the small number of resampling replications, this application tends to confirm that a positive risk premium is required by investors for bearing the expected risk while the returns are negatively correlated with the shocks on its variance. In future investigations, we will try to improve our simulations and empirical results by making use of the refinements of the bootstrap technique developed for example by Davidson and Mckinnon (2000).

Bibliography

[1] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and N. Shephard (2008a) "Realized Kernels in Practice: Trades and Quotes", *Econometrics Journal*, forthcoming.

[2] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and N. Shephard (2008b) "Multivariate Realized Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading " *Working Paper*. Available on SSRN: *http://papers.ssrn.com/sol3/papers.cfm?abstract_ id=1154144.*

[3] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and N. Shephard (2008c) "Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise", *Econometrica,* vol. 76, 1481-1536.

[4] Carrasco, M., Chernov, M., Florens, J. P. and E. Ghysels (2007) "Efficient estimation of general dynamic models with a continuum of moment conditions" *Journal of Econometrics,* 140 (2007) 529–573.

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

[6] Carrasco, M., J. P. Florens, and E. Renault (2007) "Linear Inverse Problems in Structural Econometrics: Estimation based on spectral decomposition and regularization", in the *Handbook of Econometrics*, Vol. 6, edited by J.J. Heckman and E.E. Leamer.

[7] Carrasco, M and R. Kotchoni (2008) "Efficient Estimation Using the Characteristic Function", *Working paper.* Université de Montréal.

[8] Carrasco, M. and R. Kotchoni (2009) "Assessing the Nature of Pricing Inefficiencies via Realized Measures", *Working paper.* Université de Montréal.

[9] Chen, Jacho-Chavez and Linton (2009) "An Alternative Way of Computing Efficient Instrumental Variable Estimators" *LSE STICERD Research Paper* No. EM536.

[10] Da Fonseca, J., Grasselli, M. and F. Ielpo (2008) "Estimating the Wishart Affine Stochastic Correlation Model using the Empirical Characteristic Function". *Working Paper*, Ecole Superieure d'Ingenieurs Leonard de Vinci.

[11] Davidson and Mackinnon "Bootstrap Tests: How Many Bootstraps?" *Econometric Reviews*, 19 , 55-68, 2000.

[12] Devroye, L. (1986) *"Non-Uniform Random Variate Generation"*, Spinger-Verlag.

[13] Gourieroux, C. and J. Jasiak (2005) "Autoregressive Gamma Processes". *Journal of Forecasting,* vol. 25, 129–152. Published online in Wiley InterScience (*www.interscience.wiley.com*) DOI: 10.1002/for.978.

[14] Gourieroux, C., Jasiak, J. and Sufana, R. (2005) "The Wishart Autoregressive Process". *Journal of Econometrics,* vol. 150 (2009) 167-181.

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

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

[17] Kotchoni, R. (2009) "Applications of the Characteristic Function Based Continuum GMM in Finance", *Working paper*, Université de Montréal.

[18] Yu, J. (2004) "Empirical Characteristic Function Estimation and Its Applications", *Econometric Reviews*, Vol. 23, No 2, pp. 93-123.

Proof of Proposition 1: Under Assumptions 1 to 4, $\widehat{\theta}^{(1)}(\tau)$ is a well-defined consistent estimator of θ_0. Furthermore, Assumption 6(i) ensures $\widehat{\theta}^{(1)}(\tau)$ 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 τ and θ. Hence by the implicit function theorem, $\widehat{\theta}^{(1)}(\tau)$ is continuously differentiable with respect to τ and we have:

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

where the inversibility of $\frac{\partial^2 Q_{\tau,T}}{\partial\theta\partial\theta'}$ follows from Assumption 6(i)■

Proof of Proposition 2: According to Assumptions 1 to 4, each CGMM estimator $\widehat{\theta}^{(1)}(\tau)$ is consistent for θ_0 (See for example Carrasco and Kotchoni, 2009). Hence for any choice of measure π satisfying Assumption 7, we have:

$$\begin{aligned}
Var\left(\lambda'\widehat{\theta}_\pi\right) &= \int\int \lambda' Cov\left(\widehat{\theta}^{(1)}(\tau_1), \widehat{\theta}^{(1)}(\tau_2)\right)\lambda\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2 \\
&\leq \max_\tau\left[\lambda' Var\left(\widehat{\theta}^{(1)}(\tau)\right)\lambda\right]\int\int \pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2 \\
&= \max_\tau\left[\lambda' Var\left(\widehat{\theta}^{(1)}(\tau)\right)\lambda\right].
\end{aligned}$$

The result follows from $\max_\tau\left[\lambda' Var\left(\widehat{\theta}^{(1)}(\tau)\right)\lambda\right] = O_p\left(T^{-1}\right)$ for $\tau \in \mathbf{S}^d\backslash\aleph$■

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 $g_\lambda(\tau_1,\tau_2) = \lambda' Cov\left(\widehat{\theta}^{(1)}(\tau_1), \widehat{\theta}^{(1)}(\tau_2)\right)\lambda$. 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 μ_λ 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 τ_1 in $\mathbf{S}^d\backslash\aleph$.

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

$$V_\lambda \pi(\tau_1) = \mu_\lambda I(\tau_1) \tag{4.65}$$

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

$$f(\tau_1) = \underbrace{\sum_{j=1}^{\infty}\left(\int \phi_{\lambda,j}(\tau)\,f(\tau)\,d\tau\right)\phi_{\lambda,j}(\tau_1)}_{f_0(\tau_1)} + \widetilde{f}(\tau_1),$$

where $\widetilde{f}$ is in the null set of V_λ so that $V_\lambda\widetilde{f}(\tau_1) = 0$. Hence if $f_0(\tau_1)$ solves (4.65) for $\pi(\tau_1)$, then $f(\tau_1) = f_0(\tau_1) + \widetilde{f}(\tau_1)$ also solves (4.65) for $\pi(\tau_1)$. The solution of (4.65) with minimal norm is the one in which $\widetilde{f}(\tau_1) = 0$. 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 μ_λ 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:

$$\begin{aligned} Var\left(\lambda' \widehat{\theta}_{\pi_\lambda^*}\right) &= \int \left[\int g_\lambda(\tau_1, \tau_2)\, \pi_\lambda^*(\tau_1)\, d\tau_1\right] \pi_\lambda^*(\tau_2)\, d\tau_2 \\ &= \mu_\lambda \int \left[V_\lambda V_\lambda^{-1} I(\tau_1)\, d\tau_1\right] \pi_\lambda^*(\tau_2)\, d\tau_2 \\ &= \mu_\lambda \int \pi_\lambda^*(\tau_2)\, d\tau_2 = \mu_\lambda \end{aligned}$$

■

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

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

Around a particular τ in $\mathbf{S}^d \backslash \aleph$, we have:

$$\widehat{\theta}(\tau + \tau_0) = \widehat{\theta}(\tau) + \frac{\partial \widehat{\theta}^{(1)}(\tau)}{\partial \tau}\tau_0.$$

By assumption 6(ii), $\frac{\partial^2 Q_{\tau,T}}{\partial\theta\partial\tau}$ is of full rank so that $\frac{\partial \widehat{\theta}^{(1)}(\tau)}{\partial \tau}$ is also of full rank. This implies that for $d \geq \max\{q, 2\}$, q linearly independent vectors of type $\widehat{\theta}_T(\tau + \tau_0)$ can be constructed by varying τ_0. As a consequence, the manifold $\widehat{D}_T(\theta_0)$ defined by:

$$\widehat{D}_T(\theta_0) = \left\{\theta \in \mathbb{R}^q \text{ s.t } \theta = \int \pi(\tau_1)\, \widehat{\theta}^{(1)}(\tau_1)\, d\tau_1 \text{ and } \int \pi(\tau_1)\, d\tau_1 = 1\right\}$$

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

Proof of Proposition 5: We have defined:

$$\widehat{V}_\lambda = \frac{1}{L}\left(\widehat{\Theta}_\lambda - \overline{\widehat{\Theta}}_\lambda\right)'\left(\widehat{\Theta}_\lambda - \overline{\widehat{\Theta}}_\lambda\right)$$

The (i,j) element of $\widehat{V}_\lambda$ is given by

$$\widehat{g}_\lambda(\tau_i,\tau_j)=\frac{1}{L}\sum_{l=1}^{L}\lambda'\left(\widehat{\theta}(\tau_i,l)-\overline{\widehat{\theta}}(\tau_i,l)\right)\left(\widehat{\theta}(\tau_j,l)-\overline{\widehat{\theta}}(\tau_j,l)\right)'\lambda.$$

Let $\underline{f}=(f(\tau_1),...,f(\tau_S))'$. Then as L goes to infinity, we have:

$$\left(\widehat{V}_\lambda\underline{f}\right)_i-\sum_{j=1}^{S}\lambda' Cov\left(\widehat{\theta}(\tau_i,l),\widehat{\theta}(\tau_j,l)\right)\lambda f(\tau_j)=O_p\left(L^{-1/2}\right)$$

On the other hand, we assumed that τ_j is drawn using the uniform distribution on $\mathbf{S}^d$. Hence as L and S go to infinity, we have:

$$\left(\widehat{V}_\lambda\underline{f}\right)_i-\int\lambda' Cov\left(\widehat{\theta}(\tau_i,l),\widehat{\theta}(\tau,l)\right)\lambda f(\tau)=O_p\left(L^{-1/2}\right)+O_p\left(S^{-1/2}\right)$$

This shows that

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

■

Proof of Proposition 6: We first consider (4.35):

$$\begin{aligned}\left\|(\widehat{V}_\lambda^2+\alpha I)^{-1}\widehat{V}_\lambda-(V_\lambda^2+\alpha I)^{-1}V_\lambda\right\| &\leq\\ \left\|(\widehat{V}_\lambda^2+\alpha I)^{-1}(\widehat{V}_\lambda-V_\lambda)\right\|+\left\|(\widehat{V}_\lambda^2+\alpha I)^{-1}V_\lambda-(V_\lambda^2+\alpha I)^{-1}V_\lambda\right\| &\leq\\ \underbrace{\left\|(\widehat{V}_\lambda^2+\alpha I)^{-1}\right\|}_{\leq\alpha^{-1}}\underbrace{\left\|\widehat{V}_\lambda-V_\lambda\right\|}_{=O_p\left(L^{-1/2}\right)+O_p\left(S^{-1/2}\right)}+\left\|\left[(\widehat{V}_\lambda^2+\alpha I)^{-1}-(V_\lambda^2+\alpha I)^{-1}\right]V_\lambda\right\| &\leq\end{aligned}$$

The result follows from:

$$\begin{aligned}&\left\|\left[(\widehat{V}_\lambda^2+\alpha I)^{-1}-(V_\lambda^2+\alpha I)^{-1}\right]V_\lambda\right\|\\ =\;&\left\|(\widehat{V}_\lambda^2+\alpha I)^{-1}\left(V_\lambda^2-\widehat{V}_\lambda^2\right)(V_\lambda^2+\alpha I)^{-1}V_\lambda\right\|\\ \leq\;&\underbrace{\left\|(\widehat{V}_\lambda^2+\alpha I)^{-1}\right\|}_{\leq\alpha^{-1}}\underbrace{\left\|\left(V_\lambda^2-\widehat{V}_\lambda^2\right)\right\|}_{=O_p\left(L^{-1/2}\right)+O_p\left(S^{-1/2}\right)}\underbrace{\left\|(V_\lambda^2+\alpha I)^{-1/2}\right\|}_{\leq\alpha^{-1/2}}\underbrace{\left\|(V_\lambda^2+\alpha I)^{-1/2}V_\lambda\right\|}_{\to 1}\end{aligned}$$

The difference between (4.35) and (4.36) is that in (4.36) uses the fact that $\left\|V_\lambda^{-\epsilon}f\right\|<\infty$

with $\epsilon \geq 1$. We can rewrite (4.36) as

$$\begin{aligned}\left\|\left(\widehat{V}_{\lambda,\alpha}^{-1} - V_{\lambda,\alpha}^{-1}\right) f\right\| &= \left\|\left(\widehat{V}_{\lambda,\alpha}^{-1} - V_{\lambda,\alpha}^{-1}\right) V_\lambda V_\lambda^{-1} f\right\| \\ &\leq \left\|\left(\widehat{V}_{\lambda,\alpha}^{-1} - V_{\lambda,\alpha}^{-1}\right) V_\lambda\right\| \left\|V_\lambda^{-1} f\right\|.\end{aligned}$$

We have

$$\begin{aligned}&\left(\widehat{V}_{\lambda,\alpha}^{-1} - V_{\lambda,\alpha}^{-1}\right) V_\lambda \\ =\; &(\widehat{V}_\lambda^2 + \alpha I)^{-1} \widehat{V}_\lambda V_\lambda - (V_\lambda^2 + \alpha I)^{-1} V_\lambda^2 \\ =\; &(\widehat{V}_\lambda^2 + \alpha I)^{-1} \left(\widehat{V}_\lambda - V_\lambda\right) V_\lambda \qquad (4.66)\\ &+ \left[(\widehat{V}_\lambda^2 + \alpha I)^{-1} - (V_\lambda^2 + \alpha I)^{-1}\right] V_\lambda^2. \qquad (4.67)\end{aligned}$$

The term (4.66) can be bounded in the following manner

$$\begin{aligned}\left\|(\widehat{V}_\lambda^2 + \alpha I)^{-1} \left(\widehat{V}_\lambda - V_\lambda\right) V_\lambda\right\| &\leq \left\|(\widehat{V}_\lambda^2 + \alpha I)^{-1}\right\| \left\|\widehat{V}_\lambda - V_\lambda\right\| \|V_\lambda\| \\ &= O_p\left(\alpha^{-1} L^{-1/2}\right) + O_p\left(\alpha^{-1} S^{-1/2}\right).\end{aligned}$$

For the term (4.67), 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

$$\begin{aligned}&\left\|\left[(\widehat{V}_\lambda^2 + \alpha I)^{-1} - (V_\lambda^2 + \alpha I)^{-1}\right] V_\lambda^2\right\| \\ =\; &\left\|(\widehat{V}_\lambda^2 + \alpha I)^{-1} \left(V_\lambda^2 - \widehat{V}_\lambda^2\right) (V_\lambda^2 + \alpha I)^{-1} V_\lambda^2\right\| \\ \leq\; &\underbrace{\left\|(\widehat{V}_\lambda^2 + \alpha I)^{-1}\right\|}_{\leq \alpha^{-1}} \underbrace{\left\|V_\lambda^2 - \widehat{V}_\lambda^2\right\|}_{= O_p\left(L^{-1/2}\right) + O_p\left(S^{-1/2}\right)} \underbrace{\left\|(V_\lambda^2 + \alpha I)^{-1} V_\lambda^2\right\|}_{\to 1} \\ =\; &O_p\left(\alpha^{-1} L^{-1/2}\right) + O_p\left(\alpha^{-1} S^{-1/2}\right).\end{aligned}$$

Now we turn our attention to the equation (4.37). We can write

$$\begin{aligned}(V_\lambda^2 + \alpha I)^{-1} V_\lambda f - V_\lambda^{-1} f &= \sum_{j=1}^{\infty} \left[\frac{\nu_j}{\alpha + \nu_j^2} - \frac{1}{\nu_j}\right] \langle f, \phi_j \rangle \phi_j \\ &= \sum_{j=1}^{\infty} \left(\frac{\nu_j^2}{\alpha + \nu_j^2} - 1\right) \frac{\langle f, \phi_j \rangle}{\nu_j} \phi_j.\end{aligned}$$

We now take the norm:

$$
\begin{aligned}
(4.37) &= \left\|(V_\lambda^2+\alpha I)^{-1}V_\lambda f - V_\lambda^{-1}f\right\| \\
&= \left(\sum_{j=1}^{\infty}\left(\frac{\nu_j^2}{\alpha+\nu_j^2}-1\right)^2 \frac{\left|\langle f,\phi_j\rangle\right|^2}{\nu_j^2}\right)^{1/2} \\
&= \left(\sum_{j=1}^{\infty}\nu_j^{2\epsilon-2}\left(\frac{\nu_j^2}{\alpha+\nu_j^2}-1\right)^2 \frac{\left|\langle f,\phi_j\rangle\right|^2}{\nu_j^{2\epsilon}}\right)^{1/2} \\
&\leq \left(\sum_{j=1}^{\infty}\frac{\left|\langle f,\phi_j\rangle\right|^2}{\nu_j^{2\epsilon}}\right)^{1/2} \sup_{1\leq j\leq\infty}\nu_j^{\epsilon-1}\left|\frac{\nu_j^2}{\alpha+\nu_j^2}-1\right|.
\end{aligned}
$$

Recall that as K is a compact operator, its largest eigenvalue ν_1 is bounded. We need to find an equivalent to

$$
\sup_{0\leq\nu\leq\nu_1}\nu^{\epsilon-1}\left(1-\frac{\nu^2}{\alpha+\nu^2}\right) = \sup_{0\leq\lambda\leq\nu_1^2}\lambda^{\frac{\epsilon-1}{2}}\left(1-\frac{1}{\alpha/\lambda+1}\right) \tag{4.68}
$$

Case where $\epsilon \leq 3/2$

We apply another change of variables to the objective function (4.68), $x=\alpha/\lambda$ and obtain

$$
\sup_{x\geq 0}\frac{\alpha^{\epsilon/2-1/2}}{x^{\epsilon/2-1/2}}\left(1-\frac{1}{x+1}\right).
$$

We see that an equivalent to (4.68) is $\alpha^{\epsilon/2-1/2}$ provided that

$$
\sup_{x\geq 0}\frac{1}{x^{\epsilon/2-1/2}}\left(1-\frac{1}{x+1}\right),
$$

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

$$g(x) \underset{0}{\sim} \frac{1}{\left(\frac{\epsilon-1}{2}\right) x^{\frac{\epsilon-3}{2}}} = 0,$$

provided $\epsilon < 3/2$. For $\epsilon = 3/2$, we have

$$g(x) \underset{0}{\sim} \frac{1}{\left(\frac{\epsilon-1}{2}\right)}.$$

Hence $g(x)$ is bounded on $\mathbb{R}^+$ for all $\epsilon \leq 3/2$.

Case where $\epsilon > 3/2$

We rewrite (4.68) as

$$\sup_{0\leq\lambda\leq\nu_1^2} \alpha\lambda^{\frac{\epsilon-3}{2}} \left[\frac{\left(1-\frac{1}{\alpha/\lambda+1}\right)}{\alpha/\lambda}\right].$$

The term

$$\lambda^{\frac{\epsilon-3}{2}} \left[\frac{\left(1-\frac{1}{\alpha/\lambda+1}\right)}{\alpha/\lambda}\right].$$

is the product of an increasing function of λ, namely $\lambda^{\frac{\epsilon-3}{2}}$ (which is bounded because λ is bounded) and a function of the form $\left(1-\frac{1}{x+1}\right)/x$. It is easy to show using the l'Hopital's rule that $\left(1-\frac{1}{x+1}\right)/x$ is bounded on $\mathbb{R}^+$.

$$\left(1-\frac{1}{x+1}\right)/x \underset{+\infty}{\sim} 0$$
$$\left(1-\frac{1}{x+1}\right)/x \underset{0}{\sim} 1$$

Hence the rate of (4.68) is given by α.

Finally, $f \in \Phi_\epsilon$:

$$(4.37) = O\left(\alpha^{\min\left(1,\frac{\epsilon-1}{2}\right)}\right).$$

■

Proof of Proposition 7: According to Equation (4.36) of proposition 6, $\iota'\widehat{V}_{\lambda,\alpha}^{-1}\iota$ is consistent for $\iota' V_{\lambda,\alpha}^{-1} \iota$ since we have:

$$\iota' \widehat{V}_{\lambda,\alpha}^{-1} = \iota' V_{\lambda,\alpha}^{-1} + O_p\left(\alpha^{-1} L^{-1/2}\right) + O_p\left(\alpha^{-1} S^{-1/2}\right)$$

On the other hand, we can use the Delta method to obtain:

$$\begin{aligned}
\left(\iota' \widehat{V}_{\lambda,\alpha}^{-1} \iota\right)^{-1} &\approx \left(\iota' V_{\lambda,\alpha}^{-1} \iota\right)^{-1} - \left(\iota' V_{\lambda,\alpha}^{-1} \iota\right)^{-2} \iota' \left(\widehat{V}_{\lambda,\alpha}^{-1} - V_{\lambda,\alpha}^{-1}\right) \iota \\
&= \left(\iota' V_{\lambda,\alpha}^{-1} \iota\right)^{-1} + O_p\left(\alpha^{-1} L^{-1/2}\right) + O_p\left(\alpha^{-1} S^{-1/2}\right)
\end{aligned}$$

This yields immediately:

$$\begin{aligned}
\widehat{\pi}_{\lambda,\alpha}^{*} &= \left(\iota' \widehat{V}_{\lambda,\alpha}^{-1} \iota\right)^{-1} \iota' \widehat{V}_{\lambda,\alpha}^{-1} \\
&= \left(\iota' V_{\lambda,\alpha}^{-1} \iota\right)^{-1} \iota' V_{\lambda,\alpha}^{-1} + O_p\left(\alpha^{-1} L^{-1/2}\right) + O_p\left(\alpha^{-1} S^{-1/2}\right)
\end{aligned}$$

Furthermore, using Equation (4.37) of proposition 6 yields:

$$\widehat{\pi}_{\lambda,\alpha}^{*} - \pi_{\lambda}^{*} = O\left(\alpha^{\min(1, \frac{\epsilon-1}{2})}\right) + O_p\left(\alpha^{-1} L^{-1/2}\right) + O_p\left(\alpha^{-1} S^{-1/2}\right)$$

Hence $\widehat{\pi}_{\lambda,\alpha}^{*}$ is consistent for π_{λ}^{*} provided that α, $\alpha^{-1} L^{-1/2}$ and $\alpha^{-1} S^{-1/2}$ go to zero as L and S increase to infinity.

Next, $\underline{\widehat{\theta}}^{(1)} = \left(\widehat{\theta}^{(1)}(\tau_1), ..., \widehat{\theta}^{(1)}(\tau_1)\right)'$. We have:

$$\begin{aligned}
\widehat{\theta}_{\widehat{\pi}_{\lambda,\alpha}^{*}} &= \sum_{s=1}^{S} \widehat{\pi}_{\lambda,\alpha}^{*}(\tau_s)\, \widehat{\theta}^{(1)}(\tau_s) \\
&= \sum_{s=1}^{S} \pi_{\lambda}^{*}(\tau_s)\, \widehat{\theta}^{(1)}(\tau_s) + \sum_{s=1}^{S} \left(\widehat{\pi}_{\lambda,\alpha}^{*}(\tau_s) - \pi_{\lambda}^{*}(\tau_s)\right) \widehat{\theta}^{(1)}(\tau_s) \\
&= \widehat{\theta}_{\pi_{\lambda}^{*}} + \sum_{s=1}^{S} \left(\widehat{\pi}_{\lambda,\alpha}^{*}(\tau_s) - \pi_{\lambda}^{*}(\tau_s)\right) \widehat{\theta}^{(1)}(\tau_s)
\end{aligned}$$

We first consider the second expression:

$$
\begin{aligned}
& \sum_{s=1}^{S}\left(\widehat{\pi}_{\lambda,\alpha}^{*}\left(\tau_s\right)-\pi_{\lambda}^{*}\left(\tau_s\right)\right)\widehat{\theta}^{(1)}\left(\tau_s\right) \\
= \;& \theta_0\underbrace{\sum_{s=1}^{S}\left(\widehat{\pi}_{\lambda,\alpha}^{*}\left(\tau_s\right)-\pi_{\lambda}^{*}\left(\tau_s\right)\right)}_{=0}+\sum_{s=1}^{S}\left(\widehat{\pi}_{\lambda,\alpha}^{*}\left(\tau_s\right)-\pi_{\lambda}^{*}\left(\tau_s\right)\right)\left(\widehat{\theta}^{(1)}\left(\tau_s\right)-\theta_0\right) \\
= \;& \sum_{s=1}^{S}\left(\widehat{\pi}_{\lambda,\alpha}^{*}\left(\tau_s\right)-\pi_{\lambda}^{*}\left(\tau_s\right)\right)\left(\widehat{\theta}^{(1)}\left(\tau_s\right)-\theta_0\right)
\end{aligned}
$$

which shows that the last expression goes to zero faster than $T^{-1/2}$ and only contributes to the higher order variance.

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

$$
\begin{aligned}
& E\left[\exp\left(\sum_{k=1}^{d} i\tau_k r_{t+1-k}\right) \mid \{V_{t-k}\}_{k=1}^{d}\right] \\
= \;& E\left[\exp\left(\sum_{k=1}^{d} i\tau_k\left[\mu_0+\mu_1 V_{t-k}+\delta\left(V_{t+1-k}-\beta\left(1-e^{-\kappa}\right)-e^{-\kappa}V_{t-k}\right)+\sigma_{\varepsilon}\varepsilon_t\right]\right) \mid \{V_{t-k}\}_{k=1}^{d}\right] \\
= \;& E\left[\exp\left(\sum_{k=1}^{d} i\tau_k\left(\mu_0-\delta\beta\left(1-e^{-\kappa}\right)+\left(\mu_1-\delta e^{-\kappa}\right)V_{t-k}+\delta V_{t+1-k}\right)-\frac{\sigma_{\varepsilon}^2}{2}\sum_{k=1}^{d}\tau_k^2\right) \mid \{V_{t-k}\}_{k=1}^{d}\right] \\
= \;& \exp\left[\sum_{k=1}^{d}\left(i\tau_k\left(\mu_0-\delta\beta\left(1-e^{-\kappa}\right)\right)-\frac{\sigma_{\varepsilon}^2}{2}\tau_k^2\right)\right] \\
& \times E\left[\exp\left(i\tau_d\left(\mu_1-\delta e^{-\kappa}\right)V_{t-d}+\sum_{k=1}^{d-1}\left(i\tau_k\left(\mu_1-\delta e^{-\kappa}\right)+i\tau_{k+1}\delta\right)V_{t-k}+i\tau_1\delta V_t\right) \mid \{V_{t-k}\}_{k=1}^{d}\right]
\end{aligned}
$$

Hence:

$$
\begin{aligned}
= \;& \exp\left[\sum_{k=1}^{d}\left(i\tau_k\left[\mu_0-\delta\beta\left(1-e^{-\kappa}\right)\right]-\frac{\sigma_{\varepsilon}^2}{2}\tau_k^2\right)\right] \\
& \times\exp\left[i\tau_d\left(\mu_1-\delta e^{-\kappa}\right)V_{t-d}+\sum_{k=2}^{d-1}\left[i\tau_k\left(\mu_1-\delta e^{-\kappa}\right)+i\tau_{k+1}\delta\right]V_{t-k}\right] \\
& \times\left(1-\frac{i\tau_1\delta}{c}\right)^{-q}\exp\left[\left(\frac{i\tau_1\delta e^{-\kappa}}{1-\frac{i\tau_1\delta}{c}}+i\tau_1\left(\mu_1-\delta e^{-\kappa}\right)+i\tau_2\delta\right)V_{t-1}\right]
\end{aligned}
$$

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

V_{t-1} yields:

$$
\begin{aligned}
& E\left[\exp\left(\sum_{k=1}^{d} i\tau_k r_{t+1-k}\right) \mid \{V_{t-k}\}_{k=2}^{d}\right] \\
= \;& \exp\left[\sum_{k=1}^{d}\left(i\tau_k\left[\mu_0-\delta\beta\left(1-e^{-\kappa}\right)\right]-\frac{\sigma_\varepsilon^2}{2}\tau_k^2\right)\right] \\
& \times\exp\left[i\tau_d\left(\mu_1-\delta e^{-\kappa}\right)V_{t-d}+\sum_{k=2}^{d-1}\left[i\tau_k\left(\mu_1-\delta e^{-\kappa}\right)+i\tau_{k+1}\delta\right]V_{t-k}\right] \\
& \times\left(1-\frac{iu_1}{c}\right)^{-q} E\left[\exp\left(iu_2V_{t-1}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
= \;& \exp\left[\sum_{k=1}^{d}\left(i\tau_k\left[\mu_0-\delta\beta\left(1-e^{-\kappa}\right)\right]-\frac{\sigma_\varepsilon^2}{2}\tau_k^2\right)\right] \\
& \times\exp\left[i\tau_d\left(\mu_1-\delta e^{-\kappa}\right)V_{t-d}+\sum_{k=3}^{d-1}\left[i\tau_k\left(\mu_1-\delta e^{-\kappa}\right)+i\tau_{k+1}\delta\right]V_{t-k}\right] \\
& \times\left(1-\frac{iu_1}{c}\right)^{-q}\left(1-\frac{iu_2}{c}\right)^{-q}\exp\left[iu_3V_{t-2}\right]
\end{aligned}
$$

where $u_3=\frac{u_2e^{-\kappa}}{1-\frac{iu_2}{c}}+\tau_2\left(\mu_1-\delta e^{-\kappa}\right)+\tau_3\delta$. Integrating out recursively V_{t-2} conditional on V_{t-3}, V_{t-3} conditional on V_{t-4} and so forth, we get:

$$
\begin{aligned}
& E\left[\exp\left(\sum_{k=1}^{d} i\tau_k r_{t+1-k}\right) |V_{t-d}\right] \\
= \;& \left[\prod_{k=1}^{d-1}\exp\left(i\tau_k\left[\mu_0-\delta\beta\left(1-e^{-\kappa}\right)\right]-\frac{\sigma_\varepsilon^2}{2}\tau_k^2\right)\left(1-\frac{iu_k}{c}\right)^{-q}\right] \\
& \times\exp\left[iu_{d+1}V_{t-d}\right]
\end{aligned}
$$

where:

$$
\begin{aligned}
u_1 &= \tau_1\delta \\
u_k &= \frac{u_{k-1}e^{-\kappa}}{1-\frac{iu_{k-1}}{c}}+\tau_{k-1}\left(\mu_1-\delta e^{-\kappa}\right)+\tau_k\delta,\; k=2,...,d \\
u_{d+1} &= \frac{iu_de^{-\kappa}}{1-\frac{iu_d}{c}}+i\tau_d\left(\mu_1-\delta e^{-\kappa}\right)
\end{aligned}
$$

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

$$\begin{aligned} E\left[\exp\left(\sum_{k=1}^{d} i\tau_k r_{t+1-k}\right)\right] &= \exp\left(\left[\mu_0 - \delta\beta\left(1-e^{-\kappa}\right)\right]\sum_{k=1}^{d} i\tau_k - \frac{\sigma_\varepsilon^2}{2}\sum_{k=1}^{d}\tau_k^2\right) \\ &\quad \times \left(1 - \frac{iu_{d+1}\sigma^2}{2\kappa}\right)^{-q} \prod_{k=1}^{d}\left(1 - \frac{iu_k}{c}\right)^{-q} \end{aligned}$$

Conclusion Générale

La présente thèse s'inscrit dans deux littératures voisines: l'économétrie financière et l'économétrie théorique. Le volet économétrie financière est abordé dans le chapitre 1 tandis que le volet économétrie théorique est abordé dans les chapitres 2, 3 et 4. Le chapitre 3 est une application de la théorie du chapitre 2 tandis que le chapitre 4 en est une extension. Les quatre chapitres sont passés en revue ci-après.

L'objectif du chapitre 1 est de proposer un modèle de bruit de microstructure réaliste qui servira de cadre pour étudier la qualité des estimateurs de volatilité intégré construits à partir de données à haute fréquence. Le modèle que nous proposons pour le bruit prévoit non seulement qu'il peut être autocorrélé jusqu'à l'ordre L, mais aussi qu'il peut être corrélé aux rendements efficients. En outre, nous formulons une hypothèse explicite sur la façon dont le corrélogramme du bruit varie en fonction de la fréquence des observations m. Cette hypothèse implique, entre autre, que l'autocorrélation d'ordre un du bruit tends vers un lorsque m tends vers l'infini. En formulant cette hypothèse, notre but est de traduire le fait que les observations sont de plus en plus rapprochées dans le temps lorsque m augmente. Ceci est une innovation majeure comparée aux modèles qui considèrent une structure d'autocorrélation invariante à la fréquence des observations.

Nous utilisons ce cadre pour dériver les propriétés de trois estimateurs couramment rencontrés dans cette littérature. Cet exercice préliminaire nous a permit de construire un nouvel estimateur qui combine linéairement deux estimateurs aux propriétés différentes en présence du bruit de microstructure: l'un est sans biais et se détériore à haute fréquence, l'autre est convergent dans le sens que sa variance décroît lorsque la fréquence des observations croît. Par ailleurs, l'estimateur sans biais a tendance à être le meilleur des deux lorsque la magnitude du bruit de microstructure est faible. Les poids affectés aux estimateurs dans la combinaison linéaire sont choisis de façon à minimiser la variance de l'estimateur obtenu que nous baptisons "estimateur shrinkage". Les simulations ont montré que l'estimateur shrinkage a une variance plus faible que le meilleur des deux éléments de la combinaison linéaire. Dans la partie empirique, nous testons si la mémoire maximale L de l'autocorrélation du bruit augmente avec la fréquence des observations. Les données suggèrent que si ceci est le cas, L n'augmente pas plus vite que $\sqrt{m}$.

Le chapitre 2 examine les propriétés théoriques de l'estimateur CGMM dans les modèles IID ou Markoviens, l'objectif étant de proposer un critère du choix du paramètre de régularisation α dont dépend la fonction objectif. En effet, ce paramètre de régularisation a un impact important sur l'erreur quadratique moyenne (EQM) de l'estimateur CGMM. Nous suivons une approche similaire à Newey et Smith (2004) pour dériver une expansion stochastique de l'estimateur CGMM. De ce résultat, nous déduisons une expansion de l'EQM qui permet de caractériser la façon dont celle-ci dépend de α à distance finie. Nous montrons que l'estimateur CGMM est optimal lorsque α converge vers zéro à une certaine vitesse en fonction de la taille de l'échantillon T. Nous proposons deux méthodes pour estimer le α optimal en pratique, ce dernier étant défini comme celui qui minimise l'EQM. La première exploite l'approximation analytique de l'EQM tandis que la seconde est basée sur une simulation de Monte Carlo. Nous montrons que la seconde méthode délivre un estimateur $\sqrt{T}$-convergent de l'estimateur de α. Des simulations de Monte Carlo basées sur un modèle de frontière stochastique confirment qu'il faut accorder de l'importance à la sélection optimale de α.

L'objectif du chapitre 3 est de rendre l'utilisation du CGMM accessible aux praticiens. Nous avons donc choisi d'illustrer la théorie du chapitre 2 dans le cadre de modèles Markoviens pour lesquels la dimensionnalité des intégrales de la fonction-objectif est au plus égale à 2. Dans ce cas, les quadratures de Gauss-Hermite peuvent être utilisées avec efficacité. Nous commençons par réviser la théorie du CGMM. Ensuite, nous exposons des recettes numériques utiles. Enfin, nous proposons un exercice de simulation basé sur la loi stable et une étude empirique basée sur un modèle de variance autorégressif Gamma. Certains paramètres de la loi stable ne sont pas identifiés lorsque le paramètre de stabilité est proche de 2. Lorsque le paramètre de stabilité est proche de la zone de non identification, les distributions exactes des estimateurs CGMM ne sont pas gaussiennes mais peuvent être obtenue par simulation de Monte Carlo. Dans la partie empirique, nous testons l'existence d'une liaison positive entre rendement et risque espéré ainsi que d'une liaison négative entre rendement espéré et risque non espéré. Un modèle autorégressif Gamma d'ordre 1 est utilisé pour séparer un proxy de la volatilité intégré entre son espérance conditionnelle et une innovation. Le proxy de la volatilité intégrée est obtenu à partir de données à hautes fréquence selon les méthodes du chapitre 1. Nos résultats indiquent que ces liaisons existent, mais la largeur des intervalles de confiance limitent la portée des résultats.

Le chapitre 4 vient combler un vide laissé par le chapitre 3. En effet, le temps de calcul de l'évaluation de la fonction objectif du CGMM par les quadratures augmente exponentiellement en fonction de la multiplicité des intégrales présentes dans l'expression de la fonction objectif. Dans les modèles non markoviens où l'on souhaite conditionner sur plusieurs retards, cette multiplicité augmente linéairement avec le nombre de retards et avec la dimensionnalité du modèle. Étant donné que la fonction-objectif doit être évaluée itérativement lors de l'optimisation numérique, la mise en oeuvre du CGMM devient vite laborieuse, voire impossible dans les modèles de dimensions $d \geq 3$. La solution proposée consiste à fabriquer des échantillons univariés à partir de combinaisons linéaires du vecteur initial d'observations, les poids de la combinaison linéaire étant tirés aléatoirement dans un sous-espace normalisé de $\mathbb{R}^d$. Chaque échantillon ainsi généré peut servir à estimer le paramètre d'intérêt θ_0 à moindre coût. Enfin, l'ensemble des estimateurs qui en découlent peuvent être combinés linéairement pour obtenir un estimateur final. Cette nouvelle méthode d'estimation est baptisée "*re-échantillonnage dans le domaine fréquentiel*". Dans la suite de l'article, nous discutons de la règle d'agrégation optimale et nous proposons une marche à suivre pour son implémentation. Nous conduisons une simulation de Monte Carlo basé sur un modèle à facteur autorégressif Gamma. Dans ce modèle, le rendement d'un actif est exprimé comme une fonction linéaire d'un facteur de risque latent, ce dernier étant à son tour modélisé comme un processus autorégressif Gamma d'ordre 1. Il est à noter que le facteur de risque ne se confond pas exactement avec la volatilité des rendements. Lorsqu'on marginalise le modèle, la dynamique des rendements n'est plus markovien. Pour être efficace, le CGMM doit donc se baser ici sur la fonction caractéristique d'un vecteur de L observations consécutives, pour L assez grand. Les résultats de la simulation indiquent que le re-échantillonnage dans le domaine fréquentielle est une méthode d'inférence fiable qui élargit un peu plus le champ d'utilisation du CGMM. Nous proposons une étude empirique pour clôturer ce chapitre. Dans cette étude, nous reprenons la problématique de la partie empirique du chapitre 3 et utilisons un modèle autorégressif Gamma d'ordre p à la place du modèle autorégressif d'ordre 1. Nous observons une réduction importante des intervalles de confiances des estimateurs des paramètres décrivant la liaison entre le rendement et la volatilité. En outre, nous constatons que le filtre autorégressif de la volatilité qui découle de cette modélisation pourrait être utilisé pour séparer la volatilité entre la partie continue et les sauts.

Deux avenues de recherches peuvent être explorées dans le futur. La première consisterait à explorer l'extension des idées du chapitre 1 au cas multivarié, c'est à dire à l'estimation des matrices de covariances intégrées en présence de bruit de microstructure multivariés. Ce travail exigera surtout un travail d'adaptation des hypothèses formulées ici aux cas multidimensionels. La seconde avenue de recherche est en relation avec les trois derniers chapitres de la thèse et consiste à établir une analogie théorique entre l'estimation de densité par noyaux et le CGMM. Parallèlement, un travail de promotion de l'utilisation des CGMM dans des domaines variés mérite d'être fait.