UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LA VRAISEMBLANCE EMPIRIQUE GÉNÉRALISÉE POUR UN CONTINUUM DE CONDITIONS DE MOMENTS

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN ÉCONOMIQUE

PAR

PIERRE CHAUSSÉ

FÉVRIER 2012

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

GENERALIZED EMPIRICAL LIKELIHOOD FOR A CONTINUUM OF MOMENT CONDITIONS

DISSERTATION

PRESENTED

AS PARTIAL REQUIREMENT

OF THE DOCTORATE OF ECONOMICS

BY

PIERRE CHAUSSÉ

FEBRUARY 2012

REMERCIEMENTS

J'aimerais remercier mon directeur de thèse Alain Guay pour m'avoir guidé dans la rédaction de ma thèse ainsi que pour m'avoir aidé à choisir mon sujet. Je suis également reconnaissant pour son support financier ainsi que celui des ressources humaines de l'UQÀM dans le cadre du programme de perfectionnement longue durée des chargés de cours. Je tiens également à remercier Marine Carrasco et trois arbitres anonymes choisis par les editeurs du Journal of Econometrics pour leurs commentaires et suggestions.

TABLE DES MATIÈRES

TABLE DES FIGURES

LISTE DES TABLEAUX

Tableau Page

RÉSUMÉ

Dans cette thèse, je propose une généralisation de la méthode de vraisemblance empirique généralisée (GEL) pour permettre la possibilité d'avoir soit un très grand nombre de conditions de moment ou des conditions définies sur un continuum. Cette généralisation peut permettre par exemple d'estimer des modèles de régression avec régresseurs endogènes pour lesquels le nombre d'instruments est très élevé ou encore que la relation entre les régresseurs et les variables exogènes est inconnu . Il est également possible de baser notre estimation sur des conditions de moment construites à partir de fonctions caractéristiques. Il devient alors possible d'estimer les coefficients d'une distribution quelconque ou d'un processus stochastique lorsque sa fonction de vraisemblance n'admet pas de forme analytique. C'est le cas entre autres de la distribution stable et de la plupart des processus de diffusion exprimés en temps continu. Cette généralisation a été proposée par (Carrasco and Florens, 2000) pour la méthode des moments généralisés (CGMM). Sur la base des résultats de (Newey and Smith, 2004), qui démontrent la supériorité asymptotique de GEL sur GMM, la méthode que je propose représente donc une contribution substantielle.

La thèse est divisée en trois chapitres. Le premier présente en détails la méthode de vraisemblance empirique généralisée pour un continuum de moments (CGEL), démontre la convergence en probabilité et en distribution de ses estimateurs et décrit la procédure à suivre en pratique pour estimer les coefficients du modèle à l'aide d'une approche matricielle relativement simple. De plus, je démontre l'équivalence asymptotique de CGEL et CGMM. CGEL est en fait un algorithme non-linéaire régularisé à la Tikhonov, qui permet d'obtenir l'estimateur GEL dans le cas où le nombre de conditions est très grand. Dans cette méthode, un paramètre de régularisation, α_n, permet de résoudre le problème d'optimisation mal posé qui en résulte et d'obtenir une solution unique et stable. Le paramètre α_n doit converger vers zéro lentement lorsque la taille d'échantillon augmente pour que l'estimateur soit convergent et que la solution demeure stable. Les détails du rythme de convergence de α_n sont également présentés dans ce chapitre. Finalement, le chapitre présente la façon de tester les conditions de moments en généralisant les trois tests de spécifications existants pour GEL.

Dans le chapitre 2, je présente plusieurs applications numériques. L'objectif est de voir les possibilités de CGEL, d'analyser les propriétés et ses estimateurs en échantillons finis, en comparaison avec ceux de CGMM, et de comprendre l'impact du paramètre α_n sur le biais et la variance des estimateurs. Les applications analysée sont : l'estimation d'un modèle linéaire avec endogénéité de forme inconnue, l'estimation des paramètres d'une distribution stable et l'estimation des coefficients d'un processus de diffusion. De façon générale les résultats démontrent que la dominance de CGEL sur CGMM dépend de la valeur de α_n. Cela démontre en fait la nécessité de développer une méthode de sélection de α_n.

Finalement, une méthode de sélection du paramètre α_n est proposée dans le dernier chapitre. Dans un premier temps, je démontre qu'une méthode de bootstrap simple permet difficilement de faire un choix optimal car elle produit une relation très volatile entre α_n et l'erreur quadratique moyen (MSE) du coefficient. Ensuite, je présente une approximation de second ordre du MSE de CGEL par un développement stochastique des conditions de premier ordre comme fait par (Donald and Newey, 2001) pour les double moindres carrés, (Donald, Imbens and Newey, 2010) pour GEL ainsi que (Carrasco, 2010) et (Carrasco and Kotchoni, 2010) pour CGMM. Cette approche permet d'obtenir une relation lisse entre α_n et le MSE et donc d'utiliser un algorithme d'optimisation pour obtenir la paramètre optimal. Les résultats semblent être conformes aux résultants précédents selon lesquels la méthode de vraisemblance empirique domine les autres méthodes faisant partie de la famille CGEL. Ils semblent également suggérer que α_n, pour le cas linéaire considéré, devrait être choisi aussi petit que possible car c'est de cette façon que le MSE est minimisé.

Mots-clés : Vraisemblance Généralisée, Continuum de moments, Méthode des moments généralisés, Économetrie, Variables Instrumentales

ABSTRACT

In this thesis, we propose a generalization of the generalized empirical likelihood method (GEL) to allow the possibility of a large number of moment conditions or conditions defined on a continuum. This generalization allows for example to estimate linear models with endogenous regressors for which the number of instruments is large or for cases in which the relationship between the exogenous and endogenous variables is unknown. It may also be used to estimate models for which the moment conditions are based on characteristic functions. This latter application allows us for example to estimate the parameters of a distribution or a stochastic process for which the likelihood function does not have a close form representation as it it the case with the stable distribution and many diffusion processes. A similar extension is proposed by (Carrasco and Florens, 2000) for the generalized method of moments (CGMM). Because (Newey and Smith, 2004) show that GEL dominates the generalized method of moments (GMM) asymptotically, what we propose is an important contribution.

The thesis contains three chapters. The first presents in details the generalized empirical likelihood method for a continuum (CGEL), shows the convergence in probability and in distribution of its estimator and describes the procedure to implement the method in practice using a simple matrix representation of the algorithm. Furthermore, we show that CGEL is asymptotically equivalent to CGMM. The CGEL method is in fact a regularized nonlinear algorithm a la Thikhonov which gives a stable solution to cases in which the number of conditions is so high that the system of equations that we have to solve becomes singular or nearly singular. As the sample sizes increases, the regularization parameter, α_n, that stabilizes the solution, must converge to zero at a sufficiently slow rate to keep the solution stable and the estimator convergent. The required rate of convergence of α_n is described in this chapter as well. Finally, the three tests of over-identifying restrictions known for GEL are presented for the case of a continuum of conditions.

In chapter 2, we propose some numerical examples in order to analyze the finite sample properties of CGEL, to compare it with CGMM and to see the impact of α_n on the bias and the mean squared error (MSE) of the estimators. In particular, we estimate the parameters of a stable distribution, the coefficients of a linear model with endogenous regressors for which the optimal instruments are unknown and the coefficients of a diffusion process. In general, the results show that the relative performance of CGMM and CGEL depends on the regularization parameter α_n. The method is therefore difficult to apply without a method for selecting α_n, which we succeed in finding in the last chapter of the thesis.

The last chapter proposes a data driven procedure for selecting α_n. First, we show that a simple bootstrap method does not help selecting the optimal regularization parameter because it produces a non smooth relationship between the estimated MSE and α_n. We then propose to estimate the MSE using a stochastic expansion of the first order conditions as it is done by (Donald and Newey, 2001) for two stage least squares, (Donald, Imbens and Newey, 2010) for GEL, and (Carrasco, 2010) and (Carrasco and Kotchoni, 2010) for CGMM. A numerical example shows that the approximated MSE is a smooth function of α_n which allows us to obtain the optimal value using a simple numerical algorithm. Also, the impact of α_n on the MSE that we obtain is consistent with the results of chapter 2 which suggests that it is a good approximation. Finally, it suggests that in the linear case, the parameter should be as small as possible.

Keywords : Empirical likelihood, Continuum of moments, Generalized method of moments, Econometrics, Instrumental variable

INTRODUCTION

In this thesis, we propose a generalization of the generalized empirical likelihood method (GEL) to allow the possibility of a large number of moment conditions or conditions defined on a continuum. This generalization allows for example to estimate linear models with endogenous regressors for which the number of instruments is large or for cases in which the relationship between the exogenous and endogenous variables is unknown. It may also be used to estimate models for which the moment conditions are based on characteristic functions. This latter application allows us for example to estimate the parameters of a distribution or a stochastic process for which the likelihood function does not have a closed form representation as is the case with the stable distribution and many diffusion processes. A similar extension is proposed by (Carrasco and Florens, 2000) for the generalized method of moments (CGMM).

(Newey and Smith, 2004) show that GEL dominates the two-step generalized method of moments (GMM) asymptotically by deriving their second order properties. What they find is that the second order bias of GEL is smaller than the one from GMM, and among the different GEL methods, it is the empirical likelihood (EL) estimator that possesses the smallest asymptotic bias. In particular, its bias does not increase with the number of moment conditions. As oppose to GMM, EL does not require a first step estimate, which is one of components of the second order bias of GMM. Also, two other components of the bias of GMM is absent from EL because it estimates the Jacobian and the covariance matrix of the vector of moment conditions using the empirical probabilities implied by the model. Because CGMM and CGEL are regularized versions of GMM and GEL, we expect their second order biases to behave similarly. The simulations in Chapter 2 confirms that in general CGEL is less biased than CGMM. The second important result of (Newey and Smith, 2004) is that the bias-corrected version of EL is the most efficient in terms of the root mean square error (RMSE). However, (Guggenberger, 2008) shows that the EL estimator, like the one from the limited information maximum likelihood method (LIML), fails to have finite moments. Indeed, he shows that the EL estimator has a high probability of having extreme values in small samples with weak instruments. The results of Chapter 2 confirm that the RMSE of CGEL is often higher than the RMSE of CGMM. However, the estimators of CGEL seem to have finite moments. We should therefore expect the bias-corrected version of CGEL to dominate CGMM in terms of the RMSE. The results from second order expansion of CGEL in Chapter 3 can be used derive such estimator.

The solution proposed to deal with the infinite number of conditions has such an impact on the properties of the estimator, that we can almost treat the method as being a new one. It therefore has the potential to create several ideas for future research. This thesis is an introduction to this new method and is intended to present its properties in the simple case in which the data are iid. It also illustrates how to implement it using simple applications.

CHAPITRE I

ASYMPTOTIC PROPERTIES OF THE GEL FOR A CONTINUUM

1.1 Introduction

When estimating models based on moment conditions, it is often the case that the number of conditions is so large that selecting the right ones becomes an issue. For example, in the case of linear models with endogenous regressors as considered by (Carrasco, 2010), the set of possible instruments can be countably infinite or defined on a continuum. Moment conditions can also be naturally based on a continuum when, for example, they are defined by characteristic functions or spectral densities. In these cases, methods such as instrument variables (IV) cannot be based on the whole set of moment conditions because the system of equations implied by the first order conditions becomes singular as the number of conditions increases beyond the sample size. Because it may reduce the quality of the estimators when the weak instruments are chosen, we have to be careful in the selection. (Donald and Newey, 2001) present a method for selecting the optimal number of instruments but it requires a certain ordering so that the stronger are selected and the weaker are dropped. On the other hand, (Carrasco, 2010) applies the generalized method of moment for a continuum (CGMM) of (Carrasco and Florens, 2000) in which the whole set of instruments can be used without imposing any ordering. The method is based on a Tikhonov regularization technique [1] which is comparable to

1. A common use of the Tikhonov technique in econometrics, is the ridge regression. This method regularized the inverse of $X'X$ by adding a positive number α to its diagonal.

a principal component selection procedure. The most influential moment conditions are therefore automatically selected. In this paper, we extend the generalized empirical likelihood method (GEL) of (Smith, 1997) so that it can also deal with a continuum of moment conditions (CGEL). The second order asymptotic results obtained by (Newey and Smith, 2004) and (Anatolyev, 2005) suggest that CGEL may be a good alternative to CGMM. The CGEL estimator is defined as the solution to a constrained optimization problem in which the number of constraints is infinite. Such a problem cannot easily be solved using a finite number of observations. The main contribution of the paper is to show both theoretically and practically how we can obtain a stable solution to such problems. The method can even be applied to cases in which the number of conditions is finite but large enough so that the problem becomes ill-conditioned. It offers a way to deal with the selection of moment conditions using a Tikhonov type approach similar to CGMM. Furthermore, we present the algorithms in matrix notation to simplify its implementation.

When defining the objective function of the efficient CGMM, we need the regularized solution to a linear ill-posed problem, because the optimal weighting operator cannot be continuously inverted. On the other hand, the objective function of CGEL is well defined. However, the system of equations from which we compute the Lagrange multiplier associated with the moment conditions becomes singular when the number of conditions goes to infinity. As a result, we present CGEL as a nonlinear ill-posed problem in the sense that a unique and stable solution cannot be obtained directly from the first order conditions. The literature in applied mathematics offers several ways to deal with nonlinear ill-posed problems. As a first procedure, we apply the regularized Gauss-Newton algorithm which can be compared to using ridge regression techniques to estimate a poorly conditioned nonlinear regression. We also present an alternative regularized method which is based on the singular value decomposition of the first order Taylor approximation of the solution. This method has the advantage of being less computationally demanding and asymptotically equivalent to the first procedure. We present the algorithms for the exponential tilting (CET), the empirical likelihood

(CEL) and the Euclidean empirical likelihood (CEEL) for a continuum by using a matrix notation as in (Carrasco et al., 2007) for CGMM. Moreover, in order to test the over-identifying restrictions, we present a normalized version of the three tests proposed by (Smith, 2004) so that they are all asymptotically distributed as a standardized normal distribution. We conclude the theoretical part with a brief discussion on how to implement the exponentially tilted empirical likelihood of (Schennach, 2007) for a continuum (CETEL).

We perform a numerical study in which we compare the finite sample properties of the three CGEL methods using the two proposed algorithms with CGMM. We use the example of estimating the parameters of a stable distribution using the marginal characteristic function as in (Carrasco and Florens, 2002) and (Garcia, Renault and Veredas, 2006). We also compare the empirical sizes of the three tests of over-identifying restrictions. All the results are computed for different values of the regularization parameter because no data-driven method is available to select its optimal value. What we get suggests that CGEL may outperform CGMM according to the root mean squared error criterion.

The paper is organized as follows. Section 1.2 gives an overview of GEL while section 1.3 presents the CGEL method and section 1.5 the three tests of over-identifying restrictions. Section 1.4 describes the two numerical algorithms and section 1.6 concludes.

1.2 GEL

This section presents an overview of the GEL method when there are a finite number of moment conditions. It serves as an introduction to the next section, which generalizes the method to the case of a continuum of conditions. Therefore we express the function defining the moment conditions in a way that facilitates the transition from GEL to CGEL.

We suppose that the vector $\theta_0 \in \Theta \subseteq \mathbb{R}^p$ is uniquely identified by a vector of q moment conditions. Instead of writing these conditions in the usual way as $E[g_\tau(X;\theta_0)] = 0$ for $\tau = 1, ..., q$, we incorporate the index in the function as follows :

$$E^{P_0} g(X, \tau_i; \theta_0) = 0 \;\; \forall i = 1, \cdots, q, \tag{1.1}$$

where the index i implies that τ belongs to a countable set (finite in this section), and P_0 is the true probability distribution associated with the random variable X. For example, if we are estimating a linear model using instruments, $\tau_i = i$ and defines the condition associated with the ith instrument. But it could also be an element of the function if, for example, the vector of parameters is estimated using characteristic functions. In this case, τ_i would be equal to some selected points of the function which are the most susceptible of producing good estimates.

We suppose that we can estimate the moment function from a vector of n i.i.d. realizations of the random variable X, $\{x_1, x_2, \cdots, x_n\}$. In general, we can write the $q \times 1$ vector of sample moment conditions as follows :

$$\tilde{g}(\theta) = \sum_{t=1}^{n} p_t g(x_t; \theta),$$

where p_t is the probability associated with the realization x_t.

The GMM estimator is defined as the vector of parameters that minimizes the norm of the sample moment $\bar{g}(\theta)$, which is based on the empirical density of the observations $f_n(x_t) = (1/n) \; \forall t$. Going through this optimization problem is necessary because when $q > p$, there is no solution to the sample moment conditions $\bar{g}(\theta) = 0$, in which p_t is restricted to $1/n$. On the other hand, the GEL method consists in finding the implied probabilities p_t, which are as close as possible to $1/n$ according to a certain family of discrepancies $h_n(p_t)$, that satisfy the conditions exactly. This interpretation represents the primal problem which defines the GEL estimators :

$$\hat{\theta}_{gel} = \arg\min_{\theta, p_t} \sum_{t=1}^{n} h_n(p_t) \tag{1.2}$$

subject to

$$\sum_{t=1}^{n} p_t g(x_t, \tau_i; \theta) \;\; = \;\; 0 \;\;\; \forall i = 1, \cdots q \;\; \text{and} \tag{1.3}$$

$$\sum_{t=1}^{n} p_t \quad = \quad 1, \tag{1.4}$$

as long as $h_n(p_t)$ belongs to the following Cressie-Read family of discrepancies [2].

$$h_n(p_t) = \frac{[\gamma(\gamma+1)]^{-1}[(np_t)^{\gamma+1} - 1]}{n}.$$

(Smith, 1997) shows that the empirical likelihood method (EL) of (Owen, 2001) ($\gamma = 0$) and the exponential tilting of (Kitamura and Stutzer, 1997) ($\gamma = -1$) belong to the GEL family of estimators while (Newey and Smith, 2004) show that it is also the case for the continuous updated estimator (CUE) of (Hansen, Heaton and Yaron, 1996) ($\gamma = 1$). They all have in common that we can express their dual problem as :

$$\hat{\theta}_{gel} = \arg\min_{\theta\in\Theta} \left[\max_{\lambda\in\Lambda_n} \frac{1}{n} \sum_{t=1}^{n} \rho\left(\lambda' g(x_t;\theta)\right) \right], \tag{1.5}$$

where $\rho(v)$ is a strictly concave function that depends on $h_n(p_t)$ and is normalized so that $\rho'(0) = \rho''(0) = -1$. We can show that $\rho(v) = \ln(1-v)$ corresponds to EL, $\rho(v) = -\exp(v)$ to ET and to CUE if $\rho(v)$ is quadratic. We assume that Θ is a compact set and λ, which is the $q \times 1$ vector representing the Lagrange multiplier associated with the constraint (1.3), belongs to $\Lambda_n = \{\lambda : \lambda' g(x_t;\theta) \in \mathcal{D} \ \forall \ x_t\}$, where $\mathcal{D}$ is the domain of $\rho(v)$.

(Newey and Smith, 2004) and (Anatolyev, 2005) show that the EL estimator has a lower second order asymptotic bias than ET and CUE and that its bias corrected version is higher order efficient. This performance is, to some extent, due to the fact that EL's estimators of the Jacobian and second moment matrices, as opposed to the other GEL methods, are based on the implied probabilities which carry more information than $1/n$ (see (Antoine, Bonnal and Renault, 2007)). However, because of the non negativity constraint that we need to impose on these implied probabilities, ET offers a natural way to meet this requirement, which makes it numerically more stable than EL especially in presence of model misspecification. In response to this, (Schennach, 2007) combines

2. This is a general power-divergence statistics that encompasses different tests, including the empirical likelihood ratio test (see (Baggerly, 1998)).

ET and EL in a method called the exponentially tilted empirical likelihood (ETEL). This method shares the same second order properties of EL and the stability of ET in presence of model misspecification. Although it does not belong to the GEL family, we offer below a brief discussion because its computational stability is appealing especially in the case of a continuum of conditions.

We can easily verify the equivalence of the primal and dual problems by showing that they share the same following first order conditions :

$$\sum_{t=1}^{n} p_t g(x_t, \tau_i; \theta) = 0 \quad \forall i = 1, \cdots, q,$$

$$\sum_{t=1}^{n} p_t \lambda' \left(\frac{\partial g(x_t; \theta)}{\partial \theta} \right) = 0,$$

with

$$p_t = \frac{1}{n} \rho' \left(\lambda' g(x_t; \theta) \right).$$

The following asymptotic properties of GEL are proved by (Newey and Smith, 2004). The assumptions that are required for consistency of $\hat{\lambda}_{gel}$ and $\hat{\theta}_{gel}$ are the same as for GMM plus some additional ones associated with the Lagrange multipliers. There is an identification assumption for θ_0, some boundedness conditions on higher moments of $\|g(x_t; \theta)\|$ and a non-singularity assumption of the covariance matrix Ω. The latter guarantees that the numerical solution is unique and computable at least with probability approaching one. They show that under these assumptions, $\hat{\theta}_{gel} \xrightarrow{P} \theta_0$ and $\hat{\lambda}_{gel} \xrightarrow{P} 0$. Furthermore, under some additional assumptions which allow to apply a central limit theorem, the estimators are asymptotically distributed as

$$\sqrt{n}(\hat{\theta}_{gel} - \theta_0) \xrightarrow{d} N(0, (G'\Omega^{-1}G)^{-1})$$

and

$$\sqrt{n}\hat{\lambda}_{gel} \xrightarrow{d} N\left\{0, \Omega^{-1} - \Omega^{-1}G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega^{-1}\right\},$$

where $G = E(\partial g(X; \theta_0)/\partial\theta)$ and Ω is the asymptotic covariance matrix of $n^{-1/2}\sum_t g(x_t; \theta_0)$. Therefore, GEL shares the same asymptotic properties as GMM.

1.3 CGEL

In order to illustrate how we can extend the previous results to the case in which the moment conditions are defined on a continuum, and how it affects the stability and existence of the solution, we start by assuming that τ_i, for $i = 1, ..., q$, lies in the fixed interval $[a, b]$ and is defined as $\tau_i = a + i(b-a)/q$. The space in which τ_i lies is therefore $\mathcal{T}_q \subseteq \mathbb{Q} \cap [a, b]$. As q goes to infinity, the space converges to $\mathcal{T}_\infty \equiv \mathcal{T} = [a, b]$. This representation makes sense only if τ_i is an argument of the function defining the moment conditions. For example, if we want to estimate the linear model $y_t = W_t\theta + \epsilon_t$, where $W_t = e^{-x_t^2} + u_t$ and $Cov(\epsilon_t, u_t) \neq 0$, as in (Carrasco, 2010) with $p = 1$, we can base our estimation on the following moment conditions (In the following, these three expressions will be used interchangeably : $g(x_t;\theta)$, $g_t(\theta)$ or g_t, when we refer to the function from $\mathcal{T}$ to $\mathbb{C}$. The form $g(x_t, \tau; \theta)$ or $g_t(\tau;\theta)$ will be used only when we need to specify the moment condition.) :

$$E\Big[g_t(\tau_j;\theta)\Big] = E\Big[(y_t - W_t\theta)e^{(i\tau_j x_t)}\Big] = 0, \text{ for } j = 1, ..., q \text{ and } \tau_j \in \mathcal{T}_q,$$

where the points τ_j are chosen arbitrarily unless some selection methods are used (see (Carrasco, 2010)). In the simulation below, we estimate the parameters of a stable distribution using the marginal characteristic function for which the same kind of discretization can be applied.

The objective is to define CGEL estimators as the solution to the GEL optimization problem when q goes to infinity. Therefore, we assume that the function $g_t(\tau_i;\theta)$ belongs to an Hilbert space $\mathcal{H}_q$ with inner products defined as $< g, f >_q = \sum_{i=1}^q g(\tau_i)f(\tau_i)\pi(\tau_i)\Delta\tau_i$, where $\pi(\tau)$ is an integrating density as the one introduced by (Carrasco et al., 2007). For GEL, the integrating density is the one from the uniform distribution and $\Delta\tau_i = \Delta\tau_{i-1}$, so that $< g, f >_q$ is the Euclidean inner product. If all $f()$ and $g()$ in $\mathcal{H}_q$ are square-integrable, then $\mathcal{H}_q$ converges to the Hilbert space $\mathcal{H}$ of square-integrable functions on $[a, b]$ with inner product $< f, g >= \int_a^b f(\tau)g(\tau)\pi(\tau)d\tau$.

This structure[3] implies that the estimators of GEL is defined by the primal problem :

$$\{\hat{\theta}_q, \hat{\lambda}_q, \hat{\mu}_q, \hat{p}_{qt}\} = \arg\min_{\theta,\mu,p_t,\lambda} \mathcal{L} = \sum_{t=1}^{n} h_n(p_t) + \left\langle \lambda, \sum_{t=1}^{n} p_t g_t(\theta) \right\rangle_q + \mu\left(\sum_{t=1}^{n} p_t - 1\right), \quad (1.6)$$

where the subscript q means that the estimates are based on q moment conditions. This problem converges to the following primal problem of CGEL when q goes to infinity :

$$\{\hat{\theta}, \hat{\lambda}, \hat{\mu}, \hat{p}_t\} = \arg\min_{\theta,\mu,p_t,\lambda} \mathcal{L} = \sum_{t=1}^{n} h_n(p_t) + \sum_{t=1}^{n} p_t \int_a^b \lambda(\tau) g_t(\tau;\theta)\pi(\tau)d\tau + \mu\left(\sum_{t=1}^{n} p_t - 1\right), \quad (1.7)$$

In the same way, the dual of GEL and CGEL are respectively :

$$\hat{\theta}_q = \arg\min_{\theta\in\Theta}\left[\max_{\lambda\in\Lambda_{q,n}} P_q(\lambda,\theta) = \frac{1}{n}\sum_{t=1}^{n}\rho\left(<\lambda, g_t(\theta)>_q\right)\right] \quad (1.8)$$

and

$$\hat{\theta} = \arg\min_{\theta\in\Theta}\left[\max_{\lambda\in\Lambda_n} P(\lambda,\theta) = \frac{1}{n}\sum_{t=1}^{n}\rho\left(\int_a^b \lambda(\tau) g_t(\tau;\theta)\pi(\tau)d\tau\right)\right], \quad (1.9)$$

where $\Lambda_{q,n} = \{\lambda :< \lambda, g(x_t,\theta) >_q \in \mathcal{D}\ \forall\ x_t\}$ and $\Lambda_n = \{\lambda : \int_a^b \lambda(\tau)g(x_t,\tau,\theta)\pi(\tau)d\tau \in \mathcal{D}\ \forall\ x_t\}$. Notice that the continuous updated GMM for a continuum (CCUE) is not a special case of CGEL. When $\rho(v)$ is quadratic, we will refer to the Euclidean Empirical Likelihood for a continuum (CEEL), which is based on the EEL method of (Antoine, Bonnal and Renault, 2007). In the EEL method, we minimize the euclidean distance between $1/n$ and p_t in the primal problem. We only have asymptotic equivalence between CCUE and CEEL as opposed to the case in which the number of conditions is finite (see Appendix A.4.1).

It follows that we can obtain the solution of GEL by solving the following first

3. Notice that we focus on a continuum of conditions. However, the case of a countably infinite set of conditions is implicit in this setup if we properly define $\pi(\tau)$. A typical element of the space would be $f(\tau_i)$ for $i \in \mathbb{N}$ and we would get $< f, g >= \sum_i f(\tau_i)g(\tau_i)$. Some times, in the case of a finite number of conditions, GEL fails to produce results due to poorly conditioned first order conditions. In such cases, we could use the CGEL setup to get a more stable solution. We would simply need to set $\pi(\tau_i) = 0$ for $i > q$.

order conditions :

$$\sum_{t=1}^{n} \frac{1}{n}\rho' \left(< \lambda, g_t(\theta) >_q\right) g_t(\tau_i;\theta) = 0 \quad \forall i = 1,\cdots,q, \tag{1.10}$$

$$\sum_{t=1}^{n} \frac{1}{n}\rho' \left(< \lambda, g_t(\theta) >_q\right) \left\langle \lambda, \left(\frac{\partial g_t(\theta)}{\partial\theta}\right)\right\rangle_q = 0, \tag{1.11}$$

For a given λ, solving the system of p equations (1.11) is not an issue, even if q goes to infinity, as long as the system is not singular. The problem with GEL arises when we try solving conditions (1.10) for a given θ. As q increases for a given n, the system becomes more and more poorly conditioned. Indeed, based on the Taylor expansion, we can obtain the solution by using this iterative procedure :

$$\lambda_l = \lambda_{l-1} - \left\langle \left[\frac{1}{n}\sum_{t=1}^{n} \rho''\left(\langle\lambda_{l-1}, g_t\rangle_q\right) g_t g_t'\right]^{-1}, \left[\frac{1}{n}\sum_{t=1}^{n} \rho'\left(\langle\lambda_{l-1}, g_t\rangle_q\right) g_t\right]\right\rangle_q,$$

starting with $\lambda_0 = 0$. The second term of the right hand side of this procedure is the solution to a system of q linear equations. As q increases, the system becomes singular. As a result, $\lambda(\theta)$ becomes non-computable. It is like trying to estimate a model using too many instruments. Therefore, the limiting case of equation (1.10), which implies the following continuum of conditions

$$\sum_{t=1}^{n} \frac{1}{n}\rho'\left(\int_a^b \lambda(\tau) g_t(\tau;\theta)\pi(\tau)d\tau\right) g_t(\tau,\theta) = 0 \quad \forall \tau \in [a,b], \tag{1.12}$$

is ill-posed in the sense that we cannot find a unique solution without imposing a penalty on its instability (Appendix A.4.2 shows the ill-posedness of equation (1.12) for the CEEL case even if the right-hand side is not random as required for linear ill-posed problems). The problem arises whether we are dealing with a continuum of conditions, an infinite number of countable conditions or simply when there are a finite but large number of conditions. The last two cases constitute special cases of CGEL simply by selecting the proper integrating density as suggested by (Carrasco, 2010) for CGMM.

The ill-posedness aspect of GEL in such cases was implicit in the empirical likelihood version of (Kitamura, Tripathi and Ahn, 2004) and (Donald, Imbens and Newey,

2003) since they both require a smoothing parameter. In the first paper, they use a bandwidth parameter while in the second they restrict the number of instruments which also constitute a way of smoothing the problem. (Carrasco, 2010) deals with the problem by using CGMM, which imposes a Tikhonov's type of penalization in order to make the system solvable (see appendix A.1.1 for an overview of the CGMM method). It can be seen as a method which automatically selects the most influential moment conditions among the whole set, much like a principal component procedure. A similar approach can be used to solve the ill-posedness of CGEL. Notice, however, that CGMM requires a penalization in order to define its objective function, while CGEL requires it in order to solve it. It is like a nonlinear ridge regression in the sense that the problem, which consists in minimizing $\sum_t u_t^2$ with $u_t = y_t - x(\beta)$, is well defined, but we cannot obtain a stable solution because the columns of $X(\beta) = dx(\beta)/d\beta$ are nearly collinear. In this case, as suggested by (Dagenais, 1983), we can apply the ridge regression technique to the iterative procedure and substitutes the poorly conditioned matrix $X(\beta)'X(\beta)$ by $[X(\beta)'X(\beta) + \alpha_n I]$ for some $\alpha_n > 0$. Even in Theorem 3.1 of (Newey and Smith, 2004), the uniqueness and existence of the solution are only satisfied with probability approaching one. It may not be the case in small samples, in which case a penalization as in ridge regressions would be required to obtain a stable solution.

In the rest of the paper, we use the notation from the literature on nonlinear and linear operators as the articles from which the numerical procedures used below come from. This also offers a nice and compact way to present the results, especially when working in function spaces. For example, we can rewrite the problem of ill-posedness using the notation of (Seidman and Vogel, 1989). Their definition of ill-posedness is much like the one we are facing here. Indeed, we can present the first order condition associated with the Lagrange multiplier as the problem of solving the nonlinear operator equation $L(\lambda) = 0$, where

$$L(\lambda) = E\left[\rho'\left(\int_a^b \lambda(\tau) g_t(\tau;\theta_0)\pi(\tau)d\tau\right) g_t(\theta_0)\right],$$

using the disturbed system $\hat{L}(\lambda) = 0$ defined by equation (1.12) in which θ_0 is replaced by an estimate and $E()$ by the sample mean. The solution is $\lambda = 0$ and is unique given some identification assumptions. It is however ill-posed in the sense that we cannot compute a stable and unique solution to the disturbed system without smoothing it. It is ill-posed even if the right hand side is not random as required by linear ill-posed problems. In the nonlinear case, the ill-posedness appears in the iterative procedure in which a linear ill-posed problem is solved at each iteration. Therefore, in what follows, we regard CGEL as a nonlinear ill-posed problem in function space, which implies that we need a regularized method for computing the solution.

When it is clear, we will use the linear operator notation instead of explicit integrations or inner products. For example, if we have two square-integrable functions $f(x), g(x) : \mathcal{T} \rightarrow \mathbb{C}$, we will write $fg = \int_{\mathcal{T}} f(x)g(x)\pi(x)dx$ in which case, f is an operator from $L^2(\pi) \rightarrow \mathbb{C}$. If furthermore we have a function $A(x,y) : \mathcal{T}^2 \rightarrow \mathbb{C}$, we will write $(Af)(x) = \int_{\mathcal{T}} A(x,y)f(y)\pi(y)dy$, in which case A is an operator from $L^2(\pi)$ to $L^2(\pi)$ with kernel $A(x,y)$. Using this notation, we can rewrite equation (1.9) as $P(\lambda,\theta) \equiv (1/n)\sum_{t=1}^n \rho(\lambda g_t)$, where λ is presented as a linear operator from $L^2(\pi)$ to $\mathbb{C}$. Solving the saddle point problem using this notation gives the following first order conditions[4] :

$$F_{n1}(\lambda) \equiv \frac{1}{n}\sum_{t=1}^n \rho'(\lambda g_t)g_t = 0 \tag{1.13}$$

and

$$F_{n2}(\theta) \equiv \frac{1}{n}\sum_{t=1}^n \rho'(\lambda g_t)\left[\lambda G_t\right] = 0, \tag{1.14}$$

where

$$G_t \equiv \frac{\partial g_t(\theta)}{\partial \theta}.$$

where F_{n1} is written as a function only of λ to emphasize the fact that it is the system that produces the solution $\lambda(\theta)$ for a given θ and inversely for F_{n2}. $F_{n2}(\theta)$ is the derivative of $P(\lambda,\theta)$ with respect of θ. It is therefore a vector with the same dimension as θ, which is $p \times 1$. However, $F_{n1}(\lambda)$ is the Fréchet derivative of $P(\lambda,\theta)$ with respect to the function λ. It is an operator from $L^2(\pi) \rightarrow \mathbb{C}$. That is, if $f \in L^2(\pi)$,

4. For a good review of optimization in function spaces, see (Luenberger, 1997)

then $F_{n1}(\lambda)f = \int_{\mathcal{T}} F_{n1}(\lambda,\tau)f(\tau)\pi(\tau)d\tau$. Since the Fréchet derivative is a generalization of the conventional derivatives for any vector space, we will say that $F_{n2}()$ is also a Fréchet derivative. It is an operator from $\mathbb{R}^p \to \mathbb{C}$. As a result, for a $p \times 1$ vector y, $F_{n2}(\theta)y = \sum_{i=1}^{p} F_{n2}(\theta)_i y_i$. Finally, G_t is a $p \times 1$ vector of square-integrable functions. If $h \in L^2(\pi)$, $G_t h = \int_{\mathcal{T}} h(\tau)G_t(\tau)\pi(\tau)d\tau$ while if $h \in \mathbb{R}^p$ then $G_t h = \sum_{i=1}^{p} h_i G_{ti}$ (See appendix A.1.2 from an overview of Fréchet derivatives.).

If we consider a linear ill-posed problem such as the Fredholm integral equation of the first kind $Kg = y$, we can obtain a stable and unique solution by using a Tikhonov approach which consists in solving the following (See (Carrasco, Florens and Renault, 2007) for details on how to solve linear ill-posed problems) :

$$\min_g \|Kg - y\|^2 + \alpha_n \|g\|^2,$$

where the second term imposes a penalty on the instability of the solution. The regularization parameter α_n determines the degree of penalty. We need to choose it carefully because if it is too small, the solution is more accurate but less stable and inversely if it is too large. The system $Kg = y$ is then replaced by the first order condition of the minimization problem which is :

$$K(Kg - y) + \alpha_n g = 0.$$

The system is now well-posed, given certain regularity conditions, and gives the solution $g_\alpha = (K^2 + \alpha_n I)^{-1}Ky$, where I is the identity operator and $(K^2 + \alpha_n I)^{-1}K$ is a generalized inverse of K. The ill-posedness is caused by the fact that K is a compact bounded operator and is not invertible.

When we deal with a nonlinear system such as $F(g) = y$, ill-posedness is characterized by the non-invertibility of the Fréchet derivative operator. The Fréchet derivative of $F_{n1}(\lambda)$ is an operator with kernel defined as :

$$DF_{n1}(\lambda,\tau_1,\tau_2) = \frac{1}{n}\sum_{t=1}^{n} \rho''(\lambda g_t)g_t(\tau_1)g_t(\tau_2).$$

Instead of $F_{n1}(\lambda) = 0$, we then need to solve the following minimization problem :

$$\min_\lambda \|F_{n1}(\lambda)\|^2 + \alpha_n \|\lambda\|^2.$$

In general, the penalty function can be any non-negative function satisfying certain conditions. For example, the Sobolev norm satisfies the conditions required. However, the choice of the penalty function affects only the speed of convergence of the numerical algorithms used. It does not affect the speed of convergence of the estimator to its true value as n goes to infinity, as long as we assume that the numerical solution has been reached. The choice made here is to simplify the presentation.

The first order condition of the minimization problem is

$$DF_{n1}(\lambda)F_{n1}(\lambda) + \alpha_n\lambda = 0, \tag{1.15}$$

which cannot be solved analytically, as for the linear case, because of the nonlinearity. We present the numerical method that we use for solving this system in section 1.4. The feasible CGEL is therefore defined as the vector $\hat{\theta}$ and the function $\hat{\lambda}$ which solve equations (1.14) and (1.15).

We need some assumptions for deriving the asymptotic properties of CGEL. The first set is similar to Assumption A.2 of (Carrasco et al., 2007) but for i.i.d observations.

Assumption 1. *a) The observations $\{x_1, x_2, \cdots, x_n\}$ are i.i.d, b) $L^2(\pi)$ is the Hilbert space of square integrable complex functions in which the inner product $< f, g >$ is defined as $\int f(\tau)g(\tau)\pi(\tau)d\tau$, where $\pi(\tau)$ is a density function which is absolutely continuous with respect to the Lebesgue measure, c) $g(x_t, \tau; \theta) \in L^2(\pi)$, $\forall x_t$ and θ, and d) $g(x_t, \tau, \theta)$ is continuously differentiable with respect to θ for all τ and x_t.*

The second set is similar to Assumption 1 of (Newey and Smith, 2004).

Assumption 2. *a) $\theta_0 \in \Theta$ is the unique solution to $E^{P_0}g(X;\theta) = 0$, where Θ is a compact subset of $\mathbb{R}^p$, and b) $E^{P_0}[\sup_\theta \|g(X;\theta)\|^\nu] < \infty$ for some $\nu > 2$*

The space in which τ belongs is defined by $\mathcal{T}$ instead of $[a, b]$. For example, if the moment conditions are based on the characteristic function as in (Carrasco et al., 2007), $\mathcal{T}$ is either $\mathbb{R}^2$ or $\mathbb{R}$. It is $[0,\pi]^s$, for some integer s, if the conditions are based on a spectral density as in (Berkowitz, 2001).

Assumption 1 and 2 imply that :

$$\sqrt{n}\sum_{t=1}^{n} g(x_t,\theta_0) \equiv n^{1/2}\bar{g}(\theta_0) \xrightarrow{L} N(0,K),$$

where K is a covariance operator with the following kernel [5] :

$$k(\tau_1,\tau_2) = E^{P_0}\left[g(X,\tau_1;\theta_0)g(X,\tau_2;\theta_0)\right].$$

The following assumption replaces the full rank properties of Ω imposed by (Newey and Smith, 2004). It implies that the solution of $Kf = g$ exists and is unique as long as $g \in R(K)$, where $R(K)$ is the range of K. It also implies that K can be expressed as the limit of a sequence of linear operators K_n, which is important when K needs to be estimated.

Assumption 3. *a) K is a Hilbert-Schmidt operator, which implies that it is bounded and compact. b) K has only strictly positive eigenvalues. This assumption implies that the null space of K, N(K), is* $\{0\}$*. c) The skewness operator* S *with kernel*

$$s(\tau_1,\tau_2,\tau_3) = E[g_t(\tau_1;\theta_0)g_t(\tau_2;\theta_0)g_t(\tau_3;\theta_0)]$$

is bounded and compact.

The following conditions on $G_t \equiv \partial g_t/\partial\theta$ are also required for asymptotic normality and the boundness of $\|E^{P_0}[g_t]\|_3$ guarantees that the remainder term of the Taylor expansion of the first order condition vanishes as n goes to infinity.

5. For a good review of linear operators such as covariance operators applied to econometrics, see (Carrasco, Florens and Renault, 2007)

Assumption 4. *a) rank*$(G_t) = p\ \forall t$, *b)* $E[\sup_\theta \|G_t\|] < \infty$, *c)* $E(g(\theta)) \in \mathcal{D}(K^{-1})$ *for all* θ *on a neighborhood of* θ_0 *and d)* $E^{P_0}\|g_t(\theta)\|_3 < \infty$ *for all* θ*.*

The last set of assumptions defines the properties of $\rho(v)$ that we need for the asymptotic theory.

Assumption 5. *a)* $\rho(v)$ *is strictly concave and twice continuously differentiable. b)* $\rho''(v)$ *is Lipschitz continuous at least in the neighborhood of 0, c)* $\rho'''(v)$ *is continuous in the neighborhood of 0 and d)* $\rho(v)$ *is normalized in such way that* $\rho'(0) = \rho''(0) = \rho'''(0) = -1$

These requirements are satisfied by $\rho(v)$ associated with CEL, CET and CEEL. Assumption 4 b) could be replaced by $\rho''(v)$ being everywhere differentiable since it implies Lipschitz continuity. But it is not necessary. This condition is important in order for the regularized Gauss-Newton method presented in the next section to be locally convergent as explained by (Blaschke, Neubauer and Scherzer, 1997). The proofs of the following theorems can be found in the appendix.

Theorem 1. *If assumptions 1, 2, 3 and 4 are satisfied, then* $\hat{\theta}_n \in \Theta$ *and* $\hat{\lambda}_n \in \Lambda_n$*, which are the solutions to equations (1.14) and (1.15) converge in probability to* θ_0 *and uniformly to 0 respectively as n goes to infinity,* α_n *goes to zero and* $n\alpha_n$ *goes to infinity. Furthermore, the rate of convergence is* $n^{-1/2}$*.*

Theorem 2. *If assumptions 1, to 5 are satisfied, then :*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{L} N(0, [GK^{-1}G]^{-1})$$

and

$$\sqrt{n}\hat{\lambda} \xrightarrow{L} N\left(0, \left[K^{-1} - K^{-1}G(GK^{-1}G)^{-1}GK^{-1}\right]\right)$$

as n goes to infinity, α_n goes to zero and $n\alpha_n^3$ goes to infinity.

In both theorems, α_n needs to converge to 0 not too quickly because it is necessary for the system to stay stable as n increases. As a result, CGEL shares the same asymptotic properties as CGMM.

We conclude this section by defining the exponentially tilted empirical likelihood method of (Schennach, 2007) for a continuum (CETEL). It is the vector $\hat{\theta}_{cetel}$ and the function $\hat{\lambda}_{cetel}$ which solve the following conditions :

$$\frac{1}{n}\sum_{t=1}^{n}\rho'_{EL}(\lambda g_t)\left[\lambda G_t\right] = 0 \tag{1.16}$$

and

$$DF_{nET}(\lambda)F_{nET}(\lambda) + \alpha_n\lambda = 0, \tag{1.17}$$

where $\rho_{EL}(v) = \log(1-v)$ and F_{nET} is F_{n1} with $\rho(v) = -e^v$. Since the proofs can easily be derived from the ones from theorems 1 and 2, the asymptotic results are expressed in the following corollary.

Corollary 1. *If the assumptions of the theorems 1 and 2 are satisfied, the CETEL estimator shares the same asymptotic properties as CGEL.*

1.4 Estimation procedures

In this section, we present two different estimation procedures which compete in terms of computation time and we express them in matrix form as done by (Carrasco et al., 2007) for CGMM. The first is based on the first order Taylor approximation of the solution $\lambda(\theta)$, while the second solves equation (1.15) using an iterative procedure. For the GEL case, (Guggenberger and Hahn, 2005) offer an argument for using what they call the two step empirical likelihood estimator, which is nothing more than the solution obtained from a Newton algorithm after two iterations. They show that increasing the number of iterations does not affect the third order asymptotic bias. Our first procedure approximates the solution $\lambda(\theta)$ which is then used by the numerical optimizer to compute $\hat{\theta}$. Because the second procedure is computationally demanding, it may represent a good alternative. We analyze the properties of both procedures in section 1.4 through a numerical experiment.

1.4.1 Taylor approximation and singular value decomposition

The first method follows (Carrasco and Florens, 2000) who present the singular value decomposition as a way of solving linear ill-posed problems (see also (Groetsch, 1993)). The ill-posedness arises in the first order Taylor approximation of the solution $\lambda(\theta)$ of equation (1.15), which implies (see appendix A.1.2) :

$$\hat{K}\hat{\lambda} = -\bar{g}(\theta) + o_p(1),$$

where $\hat{K}$ is the estimated covariance operator of g_t with kernel

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

Notice that this approximation is the exact solution of CEEL because in this case, $\rho(v)$ is quadratic and then $F_{n1}(\lambda)$ is linear.

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

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

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

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

where $\varphi \in N(K)$. Since $N(K) = \{0\}$, if $y \in R(K)$, the unique solution is :

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

We can obtain a stable solution from the following regularized system :

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

which implies the following solution :

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

Therefore, the solution requires an infinite number of eigenfunctions. However, when K is unknown and is replaced by $\hat{K}$, the solution is much simpler. As (Carrasco and Florens, 2000) show, the dimension of $R(\hat{K})$ is finite :

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

Therefore, $R(\hat{K})$ is spanned by $\{g(x_1; \theta), \cdots, g(x_n; \theta)\}$. It follows that the singular system of $\hat{K}$ is composed of n eigenfunctions $\nu_i^{(n)}$ and n singular values $\mu_i^{(n)}$ (they both depend on $\hat{\theta}$, but we do not write it explicitly to simplify the presentation). We can extend the previous result to our case and show that the regularized solution to $\hat{K}\tilde{\lambda} = -\bar{g}(\theta)$ is

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

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

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

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

1. We construct the $n \times n$ matrix C
2. We compute the eigenvectors β_i and eigenvalues $\mu_i^{(n)}$ for $i = 1, ..., n$
3. We compute the eigenfunctions of $\hat{K}$ as follows :
$$\nu_i^{(n)} = \frac{1}{n}\sum_{j=1}^{n} \beta_{ji} g(\theta, x_j) \qquad i = 1, ..., n$$
4. We compute $\tilde{\lambda}$:
$$\tilde{\lambda}(\theta) = -\sum_{i=1}^{n} \left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2} + \alpha_n} \right) < \bar{g}(\theta), \nu_i^{(n)} > \nu_i^{(n)}$$
5. We estimate θ_0 by solving the following problem :
$$\tilde{\theta} = arg \min_{\theta} \frac{1}{n} \sum_{t=1}^{n} \rho\Big(\tilde{\lambda}(\theta) g(x_t, \theta)\Big). \tag{1.18}$$

Because the solution to CGEL includes also an estimate of the probability distribution p_t for $t = 1, \cdots, n$ with $\sum_t p_t = 1$, which depends on λ, and that we did not obtain the exact solution to equation (1.15), we may, if we intend for example to use the implied probabilities to obtain efficient estimates of higher moments of $g_t(\theta)$, have to normalize $p_t(\tilde{\lambda})$ as follows
$$\tilde{p}_t = \frac{p_t(\tilde{\lambda})}{\sum_{t=1}^{n} p_t(\tilde{\lambda})}, \tag{1.19}$$
where
$$p_t(\tilde{\lambda}) = \frac{1}{n}\rho'\Big(\tilde{\lambda} g(x_t; \tilde{\theta})\Big).$$
For the case in which $\rho(v)$ is quadratic, which does not guarantee the non-negativity of $p_t(\tilde{\lambda})$, the latter can be transformed according to (Antoine, Bonnal and Renault, 2007).

In order to apply this method, it is convenient to rewrite the objective function in matrix notation as in (Carrasco et al., 2007). Let us define the $n \times m$ matrix β which contains the eigenvectors of C associated with its m eigenvalues different from 0, and the $m \times m$ diagonal matrix D with typical element D_{jj} :
$$D_{jj} = \frac{\mu_i^{(n)}}{\mu_i^{(n)^2} + \alpha_n}.$$

The following optimization problem is equivalent to the one given by equation (1.18) :

$$\tilde{\theta} = arg \min_{\theta} \frac{1}{n} \sum_{t=1}^{n} \rho\Big(-\frac{1}{n} \iota' C[\beta D \beta'] C_{\bullet t} \Big), \tag{1.20}$$

where ι is an $n \times 1$ vector of ones and $C_{\bullet t}$ is the t^{th} column of C (see Appendix A.3.1 for the proof).

In practice, we need to select a tolerance level in order to determine whether the eigenvalues are considered to be zero or not. Indeed, none of them will be exactly equal to zero. However, the presence of α_n in the denominator of D_{jj} makes it possible to choose $m = n$.

1.4.2 Solving a nonlinear operator equation

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

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

which converges to the fix point $g(x) = x$, where x is the solution to the initial problem. The simplest method sets $g(x) = x + \omega f(x)$. If the algorithm converges, then we have $f(x) = 0$ as required. However, this method, if it converges, is slow if we do not select a proper ω. The Newton method sets $\omega = -[f'(x)]^{-1}$ so that the algorithm becomes :

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

In order for this method to work, the inverse of the first derivative needs to be bounded. When $f'(x_{i-1})^{-1}$ is not bounded, it has to be replaced by a generalized inverse. This is similar to the problem we are facing in this section but with the exception that the solution x is a function from $L^2(\pi)$.

In the case of CGEL, we need to solve equation(1.15) which we rewrite as follows ($F_{n1}()$ has been replaced by $F()$ for simplicity) :

$$F(\lambda) \equiv \frac{1}{n} \sum_{t=1}^{n} \rho'(\lambda g_t) g(x_t; \theta) = 0. \tag{1.21}$$

So we need the solution to the general nonlinear operator equation $F(\lambda) = 0$, where F is a nonlinear operator from $L^2(\pi)$ to $L^2(\pi)$. As discussed in Section 1.3, this problem needs to be regularized in order to compute a stable solution. The second estimation procedure is therefore an iterative algorithm which converges to the solution of :

$$\min_{\lambda} \|F(\lambda)\|^2 + \alpha_n \|\lambda - \lambda_0\|^2,$$

so that $F(\lambda)$ is close to 0 and the solution $\hat{\lambda}$ sufficiently smooth. The value of α_n determines how important we consider the smoothness of the solution. If it is set too high, the solution $\hat{\lambda}$ would most likely be a constant (i.e. $\hat{\lambda}(\tau_1) = \hat{\lambda}(\tau_2)\ \forall \tau_1, \tau_2$) with $F(\hat{\lambda})$ not too close to zero, since the first term would become negligible. Conversely, a small α_n would create an unstable solution for which $F(\hat{\lambda})$ is almost zero. It is the same tradeoff that we face when solving linear ill-posed problems.

There are many algorithms that reflect this tradeoff. (Ramm, 2004b) and (Ramm, 2004a) present the continuous version of such methods and give the conditions under which they converge to the solution. The discrete algorithm presented by (Airayetpan and Ramm, 2000) is a regularized Newton method. It is a Newton method applied to a transformed equation. Indeed, the Newton method solves $F(\lambda) = 0$ with the algorithm $\lambda_i = \lambda_{i-1} - DF(\lambda_{i-1})^{-1} F(\lambda_{i-1})$ while the regularized Newton method solves $F(\lambda) + \alpha_n \lambda = 0$ which implies the following algorithm :

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

where ω_i is a sequence that we need to choose to control the speed of convergence. Another method which uses a regularized inverse which is closer to the one used for the linear case has been analyzed by (Jin, 2000). It is a regularized Gauss-Newton method which is defined as follows :

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

where the initial value λ_0 has been set equal to its asymptotic value, 0. It is the usual starting value for λ when the parameters are estimated by GEL (see for example (Guggenberger, 2008)). If the algorithm converges, the condition of equation (1.15) is satisfied. (Blaschke, Neubauer and Scherzer, 1997) show that the conditions that we impose are sufficient for the convergence of the algorithm.

In order to apply this algorithm, we will present it in matrix form. Because λ enters equation (1.15) only through $\lambda g_t(\theta) = \int_{\mathcal{T}} \lambda(\tau) g_t(\tau;\theta)\pi(\tau)d\tau$, we only need to solve for λg_t. Therefore we can obtain the result from the following iterative procedure :

$$g_t\lambda_i = g_t\lambda_{i-1} - g_t\left[\alpha_n I + DF(\lambda_{i-1})^2\right]^{-1}\left\{DF(\lambda_{i-1})F(\lambda_{i-1}) + \alpha_n\lambda_{i-1}\right\}. \tag{1.22}$$

Let us define the $n \times n$ diagonal matrix V as :

$$V_{tt} = \rho''(\lambda g_t),$$

the $n \times 1$ vector P as :

$$P_t = \rho'(\lambda g_t),$$

and the $n \times n$ matrix C as usual. The following theorem is demonstrated in Appendix A.3.2.

Theorem 3. *If the conditions of theorem 1 and 2 are satisfied, than CGEL, which is defined by the conditions of equations (1.14) and (1.15), is equivalent to the following procedure : We first iterate the following until convergence :*

$$[\lambda_i g] = \left\{[CV]^2 + \alpha_n I\right\}^{-1}\left\{[CV]^2[\lambda_{i-1}g] - [CV][CP]\right\}, \tag{1.23}$$

with the initial value :

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

where ι is an $n \times 1$ vector of ones. We then solve the following minimization problem :

$$\frac{1}{n}\sum_{t=1}^{n}\rho(\widehat{\lambda g_t}(\theta)),$$

where $\widehat{\lambda g_t}(\theta)$ is the value to which has converged the algorithm (2.1).

1.5 Over-identification tests for C-GEL

GEL offers three ways of testing the validity of the moment conditions $E^{P_0}(g_t(\theta_0)) = 0$. (Smith, 2004) summarizes them and shows that they are first order equivalent and asymptotically chi-square with $(n-q)$ degrees of freedom, where q is the number of moment conditions. The first is the J-test developed by (Hansen, 1982) which is based on the GMM criterion :

$$J_{gmm} = n\bar{g}(\hat{\theta}_{gmm})'\hat{K}^{-1}\bar{g}(\hat{\theta}_{gmm}) \equiv \|\hat{K}^{-1/2}\sqrt{n}\bar{g}(\hat{\theta}_{gmm})\|^2.$$

In the context of CGEL, two problems arise from this test. First, we need to replace $\hat{K}^{-1}$ by the generalized inverse $(\hat{K}^{\alpha_n})^{-1}$, and second, the test diverges since the number of moment conditions is infinite. (Carrasco and Florens, 2000) offer a normalized version of this test which is asymptotically $N(0,1)$. We can apply the same normalization for CGEL since it is asymptotically equivalent to CGMM. However, the tests will differ in finite sample since CGEL evaluates $\hat{K}^{\alpha_n}$ at $\hat{\theta}$ while CGMM uses a first step estimate.

The test is based on the singular value representation of the CGMM criterion described in the previous section :

$$\|(\hat{K}^{\alpha_n})^{-1/2}\sqrt{n}\bar{g}(\hat{\theta})\|^2 = \sum_{i=1}^{n}\left(\frac{\mu_i^{(n)}}{\mu_i^{(n)2}+\alpha_n}\right) < \bar{g}(\hat{\theta}), \phi_i^{(n)} >^2,$$

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

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

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

Then the first test is defined as :

$$\tilde{J} = \frac{\|(\hat{K}^{\alpha_n})^{-1/2}\sqrt{n}\bar{g}(\hat{\theta})\|^2 - p_n}{\sqrt{q_n}} \Longrightarrow N(0,1),$$

under the null that the over-identifying moment conditions are satisfied. The proof is given by (Carrasco and Florens, 2000).

The second is the Lagrange multiplier test (LM). In the dual problem, λ is the Lagrange multiplier associated with the sample moment conditions $\sum_t p_t g_t(\theta) = 0$. It should therefore be zero if the constraint is not binding. For GEL, the test is defined as follows :

$$LM = n\hat{\lambda}_{gel}'\hat{K}\hat{\lambda}_{gel}.$$

For CGEL, the same normalization is required. The second test is therefore [6] :

$$\widetilde{LM} = \frac{\|\hat{K}^{1/2}\sqrt{n}\hat{\lambda}\|^2 - p_n}{\sqrt{q_n}}.$$

The third test is based on the GEL criterion function $P_q(\lambda, \theta)$ (see equation (1.8)). We can use it for constructing a likelihood ratio test (LR) for the null hypothesis that $\lambda = 0$. It is defined as follows for GEL :

$$LR = 2n(P_q(\hat{\lambda}_{gel}, \hat{\theta}_{gel}) - \rho(0)),$$

which implies the following for CGEL :

$$\widetilde{LR} = \frac{2n(P(\hat{\lambda}, \hat{\theta}) - \rho(0)) - p_n}{\sqrt{q_n}}.$$

The following theorem shows that the three tests are first order equivalent. Moreover, it gives a way to compute them using the same matrix notations that we use above for the estimators.

Theorem 4. *If Assumptions 1 to 4 are satisfied then $\tilde{J}$, $\widetilde{LM}$ and $\widetilde{LR}$ are first order equivalent and asymptotically distributed as $N(0,1)$. Furthermore they can be computed as follows :*

$$\tilde{J} = \frac{\iota'(\beta D\beta' - D)\iota}{\sqrt{2\iota' D^2\iota}},$$

$$\widetilde{LM} = \frac{\sum_{t=1}^n (g_t(\hat{\theta})\hat{\lambda})^2 - \iota' D\iota}{\sqrt{2\iota' D^2\iota}}$$

6. To make sure that the reader is not confused with the notation, notice that $\hat{K}$ is not the same in the LM and $\widetilde{LM}$ tests. For the former, $\hat{K}\hat{\lambda}$ is $\sum_i \hat{K}_{\bullet i}\hat{\lambda}_i$, while for the latter it is $\int_{\mathcal{T}} \hat{k}(\tau, \tau_1)\hat{\lambda}(\tau_1)\pi(\tau_1)d\tau_1$.

and

$$\widetilde{LR} = \frac{2\sum_{t=1}^{n}\rho(g_t(\hat{\theta})\hat{\lambda}) - 2n\rho(0) - \iota' D\iota}{\sqrt{2\iota' D^2 \iota}}.$$

where $(g_t(\hat{\theta})\hat{\lambda})$ *comes from equation (2.1),* β *is the* $n \times n$ *matrix containing the* n *eigenvectors of* C*,* ι *is a vector of ones and* D *is an* $n \times n$ *diagonal matrix with typical element*

$$D_{ii} = \frac{\mu_i^{(n)^2}}{(\mu_i^{(n)^2} + \alpha_n)}.$$

The proof is given in the Appendix A.2.3. We can prove the asymptotic normality of the three tests by showing the first order equivalence of $\|(\hat{K}^{\alpha_n})^{-1/2}\sqrt{n}\bar{g}(\hat{\theta})\|^2$, $\|\hat{K}^{1/2}\sqrt{n}\hat{\lambda}\|^2$ and $2n(P(\hat{\lambda},\hat{\theta}) - \rho(0))$ since (Carrasco and Florens, 2000) show the result for $\tilde{J}$.

1.6 Conclusion

The CGEL method, which we apply using either the regularized Gauss-Newton algorithm or the regularized singular value decomposition of the approximated solution, is shown to be asymptotically equivalent to CGMM. We propose to extend GEL to a continuum because GEL possesses better second order asymptotic properties than GMM when the number of moment conditions is finite. The result presented in this chapter is important in the sense that it shows that the first order asymptotic properties of CGEL versus CGMM are identical just like GEL versus GMM. We should therefore expect the second order relative performance of CGMM and CGEL to be similar to the one of GEL and GMM. This hypothesis is studied in the next two chapters.

CHAPITRE II

NUMERICAL PROPERTIES OF GEL FOR A CONTINUUM

2.1 Introduction

The generalized method of moments for a continuum (CGEL), like the generalized method of moments for a continnum (CGMM) of (Carrasco and Florens, 2000), allows the number of moment conditions to be infinite. Because the number of moment conditions exceeds the sample size, the system we need solve to estimate the coefficients is singular. In the case of models based on conditional moment conditions for which the set of possible instruments is large, we need to truncate the number of instruments (see (Donald, Imbens and Newey, 2003)). CGMM and CGEL use a regularization scheme which can be applied not only to a countable set of moment conditions but also to a continuum of conditions. The regularization parameter (α) in such schemes, plays the same role as the restriction imposed on the number of instruments. However, it does not require to make any choice among the set of conditions. In fact, it affects only the smallest eigenvalues of the covariance operator associated with the vector of moment conditions. The most influencial ones are therefore automatically selected.

In all cases, we know little about how the regularization parameter affects the small sample properties of the estimators. We only know the second order asymptotic properties. (Donald and Newey, 2001) derive the second order mean square error (MSE) of the instrumental variable estimator as a function of the number of instruments, (Donald, Imbens and Newey, 2010) derive the result for GMM and GEL while (Carrasco,

2010) considers the case of CGMM. They all conclude that more instruments (or a smaller regularization parameter in the case of CGMM), improves efficiency but at the cost of a larger bias.

In order to understand the impact of α on the small sample properties, we propose three different numerical experiments. In Section (2.3) we estimate the parameters of the stable distribution using the characteristic function, in Section (2.4) we estimate a linear model with edogenous regressors and unknown optimal instruments, and we estimate the Cox-Ingersol-Ross diffusion process in Section (2.5). We compare the two different estimation procedures proposed in Chapter 1 for empirical likelihood, exponential tilting, Euclidean empirical likelihood and exponential tilted empirical likelihood for a continuum and CGMM. The results show that the relative performance of all methods depends on the model being estimated. Furthermore, α has a significant impact on the bias and MSE of the estimators and on the performance of the three tests of overidentifying restrictions, which suggests that a data driven method for selecting its value should be derived.

2.2 An overview of the CGMM and CGEL methods

The first step is to describe the two methods and emphasize how they can be implemented numerically. Lets start by arguing, as suggested by (Carrasco, 2010) and in chapter 1, that CGMM and CGEL are just generalizations of GMM and GEL respectively. To see that, we suppose that the vector of parameters $\theta_0 \in \mathbb{R}^p$ can be uniquely identified through these moment conditions :

$$E\Big(g(\theta, x_t; \tau)\Big) = 0 \ , \quad \forall \tau \in \mathcal{T}$$

where $\{x_t\}$ is a sequence of iid random variables and $\mathcal{T}$ is the space that characterizes the moment conditions' support. If $\mathcal{T}$ is $\{1, 2, \cdots, q\}$, then it is the conventional moment conditions from which the GMM and GEL estimators can be obtained. If $\mathcal{T}$ is $\mathbb{N}$, we are in the case of conditional moment conditions which induce infinitely many countable unconditional moment conditions. Finally, if $\mathcal{T}$ is an interval or simply $\mathbb{R}$, we face a continuum of moment conditions. The CGMM estimator can be defined as :

$$\hat{\theta} = \arg\min_{\theta} \left\| W^{1/2} \bar{g}(\theta) \right\|,$$

where $\bar{g}(\theta)$ is a vector with typical element $\bar{g}(\theta;\tau) = \sum_t g(\theta, x_t; \tau)/n$ and W is a linear operator. The norm can be defined generally as $\|f\|^2 = \int_{\mathcal{T}} f(\tau)^2 d\mu(\tau)$ for some measure μ. Written this way, the above three cases can be represented simply by selecting the appropriate measure. For a discrete measure, $\|f\|^2 = \sum_i f_i^2$ while for a continuum of moment, $d\mu(\tau) = \pi(\tau)d\tau$, where $\pi(\tau)$ is a density, which implies that $\|f\|^2 = \int f(\tau)^2\pi(\tau)d\tau$. The latter was first proposed by (Carrasco et al., 2007) and is described in more details by (Carrasco, 2010). The efficient CGMM is obtained by defining W as the generalized inverse of the covariance operator K, with kernel $k(\tau_1, \tau_2) = E[g_t(\theta_0, \tau_1)g_t(\theta_0, \tau_2)]$ [1]. This generalized inverse can be estimated as follows :

$$\tilde{K}_{\alpha}^{-1} = (\tilde{K}^2 + \alpha_n I)^{-1}\tilde{K}$$

where the kernel of $\tilde{K}$ is

$$\tilde{k}(\tau_1, \tau_2) = \frac{1}{n}\sum_{t=1}^{n} g_t(\tilde{\theta}, \tau_1)g_t(\tilde{\theta}, \tau_2),$$

and $\tilde{\theta}$ is a first step consistent estimate of θ_0. (Carrasco et al., 2007) show that the estimator can be defined as :

$$\hat{\theta}_{cgmm} = \arg\min_{\Theta} \tilde{v}' \left[I - C(\alpha_n I + C^2)^{-1}C\right]\tilde{v},$$

where C is an $n \times n$ matrix with typical element :

$$c_{ij} = \frac{1}{n}\int_{\mathcal{T}} g_i(\theta, \tau)g_j(\theta, \tau)d\mu(\tau),$$

with $\tilde{v} = \{\tilde{v}_1, ..., \tilde{v}_n\}$, $\tilde{v}_t = < g_t(\tilde{\theta}), \bar{g}(\theta) >$, and $\tilde{\theta} = \arg\min \|\bar{g}(\theta)\|$.

CGEL is defined as

$$\hat{\theta}_{cgel} = \arg\min_{\Theta}\left[\arg\max_{\Lambda_n} P(\lambda, \theta) = \frac{1}{n}\sum_{t=1}^{n}\rho\left(< g_t(\theta), \lambda >\right)\right],$$

---

1. For more details on why we need to regularize the inverse of K, see (Carrasco and Florens, 2000) and (Carrasco, Florens and Renault, 2007).

where $\Lambda_n = \{\lambda :< g_t(\theta), \lambda >\in \mathcal{D} \ \forall\, t = 1, ..., n\}$, with $\mathcal{D}$ being the domain of $\rho(v)$. When $\rho(v) = \log(1 - v)$ we have the empirical likelihood of (Owen, 2001) for a continuum (CEL), it is the exponential tilting of (Kitamura and Stutzer, 1997) for a continuum when $\rho(v) = -\exp(-v)$, and the Euclidean empirical likelihood of (Antoine, Bonnal and Renault, 2007) when $\rho(v) = -v - 0.5v$. As opposed to GEL, we show in chapter 1 that a quadratic $\rho(v)$ does not correspond exactly to the continuous updated CGMM.

We can compute the solution through the following iterative procedure (see chapter 1). We first iterate the following until convergence :

$$[\lambda_i g] = \left\{[CV]^2 + \alpha_n I\right\}^{-1} \left\{[CV]^2[\lambda_{i-1} g] - [CV][CP]\right\}, \tag{2.1}$$

with the initial value :

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

where ι is an $n \times 1$ vector of ones, V is an $n \times n$ diagonal matrix with $V_{tt} = \rho''(\lambda_{i-1} g_t)$, P is an $n \times 1$ vector with $P_t = \rho'(\lambda_{i-1} g_t)$, and C is the same matrix used to compute CGMM. We then solve the following minimization problem :

$$\hat{\theta} = \arg\min_{\Theta} \frac{1}{n} \sum_{t=1}^{n} \rho(\widehat{\lambda g_t}(\theta)), \tag{2.2}$$

where $\widehat{\lambda g_t}(\theta)$ is the value to which has converged the algorithm (2.1). The parameter α_n allows the algorithm to be well-conditioned. One of the objectives of the numerical study that we perform in this paper is to analyze how the choice of this parameter affects the small sample properties of $\hat{\theta}_{cgel}$. What we expect is a higher bias and a smaller variance when α_n increases. To understand why, consider the ridge regression which uses the same type of regularization. The estimate of the ridge regression in $Y = X\beta + u$ is $(X'X + \alpha_n I)^{-1} X'Y$ which implies that the bias is $[I - (X'X + \alpha_n)^{-1} X'X]\beta$. Therefore, the bias is zero when $\alpha_n = 0$ and increases as α_n diverges from zero. Its impact on the variance is not as easy to show. We use ridge regression in the case of multicolinearity which makes $X'X$ badly conditioned. A matrix is badly conditioned when its condition number, defined as its largest eigenvalue divided by its smallest, is large. If the condition number of $X'X$ is μ_{max}/μ_{min}, the condition number of $(X'X + \alpha_n)$

is $(\mu_{max} + \alpha_n)/(\mu_{min} + \alpha_n)$ which is smaller when we have multicolinearity because in that case μ_{min} is close to zero.

A last estimation procedure is worth including in our numerical analysis. (Schennach, 2007) shows that EL estimators may not be root-n consistent when the model is misspecified while ET is. It is then suggested to combine ET and EL in a method which is called the exponential tilted empirical likelihood (ETEL). The second order properties if its estimators are shown to be equivalent to EL. We can also generalize this method for the case of a continuum of conditions (CETEL). We need to use $\rho(v) = exp(-v)$ in the algorithm (2.1) and $\rho(v) = \log(1-v)$ for solving the minimization problem (2.2).

We also want to study the small sample properties of the three tests of over-identifying restrictions described by (Smith, 2004). In the case of CGEL and CGMM, we need to normalize them because the number of moment conditions is infinite. The tests converge to a $N(0,1)$ and are defined as follows :

$$\tilde{J} = \frac{\|\hat{K}_\alpha^{-1/2}\sqrt{n}\bar{g}(\hat{\theta})\|^2 - p_n}{\sqrt{q_n}}$$

$$\widetilde{LM} = \frac{\|\hat{K}^{1/2}\sqrt{n}\hat{\lambda}\|^2 - p_n}{\sqrt{q_n}}$$

and

$$\widetilde{LR} = \frac{2(nP(\hat{\lambda},\hat{\theta}) - \rho(0)) - p_n}{\sqrt{q_n}}.$$

They are computed as :

$$\tilde{J} = \frac{\iota'(\beta D\beta' - D)\iota}{\sqrt{2\iota' D^2\iota}},$$

$$\widetilde{LM} = \frac{\sum_{t=1}^n (g_t(\hat{\delta}_n)\hat{\lambda}_n)^2 - \iota' D\iota}{\sqrt{2\iota' D^2\iota}}$$

and

$$\widetilde{LR} = \frac{2\sum_{t=1}^n \rho(g_t(\hat{\delta}_n)\hat{\lambda}_n) - 2\rho(0) - \iota' D\iota}{\sqrt{2\iota' D^2\iota}}.$$

where β is the $n\times n$ matrix containing the n eigenvectors of C, ι is a vector of ones and

D is an $n \times n$ diagonal matrix with typical element

$$D_{ii} = \frac{\mu_i^{(n)^2}}{(\mu_i^{(n)^2} + \alpha_n)},$$

where $\mu_i^{(n)}$ is the ith eigenvalue of C.

Given some regularity conditions, $\sqrt{n}(\hat{\theta}_{cgmm} - \theta_0)$ and $\sqrt{n}(\hat{\theta}_{cgel} - \theta_0)$ converge to $N(0, [GK^{-1}G]^{-1})$. In order to see how the asymptotic distribution of the estimators is a good approximation of the true one, we need to estimate the covariance matrix $[GK^{-1}G]^{-1}$ by $[\bar{G}\hat{K}_\alpha^{-1}\bar{G}]^{-1}$, using the following result from (Carrasco and Florens, 2000) :

$$[\hat{K}_\alpha^{-1} f](\tau) = \sum_{t=1}^{n} \frac{\mu_t}{\mu_t^2 + \alpha_n} < \phi_t, f > \phi_t(\tau),$$

where μ_t and ϕ_t, $t = 1, ..., n$, are the eigenvalues and eigenfunctions of $\hat{K}$ defined as :

$$\phi_t(\tau) = \frac{1}{n} \sum_{i=1}^{n} \beta_{it} g_i(\hat{\theta}; \tau).$$

If we substitute f by $\bar{G} = -\sum_{t=1}^{n} w_t \exp{(ix_t\tau)}/n$, we can estimate $[GK^{-1}G]$ as follows (see Appendix (B.1)) :

$$[\bar{G}\hat{K}_\alpha^{-1}\bar{G}] = \frac{w' HB\beta D\beta' BHw}{n^2}, \tag{2.3}$$

where H is an $n \times n$ matrix with $H_{tj} = \exp{[-(x_t - x_j)^2/2]}$, and B and D are a diagonal matrices with $B_{jj} = (y_j - \hat{d}w_j)$ and $D_{jj} = (\mu_j/(\mu_j^2 + \alpha_n)$.

2.3 Estimating the coefficients of a stable distribution

As suggested by (Nolan, 2005), the family of stable distributions offers a good alternative for modeling heavy-tailed and skewed data such as stock returns. We say that a random variable follows a stable distribution if linear combinations preserve the shape of the distribution up to scale and shift, which determine respectively the variance and the expected value when they are well defined. Therefore, the normal distribution is stable because the sum of two normal random variables is also normally distributed. The Cauchy and Lévy distributions are special cases for which moments are either infinite or undefined. The notation used in this section follows (Nolan, 2009) who presents in details the properties of stable distributions. [2]

These three special cases are the only stable distributions for which the density has a closed form expression. As a result, the maximum likelihood estimation of the parameters can only be performed through numerical computation of the likelihood function. However, there is an analytical representation of its characteristic function. We can therefore base our estimation on the following continuum of moment conditions :

$$E\left[e^{i\tau x_t} - \Psi(\theta;\tau)\right] = 0 \quad \forall\ \tau \in \mathbb{R} \tag{2.4}$$

where i is the imaginary number, $\Psi(\theta;\tau)$ is the characteristic function and $\theta = \{\omega, \beta, \gamma, \delta\}$. The elements of θ are respectively the characteristic exponent [3] and the skewness, the scale and the location parameters. They are restricted to the parameter space $]0,2] \times [-1,1] \times]0,\infty[\times \mathbb{R}$. (Garcia, Renault and Veredas, 2006) estimate the parameters using indirect inference and perform a numerical study to compare it with some other methods. One of them is CGMM and was suggested by (Carrasco and Florens, 2002). We therefore use this example to compare the performance of CGEL with CGMM in small samples. We want to compare the mean-bias, median-bias and root mean squared errors (RMSE) of the estimators for different choices of α_n. We should expect CGEL, if the asymptotic results of (Newey and Smith, 2004) apply to the case of a continuum, to be less biased and more volatile than CGMM.

We need to be careful when working with stable distributions because there are more than one parametrization which implies different analytical forms for the characteristic function. In order to avoid confusions (Nolan, 2009) defines the distribution by $S(\omega, \beta, \gamma, \delta, pm)$, where $pm = 0, 1, 2$ or 3 defines the type of parametrization used. In this experiment, we follow (Garcia, Renault and Veredas, 2006) and (Carrasco and

2. See also his web site on stable distributions http ://academic2.american.edu/ jpnolan/stable/stable.html

3. In general, the characteristic exponent is defined by α_n instead of ω. But in this paper, α_n represents the regularization parameter.

Florens, 2002) by choosing $pm = 1$. Notice that when the moments exist and are finite, γ and δ are not necessarily the variance and the mean of the distribution. For example, we can represent a $N(\mu, \sigma^2)$ by $S(2, 0, \sigma/\sqrt{2}, \mu, 1)$. This parametrization implies the following characteristic function :

$$\Psi(\theta;\tau) = \begin{cases} \exp\left(-\gamma^{\omega}|\tau|^{\omega}[1 - i\beta(\tan\frac{\pi\omega}{2})(sign(\tau))] + i\delta\tau\right) & \text{for} \quad \omega \neq 1 \\ \exp\left(-\gamma|\tau|[1 + i\beta\frac{2}{\pi}(sign(\tau))\log|\tau|] + i\delta\tau\right) & \text{for} \quad \omega = 1 \end{cases},$$

where $sign(\tau) = 1$ if $\tau > 0$, -1 if $\tau < 0$ and 0 otherwise. Notice that β can be poorly identified when ω is close to 2 as the term $\tan(\pi\omega/2)$ becomes close to zero. That should reflect on the properties of the estimators of β.

We compute $\int_{\mathbb{R}} f(\tau)g(\tau)\pi(\tau)d\tau$ with $\pi(\tau)$ defined as the density of a standardized normal distribution as for (Carrasco and Florens, 2002). However, no $\tan()$ transformation is done as they do in order to transform the integrals over $\mathbb{R}$ into integrals over a finite interval. The integrals are computed directly over the interval $[-2, 2]$. Because of the integrating density, it makes almost no difference to integrate over a wider interval. Furthermore, it allows a better approximation of the integrals without being too much computationally demanding.

The regularized iterative procedure which computes the solution of $\lambda(\theta)$ is called by the numerical optimizer each time θ is updated. For some values of θ, it happens that α_n is too small to make the system well-posed. In such case, we have to increase it temporarily. More precisely, if the inverse of the condition number of $([CV]^2 + \alpha_n I)$ is less than 9.9×10^{-15}, α_n is raised by 50%. Once the procedure converges, α_n returns to its initial value for the next value of θ. The algorithm is much more stable this way.

The simulations are carried out using **R** and the random variables are generated by the **rstable** generator from the **fBasics** package. The starting values are obtained by CGMM using the identity matrix starting at the initial guess $\{\omega_0, \beta_0, \gamma_0, \delta_0\}$ $= \{1.1, 0.1, 0.1, 0\}$. The true values are not used so that we can analyze how the methods behave when little is known about the distribution. Finally, instead of reparametrizing ω and β as in (Garcia, Renault and Veredas, 2006) to restrict their parameter spaces, we use the optimizer **nlminb** which allows inequality constraints.

We perform the Monte Carlo experiment by generating 1000 samples of sizes 100 and 200. The true distribution is $S(1.7, 0.5, 0.5, 0, 1)$, which is one of the models studied by (Garcia, Renault and Veredas, 2006). The parameters are estimated using CGMM and CGEL, using both the iterative and singular value decomposition methods, with $\alpha_n = \{0.0001, 0.001, 0.005, 0.01, 0.05, 0.1\}$ for $n = 100$ and $\alpha_n = \{0.0001, 0.001, 0.01, 0.1\}$ for $n = 200$. We also compute the three tests of over-identifying restrictions for CGEL and compare their empirical sizes with the J-test of CGMM.

The properties of the estimators are presented in tables (B.1) to (B.4) [4] for $n = 100$ and in tables (B.7) to (B.10) for $n = 200$. We can see that the relative performance of CGMM and CGEL depends on which parameter is estimated and on the value of α_n. As explained by (Carrasco, 2010), we can interpret the value of α_n as a way of selecting the number of moment conditions. As the parameter goes down, more information contained in the continuum of conditions is used. Following the second order asymptotic results of (Newey and Smith, 2004), that should increase the bias of CGMM. This conclusion is verified except for γ. Its impact on the mean squared errors is more ambiguous. It seems consistent with the numerical experiment of (Carrasco and Florens, 2002) who find that the average α_n which minimizes the RMSE is around 0.05.

The impact of α_n on the bias of CGEL estimators using the iterative procedure is very similar. However, the RMSE is very stable and almost always smaller than the one of CGMM estimator. This is explained by the standard deviation which is smaller and mostly not affected by α_n. When it is affected, it tends to be positively related which suggests that CGEL uses the extra information more efficiently than CGMM. If we compare the iterative procedure with the singular value decomposition method, the

4. Notice that the results cannot be compared directly with the ones obtained by (Carrasco and Florens, 2002) even though we use the same sample size because the true distribution estimated is not the same and because they fix the parameter δ to zero. For example, the larger RMSE that we obtained is explained by the fact that β is poorly identified by the characteristic function as ω approaches 2.

latter is most of the time less biased but more volatile with also a larger RMSE.

Most of the results suggest that CGEL may outperform CGMM according to the RMSE and sometimes to the bias. It is not consistent with numerical studies on GEL like the one by (Guggenberger, 2008) who finds that GEL is much more volatile than GMM. However, his analysis is based only on linear models estimated by moment conditions constructed from weak instruments. Besides, the difference between GMM and GEL does not necessarily apply to CGEL and CGMM. A more complete numerical experiment should be performed in order to support the conclusions obtained in this section with more confidence. Furthermore, the optimal α_n for CGMM does not seem to be the same for CGEL. It would therefore be an interesting extension to derive a data-driven method based on higher order expansions to select α_n for CGEL as done by (Carrasco and Florens, 2002). We could then more easily compare the methods using their respective optimal α_n.

The sizes of the three tests are shown in tables (B.5) and (B.6) for $n = 100$ and in tables (B.11) and (B.12) for $n = 200$. We can see that α_n is negatively related to the size of all tests. They are above 50% for $\alpha_n = 0.0001$ and close to zero $\alpha_n = 0.1$. It may reflect the instability of the solution $\hat{\lambda}$, on which are based the tests, when α_n is small. As for the properties of the estimators, the value of α_n is very important which here again calls for a data-driven method to select it.

We could improve the size of the tests by using an alternative method based on (Anatolyev and Gospodinov, 2011) who suggest to modify the distribution of the tests by using a parameter that depends on the ratio number of instruments to sample size. They show that it improves the properties of the tests when the ratio is close to one. It is therefore relevant for CGEL. It could also be improved by using some bootstrap procedures to compute the finite sample critical values

2.4 Linear models with endogenous regressors

The objective of this numerical study is to compare CGMM and CGEL applied to a linear model with endogenous regressors which are related to exogenous variables through a function of unknown form. We consider a simple model in which there is only one regressor. The model that we consider is :

$$y_t = \delta W_t + \epsilon_t$$

and

$$W_t = f(x_t) + u_t,$$

where $(\epsilon_t, u_t) \sim \ iid(0, \Sigma)$. One way to estimate the model could be to approximate the function $f(x_t)$, assuming it is sufficiently smooth, by a polynomial function, since any x_t^s for $s \in \mathbb{N}$ would be a valid instrument. A method such as the one proposed by (Donald and Newey, 2001) would than be appropriate for selecting the number of instruments to include. As shown by (Carrasco, 2010), we need sufficiently rich instruments in order for $\hat{\delta}$ to reach the semi-parametric efficiency bound. The necessary conditions are satisfied by the instruments $\exp(i\tau x_t)\ \forall \tau \in \mathbb{R}$, which are the ones we will use in this paper. The continuum of moment conditions is therefore :

$$g(x_t, \tau; \delta) = (y_t - \delta W_t) \exp(i\tau x_t) \quad \forall\ \tau \in \mathbb{R}.$$

For the simulation, we follow (Carrasco, 2010) by setting the function $f(x_t) = \exp(-x_t^2)$. The density $\pi(\tau)$, which defines the norm, is the standardized Gaussian density. This choice of integrating density simplifies the computation of the matrix C because each element becomes proportional to the characteristic function of a normal distribution :

$$\begin{aligned} c_{st} &= \frac{1}{n}\int_{\mathbb{R}} g(x_s;\delta)\overline{g(x_t;\delta)}\pi(\tau)d\tau \\ &= \frac{1}{n}(y_s - \delta W_s)(y_t - \delta W_t)\int_{\mathbb{R}} \exp(i\tau(x_s - x_t))\pi(\tau)d\tau \\ &= \frac{1}{n}(y_s - \delta W_s)(y_t - \delta W_t)\exp(-0.5(x_s - x_t)^2). \end{aligned}$$

For the purpose of the simulation, we set $\delta = 0.1$ and

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

Since CGEL and CGMM can be seen as being a generalization of GEL and GMM, the differences between the properties of $\hat{\theta}_{cgmm}$ and $\hat{\theta}_{cgel}$ should be similar to the case in which the number of moment conditions is finite. (Newey and Smith, 2004) derive the second order asymptotic bias of GEL and GMM which constitutes a starting point for comparing small sample properties. It is shown that the asymptotic bias of GEL is smaller than GMM and that the asymptotic bias of EL is the same as the bias of GMM when the optimal linear combination of the sample moment conditions $G'\Omega^{-1}\bar{g}(\theta) = 0$ is used, where $\Omega = \lim_{n\to\infty} E[n\bar{g}(\theta_0)\bar{g}(\theta_0)']$, and $G = E(\partial g_t(\theta_0)/\partial\theta)$. However, very little is known about their exact relative properties in small samples. (Guggenberger, 2008) makes such comparison through a Monte Carlo experiment and concludes that GEL should not be used since its gain in terms of bias does not compensate for the extremely heavy tails (simulated) of the distribution of its estimators. However, the analysis is made for $f(x_t) = \Pi x_t$ with a very low concentration parameter, which implies that the instruments are known but weak.

In order to analyze the results below and see whether the second order differences of GMM and GEL applies to CGMM and CGEL in small samples, let us review the results of (Newey and Smith, 2004). The second order bias of GMM can be expressed as $(B_I + B_G + B_\Omega + B_W)$, where B_I is the bias of GMM using the exact optimal linear combination of the sample moment conditions, B_G comes from the estimation of the expected value of the Jacobian G, B_Ω from the estimation of the covariance matrix Ω, and B_W from the estimation of the preliminary estimate $\tilde{\theta}$. For GEL, it is $(B_I + (1 + \rho'''(0)/2)B_\Omega$. The term B_W does not appear as GEL does not require a preliminary estimate, and the absence of B_G results from the fact that the GEL uses an efficient estimate of G using implied probabilities instead of $1/n$. However, the impact of B_Ω on the asymptotic bias depends on $\rho(v)$. Since $\rho'''(0) = -2$, -1 and 0 for EL, ET and CUE respectively, B_Ω does not affect the asymptotic bias of EL. For ET, CUE and

GMM B_Ω appears unless the third moment of the residuals of the model, $E(\epsilon_t^3)$, is zero. In order to analyze these properties, we consider two cases : (i) $\{\epsilon, u_t\} \sim iidN(0, \Sigma)$, and (ii) $\{\epsilon_t, u_t\} = (\eta_t^2 - E(\eta_t^2))$, where $\eta_t \sim N(0, \Sigma)$.

In order to construct a data driven method to select the optimal α_n, (Carrasco, 2010) derives the second order mean squared errors of CGMM estimators. The result is similar to the case in which the number of moment conditions is discrete. Indeed, as α_n decreases, which is similar to choosing more instruments, the bias goes up and variance goes down, and inversely if α_n increases. Since the number of instruments does not affect the bias of EL, we should expect the same result for CEL.

As seen above, the three tests of over-identifying restrictions are based on the asymptotic normality of the normalized J, LM and LR tests. We cannot use the asymptotic distribution of the original tests since the degree of freedom goes to infinity. However, we should expect the true distribution in small samples be skewed. The distribution of the three tests may therefore be better approximated by the Gamma distribution. The Gamma distribution is better in this case than the Chi-square distribution since it allows the mean and the variance to be real numbers instead of integers. The above tests are normalized as $(Stats - p_n)/\sqrt{q_n}$. Therefore, a Gamma distribution with mean and variance equal to p_n and q_n is considered as an alternative to the asymptotic distribution. Since the mean and variance of a $Gamma(k, \nu)$ is $k\nu$ and $k\nu^2$ respectively, we will compute the size of the three tests using the $Gamma[(p_n^2/q_n), (q_n/p_n)]$.

Notice that $\hat{\lambda}$ is not directly computed in the algorithm (2.1). Instead, it computes $< g_t(\theta), \lambda >$ since it is all we need to solve equation (2.2). In the discrete case, $\hat{\lambda}$ can help testing which moment condition is not satisfied when we reject the test of over-identifying restrictions. Being a Lagrange multiplier associated with the constraint $\sum_{t=1}^n p_t g_t(\theta) = 0$, $\hat{\lambda}$ is different from zero when the constraint is binding. We can compute it by using the first Taylor approximation evaluated at the solution $\hat{\theta}_{cgel}$, which is :

$$\hat{\lambda} \approx -\hat{K}_\alpha^{-1}\bar{g}(\hat{\theta}_{cgel}),$$

and can be computed as :

$$\hat{\lambda}(\tau) \approx -\frac{1}{n^2}\sum_{t=1}^{n}\frac{\mu_t^2}{\mu_t^2+\alpha_n}\iota'\beta_{\bullet t}\beta_{\bullet t}'g(\hat{\delta};\tau).$$

An example of $\hat{\lambda}(\tau)$ is shown in figure (B.7). Such computation could be useful for example if we estimate a model in the frequency domain and want to verify which frequencies are not suitable for the model being estimated.

For each method, we perform 2000 replications for sample sizes of 100, 200 and 300. We also consider $\alpha_n = \{.00005, .0001, .0002, .0005, .0008, .001, .002, .01, .05, .1\}$. When α_n is too small, which makes the algorithm (2.1) ill-conditioned, it is slightly increased until the inverse of the condition number of $([CV]^2+\alpha_n I)$ reaches at least 5e-16, which is a little larger than the smallest floating-point number. It happens sometimes in the process of solving (2.2). Since the dimension of θ is one, the algorithm "optimize()" of R is used, which is a combination of golden section search and successive parabolic interpolation. It is not based on numerical derivatives, it is super-linear convergent, and does not require to set a starting value. We only need to set the interval of possible solutions. Intervals from $[-2,2]$ to $[-10,10]$ have been tried and produced the same results. We therefore used the smallest in order to increase the speed of convergence.

2.4.1 Mean and median biases, and RMSE

Tables (B.13) to (B.24) summarize the results on the mean and median biases for all methods, sample sizes, α_n and for symmetric and skewed error terms. Figures (B.1) to (B.6) plot them for a better comparison. The result is consistent with the second order bias derived by (Carrasco, 2010). Indeed, we can see that the median and mean biases increase when α_n decreases. It is also consistent with the GMM second order biases if we interpret $1/\alpha_n$ as being the number of instruments. On the other hand, the impact of α_n on the biases of all CGEL methods when the error terms are normal is a lot less important and it goes in the opposite direction. There may be two opposite forces affecting the bias in the case of CGEL. One is related to the interpretation of $1/\alpha_n$ being the number of instruments and the other comes from fact that the first order condition for λ is approximated when α_n is different from zero. The former should, according to (Newey and Smith, 2004), make α_n be negatively related to the bias except for CEL. However, the second force should make it positively related since a larger α_n implies a bad approximation of the first order condition. In the case of CGMM, α_n affects only the matrix of weights, not the first order condition of the optimization problem. Therefore, only the first force should affect the bias. The story is different when the error terms are skewed. In that case, the impact of α_n on the biases is very similar among CGMM and CGEL methods except for CETEL and CEL for which the relationship is similar to the case of symmetric error terms. This result is consistent with the second order bias of GEL which includes a term (B_Ω) when the error terms are skewed, except when $\rho'''(0) = -2$ as it is the case for EL and ETEL.

If we compare the biases of all methods, in the case of normal errors, all CGEL outperforms CGMM except for large α_n. If we only look at the CGEL methods, none of them seems to consistently outperform the others. CEL has a small advantage over the others for the mean bias but not for the median bias. When the error terms are skewed, CEEL produces the most biased estimators while CETEL outperforms the others followed by CEL.

As opposed to the biases, the behavior of the RMSE is more stable. First, CGMM outperforms all CGEL methods at least for small α_n. The relative better performance of CGMM becomes stronger when the error terms are skewed. This lack of efficiency of CGEL when α_n is small is caused by the instability of the algorithm (2.1) which becomes poorly conditioned in that case. In fact, while performing the simulations, α_n needed to be increased more often when the error terms were skewed in order to keep the algorithm stable. However, the relative performance of CGMM, in terms of the RMSE, is not as important as it has been suggested for the discrete case. Furthermore, it is very small when the error term is symmetric.

In figures (B.1) to (B.6), we can see that the RMSE has a U shape which implies that the optimal α_n is somewhere between 5e-5 and 0.1. However, the choice does not only depend on the sample size. For example, in the Monte Carlo experiment of (Carrasco, 2010) (with $n = 500$), in which the choice of α_n is based on the second order asymptotic RMSE, its optimal value is in average 0.03 with standard deviation of 0.01. Therefore, we cannot compare CGMM and CGEL until we have a way to select the optimal α_n, which would be possible if we do a second order expansion of CGEL estimators. An alternative method could be to use a bootstrap method for selecting α_n. We can perform a small simulation for analyzing how we can improve the properties of the estimators by choosing the optimal α_n at each iteration. If we re-sample the data 20 times at each iteration and select the α_n which minimizes the RMSE for CEL, n=100, normal errors and 400 iterations, we obtain a RMSE of 0.14 which beats all cases. In the simulation, the mean of the optimal α_n is 0.02 and its standard error is 0.03. However, this method is infeasible in practice since we do not know the true value of δ which prevents us from computing the RMSE. But it gives some insights for future research.

We conclude the analysis of the bias and RMSE by comparing the properties of CGMM and CGEL with GMM based on a finite number of instruments. Tables (B.31) to (B.34) show the properties of GMM when the instruments are $\{x_t, x_t^2, x_t^3, x_t^4\}$. Both CGEL and CGMM perform better than GMM with respect to the bias and the RMSE. It suggests that using a continuum may be better when the relationship between the regressors and the instruments are of unknown form.

2.4.2 Tests of over-identifying restrictions

Table (B.35) to (B.58) give the size of the three tests for each α_n, sample sizes, and type of error terms. There is no clear relationship between the size and α_n. For example, most of the time the size of the J-test based on CGMM goes down as α_n goes up when the error terms are normally distributed and inversely when they are skewed. We obtain the same result for the J-test based on CGEL. However, for the LR and LM tests, the relationship is stable. The size always goes down when α_n goes up. It is therefore difficult to compare the properties of the tests using the asymptotic distribution and the Gamma distribution. We can select the α_n which produces the smallest RMSE, since that is the logical choice, and then compare the different tests for this value only. This is done in tables (B.59) and (B.60). The Gamma approximation only improves the sizes of the LM and LR tests, when the sample sizes is 200 or 300 and the error terms are skewed. In the other cases, it often makes it worse. In the case of normal errors, the sizes of LM and LR tests are not too bad. However, the size of the J-test can be very bad even if the sample size is 300.

The choice of α_n based on the RMSE seems to be a good one at least for the LR and LM tests. However, the tests would probably improve if we had an optimal α_n for each iteration instead of fixing it. The impact of α_n is so important, it would probably make a big difference.

2.4.3 Distribution of $\hat{\delta}$

Table (B.61) shows the probabilities of rejecting the null $H0 : \delta = 0.1$ using the critical values of the asymptotic distribution of $\sqrt{n}(\hat{\delta} - 0.1)/sd(\hat{\delta})$, which is $N(0, 1)$, for CEL, $n = 300$, and normal errors. The other cases are very similar and therefore not reported. In each case, the probability of rejecting the null is underestimated, and the smaller α_n the better. Therefore, choosing α_n which minimizes the RMSE will not in this case improve the properties of these tests. However, the normal distribution has the right shape as we can see on figure (B.8), which shows the QQ-plot of $\sqrt{n}(\hat{\delta} - 0.1)/sd(\hat{\delta})$ for $\alpha_n = 0.00005$. The problem is therefore caused by the estimated standard deviation of $\hat{\delta}$ which is too large.

2.5 The Cox-Ingersoll-Ross process

The example is taken from (Carrasco et al., 2007) [5]. We want to estimate the parameters of the Cox-Ingersoll-Ross (CIR) diffusion process. The moment conditions are based on the characteristic function derived by (Singleton, 2001). What makes the

5. This example is not appropriate for CGEL presented in this paper because the moment conditions are not iid. However, it is included because it presents another potential application.

characteristic function appealing is the existence of its analytic form for many continuous time processes for which the analytic form of the likelihood does not exist. But to take advantage of the whole set of moment conditions, we need methods like CGMM or CGEL.

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

$$dr_t = (\gamma - \kappa r_t)dt + \sigma\sqrt{r_t}dW_t. \tag{2.5}$$

The conditional characteristic function for this process is :

$$E\left(e^{i\tau r_{t+1}}\middle| r_t\right) \equiv \psi(\tau|r_t) = \left(1 - \frac{i\tau}{c}\right)^{-2\gamma/\sigma^2} \exp\left[\frac{i\tau e^{-\kappa}}{1-\frac{i\tau}{c}} r_t\right], \tag{2.6}$$

with

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

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

$$g(x_t, \tau, \theta) = m(\tau_1, Y_t)\left[e^{i\tau_2 r_{t+1}} - \psi(\tau_2|r_t)\right]. \tag{2.7}$$

This is the double index version of the moment conditions. As (Carrasco et al., 2007) explain, the estimators reach the Cramer-Rao lower bound if the instruments are $m(\tau_1, Y_t) = e^{i\tau_1 r_t}$. Moreover, the moment function $g_t(\theta)$ defined in such a way is a martingale difference sequence which implies that we don't need to smooth the function as would be necessary if it were weakly dependent. It follows that the moment conditions are :

$$E^{P_0} g(x_t, \tau, \theta) = E^{P_0}\left(e^{i\tau_1 r_t}\left[e^{i\tau_2 r_{t+1}} - \psi(\tau_2|r_t)\right]\right) \quad \forall\, \tau \in \mathbb{R}^2, \tag{2.8}$$

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

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

where the integrating density $\pi(\tau_1, \tau_2) = \pi(\tau_1)\pi(\tau_2)$ with $\pi(\tau)$ being the density of a standard normal distribution. The effect of the choice of integrating density on the properties of the estimators is unknown for now. The choice is driven by the fact that it allows us to get an analytic representation of one of the double integrals that we need to compute for each $C_{ij},\ j,i = 1, ..., n$. The simplification is possible only because of this specific moment function. For more details on how C_{ij} can be computed, see the appendix of (Carrasco et al., 2007).

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

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

In order to be able to compare our results with (Carrasco et al., 2007), the first study is the same as the one that produced the results they reported in their paper. The sample size is 500, the number of iterations is 100 and $\alpha = 0.02$. The results are reported in Table B.4. For γ and κ, C-GEL seems the have the smallest bias while C-CUE stands between C-GMM and C-GEL. This result is consistent with what we find in the literature. We know from (Newey and Smith, 2004) and (Anatolyev, 2005) that the second order asymptotic bias of GMM processes two more terms if we compare it with CUE and the latter has one more term if we compare it with EL. The fact that we use a continuum of moment conditions does not seem to affect this result. What is most surprising is how the mean square error (MSE) of C-GEL and C-CUE performs relative to C-GMM. They are in fact much smaller which contradicts the results of (Guggenberger, 2008) who gives evidence that the distribution of GEL estimator has heavy tails. However, as for C-GMM, C-GEL and C-CUE have a hard time estimating the volatility parameter. Both the bias and the MSE are by far greater than the ones obtained using MLE which remains for now a mystery. Finally, there is no significant difference between C-GELsv and C-GELgn. Since the latter has a higher computational cost, the former may be preferred. But for now it is pure speculation.

The second Monte-Carlo study tries to see how the choice of α affects the bias and the MSE. To do so, the previous experiment has been repeated for $\alpha = \{0.001, 0.005, 0.01,$

$0.02, 0.05, 0.1\}$ and for n=200 and 500. The results are shown in Figures B.9 to B.12. It does not seem to have a clear tendency. However, it seems that if α is too small the MSE is larger. This is caused by the fact that if we do not regularize enough the system, it becomes unstable. Otherwise, the MSE does not seem to be too sensitive. Similar comments can be made with respect to the bias. It seems however that the choice of α has more impact in larger sample sizes. In such cases, a judicious choice could reduce the bias substantially. But further research needs to be done in order to know what is a judicious choice.

2.6 Conclusion

We have analyzed different properties of CGEL in the context of three different models. In the linear model with endogenous regressors and unknown optimal instruments, we can conclude that CGMM outperforms CGEL according to the RMSE, but the difference is sometimes negligible. However, the bias of CGMM estimators depends on the parameter α_n and is in general larger than the bias of CGEL estimators. For nonlinear models, in particular the estimation of the parameters of a stable distribution, CEL seems to dominate CGMM both in terms of bias and RMSE. This finding contradicts previous results in the discrete case suggesting that the distribution of empirical likelihood estimators have heavy tails.

We know more about the properties of CGEL, but there is still much we can learn. Developing methods for selecting the optimal α_n should be a priority before using it in applied research. Until then, we can hardly compare it with CGMM. This method could be based on second order expansion or on bootstrap methods. This selection method would probably improve the size of the three tests of over-identifying restrictions which seem to perform better around the optimal α_n. As an alternative to improve the size of the over-identifying restriction, we could follow the advice of (Arellano, Hansen and Sentana, 2011) and apply the algorithm of (Imhof, 1961) to compute the distribution of quadratic forms in normal variables. The J-test can be seen as being a weighted sum of chi-square variables with the weights equal to $\mu_i^2/(\mu_i^2+\alpha_n)$ and (Imhof, 1961) proposes a way to compute the critical values of such random process.

CHAPITRE III

SELECTING THE REGULARIZATION PARAMETER

3.1 Introduction

We saw in the previous chapter through numerical experiments that the relative performance of CGMM and CGEL depends on α. For some values, CGEL dominates CGMM and for others the result is reversed. In order to really be able to compare the different methods, we need to have a data driven procedure to select the parameter as is proposed by (Carrasco, 2010) and (Carrasco and Kotchoni, 2010) for CGMM. The main objective of this chapter is to obtain an expression for the higher order MSE of the CGEL estimator using the stochastic expansion of the first order conditions. This approximated MSE can serve as criterion for the selection of the optimal regularization parameter. We limit our analysis to the linear model of Section 2.4 with θ being a scalar. The study of this simple case will act as a starting point for other more general models that can be analyzed in future research.

But before we derive the higher order MSE, we want to see if α can easily be selected by using a bootstrap method. It is a natural way of doing inference when the exact distribution is unknown. The next section analyze this possibility briefly.

3.2 The model

We consider the linear model of Section (2.4) :

$$\begin{aligned} y_t =& \theta W_t + \epsilon_t \\ W_t =& f(x_t) + u_t, \end{aligned} \tag{3.1}$$

where $(\epsilon_t, u_t) \sim \ iid(0, \Sigma)$. The model is estimated using the continuum of instruments $Z_t(\tau) = \exp{(ix_t\tau)}$, and the moment conditions

$$E[g_t(\theta;\tau)] = 0 \ \forall \tau \in \mathbb{R},$$

where $g_t(\theta;\tau) = (y_t - \theta W_t)Z_t(\tau)$. We assume that Assumptions 1 to 5 are satisfied so that the results of Chapter 1 hold. As for the previous chapters, we define the covariance operator K by its kernel $k(\tau_1,\tau_2) = E[g_t(\tau_1)\overline{g_t(\tau_2)}]$, where $\overline{g_t(\tau_2)} = (y_t - \theta W_t)Z_t(-\tau_2)$, is the complexe conjugate of $g_t(\tau_2)$. We also need to define higher moments operators that are needed in Section 3.4. We know that the covariance operator is used to obtain the variance of the random variable $< g_t, f >$ for any non random function $f \in L^2(\pi)$. That variance is $\|K^{1/2}f\|^2 =< f, Kf >= \int k(\tau_1,\tau_2)f(\tau_2)\overline{f(\tau_1)}d\pi(\tau_1)d\pi(\tau_2)$. We can also define the skewness operator S defined by its kernel $s(\tau_1,\tau_2,\tau_3) = E[g_t(\tau_1)\overline{g_t(\tau_2)g_t(\tau_3)}]$. This is an operator from $L^2(\pi)$ to $L^2(\pi)\otimes L^2(\pi)$, with

$$(Sf)(\tau_1,\tau_2) = \int s(\tau_1,\tau_2,\tau_3)f(\tau_3)d\pi(\tau_3).$$

The skewness operator defines the skewness of the random variable $< g_t, f >$ as :

$$Sf^3 = \int\int\int s(\tau_1,\tau_2,\tau_3)\overline{f(\tau_1)}f(\tau_2)f(\tau_3)d\pi(\tau_1)d\pi(\tau_2)d\pi(\tau_3).$$

Similarly, the kurtosis operator H, with kernel $h(\tau_1,\tau_2,\tau_3,\tau_4)$ defines the kurtosis of $< g_t, f >$ as Hf^4. As for the covariance operator, we say that the i^{th} moment operator, M_i, is bounded if :

$$\|M_i\|^2 = \sup_{\|\phi\|\leq 1} |M_i\phi^i| < \infty$$

The estimate is obtained by solving the following first order conditions :

$$\frac{1}{n}\sum_{t=1}^n \rho'(\lambda g_t(\theta))\left[\lambda G_t\right] = 0$$

and

$$\left[\frac{1}{n}\sum_{t=1}^{n}\rho''[\lambda g_t(\theta)][g_t(\theta)\overline{g_t(\theta)}]\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho'(\lambda g_t)g_t\right]+\lambda\alpha=0,$$

where $G_t(\tau) = -W_t Z_t(\tau)$. We suppose for this chapter that the integrating density $\pi(\tau)$ is the density of the standardized normal distribution.

3.3 A Monte Carlo approach

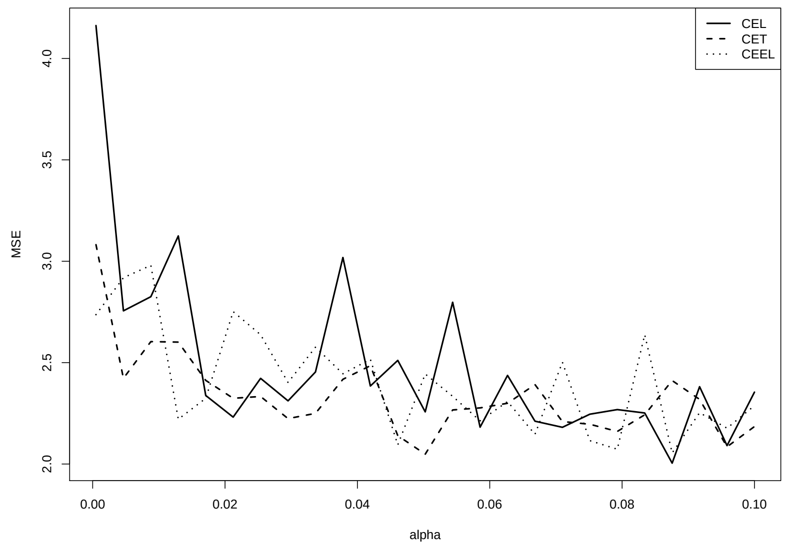
First we consider using a Monte Carlo simulation to compute the mean square error (MSE). A similar approach is proposed by (Carrasco and Kotchoni, 2010) for CGMM but in the context of nonlinear moment conditions based on characteristic functions. They conclude that the Monte Carlo method dominates the one based on an approximated MSE. However, the latter is computed using simulations as well because the non-linearity of the moment conditions makes it difficult to obtain a close form representation of the covariances between the higher order terms of the estimator. Since we are considering linear models, we won't face the same difficulty in the next section.

Because we suppose that $\{x_t, W_t, y_t\}$ are iid realizations, we can use the following simple procedure. We consider the first estimate of the coefficient θ for a given α, say $\hat{\theta}_\alpha$, to be the population value of θ and compute the MSE by resampling $\{x_t, W_t, y_t\}$ jointly N times with replacement. Let $\hat{\theta}^i_\alpha$ be the estimate of the i^{th} sample. Then, the MSE is defined as :

$$MSE(\alpha) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}^i_\alpha - \hat{\theta}_\alpha)^2 \tag{3.2}$$

A grid search can then determined the optimal value of α. Also, if $RMSE(\alpha)$ is sufficiently smooth, we can apply a bracketing method such as the golden section search Method. For the simulation, we set $N = 500$, $n = 300$, $\theta_0 = 0$, $E(u_t\epsilon_t) \equiv \sigma_{\epsilon u} = 0.5$ and $\sigma^2_\epsilon = \sigma^2_u = 1$. Figure 3.1 show the result for a grid of 100 α's between 0.0005 and 0.1. The method seems to capture well the magnitude of the MSE if we compare it with the value based on the stochastic expansion of Section 3.4 below. However, the function is not smooth enough to make it possible for us to find the minimum using a numerical

Figure 3.1 Estimated MSE using the bootstrap method (n=300, N=500)



optimizer. We could use a fine grid, but it is very time consuming and the volatility of the function makes it hard to identify the minimum. Notice that the volatility of the function can be reduced by increasing N, but even with 2,000 resamplings, the function remains too volatile. Another possibility would be to smooth the function and choose the α that minimizes it, but we would need to run some simulations in order to see if that approach is reliable. Also, we need to verify that the moments of the estimators exist to validate the use of a bootstrapping method. The results of Chapter 2 suggest that they do, but we need a more robust proof before developing such procedure. We leave it for future research.

3.4 Second order expansion : the linear model case

The method presented in the previous section is very time consuming and makes it hard to select α because the function $MSE(\alpha)$ that we obtain is not smooth enough. Another alternative is to derive the higher order MSE of the estimator of θ using the same approach used by (Donald and Newey, 2001) for 2SLS, (Donald, Imbens and

Newey, 2010) for GEL and (Carrasco, 2010) for CGMM. We proceed with the simplest case in which θ is a scalar. Therefore, we don't need to compute the MSE of some linear combination of the vector of coefficients. We are just interested in obtaining an expression for $E[n(\hat{\theta}-\theta_0)^2]$.

For the expansion, we need Assumptions 1 to 5 from Chapter 1 and the following additional ones :

Assumption 6. *a) $\rho(v)$ is five times continuously differentiable around zero, b) the first four moment operators of $Z_t(\tau)$ are bounded, and c) $E|y_t-\theta_0 W_t|^a<\infty,\ \forall a=1,...,5$.*

The first assumption is satisfied by CEL ($\rho(v)=\log(1-v)$), CET ($\rho(v)=\exp(-v)$) and CEEL ($\rho=-(v+v/2)$). The second and third imply that the first four moment operators of $g_t(\theta_0;\tau)$ are bounded because $E(\epsilon_t|x_t)=0$ implies $E[g_t(\theta_0)^a](\tau_1,...,\tau_a)=E[(y_t-\theta_0W_t)^a]E[Z_t(\tau_1),...,Z_t(\tau_a)]$. Since $E[Z_t(\tau_1),...,Z_t(\tau_a)]$ is the characteristic function of x_t evaluated at $(\tau_1-\tau_2-...-\tau_a)$, the last assumption imposes restrictions on the probability density of x_t.

The proof of the following theorem is presented in Appendix C.

Theorem 5. *If Assumptions 1 to 6 are satisfied, then the CGEL estimator of θ in the model (3.1) can be written as :*

$$
\begin{aligned}
(\hat{\theta}-\theta_0) =& \Omega_\alpha\overline{G}(K^2+\alpha I)^{-1}\Bigg\{-K\hat{g}-\rho_3[S\hat{g}][\Sigma\hat{g}] \\
&+\frac{(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)}{\sigma_\epsilon^4}K^2[\Sigma\hat{g}]+\frac{4\sigma_{\epsilon u}}{\sigma_\epsilon^2}K^2\Omega_\alpha[\overline{G}K_\alpha^{-1}\hat{g}][\Sigma\hat{g}] \\
&+\rho_3[SG]\Omega_\alpha[\overline{G}K_\alpha^{-1}\hat{g}][\Sigma\hat{g}]+\frac{3\rho_3}{2}[SK][\Sigma\hat{g}][\Sigma\hat{g}]\Bigg\} \\
&+o_p\left(\frac{1}{\alpha n}\right),
\end{aligned}
\tag{3.3}
$$

where $G(\tau)=-E[W_tZ_t(\tau)]$, $\hat{g}(\tau)=\sum_{t=1}^n g_t(\theta_0;\tau)/n$, $\Sigma=[K_\alpha^{-1}-K_\alpha^{-1}G\Omega_\alpha\overline{G}K_\alpha^{-1}]$, S is the skewness operator of $g_t(\tau)$, $\Omega_\alpha=[\overline{G}K_\alpha^{-1}G]^{-1}$, and $\rho_3=\rho'''(0)$.

The subscript n is omitted from α for simplicity. The assumptions from Chapter 1 imply that the last term goes to zero even if we multiply $\hat{\theta}$ by $\sqrt{n}$ since it is assumed that $\alpha^2 n \to \infty$. The notation used in Theorem 5 implies for example that

$$\{[S\hat{g}][\Sigma\hat{g}]\}(\tau) = \int \left[\int s(\tau,\tau_1,\tau_2)\hat{g}(\tau_2)d\pi(\tau_2)\right](\tau,\tau_1)\left[\int \Sigma(\tau_1,\tau_3)\hat{g}(\tau_3)d\pi(\tau_3)\right](\tau_1)d\pi(\tau_1),$$

and then

$$\overline{G}(K^2+\alpha I)^{-1}[S\hat{g}][\Sigma\hat{g}] = \sum_{i=1}^{\infty}\frac{1}{\mu_i^2+\alpha} < G,\phi_i >< \{[S\hat{g}][\Sigma\hat{g}]\},\phi_i >,$$

where μ_i's and ϕ_i's are the eigenvalues and orthonormalized eigenfunctions of K. For the other terms the integrals must be performed within each bracket first and then from left to right across the brackets. For example, in the last term, we compute $[SK](\tau_1,\tau_2,\tau_3)$, $[\Sigma\hat{g}](\tau_3)$, and $[\Sigma\hat{g}](\tau_2)$ first. Then, we can compute successively $[SK\Sigma\hat{g}](\tau_1,\tau_2)$ and $[SK\Sigma\hat{g}\Sigma\hat{g}](\tau_1)$.

We can then obtain the MSE $E[n(\hat{\theta}-\theta_0)^2]$ up to a term that is $o(1/(n\alpha^2))$. The proof of the following theorem is also presented in Appendix C.

Theorem 6. *If Assumptions 1 to 6 are satisfied, then the MSE of the CGEL estimator of θ in the model (3.1) can be written as :*

$$E[n(\hat{\theta}-\theta_0)^2](\alpha) = \Omega_\alpha^2\overline{G}(K^2+\alpha I)^{-1}\Xi(K^2+\alpha I)G + o\left(\frac{1}{n\alpha^2}\right), \tag{3.4}$$

with

$$\Xi = K^3+\frac{\rho_3^2}{n}[S\Sigma\Delta\Sigma S] + \frac{4\Omega_\alpha^2(k_\epsilon-1)}{n}K^2\Sigma K\Sigma K^2 + \frac{16\Omega_\alpha^2\sigma_{\epsilon u}^2}{\sigma_\epsilon^4 n}K^2\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma K^2$$
$$+\frac{\rho_3^2\Omega_\alpha^2}{n}SG\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma\overline{G}S + \frac{9\rho_3^2}{4n}SK\Sigma^2\Delta\Sigma^2KS + \frac{\rho_3}{n}KS\Sigma S$$
$$+\frac{3(k_\epsilon-1)}{n}K^2\Sigma K^2 - \frac{4\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^2 n}KSK_\alpha^{-1}G\Sigma K^2 - \frac{\rho_3\Omega_\alpha}{n}[KSK_\alpha^{-1}G\Sigma\overline{G}S]$$
$$-\frac{3\rho_3}{2n}KS\Sigma^2KS + \frac{2\rho_3S_\epsilon}{\sigma_\epsilon^2 n}S\Sigma\Psi\Sigma K^2 - \frac{4\rho_3\Omega_\alpha\sigma_{\epsilon u}}{n\sigma_\epsilon^2}S\Sigma\Delta K_\alpha^{-1}G\Sigma K^2$$
$$-\frac{\rho_3^2\Omega_\alpha}{n}S\Sigma\Delta K_\alpha^{-1}G\Sigma\overline{G}S - \frac{3\rho_3^2}{2n}S\Sigma\Delta\Sigma\Sigma KS$$
$$-\frac{8\sigma_{\epsilon u}\Omega_\alpha S_\epsilon}{n\sigma_\epsilon^4}K^2\Sigma\Psi K_\alpha^{-1}G\Sigma K^2$$
$$-\frac{2\rho_3S_\epsilon\Omega_\alpha}{\sigma_\epsilon^2 n}K^2\Sigma\Psi K_\alpha^{-1}G\Sigma\overline{G}S - \frac{3\rho_3S_\epsilon}{\sigma_\epsilon^2 n}K^2\Sigma\Psi\Sigma^2KS$$
$$+\frac{4\sigma_{\epsilon u}\rho_3\Omega_\alpha^2}{\sigma_\epsilon^2 n}K^2\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma\overline{G}S + \frac{6\sigma_{\epsilon u}\rho_3\Omega_\alpha}{\sigma_\epsilon^2 n}K^2\Sigma\overline{G}K_\alpha^{-1}\Delta\Sigma^2KS$$
$$+\frac{3\rho_3^2\Omega_\alpha}{2n}SG\Sigma\overline{G}K_\alpha^{-1}\Delta\Sigma^2KS \tag{3.5}$$

See Appendix C for a clear definition of the four dimensional operators Δ and Ψ. It is also important to refer to the Appendix for the order of integration. As opposed to Theorem 5, we can not easily find a notation that can make it clear how to compute each term. For example, $\Delta(\tau_1,\tau_2,\tau_3,\tau_4) = [k(\tau_1,\tau_2)k(\tau_3,\tau_4) + k(\tau_1,\tau_3)k(\tau_2,\tau_4) + k(\tau_1,\tau_4)k(\tau_2,\tau_3)]$. We need to know which τ is associated with the ones from the other operators.

It is not trivial to compare Equation (3.4) and the one derived by (Donald, Imbens and Newey, 2010) for GEL. The complexity of our expression comes from the complexity of the first order conditions for CGEL which is very different from GEL. For GEL, the expansion is based on the condition :

$$\frac{1}{n}\sum_{t=1}^n\begin{pmatrix}\rho'(\lambda'g_t)[G_t'\lambda]\\ \rho'(\lambda'g_t)g_t\end{pmatrix} = 0,$$

while the second condition of CGEL is premultiplied by a covariance operator and the term αI is added. As a result, ρ_3, which is the value that makes the higher order MSE different across the different CGEL's, appears in the expansion much sooner. We can see that the expression is simpler for CEEL since in that case $\rho_3 = 0$, but it is not clear whether it makes the MSE smaller or larger as in (Donald, Imbens and Newey, 2010) and (Newey and Smith, 2004). For this chapter we will only rely on a numerical computation of the MSE to analyze its behavior.

In section C.1.5, a matrix representation of each term is given. For example, the first term of the MSE is $\overline{G}K_\alpha^{-1}KK_\alpha^{-1}G$, which can be written as $w'HB\beta D_{32}\beta' BHw/n^2$, where w is the $n\times 1$ vector of W_t, H is an $n\times n$ matrix with $H_{ts} =< Z_t, Z_s > /n = \exp\left[-(x_t-x_s)^2/2\right]/n$, B is a diagonal matrix with $B_{tt}=\epsilon_t$, D_{ab} is a diagonal matrix with the i^{th} diagonal being $\mu_i^a/(\mu_i^2+\alpha)^b$, and the μ_i's are the eigenvalues of the matrix C defined in Chapter 1. The matrix β requires some precision. We saw in Chapter 1 that we can construct the eigenfunctions of $\hat{K}$ using the following :

$$\hat{\phi}_i = \frac{1}{n}\sum_{l=1}^n \beta_{li} g_l(\hat{\theta}) = \beta_i' g(\hat{\theta})/n = \beta_i' BZ,$$

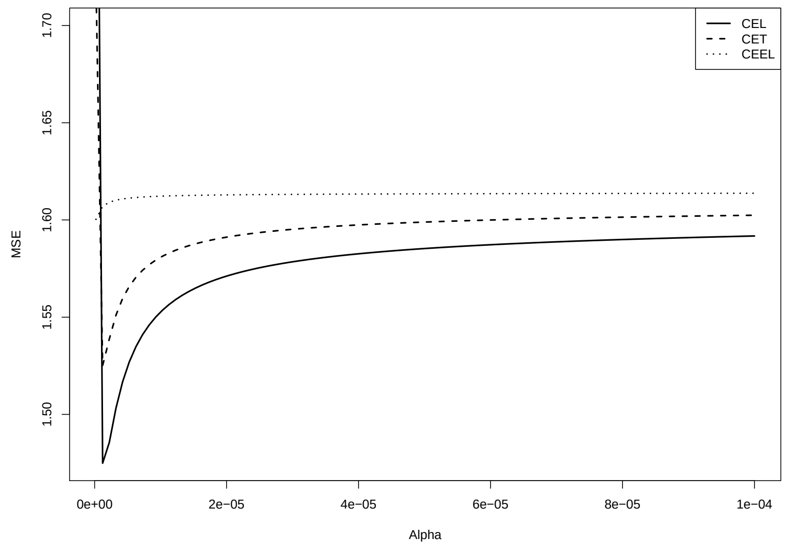
where β_i is the i^{th} eigenvector of the matrix C. However, as it is well explained by (Carrasco, Florens and Renault, 2007), the eigenfunctions need to be normalized. In order to keep the same matrix expression, we choose to redefine the eigenvectors β_i. The necessary normalization can be obtained by computing the norm of $\hat{\phi}_i$:

$$\|\hat{\phi}_i\|^2 = \frac{1}{n^2}\sum_{t,s=1}^n \beta_{ti}\beta_{si} < g_t, g_s >= \frac{1}{n}\beta_i' C\beta_i = \frac{\mu_i}{n}$$

Therefore, let us rewrite the i^{th} eigenvector of C as $\tilde{\beta}_i$. Then, the i^{th} column of β, β_i, in the above matrix representation of the first term of the MSE and in the definition of $\hat{\phi}_i$ is $\sqrt{n}\tilde{\beta}_i/\sqrt{\mu_i}$. Of course, one needs to be careful in practice because many eigenvalues may be very close to zero and when it is the case, they are computed with very little precision. The norm of $\hat{\phi}_i$ may even fail to be equal to one after normalization. A good practice to avoid numerical problems is therefore to drop the β_i and μ_i for which $\|\hat{\phi}_i\| \neq 1$. If only N eigenvalues respect the condition, we can still use the same matrix representation with β being $n\times N$ and D being $N\times N$.

In order to analyze the MSE for the different CGEL methods, we can plot the

Figure 3.2 Estimated MSE as function of α (n=500)



function using simulated data from Model 3.1. But first, we need a first step estimate of θ to compute the matrices B, β and D. Any root-n consistent estimate, $\tilde{\theta}$, will produce the desired result. We therefore choose to obtain that estimate using the CGMM estimator with the identity operator, which is defined as :

$$\tilde{\theta} = \min_{\theta} \iota' C \iota$$

Figures 3.2 and 3.3 show the results with $n = 500$ and 300 for the normal ϵ_t case. The result is compatible with Figures B.1 to B.3 which show that the minimum is reached at a very low α. Furthermore, the result shows that CEL dominates the two other CGEL methods for any α, and CET dominates CEEL except for α very close to zero. In fact the MSE of CEEL seems to never increase as α approaches zero like the other two methods. That result is compatible with the simulations of Chapter 2. In fact CEEL is the most stable method when α is close to zero. We rarely obtain a singular system with CEEL while it is frequent with the others especially with CEL. The case in which the distribution of ϵ_t is skewed is presented in Figure 3.4 for $n = 500$. According

Figure 3.3 Estimated MSE as function of α (n=300)

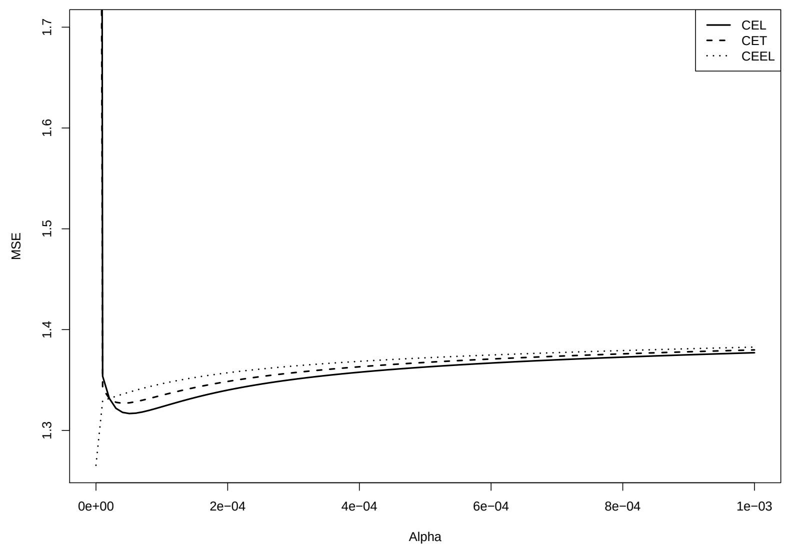
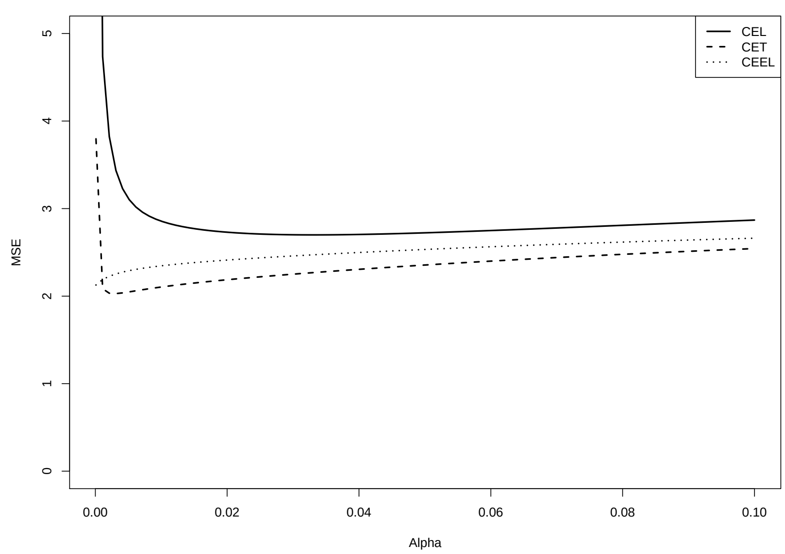


Figure 3.4 Estimated MSE as function of α (n=500, Skewed error term)



to this graph, CEL no longer dominates the two other methods. It is now the worst and CET is the best. This result differs from the one obtained by (Newey and Smith, 2004) who find that an extra term is added to the asymptotic bias of GMM, ET and CUE when the error terms are skewed, while the same term is absent from the asymptotic bias of EL whether the error term is skewed or not. This comes from the fact that ρ_3 affects the higher order properties of the estimator of CGEL in a very different way as discussed above. Notice also that the optimal α is not as small as for the normal case, which is compatible with what we found in Chapter 2. Since the bias tends to be negatively related to α, it is probably because the bias term is more important with skewed errors.

3.5 Conclusion

The evidence that we found in this section, which is also compatible with the simulations of Section 2.4 suggests that the parameter α should be as small as possible. If the algorithm fails, we just increase it a little at a time until the system becomes non singular. The fact that we do not get the same results as CGMM is, to a certain extend, compatible with (Newey and Smith, 2004). What dominates the MSE of CGEL seems to be its variance which is an increasing function of α. The bias does not seem to be an important factor as opposed to CGMM.

However, it is possible that omitting the terms that are $o_p(1/(\alpha n)$ implies that we miss some important factors. The derivation in Appendix C makes it possible to derive those terms. The next step should include the derivation of those terms and a numerical analysis to compare the CGEL using the optimal α and CGMM using the selection procedure proposed by (Carrasco, 2010).

ANNEXE A

APPENDIX : CHAPTER 1

A.1 Overview of some concepts

A.1.1 CGMM

We present here a brief overview of CGMM. It is developed by (Carrasco and Florens, 2000), summarized by (Carrasco, Florens and Renault, 2007) and (Carrasco et al., 2007) show how it can be implemented by expressing the objective function in matrix form. The estimator is defined as :

$$\hat{\theta}_{cgmm} = \arg\min_{\Theta} \|B_n \bar{g}(\theta)\|,$$

where B_n is a sequence of random operators from $L^2(\pi) \to L^2(\pi)$ which converges to the linear bounded operator B. It plays the same role as the weighting matrix of GMM. In order to achieve efficiency, the operator B must be defined as the inverse of the square root of the asymptotic covariance operator of $\sqrt{n}\bar{g}(\theta_0)$, K. Because of its properties, the inverse of K is unbounded. As a result, the objective function is ill-posed because it can be written as $< \bar{g}(\theta), K^{-1}\bar{g}(\theta) >$, where the second term of the inner product is the solution to $Kx = \bar{g}(\theta)$. A stable and unique solution can be computed using the Tikhonov approach for which the inverse of the linear operator is substituted by the regularized inverse

$$(K^{\alpha_n})^{-1} = (\alpha_n I + K^2)^{-1} K.$$

The feasible optimal CGMM estimator, in which K is replaced by the consistent estimate $\hat{K}$, is therefore defined as :

$$\hat{\theta}_{cgmm} = \arg\min_{\Theta} \|(\hat{K}^{\alpha_n})^{-1/2}\bar{g}(\theta)\|.$$

In order for $\hat{\theta}_{cgmm}$ to be consistent, certain conditions are required. One of them imposes a rate of convergence for α_n which must satisfy $n\alpha_n^{3/2} \to \infty$ as α_n goes to zero, which implies that $\alpha_n = O(n^{-2/3+\eta})$ for $0 < \eta < 2/3$. The condition on α_n is required in order for $\|(\hat{K}^{\alpha_n})^{-1/2}f_n - K^{-1/2}f\|$ to be $o_p(1)$ for any f_n converging to f. To prove asymptotic normality, the required rate of convergence of α_n is different. We need $n\alpha_n^3 \to \infty$ as α_n goes to zero because we need $\|(\hat{K}^{\alpha_n})^{-1}f_n - K^{-1}f\|$ to be $o_p(1)$. The latter implies that $\alpha_n = O(n^{-1/3+\eta})$ for $0 < \eta < 1/3$. Given these conditions (Carrasco and Florens, 2000) show that $\sqrt{n}(\hat{\theta}_{cgmm} - \theta_0)$ is asymptotically distributed as $N(0, [GK^{-1}G]^{-1})$.

In order to compute the two step CGMM estimator, we first solve :

$$\tilde{\theta} = \arg\min_{\Theta} \|\bar{g}(\theta)\|,$$

in which the identity operator has been used instead of $(\hat{K}^{\alpha_n})^{-1}$, and then :

$$\hat{\theta}_{cgmm} = \arg\min_{\Theta} \tilde{v}' \left[I - C(\alpha_n I + C^2)^{-1}C\right] \tilde{v},$$

where C is the same matrix defined in Section 1.4.1 and $\tilde{v} = \{\tilde{v}_1, ..., \tilde{v}_n\}$ with $\tilde{v}_t = g_t(\tilde{\theta})\bar{g}(\theta)$.

A.1.2 Fréchet derivative

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

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

If $F(\lambda) =< g_t(\theta), \lambda >$, the Fréchet derivative DF_λ is $g_t(\theta)$. It is an operator from $L^2(\pi)$ to $\mathbb{R}$ defined as $DF_\lambda h = g_t(\theta)h =< g_t(\theta), h > \quad \forall h \in L^2(\pi)$. The proof is straightforward since $F(\lambda)$ is linear :

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

Fréchet derivative of the first order condition of CGEL with respect to λ

We need to solve the following system :

$$F_{n1}(\hat{\lambda}) \equiv \frac{1}{n}\sum_{t=1}^{n}\rho'(\hat{\lambda}g_t)g_t = 0,$$

where g_t is evaluated at a consistent estimate $\tilde{\theta}$ of θ_0. If Assumptions 4 and 5 are satisfied, the second order Taylor expansion about $\hat{\lambda}=0$ with remainder exists and is given by

$$F_{n1}(\hat{\lambda}g_t) = F_{n1}(0) + DF_{n1}(0)\hat{\lambda} + \int_0^1 (1-\delta)D^2F_{n1}(\delta\hat{\lambda}g_t)\hat{\lambda}^2 d\delta,$$

where

$$\begin{aligned}D^2F_{n1}(\delta\hat{\lambda}) &= \frac{1}{n}\sum_{t=1}^{n}\rho'''(\delta\hat{\lambda}g_t)g_tg_tg_t \\ &= -\frac{1}{n}\sum_{t=1}^{n}g_tg_tg_t + o_p(1) \\ &\equiv -\hat{S} + o_p(1)\end{aligned}$$

by the continuity of $\rho'''(v)$, $\rho'''(0)=-1$, $\hat{\lambda}=O_p(n^{-1/2})$ and Lemma 1. See (Li and He, 2005) for more details on Taylor remainders in function spaces. $\hat{S}$, which represents the estimator of the skewness operator of g_t with kernel $\hat{s}(\tau_1,\tau_2,\tau_3) = 1/n\sum_t g_t(\tau_1)g_t(\tau_2)g_t(\tau_3)$, is bounded by assumption. Therefore,

$$\begin{aligned}D^2F_{n1}(\delta\hat{\lambda})\hat{\lambda}^2 &= -\hat{S}\hat{\lambda}^2 + o_p(1) \\ &= -\int_{\mathcal{T}}\int_{T}\hat{s}(\tau_1,\tau_2,\tau_3)\hat{\lambda}(\tau_1)\hat{\lambda}(\tau_2)\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2 \\ &= -\int_{\mathcal{T}}\int_{T}\hat{s}(\tau_1,\tau_2,\tau_3)O_p(n^{-1})\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2 \\ &= O_p(n^{-1})\end{aligned}$$

by the dominated convergence theorem. It follows that :

$$0 = F_{n1}(\hat{\lambda}) = F_{n1}(0) + DF_{n1}(0)\hat{\lambda} + O_p(\|\hat{\lambda}\|^2).$$

$F_{n1}(0)$ is the sample average of the moment function $\bar{g}(\theta)$ and $DF_{n1}(0)$ is the Fréchet derivative of $F_{n1}(\hat{\lambda})$ evaluated at $\lambda = 0$. We will prove that the following is indeed the Fréchet derivative of $F_{n1}(\hat{\lambda})$. Since it is an operator, it is defined by its kernel :

$$DF_{n1}(\hat{\lambda})(\tau_1, \tau_2) = \frac{1}{n}\sum_{t=1}^{n} \rho''(\hat{\lambda}g_t)g_t(\tau_1)g_t(\tau_2).$$

We want to show that :

$$\frac{\|F_{n1}(\hat{\lambda} + h) - F_{n1}(\hat{\lambda}) - DF_{n1}(\hat{\lambda})h\|}{\|h\|} \longrightarrow 0,$$

as h goes to zero for a fixed n. First we have :

$$\begin{aligned}(DF_{n1}(\hat{\lambda})h)(\tau) &= \frac{1}{n}\sum_{t=1}^{n} \rho''(\hat{\lambda}g_t)\int_{\mathcal{T}} g_t(\tau)g_t(\tau_1)h(\tau_1)\pi(\tau_1)d\tau_1 \\ &= \frac{1}{n}\sum_{t=1}^{n} \rho''(\hat{\lambda}g_t)[hg_t]g_t(\tau)\end{aligned}$$

and

$$F_{n1}(\hat{\lambda} + h) = \frac{1}{n}\sum \rho'(\hat{\lambda}g_t + hg_t)g_t.$$

Then,

$$\begin{aligned}&\frac{\|\frac{1}{n}\sum_{t=1}^{n} g_t\left(\rho'(\hat{\lambda}g_t + hg_t) - \rho'(\hat{\lambda}g_t) - \rho''(\hat{\lambda}g_t)hg_t\right)\|}{\|h\|} \\ \leq\; & \frac{1}{n}\sum_{t=1}^{n} \frac{\|g_t\left(\rho'(\hat{\lambda}g_t + hg_t) - \rho'(\hat{\lambda}g_t) - \rho''(\hat{\lambda}g_t)hg_t\right)\|}{\|h\|} \\ =\; & \frac{1}{n}\sum_{t=1}^{n} \left|\rho'(\hat{\lambda}g_t + hg_t) - \rho'(\hat{\lambda}g_t) - \rho''(\hat{\lambda}g_t)hg_t\right| \frac{\|g_t\|}{\|h\|} \\ =\; & \frac{1}{n}\sum_{t=1}^{n} o(h)\frac{\|g_t\|}{\|h\|} \\ & \longrightarrow 0,\end{aligned}$$

where the third equality comes from the continuity of $\rho'()$ and the theory of calculus for scalar functions. $DF_{n1}(\hat{\lambda})$ is therefore the Fréchet derivative of $F_{n1}(\hat{\lambda})$. Therefore, using Lemma 1 and Lemma 2, the Taylor approximation is :

$$0 = -\bar{g}(\theta) - \left(\frac{1}{n}\sum_{t=1}^{n} g_t(\theta)g_t(\theta)\right)\hat{\lambda} + O_p(n^{-1}) = -\bar{g}(\theta) + -\hat{K}\hat{\lambda} + O_p(n^{-1}).$$

A.2 Proofs

A.2.1 Theorem 1

The steps are similar to the proof of Theorem 3.1 of (Newey and Smith, 2004). They show the following lemma which applies also to our case.

Lemma 1. *If Assumption 2 is satisfied, then for any* ζ *with* $1/\nu < \zeta < 1/2$ *and* $\Lambda_\zeta = \{\lambda : \|\lambda\| < n^{-\zeta}\}$

$$\sup_{t,\lambda\in\Lambda_\zeta,\theta\in\Theta} |\lambda g_t(\theta)| \xrightarrow{P} 0,$$

and $\Lambda_\zeta \subseteq \Lambda_n$ *w.p.a.1.*

However, the proof of the following lemma is different because $\hat{\lambda}$ is a regularized solution. Therefore, we cannot define it as $\arg\max P(\lambda, \bar{\theta})$ for some convergent $\bar{\theta}$.

Lemma 2. *If Assumptions 1 to 4 are satisfied, and if* $\bar{\theta} \xrightarrow{P} \theta_0$ *and* $\bar{g}(\bar{\theta}) = O_p(n^{-1/2})$*, then*

$$\bar{\lambda} = \arg\min_{\Lambda_n} V(\lambda, \bar{\theta}) \equiv \left(\|F_{n1}(\lambda, \bar{\theta})\|^2 + \alpha_n\|\lambda\|^2\right)$$

exists w.p.a.1, $\bar{\lambda} = O_p(n^{-1/2})$ *and* $P(\bar{\lambda}, \bar{\theta}) \leq \rho_0 + O_p(n^{-1})$*, where* $\rho_0 = \rho(0)$*.*

Proof. *Let us define* $\tilde{\lambda} = \arg\min_{\Lambda_n} V(\lambda, \bar{\theta})$*. Then :*

$$\begin{aligned} V(0, \bar{\theta}) &\geq V(\tilde{\lambda}, \bar{\theta}) \\ &= V(0, \bar{\theta}) + V'(0, \bar{\theta})\tilde{\lambda} + \left[\int_0^1 (1-\delta)V''(\delta\tilde{\lambda}, \bar{\theta})d\delta\right]\tilde{\lambda}^2, \end{aligned}$$

where

$$V(0, \bar{\theta}) = \|F_{n1}(0, \bar{\theta})\|^2 = \|\bar{g}(\bar{\theta})\|^2,$$

$$V'(0, \bar{\theta}) = 2DF_{n1}(0, \bar{\theta})F_{n1}(0, \bar{\theta}) = 2\hat{K}(\bar{\theta})\bar{g}(\bar{\theta})$$

and

$$\begin{aligned}V''(\delta\tilde{\lambda},\bar{\theta}) &= 2\left[D^2F_{n1}(\delta\tilde{\lambda},\bar{\theta})F_{n1}(\delta\tilde{\lambda},\bar{\theta}) + DF_{n1}(\delta\tilde{\lambda},\bar{\theta})^2 + \alpha_n I\right] \\ &= 2\left[\frac{1}{n}\sum_{t=1}^{n}\rho'''(\delta\tilde{\lambda}g_t(\bar{\theta}))g_t(\bar{\theta})g_t(\bar{\theta})g_t(\bar{\theta})\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho'(\delta\tilde{\lambda}g_t(\bar{\theta}))g_t(\bar{\theta})\right] \\ &\quad +2\left[\sum_{t=1}^{n}\rho''(\delta\tilde{\lambda}g_t(\bar{\theta}))g_t(\bar{\theta})g_t(\bar{\theta})\right]^2 + 2\alpha_n I \\ &= 2\left[\hat{S}(\bar{\theta})\bar{g}(\bar{\theta}) + \hat{K}(\bar{\theta})^2 + \alpha_n I + o_p(1)\right],\end{aligned}$$

by Lemma 1 since it implies, with the properties of $\rho(v)$, that $\rho'(\delta\tilde{\lambda}g_t(\bar{\theta}))$, $\rho''(\delta\tilde{\lambda}g_t(\bar{\theta}))$ and $\rho'''(\delta\tilde{\lambda}g_t(\bar{\theta}))$ converge in probability to -1. The term $\hat{S}(\bar{\theta})\bar{g}(\bar{\theta})$ is a linear operator with kernel $\int_{\mathcal{T}}\hat{s}(\tau_1,\tau_2,\tau)\bar{g}(\tau;\bar{\theta})\pi(\tau)d\tau$, where $\hat{s}(\tau_1,\tau_2,\tau) = 1/n\sum_t g_t(\tau_1;\bar{\theta})g_t(\tau_2;\bar{\theta})g_t(\tau;\bar{\theta})$. It is $O_p(n^{-1/2})$ by the assumption on $\bar{g}(\bar{\theta})$, boundness of the skewness operator and dominated convergence theorem. It follows that

$$V''(\delta\tilde{\lambda},\bar{\theta}) = \hat{K}(\bar{\theta})^2 + \alpha_n I + o_p(1),$$

and then,

$$\begin{aligned}\int_0^1(1-\delta)V''(\delta\tilde{\lambda},\bar{\theta})d\delta &= \left[\hat{K}(\bar{\theta})^2 + \alpha_n I + o_p(1)\right]\int_0^1(1-\delta)d\delta \\ &= \frac{1}{2}\left[\hat{K}(\bar{\theta})^2 + \alpha_n I + o_p(1)\right],\end{aligned}$$

where $\hat{K}(\bar{\theta})^2 + \alpha_n I$ is a strictly positive definite linear operator since $\alpha_n > 0$. It follows that

$$\begin{aligned}\|\bar{g}(\bar{\theta})\|^2 &\geq \|\bar{g}(\bar{\theta})\|^2 + 2\hat{K}(\bar{\theta})\bar{g}(\bar{\theta})\tilde{\lambda} + [\hat{K}(\bar{\theta})^2 + \alpha_n I + o_p(1)]\tilde{\lambda}^2 \\ 0 &\geq 2\hat{K}(\bar{\theta})\bar{g}(\bar{\theta})\tilde{\lambda} + [\hat{K}(\bar{\theta})^2 + \alpha_n I + o_p(1)]\tilde{\lambda}^2 \\ 0 &\leq -2\hat{K}(\bar{\theta})\bar{g}(\bar{\theta})\tilde{\lambda} - [\hat{K}(\bar{\theta})^2 + \alpha_n I + o_p(1)]\tilde{\lambda}^2 \\ &\leq C_1\|\bar{g}(\bar{\theta})\|\|\tilde{\lambda}\| - [\hat{K}(\bar{\theta})^2 + \alpha_n I + o_p(1)]\tilde{\lambda}^2 \\ &\leq C_1\|\bar{g}(\bar{\theta})\|\|\tilde{\lambda}\| - C_2\|\tilde{\lambda}\|^2 - o_p(\|\tilde{\lambda}\|^2),\end{aligned}$$

where $C1 > 0$ and C_2 is the smallest eigenvalue of $(\hat{K}(\bar{\theta})^2 + \alpha_n I)$. Therefore, we have

$$C_2\|\tilde{\lambda}\|^2 + o_p(\|\tilde{\lambda}\|^2) \leq C_1\|\bar{g}(\bar{\theta})\|\|\tilde{\lambda}\|$$

$$C_2\|\tilde{\lambda}\| + o_p(\|\tilde{\lambda}\|) \leq C_1\|\bar{g}(\bar{\theta})\| = O_p(n^{-1/2}),$$

which implies $\|\tilde{\lambda}\| = O_p(n^{-1/2})$. *Notice that without* α_n, *the rate of convergence of* $\tilde{\lambda}$ *is undefined since nothing guarantees that the smallest eigenvalue of* $\hat{K}^2$ *is strictly positive. However, the eigenvalues of* K *are strictly positive. Therefore,* α_n *must go to zero at a speed slower than* $O_p(\|\hat{K}^2 - K^2\|) = O_p(n^{-1})$ *(see (Carrasco and Florens, 2000)). Because* $\Lambda_\zeta \subseteq \Lambda_n$ *w.p.a.1. The second result follows.*

Notice that if it was not for the restriction imposed by the domain of $\rho(v)$, *the solution would exist in small sample as well because* α_n *guarantees that the problem is well-posed. This restriction applies only to CEL because in this case the domain of* $\rho(v)$ *is* $]-\infty, 1[$. *For the other CGEL methods considered here, the solution exists always. This is shown using Theorem 1 of (Seidman and Vogel, 1989).*

If we substitute the solution in the objective function of CGEL we obtain :

$$\begin{aligned} P(\bar{\lambda},\bar{\theta}) &= \rho_0 - \bar{g}(\bar{\theta})\bar{\lambda} + \left[\int_0^1 (1-\delta)\left(\frac{1}{n}\sum_{t=1}^n \rho''(\delta\bar{\lambda}g_t(\bar{\theta}))g_t(\bar{\theta})g_t(\bar{\theta})\right)d\delta\right]\bar{\lambda}^2 \\ &\leq \rho_0 + \|\bar{g}(\bar{\theta})\|\|\bar{\lambda}\| + C\|\bar{\lambda}\|^2 \\ &= \rho_0 + O_p(n^{-1}) \end{aligned}$$

by Lemma 1 and the above results. □

Lemma 3. *If Assumptions 1 to 4 are satisfied, then* $\bar{g}(\hat{\theta}) = O_p(n^{-1/2})$.

Proof. *All we need is to show that we can obtain the same inequality as in Lemma A3 of (Newey and Smith, 2004). Let* $\tilde{\lambda} = -n^{-\zeta}\bar{g}(\hat{\theta})/\|\bar{g}(\hat{\theta})\|$, *then by Lemma A3 of (Newey and Smith, 2004) :*

$$P(\tilde{\lambda},\hat{\theta}) \geq \rho_0 + n^{-\zeta}\|\bar{g}(\hat{\theta})\| - Cn^{-2\zeta}$$

Because $\hat{\lambda}$ *solves the regularized first order condition, we cannot say that* $P(\hat{\lambda},\hat{\theta}) \geq P(\tilde{\lambda},\hat{\theta})$. *But it holds w.p.a.1 because as* α_n *goes to zero, the first order condition for* λ *converges to zero. Therefore, we have, w.p.a.1,*

$$\rho_0 + n^{-\zeta}\|\bar{g}(\hat{\theta})\| - Cn^{-2\zeta} \leq P(\tilde{\lambda},\hat{\theta}) \leq P(\hat{\lambda},\hat{\theta}) \leq P(\hat{\lambda},\theta_0) \leq \rho_0 + O_p(n^{-1})$$

□

Proof (Proof of Theorem 1)**.** *The proof is straightforward using Lemma 1 to 3 and using the same arguments as (Newey and Smith, 2004) for Theorem 3.1.* □

A.2.2 Theorem 2

In order to prove asymptotic normality we first recall the regularized first order conditions :

$$DF(\lambda,\theta)F(\lambda,\theta)+\alpha_n\lambda=0, \tag{A.1}$$

$$\frac{1}{n}\sum_{t=1}^{n}\rho'\left(\lambda g_t\right)\lambda G_t=0, \tag{A.2}$$

where θ has been explicitly included in $F()$ because we will have to expand it around $\lambda=0$ and θ_0. Notice that the subscript of $F_{n1}()$ has been omitted for notational convenience. $DF()$ is the Fréchet derivative of $F()$. It is an integral operator with kernel :

$$DF(\tau_1,\tau_2)=\frac{1}{n}\sum_{t=1}^{n}\rho''(\lambda g_t)g_t(\tau_1)g_t(\tau_2).$$

It follows that :

$$[DF(\lambda,\theta)F(\lambda,\theta)](\tau)=\int_{\mathcal{T}}\left\{\frac{1}{n}\sum_{t=1}^{n}\rho''(\lambda g_t)g_t(\tau)g_t(\tau_2)\right\}\left\{\frac{1}{n}\sum_{s=1}^{n}\rho'(\lambda g_s)g_s(\tau_2)\right\}\pi(\tau_2)d\tau_2.$$

We will denote $F'()$ as the derivative of $F()$ with respect to θ. It is an operator from $L^2(\pi)$ to $\mathbb{R}^p$ or from $\mathbb{R}^p$ to $L^2(\pi)$ depending on what turns out to be in front of it. It should always be clear from the context. We first expand equation (A.1) about $\lambda=0$ and $\theta=\theta_0$. We denote F_0, DF_0, F_0' and so on, as the operators evaluated at the true value :

$$\begin{aligned}0 &= DF(\hat{\lambda},\hat{\theta})F(\hat{\lambda},\hat{\theta})+\alpha_n\hat{\lambda}\\ &= DF_0F_0+\left[D^2F_0F_0+DF_0DF_0+\alpha_nI\right]\hat{\lambda}\\ &\quad+\left[DF_0'F_0+DF_0F_0'\right](\hat{\theta}-\theta_0)+O_p(\|\hat{\lambda}\|^2+\|\hat{\theta}-\theta_0\|^2),\end{aligned}$$

where D^2F is the Fréchet derivative of DF and DF' if the derivative of DF with respect to θ. Let us develop each term one by one (recall that $\rho'(0)=\rho''(0)=\rho'''(0)=-1$:

$$DF_0=\frac{1}{n}\sum_{t=1}^{n}\rho''(0)g_tg_t=-\hat{K}(\theta_0)$$

$$F_0 = \frac{1}{n}\sum_{t=1}^{n}\rho'(0)g_t = -\bar{g}(\theta_0).$$

It follows that :

$$DF_0F_0 = \hat{K}(\theta_0)\bar{g}(\theta_0)$$

and

$$D^2F_0 = \frac{1}{n}\sum_{t=1}^{n}\rho'''(0)g_tg_tg_t = -\hat{S}(\theta_0),$$

where $\hat{S}$ if the estimated skewness operator. It follows that :

$$D^2F_0F_0 = \hat{S}(\theta_0)\bar{g}(\theta_0).$$

The other terms can be obtained in the same way. The expansion of the regularized first order conditions is then :

$$\begin{aligned} 0 &= \hat{K}_0\bar{g}_0 + \left\{\hat{K}_0^2 + \hat{S}_0\bar{g}_0 + \alpha_n I\right\}\hat{\lambda} \\ &\quad + \left\{[\bar{G}g_0 + g\bar{G}_0]\bar{g}_0 + \hat{K}_0\bar{G}_0\right\}(\hat{\theta} - \theta_0) + O_p(n^{-1}). \end{aligned}$$

The second equation can be expanded in the same way :

$$0 = -\bar{G}_0\hat{\lambda} + O_p(n^{-1}).$$

We can rewrite the above equations in the following compact representation :

$$0 = -B + A_1\hat{\lambda} + A_2(\hat{\theta} - \theta_0) + O_p(n^{-1}) \qquad \text{(A.3)}$$

$$0 = A_3\hat{\lambda} + O_p(n^{-1}), \qquad \text{(A.4)}$$

where :

$$B = -\hat{K}_0\bar{g}_0,$$

$$A_1 = (\hat{K}_0^2 + \alpha_n I) + O_p(n^{-1}),$$

because $\hat{S}_0\bar{g}_0$ and $\hat{\lambda}_n$ are $O_p(n^{-1/2})$,

$$A_2 = \hat{K}_0\bar{G}_0 + O_p(n^{-1/2})$$

and

$$A_3 = \bar{G}_0.$$

We can solve the system to obtain the following :

$$\sqrt{n}\hat{\lambda} = [I - A_1^{-1}A_2(A_3A_1^{-1}A_2)^{-1}A_3]A_1^{-1}\sqrt{n}B + o_p(1)$$

and

$$\sqrt{n}(\hat{\theta} - \theta_0) = (A_3A_1^{-1}A_2)^{-1}A_3A_1^{-1}\sqrt{n}B + o_p(1).$$

If we analyze the last term of each equation, we have :

$$\begin{aligned} -A_1^{-1}\sqrt{n}B &= (\hat{K}_0^2 + \alpha_n I)^{-1}\hat{K}_0[\sqrt{n}\bar{g}_0] + o_p(n^{-1}) \\ &= (\hat{K}_0^{\alpha_0})^{-1}[\sqrt{n}\bar{g}_0] + o_p(n^{-1}) \\ &= K^{-1}g + \left\{(\hat{K}_0^{\alpha_n})^{-1}[\sqrt{n}\bar{g}_0] - K^{-1}g\right\} + o_p(n^{-1}) \\ &= K^{-1}g + o_p(1) \end{aligned}$$

as n goes to infinity, α_n goes to zero and $n\alpha_n^3 \to \infty$ by theorem 7 (ii) of (Carrasco and Florens, 2000), where $g \sim N(0, K)$. Therefore, $A_1^{-1}\sqrt{n}B$ converges to $N(0, K^{-1})$ (See appendix A.1.1 for details on $\hat{K}^{\alpha_n}$). Using the convergence properties of A_1, A_2 and A_3, we obtain :

$$\sqrt{n}\hat{\lambda} \xrightarrow{L} \left[I - K^{-1}G(GK^{-1}G)^{-1}G\right] N(0, K^{-1})$$

and

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{L} \left[(GK^{-1}G)^{-1}G\right] N(0, K^{-1}).$$

The rest of the proof follows by simple manipulations.

A.2.3 Theorem 4

In Appendix A.1.2, it is shown that

$$\hat{K}\hat{\lambda} = -\bar{g}(\hat{\theta}) + o_p(1).$$

It follows that

$$(\hat{K}^{\alpha_n})^{-1/2}\hat{K}\hat{\lambda} = -(\hat{K}^{\alpha_n})^{-1/2}\bar{g}(\hat{\theta}) + o_p(1).$$

Theorem 7 (i) of (Carrasco and Florens, 2000) implies that

$$(\hat{K}^{\alpha_n})^{-1/2}\hat{K}f_n = \hat{K}^{1/2}f_n + o_p(1).$$

which proves the first order equivalence of $\|\sqrt{n}\hat{K}^{1/2}\hat{\lambda}\|$ and $\|\sqrt{n}(\hat{K}^{\alpha_n})^{-1/2}\bar{g}(\hat{\theta})\|$.

In order to show the first order equivalence of $\widetilde{LR}$, we expand the CGEL objective function about $\lambda = 0$:

$$\begin{aligned}
2nP(\hat{\lambda},\hat{\theta}) &= 2nP(0,\hat{\theta}) + 2nP_{\lambda}(0,\hat{\theta})\hat{\lambda} + 2n\hat{\lambda}P_{\lambda\lambda}(\tilde{\lambda},\hat{\theta})\hat{\lambda} \\
&= 2\rho(0) - 2n\bar{g}(\hat{\theta})\hat{\lambda} + n\hat{\lambda}\left(\frac{1}{n}\sum_{t=1}^{n}\rho''(\tilde{\lambda}g_t(\hat{\theta}))g_t(\hat{\theta})g_t(\hat{\theta})\right)\hat{\lambda} \\
&= 2\rho(0) - n\hat{\lambda}\left(\frac{1}{n}\sum_{t=1}^{n}g_t(\hat{\theta})g_t(\hat{\theta})\right)\hat{\lambda} + o_p(1) \\
&= 2\rho(0) - n\hat{\lambda}\hat{K}\hat{\lambda} + o_p(1),
\end{aligned}$$

where $\tilde{\lambda} \in [0,\hat{\lambda}]$.

The second part of the theorem follows by simple manipulation using the singular value representation of the inverse problem solution. The CGMM objective function can be written as :

$$\|(\hat{K}^{\alpha_n})^{-1/2}\sqrt{n}\bar{g}(\hat{\theta})\|^2 = \sum_{i=1}^{n}\left(\frac{\mu_i^{(n)^2}}{\mu_i^{(n)^2}+\alpha_n}\right)\frac{<\bar{g}(\theta),\phi_i^{(n)}>^2}{\mu_i^{(n)}},$$

where

$$\phi_i^{(n)} = \frac{\nu_i^{(n)}}{\|\nu_i^{(n)}\|},$$

and

$$\begin{aligned}
\|\nu_i^{(n)}\|^2 &= \left\langle \frac{1}{n}\sum_{j=1}^{n}\beta_{ji}g_j, \frac{1}{n}\sum_{j=1}^{n}\beta_{ji}g_j \right\rangle \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}\beta_{ji}\beta_{li}\int g_j(\tau)g_j(\tau)\pi(\tau)d\tau \\
&= \frac{1}{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\beta_{ji}\beta_{li}C_{jl} \\
&= \frac{1}{n}\sum_{j=1}^{n}\beta_{ji}\left(C_{j\bullet}\beta_i\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\beta_{ji}\left(\mu_i^{(n)}\beta_{ji}\right)
\end{aligned}$$

$$= \frac{\mu_i^{(n)}}{n}\|\beta_i\|^2 = \frac{\mu_i^{(n)}}{n}.$$

It follows that

$$\begin{aligned}
< \sqrt{n}\bar{g}(\hat{\theta}), \phi_i^{(n)} > &= \left\langle \frac{1}{\sqrt{n}}\sum_{t=1}^n g_t, \frac{1}{\sqrt{n\mu_i^{(n)}}}\sum_{j=1}^n \beta_{ji} g_j \right\rangle \\
&= \frac{1}{n\sqrt{\mu_i^{(n)}}}\sum_{t=1}^n\sum_{j=1}^n \beta_{ji}\int g_t(\tau)g_j(\tau)\pi(\tau)d\tau \\
&= \frac{1}{\sqrt{\mu_i^{(n)}}}\sum_{t=1}^n\sum_{jk=1}^n \beta_{ji}C_{tj} \\
&= \frac{1}{\sqrt{\mu_i^{(n)}}}\sum_{t=1}^n C_{t\bullet}\beta_i = \frac{1}{\sqrt{\mu_i^{(n)}}}\sum_{t=1}^n \mu_i^{(n)}\beta_{ti} \\
&= \sqrt{\mu_i^{(n)}}\iota'\beta_i.
\end{aligned}$$

Therefore, the CGMM objective function becomes :

$$\|(\hat{K}^\alpha)^{-1/2}\sqrt{n}\bar{g}(\hat{\theta})\|^2 = \sum_{i=1}^n \left(\frac{\mu_i^{(n)^2}}{\mu_i^{(n)^2}+\alpha_n}\right)(\iota'\beta_i)^2,$$

which concludes the proof for $\tilde{J}$. The proof of the $\widetilde{LM}$ representation is much simpler :

$$\begin{aligned}
\hat{\lambda}\hat{K}\hat{\lambda} &= \int\int \hat{\lambda}(\tau_1)\hat{\lambda}(\tau_2)\left(\frac{1}{n}\sum_{t=1}^n g_t(\tau_1)g_t(\tau_2)\right)\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2 \\
&= \frac{1}{n}\sum_{t=1}^n \int \hat{\lambda}(\tau_1)g_t(\tau_1)\pi(\tau_1)d\tau_1 \int \hat{\lambda}(\tau_2)g_t(\tau_2)\pi(\tau_2)d\tau_2 \\
&= \frac{1}{n}\sum_{t=1}^n (\hat{\lambda}g_t)^2.
\end{aligned}$$

The result follows.

A.3 Computation of CGEL

A.3.1 Computation using the singular value decomposition.

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

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

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

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

where :

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

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

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

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

$$< g_t, \tilde{\lambda} > = -\sum_{i=1}^{m}\left(\frac{\mu_i^{(n)}}{\mu_i^{(n)^2}+\alpha_n}\right)\left[\frac{1}{n}\iota' C\beta_i\right][C_{t\bullet}\beta_i]$$

$$= -\frac{1}{n}\left[\iota' C\right]\sum_{i=1}^{m}\left(\frac{\mu_i^{(n)}}{\mu_i^{(n)2}+\alpha_n}\right)\left[\beta_i\beta_i'\right]C_{\bullet t}$$
$$= -\frac{1}{n}\iota' C\left(\beta D\beta'\right)C_{\bullet t},$$

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

$$\frac{1}{n}\sum_{t=1}^{n}\rho\left(-\frac{1}{n}\iota' C\left(\beta D\beta'\right)C_{\bullet t}\right).$$

A.3.2 Computation using the regularized Gauss-Newton method

To simplify the notation, we will set $g_t \equiv g_t(\theta)$, $\lambda \equiv \lambda_{i-1}$, $\lambda' \equiv \lambda_i$, $p_t = \rho'(\lambda g_t)$, $p_t' = \rho'(\lambda' g_t)$, $p_t^2 = \rho''(\lambda g_t)$ and $p_t'^2 = \rho''(\lambda' g_t)$. We want to rewrite the following algorithm :

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

which can be written as :

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

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

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

Once we apply the transformation, the term becomes :

$$\begin{aligned}\left[DF(\lambda)^2\lambda'\right]g_s &= \int_{\mathcal{T}}\frac{1}{n}\sum_t\sum_l p_t^2 p_l^2 g_t(\tau_1)C_{tl} < g_l, \lambda' > g_s(\tau_1)\pi(\tau_1)d\tau_1 \\ &= \frac{1}{n}\sum_t\sum_l p_t^2 p_l^2 C_{tl} < g_l, \lambda' > \int_{\mathcal{T}} g_t(\tau_1)g_s(\tau_1)\pi(\tau_1)d\tau_1\end{aligned}$$

$$\begin{aligned} &= \sum_t \sum_l p_t^2 p_l^2 C_{tl} < g_l, \lambda' > C_{ts} \\ &= \sum_t C_{ts} p_t^2 \left[C_{t\bullet} V < g, \lambda' > \right] \\ &= C_{s\bullet} VCV < g, \lambda' >, \end{aligned}$$

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

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

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

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

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

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

The second term on the right hand side is :

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

If we apply the transformation it becomes :

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

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

$$CVCP,$$

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

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

which implies

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

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

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

A.4 CCUE and CEEL

A.4.1 Note on CCUE

In Section 1.4.1, we argue that the exact solution of $\hat{\lambda}(\theta)$, in the case of CEEL, can be obtained from the linear ill-posed problem $\hat{K}\hat{\lambda} = -\bar{g}(\theta)$. In this case, the iterative procedure stops after the first iteration and the solution is $\hat{\lambda}(\theta) = -(\hat{K}^{\alpha_n})^{-1}\bar{g}(\theta)$. Because $\rho()$ is quadratic, we can write the objective function as :

$$\begin{aligned} P(\hat{\lambda}(\theta),\theta) &= \rho(0)+\bar{g}(\hat{K}^{\alpha_n})^{-1}\bar{g}-\bar{g}\hat{K}(\hat{K}^{\alpha_n})^{-2}\bar{g}/2 \\ &= \rho(0)+\frac{1}{2}\bar{g}(\hat{K}^{\alpha_n})^{-1}\bar{g}+o_p(1), \end{aligned}$$

because $\hat{K}(\hat{K}^{\alpha_n})^{-2} = (K_n^{\alpha_n})^{-1}+o_p(1)$. Therefore, CEEL is equivalent to CCUE, defined as CGMM in which $\hat{K}^{\alpha_n}(\tilde{\theta})$ is replaced by $\hat{K}^{\alpha_n}(\theta)$, only asymptotically.

A.4.2 CEEL and the ill-posedness of CGEL

The case in which $\rho(v)$ is quadratic offers a way to show that linear and nonlinear ill-posed problems are very different. If we consider the following system of n linear equations $Ax = y$, in which the matrix A is poorly conditioned, the stability of the solution is an issue only if the right-hand side is random. In CGMM, we need the solution to $Kx = \bar{g}$ in order to compute the objective function. Because $\bar{g}$ is random, the properties of K imply that the system is ill-posed [1]. In nonlinear system of equations, the problem can be ill-posed even if the right-hand side is not random as in equation (1.13). For example, when the number of conditions is finite and $\rho(v) = -v - 0.5v^2$, the equation is :

$$F_{n1}(\lambda) \equiv \frac{1}{n}\sum_{t=1}^{n}(-1 - g_t'\lambda)g_t = 0,$$

which implies that $\hat{\lambda}(\theta)$ is the solution to the following system of linear equations :

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

Since $\bar{g}$ is random, the solution is unstable if $\hat{K}$ is poorly conditioned. For the case of a continuum, it is ill-posed by the properties of the covariance operator. The randomness of the left-hand side $F_{n1}(\lambda)$ is therefore as important as the randomness of the right-hand side for the stability of the solution, as opposed to linear ill-posed problem. Equation (1.13) is therefore ill-posed.

1. Notice that we can have ill-posed problems even if the right hand side is not random. See (Carrasco, Florens and Renault, 2007) for some examples.

ANNEXE B

APPENDIX : CHAPTER 2

B.1 Computation of the covariance matrix

The estimate of $[GK^{-1}G]$ is

$$\begin{aligned}
[\bar{G}\hat{K}_\alpha^{-1}\bar{G}] &= \left\langle \bar{G}, \left[\sum_{t=1}^{n} \frac{\mu_t}{\mu_t^2 + \alpha_n} \left\langle \phi_t, \bar{G} \right\rangle \phi_t \right] \right\rangle \\
&= \sum_{t=1}^{n} \frac{\mu_t}{\mu_t^2 + \alpha_n} \left\langle \phi_t, \bar{G} \right\rangle^2 ,
\end{aligned}$$

where μ_t are the eigenvalues of the matrix C. Since β is the matrix of eigenvectors of C, we have :

$$\begin{aligned}
< \phi_t, \bar{G} > &= \left\langle \frac{1}{n} \sum_{l=1}^{n} \beta_{lt} g_l(\hat{\theta}), \bar{G} \right\rangle \\
&= \int_{\mathbb{R}} \left(\frac{1}{n} \sum_{l=1}^{n} \beta_{lt} g_l(\hat{\theta}; \tau) \bar{G}(\tau) d\pi(\tau) \right) \\
&= \frac{1}{n} \sum_{l=1}^{n} \beta_{lt} \int_{\mathbb{R}} g_l(\hat{\theta}; \tau) \overline{\bar{G}(\tau)} d\pi(\tau) \\
&= \frac{1}{n} \sum_{l=1}^{n} \beta_{lt} \int_{\mathbb{R}} (y_l - \hat{d} w_l) e^{ix_l\tau} \left(\frac{-1}{n} \sum_{s=1}^{n} w_s e^{-ix_s\tau} \right) d\pi(\tau) \\
&= \frac{-1}{n^2} \sum_{l=1}^{n} \beta_{lt} (y_l - \hat{d} w_l) \sum_{s=1}^{n} w_s \int_{\mathbb{R}} e^{i(x_l - x_s)\tau} d\pi(\tau) \\
&= \frac{-1}{n^2} \sum_{l=1}^{n} \beta_{lt} (y_l - \hat{d} w_l) \sum_{s=1}^{n} w_s \left(e^{-(x_l - x_s)^2/2} \right) ,
\end{aligned}$$

where the last result comes from the fact that $\pi(\tau)$ is the N(0,1) distribution. Defining the $n \times n$ symetric matrix H as $H_{tl} = e^{-(x_t - x_l)^2/2}/n$ and the $n \times 1$ vector $w = \{w_1, w_2, ..., w_n\}'$ allows us to simplify the above result as :

$$< \phi_t, \bar{G} >= -\frac{1}{n}\sum_{l=1}^{n} \beta_{lt}(y_l - \hat{d}w_l)\left(w'H_{\bullet l}\right)$$

We can simplify further by defining the diagonal matrices B and D as $B_{jj} = (y_j - \hat{d}w_j)$ and $D_{jj} = (\mu_j/(\mu_j^2 + \alpha_n)$ respectively. It follows that :

$$< \phi_t, \bar{G} >= -\frac{1}{n}w'H(B\beta_{\bullet t}),$$

and then

$$[\bar{G}\hat{K}_\alpha^{-1}\bar{G}] = \frac{w'HB\beta D\beta' BHw}{n^2}. \tag{B.1}$$

Notice that before applying the above formula, the eigenvectors β_i must be normalized to make $\|\phi_i\| = 1$. In fact, if we use directly the eigenvectors of the matrix C and apply the above definition of ϕ_i, we obtain :

$$\|\phi_i\|^2 = \frac{1}{n^2}\sum_{t,s=1}^{n} \beta_{ti}\beta si \int g_t(\tau)g_s(\tau)\pi(\tau) = \frac{1}{n}\beta_i' C\beta_i = \frac{\mu_i}{n}$$

We should therefore redefine β_i as $\sqrt{n}\beta_i/\sqrt{\mu_i}$. However, one needs to be careful with that transformation because if the eigenvalues are very close to zero, nothing guaranties that $\|\phi_i\| = 1$. One approach is to drop all eigenvalues for which the transformed $\|\phi_i\|$ is sufficiently close to 1.

B.2 Results : Stable distribution

Statistics		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
Mean	$\alpha = 0.1$	1.69489	1.70224	1.70515	1.72307	1.64344	1.63873	1.63694
	$\alpha = 0.05$	1.69849	1.71083	1.72337	1.73850	1.64147	1.63445	1.64546
	$\alpha = 0.01$	1.70748	1.74408	1.76147	1.77918	1.66258	1.67036	1.66967
	$\alpha = 0.005$	1.70967	1.74968	1.77106	1.79081	1.66614	1.66380	1.66497
	$\alpha = 0.001$	1.70975	1.79151	1.81687	1.83675	1.67559	1.67731	1.67603
	$\alpha = 0.0001$	1.68599	1.84470	1.86625	1.88307	1.69153	1.69185	1.68991
Median	$\alpha = 0.1$	1.69873	1.71266	1.71561	1.73033	1.64526	1.63859	1.63858
	$\alpha = 0.05$	1.71026	1.71683	1.73613	1.74747	1.64416	1.63208	1.65248
	$\alpha = 0.01$	1.71957	1.75064	1.76343	1.78146	1.67085	1.67324	1.67297
	$\alpha = 0.005$	1.72058	1.75137	1.77142	1.79856	1.67642	1.66637	1.67456
	$\alpha = 0.001$	1.72185	1.79419	1.82366	1.84109	1.67997	1.67684	1.67783
	$\alpha = 0.0001$	1.70995	1.85933	1.88253	1.89608	1.70213	1.70212	1.69980
S-dev	$\alpha = 0.1$	0.15966	0.14773	0.14065	0.13830	0.15817	0.16282	0.16313
	$\alpha = 0.05$	0.16467	0.13746	0.13515	0.13250	0.16334	0.16181	0.16622
	$\alpha = 0.01$	0.16845	0.13519	0.12252	0.11757	0.15503	0.15594	0.15351
	$\alpha = 0.005$	0.16564	0.12588	0.11562	0.11541	0.16505	0.16236	0.15699
	$\alpha = 0.001$	0.17789	0.11064	0.10065	0.09672	0.14991	0.15441	0.15451
	$\alpha = 0.0001$	0.19474	0.09151	0.08640	0.08385	0.14382	0.14464	0.15203
Mean-bias	$\alpha = 0.1$	0.00511	0.00224	0.00515	0.02307	0.05656	0.06127	0.06306
	$\alpha = 0.05$	0.00151	0.01083	0.02337	0.03850	0.05853	0.06555	0.05454
	$\alpha = 0.01$	0.00748	0.04408	0.06147	0.07918	0.03742	0.02964	0.03033
	$\alpha = 0.005$	0.00967	0.04968	0.07106	0.09081	0.03386	0.03620	0.03503
	$\alpha = 0.001$	0.00975	0.09151	0.11687	0.13675	0.02441	0.02269	0.02397
	$\alpha = 0.0001$	0.01401	0.14470	0.16625	0.18307	0.00847	0.00815	0.01009
Median-bias	$\alpha = 0.1$	0.00127	0.01266	0.01561	0.03033	0.05474	0.06141	0.06142
	$\alpha = 0.05$	0.01026	0.01683	0.03613	0.04747	0.05584	0.06792	0.04752
	$\alpha = 0.01$	0.01957	0.05064	0.06343	0.08146	0.02915	0.02676	0.02703
	$\alpha = 0.005$	0.02058	0.05137	0.07142	0.09856	0.02358	0.03363	0.02544
	$\alpha = 0.001$	0.02185	0.09419	0.12366	0.14109	0.02003	0.02316	0.02217
	$\alpha = 0.0001$	0.00995	0.15933	0.18253	0.19608	0.00213	0.00212	0.00020
RMSE	$\alpha = 0.1$	0.15966	0.14768	0.14067	0.14014	0.16790	0.17389	0.17482
	$\alpha = 0.05$	0.16460	0.13782	0.13709	0.13792	0.17343	0.17451	0.17486
	$\alpha = 0.01$	0.16853	0.14213	0.13702	0.14170	0.15941	0.15865	0.15640
	$\alpha = 0.005$	0.16584	0.13527	0.13566	0.14681	0.16840	0.16627	0.16078
	$\alpha = 0.001$	0.17807	0.14354	0.15421	0.16747	0.15181	0.15599	0.15628
	$\alpha = 0.0001$	0.19515	0.17119	0.18734	0.20134	0.14400	0.14480	0.15229

Table B.1 Properties of the estimator of ω for a sample size of 100

Statistics		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
Mean	α = 0.1	0.36000	0.47102	0.46902	0.47447	0.45157	0.47250	0.47053
	α =0.05	0.37133	0.46180	0.48694	0.51533	0.47904	0.47332	0.47482
	α = 0.01	0.31670	0.48323	0.47320	0.48112	0.46299	0.46763	0.46528
	α = 0.005	0.30179	0.46385	0.47390	0.49016	0.50630	0.48215	0.48258
	α = 0.001	0.27162	0.45651	0.42992	0.48261	0.49868	0.50548	0.49953
	α = 0.0001	0.22802	0.47037	0.50750	0.49584	0.52892	0.53156	0.50761
Median	α = 0.1	0.43002	0.54613	0.54240	0.54897	0.45479	0.46427	0.45676
	α = 0.05	0.48171	0.53456	0.57287	0.61374	0.45813	0.46615	0.47738
	α = 0.01	0.41724	0.58898	0.57669	0.61146	0.45968	0.48165	0.47150
	α = 0.005	0.41978	0.58246	0.59298	0.61495	0.50346	0.49615	0.49762
	α = 0.001	0.35726	0.59316	0.55920	0.62638	0.53139	0.52833	0.52550
	α = 0.0001	0.28476	0.65297	0.68696	0.68784	0.57053	0.57520	0.55335
S-dev	α = 0.1	0.52951	0.48795	0.48368	0.49725	0.39350	0.36492	0.37242
	α = 0.05	0.54823	0.48049	0.48479	0.48862	0.38013	0.38463	0.37718
	α = 0.01	0.54713	0.51383	0.52276	0.54408	0.39364	0.38819	0.39569
	α = 0.005	0.55005	0.52064	0.53068	0.54790	0.38625	0.38866	0.39654
	α = 0.001	0.53292	0.54423	0.56013	0.56475	0.39271	0.40288	0.41325
	α = 0.0001	0.48979	0.57087	0.55963	0.58086	0.38895	0.38313	0.41669
Mean-bias	α = 0.1	0.14000	0.02898	0.03098	0.02553	0.04843	0.02750	0.02947
	α = 0.05	0.12867	0.03820	0.01306	0.01533	0.02096	0.02668	0.02518
	α = 0.01	0.18330	0.01677	0.02680	0.01888	0.03701	0.03237	0.03472
	α = 0.005	0.19821	0.03615	0.02610	0.00984	0.00630	0.01785	0.01742
	α = 0.001	0.22838	0.04349	0.07008	0.01739	0.00132	0.00548	0.00047
	α = 0.0001	0.27198	0.02963	0.00750	0.00416	0.02892	0.03156	0.00761
Median-bias	α = 0.1	0.06998	0.04613	0.04240	0.04897	0.04521	0.03573	0.04324
	α = 0.05	0.01829	0.03456	0.07287	0.11374	0.04187	0.03385	0.02262
	α = 0.01	0.08276	0.08898	0.07669	0.11146	0.04032	0.01835	0.02850
	α = 0.005	0.08022	0.08246	0.09298	0.11495	0.00346	0.00385	0.00238
	α = 0.001	0.14274	0.09316	0.05920	0.12638	0.03139	0.02833	0.02550
	α = 0.0001	0.21524	0.15297	0.18696	0.18784	0.07053	0.07520	0.05335
RMSE	α = 0.1	0.54745	0.48857	0.48443	0.49766	0.39627	0.36578	0.37339
	α = 0.05	0.56286	0.48177	0.48472	0.48862	0.38051	0.38536	0.37783
	α = 0.01	0.57676	0.51385	0.52319	0.54414	0.39518	0.38935	0.39701
	α = 0.005	0.58441	0.52163	0.53106	0.54771	0.38610	0.38887	0.39672
	α = 0.001	0.57955	0.54570	0.56422	0.56473	0.39252	0.40272	0.41305
	α = 0.0001	0.56002	0.57135	0.55940	0.58058	0.38983	0.38423	0.41655

Table B.2 Properties of the estimator of β for a sample size of 100

Statistics		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
Mean	$\alpha = 0.1$	0.48793	0.48720	0.48898	0.49120	0.48998	0.48635	0.48577
	$\alpha = 0.05$	0.48761	0.48813	0.48859	0.49094	0.48629	0.48555	0.48858
	$\alpha = 0.01$	0.48939	0.48802	0.48997	0.49157	0.49313	0.49051	0.49054
	$\alpha = 0.005$	0.49383	0.48723	0.48748	0.48925	0.49055	0.48862	0.49107
	$\alpha = 0.001$	0.49380	0.49371	0.49455	0.48968	0.49197	0.49416	0.49384
	$\alpha = 0.0001$	0.49750	0.49603	0.49416	0.48370	0.49296	0.49314	0.49494
Median	$\alpha = 0.1$	0.48590	0.48656	0.48746	0.49019	0.49072	0.48693	0.48740
	$\alpha = 0.05$	0.48581	0.48723	0.48733	0.49093	0.48779	0.48523	0.48779
	$\alpha = 0.01$	0.48998	0.48881	0.48806	0.49008	0.49315	0.48954	0.48954
	$\alpha = 0.005$	0.49332	0.48662	0.48760	0.48635	0.49083	0.48773	0.49071
	$\alpha = 0.001$	0.49392	0.49322	0.49482	0.48736	0.49084	0.49412	0.49390
	$\alpha = 0.0001$	0.49265	0.49411	0.49187	0.48345	0.49287	0.49309	0.49290
S-dev	$\alpha = 0.1$	0.04960	0.04650	0.04681	0.04691	0.05341	0.05097	0.05314
	$\alpha =0.05$	0.05148	0.04610	0.04723	0.04578	0.05256	0.05241	0.05341
	$\alpha =0.01$	0.04915	0.04575	0.04699	0.04405	0.05143	0.05155	0.05131
	$\alpha =0.005$	0.04891	0.04721	0.04662	0.04592	0.05037	0.04953	0.05176
	$\alpha =0.001$	0.05012	0.04655	0.04748	0.05314	0.04960	0.04985	0.04972
	$\alpha =0.0001$	0.13789	0.04964	0.04877	0.06163	0.04819	0.04825	0.04931
Mean-bias	$\alpha = 0.1$	0.01207	0.01280	0.01102	0.00880	0.01002	0.01365	0.01423
	$\alpha = 0.05$	0.01239	0.01187	0.01141	0.00906	0.01371	0.01445	0.01142
	$\alpha = 0.01$	0.01061	0.01198	0.01003	0.00843	0.00687	0.00949	0.00946
	$\alpha = 0.005$	0.00617	0.01277	0.01252	0.01075	0.00945	0.01138	0.00893
	$\alpha = 0.001$	0.00620	0.00629	0.00545	0.01032	0.00803	0.00584	0.00616
	$\alpha = 0.0001$	0.00250	0.00397	0.00584	0.01630	0.00704	0.00686	0.00506
Median-bias	$\alpha = 0.1$	0.01410	0.01344	0.01254	0.00981	0.00928	0.01307	0.01260
	$\alpha = 0.05$	0.01419	0.01277	0.01267	0.00907	0.01221	0.01477	0.01221
	$\alpha = 0.01$	0.01002	0.01119	0.01194	0.00992	0.00685	0.01046	0.01046
	$\alpha = 0.005$	0.00668	0.01338	0.01240	0.01365	0.00917	0.01227	0.00929
	$\alpha = 0.001$	0.00608	0.00678	0.00518	0.01264	0.00916	0.00588	0.00610
	$\alpha = 0.0001$	0.00735	0.00589	0.00813	0.01655	0.00713	0.00691	0.00710
RMSE	$\alpha = 0.1$	0.05102	0.04821	0.04806	0.04771	0.05432	0.05274	0.05499
	$\alpha = 0.05$	0.05293	0.04758	0.04857	0.04665	0.05429	0.05434	0.05459
	$\alpha = 0.01$	0.05026	0.04727	0.04803	0.04483	0.05186	0.05239	0.05215
	$\alpha = 0.005$	0.04927	0.04889	0.04825	0.04714	0.05123	0.05080	0.05250
	$\alpha = 0.001$	0.05048	0.04695	0.04777	0.05411	0.05022	0.05016	0.05008
	$\alpha = 0.0001$	0.13785	0.04977	0.04909	0.06372	0.04868	0.04871	0.04955

Table B.3 Properties of the estimator of γ for a sample size of 100

Statistics		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
Mean	$\alpha = 0.1$	0.00037	0.00337	0.00359	-0.01170	0.00701	0.01638	0.01658
	$\alpha = 0.05$	0.00265	0.00067	-0.00393	-0.00725	0.02034	0.01526	0.01499
	$\alpha = 0.01$	-0.00108	-0.00771	-0.01653	-0.01865	0.00850	0.00123	0.00173
	$\alpha = 0.005$	-0.00277	-0.01275	-0.02290	-0.02108	0.01874	0.01233	0.01401
	$\alpha = 0.001$	-0.00692	-0.02253	-0.03106	-0.03787	0.01047	0.01361	0.01306
	$\alpha = 0.0001$	-0.01282	-0.03692	-0.03869	-0.04615	0.01328	0.01363	0.00760
Median	$\alpha = 0.1$	-0.01578	-0.00196	-0.00006	-0.00777	-0.00128	0.00413	0.00468
	$\alpha = 0.05$	-0.00959	-0.00344	-0.00860	-0.01232	0.00786	0.00019	0.00670
	$\alpha = 0.01$	-0.01437	-0.01001	-0.01543	-0.02166	0.00345	-0.00962	-0.00938
	$\alpha = 0.005$	-0.02217	-0.01697	-0.02435	-0.02592	0.00831	0.00227	0.00404
	$\alpha = 0.001$	-0.02914	-0.02270	-0.03527	-0.03688	0.00654	0.00627	0.00613
	$\alpha = 0.0001$	-0.02096	-0.04269	-0.04271	-0.05091	0.00943	0.01028	-0.00112
S-dev	$\alpha =0.1$	0.12809	0.10768	0.10648	0.22115	0.11866	0.13988	0.14033
	$\alpha =0.05$	0.13428	0.10429	0.10370	0.09657	0.13096	0.12728	0.12029
	$\alpha =0.01$	0.13369	0.09920	0.09243	0.09879	0.11626	0.11665	0.11632
	$\alpha =0.005$	0.13077	0.09638	0.09701	0.09356	0.12783	0.11883	0.14382
	$\alpha =0.001$	0.15342	0.10106	0.09103	0.08744	0.11219	0.13997	0.14011
	$\alpha =0.0001$	0.12609	0.08938	0.08498	0.16146	0.11009	0.10996	0.12046
Mean-bias	$\alpha = 0.1$	0.00037	0.00337	0.00359	0.01170	0.00701	0.01638	0.01658
	$\alpha = 0.05$	0.00265	0.00067	0.00393	0.00725	0.02034	0.01526	0.01499
	$\alpha = 0.01$	0.00108	0.00771	0.01653	0.01865	0.00850	0.00123	0.00173
	$\alpha = 0.005$	0.00277	0.01275	0.02290	0.02108	0.01874	0.01233	0.01401
	$\alpha = 0.001$	0.00692	0.02253	0.03106	0.03787	0.01047	0.01361	0.01306
	$\alpha = 0.0001$	0.01282	0.03692	0.03869	0.04615	0.01328	0.01363	0.00760
Median-bias	$\alpha = 0.1$	0.01578	0.00196	0.00006	0.00777	0.00128	0.00413	0.00468
	$\alpha = 0.05$	0.00959	0.00344	0.00860	0.01232	0.00786	0.00019	0.00670
	$\alpha = 0.01$	0.01437	0.01001	0.01543	0.02166	0.00345	0.00962	0.00938
	$\alpha = 0.005$	0.02217	0.01697	0.02435	0.02592	0.00831	0.00227	0.00404
	$\alpha = 0.001$	0.02914	0.02270	0.03527	0.03688	0.00654	0.00627	0.00613
	$\alpha = 0.0001$	0.02096	0.04269	0.04271	0.05091	0.00943	0.01028	0.00112
RMSE	$\alpha = 0.1$	0.12803	0.10768	0.10649	0.22135	0.11880	0.14077	0.14123
	$\alpha = 0.05$	0.13424	0.10424	0.10373	0.09680	0.13247	0.12813	0.12116
	$\alpha = 0.01$	0.13363	0.09945	0.09385	0.10049	0.11651	0.11659	0.11627
	$\alpha = 0.005$	0.13073	0.09717	0.09963	0.09586	0.12913	0.11941	0.14442
	$\alpha = 0.001$	0.15350	0.10349	0.09614	0.09525	0.11262	0.14056	0.14065
	$\alpha = 0.0001$	0.12668	0.09667	0.09333	0.16785	0.11083	0.11075	0.12064

Table B.4 Properties of the estimator of δ for a sample size of 100

Tests		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
J-Test	$\alpha = 0.1$	0.001	0.040	0.030	0.023	0.042	0.058	0.059
	$\alpha = 0.05$	0.003	0.023	0.030	0.019	0.083	0.076	0.069
	$\alpha = 0.01$	0.026	0.078	0.068	0.048	0.129	0.117	0.113
	$\alpha = 0.005$	0.036	0.095	0.084	0.083	0.185	0.184	0.196
	$\alpha = 0.001$	0.142	0.157	0.153	0.154	0.322	0.336	0.337
	$\alpha = 0.0001$	0.540	0.582	0.530	0.516	0.782	0.784	0.783
LM-Test	$\alpha = 0.1$		0.004	0.000	0.000	0.004	0.000	0.001
	$\alpha = 0.05$		0.004	0.000	0.000	0.010	0.000	0.000
	$\alpha = 0.01$		0.053	0.011	0.000	0.061	0.009	0.002
	$\alpha = 0.005$		0.129	0.020	0.000	0.080	0.009	0.012
	$\alpha = 0.001$		0.366	0.087	0.003	0.256	0.088	0.071
	$\alpha = 0.0001$		0.517	0.215	0.012	0.497	0.304	0.308
LR-Test	$\alpha = 0.1$		0.000	0.000	0.001	0.008	0.008	0.010
	$\alpha = 0.05$		0.000	0.000	0.000	0.026	0.018	0.023
	$\alpha = 0.01$		0.002	0.003	0.001	0.107	0.089	0.095
	$\alpha = 0.005$		0.006	0.001	0.001	0.160	0.149	0.168
	$\alpha = 0.001$		0.013	0.007	0.005	0.403	0.358	0.367
	$\alpha = 0.0001$		0.041	0.013	0.015	0.726	0.694	0.718

Table B.5 Sizes of tests of overidentifying rectrictions (level=0.05,sample size=100

Tests		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
J-test	$\alpha = 0.1$	0.001	0.028	0.017	0.015	0.029	0.038	0.039
	$\alpha = 0.05$	0.003	0.009	0.016	0.010	0.054	0.060	0.043
	$\alpha = 0.01$	0.019	0.050	0.038	0.027	0.099	0.085	0.081
	$\alpha = 0.005$	0.028	0.065	0.055	0.070	0.131	0.137	0.149
	$\alpha = 0.001$	0.116	0.116	0.114	0.124	0.254	0.255	0.256
	$\alpha = 0.0001$	0.498	0.499	0.456	0.446	0.724	0.727	0.723
LM-test	$\alpha = 0.1$		0.003	0.000	0.000	0.004	0.000	0.001
	$\alpha = 0.05$		0.003	0.000	0.000	0.009	0.000	0.000
	$\alpha = 0.01$		0.034	0.010	0.000	0.055	0.007	0.002
	$\alpha = 0.005$		0.086	0.012	0.000	0.074	0.007	0.009
	$\alpha = 0.001$		0.323	0.060	0.003	0.235	0.060	0.049
	$\alpha = 0.0001$		0.489	0.169	0.009	0.455	0.254	0.264
LR-test	$\alpha = 0.1$		0.000	0.000	0.001	0.006	0.005	0.007
	$\alpha = 0.05$		0.000	0.000	0.000	0.018	0.012	0.013
	$\alpha = 0.01$		0.001	0.000	0.001	0.089	0.062	0.067
	$\alpha = 0.005$		0.002	0.000	0.000	0.131	0.108	0.126
	$\alpha = 0.001$		0.006	0.004	0.003	0.360	0.309	0.321
	$\alpha = 0.0001$		0.023	0.007	0.011	0.677	0.637	0.682

Table B.6 Sizes of tests of overidentifying rectrictions (level=0.01,sample size=100

Statistics		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
Mean	$\alpha = 0.1$	1.71696	1.70282	1.71877	1.72734	1.88559	1.89430	1.87629
	$\alpha = 0.01$	1.71670	1.72835	1.75559	1.76642	1.67619	1.89172	1.67708
	$\alpha = 0.001$	1.70269	1.77028	1.80992	1.82074	1.68399	1.89607	1.69003
	$\alpha = 0.0001$	1.65994	1.82094	1.85877	1.85625	1.69726	1.69461	1.69340
Median	$\alpha = 0.1$	1.71194	1.71371	1.72601	1.73958	2.00000	2.00000	2.00000
	$\alpha = 0.01$	1.71836	1.74913	1.76422	1.78336	1.68408	2.00000	1.68140
	$\alpha = 0.001$	1.70906	1.79705	1.81690	1.84054	1.68471	2.00000	1.69168
	$\alpha = 0.0001$	1.67003	1.84869	1.87036	1.88178	1.70069	1.69849	1.69830
S-dev	$\alpha = 0.1$	0.12555	0.14470	0.14715	0.14843	0.18078	0.17669	0.18063
	$\alpha = 0.01$	0.12580	0.16688	0.11501	0.15545	0.12025	0.17309	0.11672
	$\alpha = 0.001$	0.13472	0.18415	0.09441	0.14207	0.11242	0.17167	0.11521
	$\alpha = 0.0001$	0.14754	0.18079	0.11203	0.16719	0.10697	0.11244	0.11367
Mean-bias	$\alpha = 0.1$	0.01696	0.00282	0.01877	0.02734	0.18559	0.19430	0.17629
	$\alpha = 0.01$	0.01670	0.02835	0.05559	0.06642	0.02381	0.19172	0.02292
	$\alpha = 0.001$	0.00269	0.07028	0.10992	0.12074	0.01601	0.19607	0.00997
	$\alpha = 0.0001$	0.04006	0.12094	0.15877	0.15625	0.00274	0.00539	0.00660
Median-bias	$\alpha = 0.1$	0.01194	0.01371	0.02601	0.03958	0.30000	0.30000	0.30000
	$\alpha = 0.01$	0.01836	0.04913	0.06422	0.08336	0.01592	0.30000	0.01860
	$\alpha = 0.001$	0.00906	0.09705	0.11690	0.14054	0.01529	0.30000	0.00832
	$\alpha = 0.0001$	0.02997	0.14869	0.17036	0.18178	0.00069	0.00151	0.00170
RMSE	$\alpha = 0.1$	0.12663	0.14465	0.14827	0.15086	0.25902	0.26257	0.25234
	$\alpha = 0.01$	0.12684	0.16918	0.12769	0.16897	0.12252	0.25824	0.11889
	$\alpha = 0.001$	0.13468	0.19702	0.14486	0.18639	0.11350	0.26055	0.11558
	$\alpha = 0.0001$	0.15281	0.21744	0.19429	0.22877	0.10695	0.11251	0.11381

Table B.7 Properties of the estimator of ω for a sample size of 200

Statistics		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
Mean	$\alpha = 0.1$	0.39542	0.51898	0.54702	0.56194	-0.63630	-0.67003	-0.59104
	$\alpha = 0.01$	0.33940	0.52639	0.52254	0.57207	0.50524	-0.64819	0.49342
	$\alpha = 0.001$	0.34246	0.55086	0.55161	0.57682	0.49168	-0.65945	0.52560
	$\alpha = 0.0001$	0.30877	0.58925	0.61615	0.64998	0.51096	0.51468	0.50963
Median	$\alpha = 0.1$	0.48193	0.53398	0.56083	0.58689	-0.99529	-0.99598	-0.98442
	$\alpha = 0.01$	0.42145	0.54511	0.53358	0.60291	0.49444	-0.99611	0.49727
	$\alpha = 0.001$	0.41107	0.59526	0.60679	0.65944	0.47592	-0.99629	0.51938
	$\alpha = 0.0001$	0.33218	0.65958	0.70821	0.78445	0.49686	0.52672	0.52056
S-dev	$\alpha =0.1$	0.45157	0.32674	0.32811	0.32952	0.53526	0.51577	0.55171
	$\alpha = 0.01$	0.45127	0.34354	0.34340	0.35339	0.29731	0.53252	0.29903
	$\alpha = 0.001$	0.41001	0.36606	0.37297	0.39335	0.30857	0.52585	0.30096
	$\alpha = 0.0001$	0.36081	0.38383	0.39542	0.41388	0.31186	0.30976	0.31454
Mean-bias	$\alpha = 0.1$	0.10458	0.01898	0.04702	0.06194	1.13630	1.17003	1.09104
	$\alpha = 0.01$	0.16060	0.02639	0.02254	0.07207	0.00524	1.14819	0.00658
	$\alpha = 0.001$	0.15754	0.05086	0.05161	0.07682	0.00832	1.15945	0.02560
	$\alpha = 0.0001$	0.19123	0.08925	0.11615	0.14998	0.01096	0.01468	0.00963
Median-bias	$\alpha = 0.1$	0.01807	0.03398	0.06083	0.08689	1.49529	1.49598	1.48442
	$\alpha = 0.01$	0.07855	0.04511	0.03358	0.10291	0.00556	1.49611	0.00273
	$\alpha = 0.001$	0.08893	0.09526	0.10679	0.15944	0.02408	1.49629	0.01938
	$\alpha = 0.0001$	0.16782	0.15958	0.20821	0.28445	0.00314	0.02672	0.02056
RMSE	$\alpha = 0.1$	0.46330	0.32713	0.33130	0.33513	1.25594	1.27856	1.22248
	$\alpha = 0.01$	0.47879	0.34438	0.34397	0.36049	0.29721	1.26556	0.29895
	$\alpha = 0.001$	0.43905	0.36940	0.37634	0.40059	0.30853	1.27301	0.30190
	$\alpha = 0.0001$	0.40819	0.39389	0.41194	0.44002	0.31189	0.30995	0.31453

Table B.8 Properties of the estimator of β for a sample size of 200

Statistics		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
Mean	$\alpha = 0.1$	0.49599	0.49398	0.49056	0.48604	0.52006	0.52114	0.51991
	$\alpha = 0.01$	0.49755	0.49326	0.49347	0.48530	0.49467	0.52141	0.49479
	$\alpha = 0.001$	0.49807	0.49781	0.49628	0.48388	0.49566	0.52134	0.49555
	$\alpha = 0.0001$	0.49197	0.50552	0.49609	0.48464	0.49882	0.49740	0.49720
Median	$\alpha = 0.1$	0.49534	0.49398	0.49377	0.49406	0.52354	0.52417	0.52079
	$\alpha = 0.01$	0.49785	0.49694	0.49349	0.49399	0.49278	0.52227	0.49414
	$\alpha = 0.001$	0.49855	0.49996	0.49538	0.49700	0.49522	0.52437	0.49433
	$\alpha = 0.0001$	0.49184	0.50465	0.50004	0.49685	0.49770	0.49725	0.49722
S-dev	$\alpha = 0.1$	0.03478	0.06682	0.04999	0.06956	0.04021	0.03957	0.03913
	$\alpha = 0.01$	0.03590	0.05722	0.03967	0.07164	0.03644	0.03874	0.03524
	$\alpha = 0.001$	0.03639	0.06906	0.04115	0.08901	0.03494	0.03913	0.03490
	$\alpha = 0.0001$	0.03741	0.06603	0.06493	0.08869	0.03475	0.03508	0.03512
Mean-bias	$\alpha = 0.1$	0.00401	0.00602	0.00944	0.01396	0.02006	0.02114	0.01991
	$\alpha = 0.01$	0.00245	0.00674	0.00653	0.01470	0.00533	0.02141	0.00521
	$\alpha = 0.001$	0.00193	0.00219	0.00372	0.01612	0.00434	0.02134	0.00445
	$\alpha = 0.0001$	0.00803	0.00552	0.00391	0.01536	0.00118	0.00260	0.00280
Median-bias	$\alpha = 0.1$	0.00466	0.00602	0.00623	0.00594	0.02354	0.02417	0.02079
	$\alpha = 0.01$	0.00215	0.00306	0.00651	0.00601	0.00722	0.02227	0.00586
	$\alpha = 0.001$	0.00145	0.00004	0.00462	0.00300	0.00478	0.02437	0.00567
	$\alpha = 0.0001$	0.00816	0.00465	0.00004	0.00315	0.00230	0.00275	0.00278
RMSE	$\alpha = 0.1$	0.03500	0.06705	0.05085	0.07092	0.04492	0.04484	0.04389
	$\alpha = 0.01$	0.03597	0.05759	0.04018	0.07310	0.03681	0.04425	0.03560
	$\alpha = 0.001$	0.03642	0.06906	0.04130	0.09042	0.03519	0.04455	0.03516
	$\alpha = 0.0001$	0.03825	0.06623	0.06502	0.08996	0.03476	0.03516	0.03521

Table B.9 Properties of the estimator of γ for a sample size of 200

Statistics		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
Mean	$\alpha = 0.1$	-0.00291	-0.00599	-0.09547	-0.13376	-0.06758	-0.06912	-0.06709
	$\alpha = 0.01$	-0.01667	-0.02324	-0.07771	-0.10108	0.00945	-0.06889	0.00781
	$\alpha = 0.001$	-0.01076	-0.03390	-0.03524	-0.10551	0.00161	-0.06848	0.00535
	$\alpha = 0.0001$	-0.00813	-0.04288	-0.04943	-0.07804	0.00219	0.00628	0.00619
Median	$\alpha = 0.1$	-0.00749	-0.00847	-0.01149	-0.01513	-0.07148	-0.07334	-0.06942
	$\alpha = 0.01$	-0.02176	-0.00955	-0.01874	-0.02592	0.00300	-0.07125	0.00577
	$\alpha = 0.001$	-0.01419	-0.02129	-0.02906	-0.03638	-0.00275	-0.07176	0.00058
	$\alpha = 0.0001$	-0.01682	-0.02873	-0.03525	-0.04217	-0.00133	0.00353	0.00376
S-dev	$\alpha = 0.1$	0.07952	0.22553	0.91299	1.15902	0.06208	0.06200	0.06301
	$\alpha = 0.01$	0.07921	0.33921	0.88109	0.88556	0.07952	0.06241	0.07511
	$\alpha = 0.001$	0.08371	0.27004	0.32871	0.78683	0.07772	0.06132	0.07785
	$\alpha = 0.0001$	0.11880	0.34269	0.46714	0.44679	0.07833	0.07786	0.07801
Mean-bias	$\alpha = 0.1$	0.00291	0.00599	0.09547	0.13376	0.06758	0.06912	0.06709
	$\alpha = 0.01$	0.01667	0.02324	0.07771	0.10108	0.00945	0.06889	0.00781
	$\alpha = 0.001$	0.01076	0.03390	0.03524	0.10551	0.00161	0.06848	0.00535
	$\alpha = 0.0001$	0.00813	0.04288	0.04943	0.07804	0.00219	0.00628	0.00619
Median-bias	$\alpha = 0.1$	0.00749	0.00847	0.01149	0.01513	0.07148	0.07334	0.06942
	$\alpha = 0.01$	0.02176	0.00955	0.01874	0.02592	0.00300	0.07125	0.00577
	$\alpha = 0.001$	0.01419	0.02129	0.02906	0.03638	0.00275	0.07176	0.00058
	$\alpha = 0.0001$	0.01682	0.02873	0.03525	0.04217	0.00133	0.00353	0.00376
RMSE	$\alpha = 0.1$	0.07953	0.22549	0.91752	1.16613	0.09174	0.09284	0.09201
	$\alpha = 0.01$	0.08091	0.33983	0.88408	0.89087	0.08004	0.09293	0.07547
	$\alpha = 0.001$	0.08436	0.27203	0.33043	0.79348	0.07770	0.09190	0.07800
	$\alpha = 0.0001$	0.11901	0.34519	0.46952	0.45334	0.07833	0.07807	0.07822

Table B.10 Properties of the estimator of δ for a sample size of 200

Tests		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
J-Test	$\alpha = 0.1$	0.002	0.049	0.040	0.047	0.515	0.523	0.475
	$\alpha = 0.01$	0.046	0.100	0.075	0.101	0.166	0.587	0.159
	$\alpha = 0.001$	0.253	0.349	0.296	0.302	0.494	0.726	0.504
	$\alpha = 0.0001$	0.758	0.687	0.660	0.640	0.840	0.843	0.843
LM-Test	$\alpha = 0.1$		0.035	0.012	0.030	0.612	0.563	0.127
	$\alpha = 0.01$		0.135	0.038	0.039	0.170	0.624	0.008
	$\alpha = 0.001$		0.509	0.131	0.043	0.244	0.691	0.100
	$\alpha = 0.0001$		0.738	0.381	0.053	0.722	0.363	0.363
LR-Test	$\alpha = 0.1$		0.026	0.014	0.034	0.493	0.361	0.193
	$\alpha = 0.01$		0.039	0.011	0.040	0.222	0.410	0.132
	$\alpha = 0.001$		0.073	0.009	0.044	0.241	0.468	0.458
	$\alpha = 0.0001$		0.138	0.046	0.055	0.848	0.355	0.355

Table B.11 Sizes of tests of overidentifying rectrictions (level=0.05,sample size=200

Tests		CGMM	iter-EL	iter-ET	iter-EEL	sv-EL	sv-ET	sv-EEL
J-Test	α =0.1	0.002	0.030	0.029	0.040	0.420	0.420	0.394
	α =0.01	0.031	0.070	0.049	0.078	0.101	0.503	0.110
	α =0.001	0.198	0.290	0.223	0.249	0.417	0.647	0.439
	α =0.0001	0.711	0.633	0.600	0.576	0.800	0.795	0.795
LM-Test	α = 0.1		0.030	0.010	0.029	0.588	0.510	0.074
	α = 0.01		0.111	0.024	0.039	0.156	0.598	0.005
	α = 0.001		0.460	0.086	0.042	0.001	0.666	0.078
	α = 0.0001		0.700	0.294	0.053	0.699	0.000	0.001
LR-Test	α = 0.1		0.026	0.013	0.032	0.410	0.262	0.110
	α = 0.01		0.038	0.010	0.039	0.192	0.276	0.098
	α = 0.001		0.063	0.006	0.043	0.001	0.344	0.408
	α = 0.0001		0.096	0.031	0.053	0.830	0.000	0.001

Table B.12 Sizes of tests of overidentifying rectrictions (level=0.01,sample size=200

B.3 Results : Linear models

B.3.1 Properties of $\hat{d}$

Table B.13 Mean Bias : Sample size n = 100 (Normal errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.02562478	0.0029303	0.00476486	0.00666785	0.00950619
α_n = 1e-04	0.02305654	0.00359117	0.00451997	0.00621626	0.00700388
α_n = 2e-04	0.02034771	0.00365248	0.00425565	0.00563814	0.00794489
α_n = 5e-04	0.01656608	0.00373696	0.00398567	0.00503249	0.00777986
α_n = 8e-04	0.01454939	0.00384274	0.00393383	0.0048457	0.00704344
$\alpha_n = 0.001$	0.0135756	0.00390114	0.00392627	0.00479657	0.00613665
$\alpha_n = 0.002$	0.01048398	0.00427903	0.00400401	0.00476242	0.00761083
$\alpha_n = 0.01$	0.00275513	0.00418099	0.00429176	0.0045415	0.00681726
$\alpha_n = 0.05$	0.00556384	0.00520145	0.00497277	0.00489503	0.00497213
$\alpha_n = 0.1$	0.0087758	0.00574972	0.0055919	0.00542786	0.00679606

Table B.14 Mean Bias : Sample size n = 200 (Normal errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.01643196	0.00060746	3.667e-05	0.00097735	0.00048811
α_n = 1e-04	0.01483582	0.00057837	0.00013939	0.00076589	0.00027424
α_n = 2e-04	0.01316262	0.00043379	0.00021242	0.00054079	0.00069239
α_n = 5e-04	0.01085107	0.00036234	0.00028189	0.00026513	0.00010614
α_n = 8e-04	0.00963365	0.00037541	0.00031631	0.00012347	0.00063212
$\alpha_n = 0.001$	0.00905095	0.00036858	0.00032129	6.482e-05	0.00042743
$\alpha_n = 0.002$	0.00723056	0.00025512	0.00024888	2.008e-05	0.00048475
$\alpha_n = 0.01$	0.00290905	0.00015675	0.00013184	0.00014777	9.732e-05
$\alpha_n = 0.05$	0.00146529	0.00057056	0.00052801	0.00047213	0.00106267
$\alpha_n = 0.1$	0.00306491	0.00102575	0.0009671	0.00090554	0.00096965

Table B.15 Mean Bias : Sample size n = 300 (Normal errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.01030626	0.00032195	0.00492449	0.00146567	0.00104466
α_n = 1e-04	0.00914862	0.00036766	0.00486859	0.00133567	0.00116418
α_n = 2e-04	0.0079379	0.00051051	0.00481917	0.00121945	0.00085572
α_n = 5e-04	0.00627688	0.00070713	0.00480013	0.00112774	0.00090747
α_n = 8e-04	0.00540989	0.00078485	0.00479021	0.00107928	0.00081149
α_n = 0.001	0.00499735	0.00081872	0.00478823	0.00105971	0.00083083
α_n = 0.002	0.00372073	0.00097175	0.00480767	0.0010637	0.00094298
α_n = 0.01	0.0007127	0.00125448	0.0046313	0.00119295	0.00127607
α_n = 0.05	0.00237256	0.00149975	0.00446951	0.00144483	0.00169447
α_n = 0.1	0.00347874	0.00192621	0.0047061	0.00185485	0.00189624

Table B.16 Median Bias : Sample size n = 100 (Normal errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.03225928	0.00827077	0.00689274	0.00391925	0.00654202
α_n = 1e-04	0.03089027	0.00792106	0.00656623	0.00434972	0.00664073
α_n = 2e-04	0.02956135	0.00706166	0.00689963	0.00553863	0.00536074
α_n = 5e-04	0.02663834	0.00733173	0.00691028	0.00561828	0.0066398
α_n = 8e-04	0.02457663	0.00685773	0.0075753	0.00497329	0.0064235
$\alpha_n = 0.001$	0.02428623	0.00666248	0.00792237	0.00626684	0.00717561
$\alpha_n = 0.002$	0.02137248	0.00583795	0.00573109	0.00503184	0.00588949
$\alpha_n = 0.01$	0.0084727	0.00477192	0.00473053	0.00491839	0.00458089
$\alpha_n = 0.05$	0.00043196	0.00209412	0.00197533	0.00197011	0.0021125
$\alpha_n = 0.1$	0.00142966	0.00102143	0.00136199	0.00196907	0.00101974

Table B.17 Median Bias : Sample size n = 200 (Normal errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.01820424	0.00412774	0.00252327	0.00164779	0.00406526
α_n = 1e-04	0.01694923	0.00372064	0.00273001	0.00136629	0.00388569
α_n = 2e-04	0.0158523	0.00313841	0.0039954	0.00289202	0.00321027
α_n = 5e-04	0.0139417	0.00396518	0.00388867	0.00404664	0.00433995
α_n = 8e-04	0.01262123	0.00409999	0.00410485	0.00346443	0.00418192
$\alpha_n = 0.001$	0.0116871	0.00432062	0.00438493	0.0034514	0.00528388
$\alpha_n = 0.002$	0.00983711	0.0047449	0.00484008	0.00412918	0.00465816
$\alpha_n = 0.01$	0.00610072	0.0047082	0.00429256	0.00439103	0.00422786
$\alpha_n = 0.05$	0.00354922	0.00440362	0.0045653	0.0044884	0.00442371
$\alpha_n = 0.1$	0.00211171	0.00423591	0.00424953	0.00424736	0.00424707

Table B.18 Median Bias : Sample size n = 300 (Normal errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.01329669	0.00368956	0.00130036	0.00282369	0.00350953
α_n = 1e-04	0.0123709	0.00378608	0.00170928	0.00285106	0.00348109
α_n = 2e-04	0.01120362	0.00399986	0.00107752	0.00320232	0.00414839
α_n = 5e-04	0.00929967	0.00332517	0.00098975	0.00318743	0.00338514
α_n = 8e-04	0.00869767	0.00328853	0.00124175	0.00322884	0.00311073
α_n = 0.001	0.00817675	0.00278442	0.00114809	0.00311546	0.00280582
α_n = 0.002	0.00762698	0.00322569	0.00164885	0.00280503	0.00323316
α_n = 0.01	0.00464424	0.00264822	0.00262712	0.00200748	0.00229234
α_n = 0.05	0.00115297	0.00201692	0.0014356	0.00173731	0.00198331
α_n = 0.1	0.00013905	0.0010413	0.00178178	0.00113627	0.00110638

Table B.19 Mean Bias : Sample size n = 100 (Skewed errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.04322024	0.03175316	0.04679072	0.06848327	0.00457451
α_n = 1e-04	0.04216741	0.03000321	0.04435897	0.06497181	0.00692117
α_n = 2e-04	0.04130756	0.0296538	0.04207727	0.06131852	0.00708188
α_n = 5e-04	0.04057559	0.02962996	0.04011477	0.05607207	0.00608238
α_n = 8e-04	0.04032414	0.02901768	0.04055521	0.05380342	0.00892397
α_n = 0.001	0.04020069	0.0290943	0.04021171	0.05298205	0.00905509
α_n = 0.002	0.03958338	0.0290952	0.03936717	0.05058715	0.01003724
α_n = 0.01	0.03395415	0.02952588	0.03665696	0.04500633	0.01876628
α_n = 0.05	0.02014521	0.02443736	0.02679563	0.03056566	0.01731632
α_n = 0.1	0.01321143	0.01969064	0.02012405	0.02163411	0.01428457

Table B.20 Mean Bias : Sample size n = 200 (Skewed errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.01889345	0.0109295	0.01890925	0.03055619	0.00467005
α_n = 1e-04	0.01817991	0.00994669	0.01782635	0.02849602	0.00503913
α_n = 2e-04	0.0176694	0.0100513	0.01675306	0.02663142	0.00272308
α_n = 5e-04	0.01742609	0.0096354	0.01585899	0.0243685	0.0021506
α_n = 8e-04	0.01748824	0.00948867	0.01563332	0.02320865	0.00263747
$\alpha_n = 0.001$	0.01754674	0.00952156	0.01559871	0.02283159	0.00147498
$\alpha_n = 0.002$	0.01771697	0.00985705	0.01579779	0.0220299	0.00025167
$\alpha_n = 0.01$	0.01592092	0.01227258	0.01641596	0.02092336	0.00754097
$\alpha_n = 0.05$	0.00880415	0.01162413	0.01256642	0.01410954	0.01001664
$\alpha_n = 0.1$	0.00512495	0.00941123	0.00910306	0.00921796	0.00764679

Table B.21 Mean Bias : Sample size n = 300 (Skewed errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.01366358	0.00805018	0.016824	0.02186628	0.00320828
α_n = 1e-04	0.01310266	0.00766555	0.01600047	0.02037577	0.00311773
α_n = 2e-04	0.01267714	0.00728759	0.01527274	0.01900933	0.00247926
α_n = 5e-04	0.01240708	0.00710007	0.01461603	0.01739176	0.0014964
α_n = 8e-04	0.01238456	0.00714278	0.01445243	0.01667555	0.0005778
$\alpha_n = 0.001$	0.0123889	0.00718611	0.01442016	0.01638163	0.00118916
$\alpha_n = 0.002$	0.01237205	0.00720134	0.01449851	0.0157096	0.00155267
$\alpha_n = 0.01$	0.01067736	0.00898693	0.01477217	0.01457895	0.00543501
$\alpha_n = 0.05$	0.0051099	0.00857654	0.01190679	0.00913674	0.00564305
$\alpha_n = 0.1$	0.00233356	0.00688051	0.0093768	0.00535973	0.0036136

Table B.22 Median Bias : Sample size n = 100 (Skewed errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.04117114	0.03141106	0.04672733	0.06448001	0.01385994
α_n = 1e-04	0.04158926	0.02950671	0.04422634	0.06059295	0.01497358
α_n = 2e-04	0.03984219	0.03000596	0.04081841	0.05705274	0.01258922
α_n = 5e-04	0.03793008	0.03009771	0.03845247	0.05417645	0.0132187
α_n = 8e-04	0.03965221	0.02938114	0.03919571	0.05154589	0.01482718
$\alpha_n = 0.001$	0.03944491	0.02862093	0.03934487	0.05018109	0.01514093
$\alpha_n = 0.002$	0.0379971	0.02961347	0.04008613	0.04836142	0.01789673
$\alpha_n = 0.01$	0.03299808	0.02843137	0.03624748	0.04384102	0.02481972
$\alpha_n = 0.05$	0.01972695	0.02480841	0.02861515	0.03217973	0.02219103
$\alpha_n = 0.1$	0.01303754	0.02016095	0.0204709	0.02300209	0.016346

Table B.23 Median Bias : Sample size n = 200 (Skewed errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.02046125	0.01266095	0.02040266	0.03067375	0.00447502
α_n = 1e-04	0.02050881	0.01192357	0.01937809	0.02923229	0.003289
α_n = 2e-04	0.01959556	0.01204651	0.01865755	0.02755621	0.00334141
α_n = 5e-04	0.01929613	0.0122162	0.01870116	0.02519239	0.00478586
α_n = 8e-04	0.0194906	0.01155982	0.01841611	0.02462356	0.00414479
$\alpha_n = 0.001$	0.01931678	0.01168096	0.01812904	0.02443748	0.00562058
$\alpha_n = 0.002$	0.01984521	0.01076552	0.01757607	0.02294144	0.00748371
$\alpha_n = 0.01$	0.01786845	0.01301108	0.01817815	0.02262484	0.01199842
$\alpha_n = 0.05$	0.00934567	0.01333073	0.01417924	0.01577192	0.01142483
$\alpha_n = 0.1$	0.00566892	0.01058475	0.00953331	0.00954045	0.00894172

Table B.24 Mean Bias : Sample size n = 300 (Skewed errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.01447937	0.01050093	0.01667783	0.02205584	1.48e-05
α_n = 1e-04	0.01417104	0.00907986	0.01578155	0.02085283	0.0006837
α_n = 2e-04	0.01421882	0.00925079	0.01505789	0.02007586	0.00107424
α_n = 5e-04	0.01345788	0.00976065	0.01472985	0.01881458	0.00187952
α_n = 8e-04	0.01360723	0.00981123	0.0144755	0.01798665	0.00239983
$\alpha_n = 0.001$	0.01329835	0.01003843	0.01444399	0.01740959	0.0029494
$\alpha_n = 0.002$	0.01322571	0.01003538	0.01475738	0.01692871	0.00375411
$\alpha_n = 0.01$	0.0106786	0.01260907	0.0146392	0.01488981	0.00751409
$\alpha_n = 0.05$	0.00565948	0.01080076	0.01259701	0.0092827	0.00644142
$\alpha_n = 0.1$	0.00213396	0.00881337	0.00986448	0.00599044	0.00444412

Table B.25 RMSE : Sample size n = 100 (Normal errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.1472893	0.1645179	0.1629718	0.162406	0.1838469
α_n = 1e-04	0.1472906	0.1639015	0.1617745	0.1613375	0.1761081
α_n = 2e-04	0.1474001	0.1623875	0.1605506	0.1601946	0.1802455
α_n = 5e-04	0.1477212	0.1606101	0.1592105	0.1589459	0.1752293
α_n = 8e-04	0.1479652	0.1598319	0.1586551	0.1584652	0.1734071
$\alpha_n = 0.001$	0.1481028	0.1594629	0.1583971	0.1582774	0.1695996
$\alpha_n = 0.002$	0.1486421	0.1589260	0.1577284	0.1577736	0.1725939
$\alpha_n = 0.01$	0.1509631	0.1566371	0.1566613	0.1568014	0.1661098
$\alpha_n = 0.05$	0.1554472	0.1572535	0.1569965	0.1573525	0.1570212
$\alpha_n = 0.1$	0.1578245	0.1571878	0.1571556	0.1572816	0.1602877

Table B.26 RMSE : Sample size n = 200 (Normal errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.1056653	0.1104891	0.1104701	0.1108264	0.1141015
α_n = 1e-04	0.1057093	0.1101518	0.1102073	0.1105651	0.1142013
α_n = 2e-04	0.1058043	0.1099186	0.1099246	0.1102480	0.1146294
α_n = 5e-04	0.1060086	0.1095790	0.1095381	0.1097709	0.1110587
α_n = 8e-04	0.1061498	0.1093782	0.1093298	0.1095121	0.1141491
$\alpha_n = 0.001$	0.1062272	0.1092967	0.1092345	0.1093880	0.1138947
$\alpha_n = 0.002$	0.1065195	0.1090565	0.1089965	0.1090480	0.1127140
$\alpha_n = 0.01$	0.107679	0.1091229	0.1090116	0.1089486	0.1090317
$\alpha_n = 0.05$	0.1097850	0.1097538	0.1096917	0.1096852	0.1127007
$\alpha_n = 0.1$	0.1107953	0.1100556	0.1100334	0.1100415	0.1100438

Table B.27 RMSE : Sample size n = 300 (Normal errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.08632328	0.08851585	0.09112661	0.08850475	0.0897106
α_n = 1e-04	0.08630727	0.08830509	0.09088857	0.08828955	0.09009723
α_n = 2e-04	0.08634162	0.08816808	0.09062462	0.08812276	0.08916151
α_n = 5e-04	0.0864709	0.08811736	0.09038497	0.08801503	0.08789636
α_n = 8e-04	0.08658185	0.08810503	0.09032929	0.0879944	0.08808515
$\alpha_n = 0.001$	0.08664873	0.08810136	0.09031399	0.08798697	0.08814933
$\alpha_n = 0.002$	0.08693269	0.08811646	0.09029948	0.08798475	0.08823315
$\alpha_n = 0.01$	0.08825062	0.08857116	0.09050838	0.08841588	0.08863802
$\alpha_n = 0.05$	0.09027984	0.08977714	0.09144723	0.08976306	0.0906235
$\alpha_n = 0.1$	0.0910263	0.09033807	0.0918325	0.0903575	0.09034298

Table B.28 RMSE : Sample size n = 100 (Skewed errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.1545244	0.1725079	0.1692036	0.1879850	0.1941793
α_n = 1e-04	0.1539386	0.1707775	0.1678388	0.1841124	0.1827317
α_n = 2e-04	0.1536431	0.1654014	0.1675085	0.1804877	0.1839330
α_n = 5e-04	0.1537159	0.1605256	0.1660802	0.1697099	0.1825450
α_n = 8e-04	0.1539466	0.1595038	0.1599031	0.1666803	0.1744496
$\alpha_n = 0.001$	0.1540935	0.1590071	0.1593295	0.1655371	0.1741122
$\alpha_n = 0.002$	0.1546340	0.1575872	0.1579754	0.1624816	0.1755343
$\alpha_n = 0.01$	0.1552365	0.1566357	0.1564866	0.1585374	0.1701071
$\alpha_n = 0.05$	0.1543037	0.1550728	0.1544885	0.1548934	0.1652971
$\alpha_n = 0.1$	0.1544683	0.1541577	0.1536405	0.1535915	0.1605812

Table B.29 RMSE : Sample size n = 200 (Skewed errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.1074161	0.1149413	0.1118039	0.1173683	0.1244314
α_n = 1e-04	0.1070558	0.1129433	0.1107705	0.1156403	0.1236196
α_n = 2e-04	0.1068472	0.1118347	0.1095448	0.1140201	0.1185880
α_n = 5e-04	0.106875	0.1096539	0.1087461	0.1120292	0.1193718
α_n = 8e-04	0.1070416	0.1091260	0.1084729	0.1105477	0.1188913
$\alpha_n = 0.001$	0.1071561	0.1087878	0.1083786	0.1102224	0.1178926
$\alpha_n = 0.002$	0.1076262	0.1080663	0.1082513	0.1095102	0.1181819
$\alpha_n = 0.01$	0.1086749	0.1090069	0.1087280	0.1093348	0.1143429
$\alpha_n = 0.05$	0.1089568	0.1092464	0.1088358	0.1087537	0.1102783
$\alpha_n = 0.1$	0.1093078	0.1089747	0.1087197	0.1085484	0.1100196

Table B.30 RMSE : Sample size n = 300 (Skewed errors)

	Method				
	CGMM	CEL	CET	CEEL	CETEL
α_n = 5e-05	0.08624627	0.09098715	0.0896867	0.09155394	0.09729053
α_n = 1e-04	0.08609531	0.08992405	0.08907391	0.09068455	0.09697694
α_n = 2e-04	0.08604253	0.08920539	0.08855415	0.08991349	0.09484127
α_n = 5e-04	0.08614867	0.08804619	0.08803973	0.0889454	0.09531481
α_n = 8e-04	0.08628651	0.08767901	0.08787182	0.08847695	0.0935051
α_n = 0.001	0.08637038	0.0877428	0.08781966	0.08828041	0.09574633
α_n = 0.002	0.08668675	0.08776286	0.08777866	0.08783789	0.09173923
α_n = 0.01	0.08740295	0.08853751	0.088278	0.08776926	0.09130512
α_n = 0.05	0.08784766	0.08907745	0.08846469	0.08754575	0.09015639
α_n = 0.1	0.08823282	0.0888508	0.08825112	0.08755756	0.09048149

Table B.31 Bias : Estimation using the four instruments : x_t, x_t^2, x_t^3, x_t^4 (Normal errors)

	n= 100	n= 200	n= 300
GMM	0.05538244	0.025916915	0.015111610
EL	0.02405452	0.007473711	0.005498350
ET	0.01990145	0.006186404	0.005267829
CUE	0.01663181	0.006124326	0.005178538

Table B.32 RMSE : Estimation using the four instruments : x_t, x_t^2, x_t^3, x_t^4 (Normal errors)

	n= 100	n= 200	n= 300
GMM	0.1940853	0.1192363	0.09217319
EL	0.2100733	0.1228505	0.09294868
ET	0.1836080	0.1158240	0.09099490
CUE	0.1691930	0.1132681	0.09047816

Table B.33 Bias : Estimation using the four instruments : x_t, x_t^2, x_t^3, x_t^4 (Skewed errors)

	n= 100	n= 200	n= 300
GMM	0.011433856	0.000846777	0.001218036
EL	0.003787733	0.001650563	0.002289558
ET	0.029538728	0.011496465	0.009227936
CUE	0.051105505	0.022305443	0.014860183

Table B.34 RMSE : Estimation using the four instruments : x_t, x_t^2, x_t^3, x_t^4 (Skewed errors)

	n= 100	n= 200	n= 300
GMM	0.2151296	0.1201280	0.09152139
EL	0.2394605	0.1406258	0.10732795
ET	0.2007928	0.1230824	0.09503599
CUE	0.1875064	0.1174230	0.09154561

B.3.2 Properties of the over-identifying restriction tests

Table B.35 J-test (CGMM) : Sample size n = 100 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.001	0.006	0.0135	0	0.006	0.0265
α_n = 1e-04	0.0015	0.007	0.0185	0	0.0075	0.0365
α_n = 2e-04	0.0015	0.009	0.0205	0	0.012	0.0455
α_n = 5e-04	0.002	0.017	0.0315	0	0.0175	0.063
α_n = 8e-04	0.0035	0.018	0.0355	5e-04	0.019	0.0755
$\alpha_n = 0.001$	0.0045	0.0195	0.0385	5e-04	0.021	0.081
$\alpha_n = 0.002$	0.007	0.0275	0.0565	0.0015	0.0305	0.099
$\alpha_n = 0.01$	0.025	0.073	0.1095	0.0035	0.076	0.1735
$\alpha_n = 0.05$	0.0725	0.1475	0.2005	0.022	0.145	0.273
$\alpha_n = 0.1$	0.0985	0.1795	0.2405	0.0355	0.1755	0.305

Table B.36 J-test (CGMM) : Sample size n = 200 (Normal errors)

	Asymptotic distribution			Gamma approximation		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.0025	0.013	0.0275	5e-04	0.014	0.0505
α_n = 1e-04	0.003	0.0135	0.0315	5e-04	0.0155	0.062
α_n = 2e-04	0.0045	0.0185	0.0365	5e-04	0.0205	0.074
α_n = 5e-04	0.0065	0.025	0.0465	5e-04	0.0265	0.0925
α_n = 8e-04	0.0075	0.028	0.06	5e-04	0.031	0.1025
$\alpha_n = 0.001$	0.0075	0.031	0.0635	5e-04	0.033	0.1065
$\alpha_n = 0.002$	0.01	0.038	0.0745	0.0015	0.0395	0.12
$\alpha_n = 0.01$	0.0155	0.0465	0.0755	0.0035	0.0455	0.1175
$\alpha_n = 0.05$	0.0185	0.039	0.0595	0.004	0.037	0.098
$\alpha_n = 0.1$	0.0195	0.0355	0.0545	0.004	0.0345	0.0875

Table B.37 J-test (CGMM) : Sample size n = 300 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.026	0.078	0.132	0.0115	0.084	0.237
α_n = 1e-04	0.0315	0.091	0.1645	0.0145	0.0955	0.279
α_n = 2e-04	0.0405	0.1115	0.1935	0.0165	0.119	0.333
α_n = 5e-04	0.058	0.147	0.229	0.02	0.151	0.408
α_n = 8e-04	0.064	0.159	0.246	0.0215	0.164	0.4325
$\alpha_n = 0.001$	0.0655	0.1625	0.2495	0.022	0.167	0.434
$\alpha_n = 0.002$	0.067	0.1605	0.241	0.0235	0.1615	0.395
$\alpha_n = 0.01$	0.042	0.101	0.142	0.0175	0.1	0.2
$\alpha_n = 0.05$	0.0325	0.063	0.088	0.014	0.0605	0.125
$\alpha_n = 0.1$	0.0335	0.0635	0.087	0.0135	0.0595	0.1165

Table B.38 J-test (CGMM) : Sample size n = 100 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.0015	0.021	0.038	0	0.022	0.071
α_n = 1e-04	0.0015	0.021	0.0415	0	0.024	0.0745
α_n = 2e-04	0.002	0.0235	0.0455	0	0.025	0.082
α_n = 5e-04	0.0035	0.0285	0.054	5e-04	0.032	0.096
α_n = 8e-04	0.0035	0.033	0.058	5e-04	0.0355	0.1085
$\alpha_n = 0.001$	0.004	0.0355	0.062	5e-04	0.036	0.116
$\alpha_n = 0.002$	0.0075	0.0445	0.0725	5e-04	0.0465	0.128
$\alpha_n = 0.01$	0.0235	0.0685	0.107	0.004	0.069	0.158
$\alpha_n = 0.05$	0.051	0.1125	0.151	0.011	0.111	0.2035
$\alpha_n = 0.1$	0.0745	0.132	0.1795	0.022	0.1315	0.2325

Table B.39 J-test (CGMM) : Sample size n = 200 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.0085	0.0195	0.0255	0.002	0.0205	0.042
α_n = 1e-04	0.0075	0.017	0.026	0.002	0.018	0.034
α_n = 2e-04	0.0075	0.017	0.026	0.0015	0.0185	0.0345
α_n = 5e-04	0.008	0.0175	0.028	0.0025	0.0195	0.038
α_n = 8e-04	0.009	0.0175	0.029	0.0025	0.0175	0.0405
$\alpha_n = 0.001$	0.009	0.0175	0.029	0.003	0.0185	0.0425
$\alpha_n = 0.002$	0.0095	0.018	0.0325	0.0045	0.018	0.0485
$\alpha_n = 0.01$	0.0125	0.0235	0.0425	0.0045	0.023	0.062
$\alpha_n = 0.05$	0.018	0.036	0.05	0.0045	0.0345	0.068
$\alpha_n = 0.1$	0.0205	0.0405	0.052	0.0055	0.0395	0.0725

Table B.40 J-test (CGMM) : Sample size n = 300 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0	0.001	0.0025	0	0.0015	0.0075
α_n = 1e-04	0	0.0015	0.0025	0	0.002	0.008
α_n = 2e-04	5e-04	0.002	0.0025	0	0.002	0.01
α_n = 5e-04	5e-04	0.002	0.0065	0	0.002	0.015
α_n = 8e-04	5e-04	0.002	0.008	0	0.002	0.019
$\alpha_n = 0.001$	5e-04	0.0015	0.0095	0	0.0015	0.023
$\alpha_n = 0.002$	5e-04	0.0055	0.016	0	0.0055	0.0285
$\alpha_n = 0.01$	0.005	0.017	0.0275	0	0.0165	0.0475
$\alpha_n = 0.05$	0.013	0.031	0.048	0.0015	0.029	0.077
$\alpha_n = 0.1$	0.015	0.0345	0.059	0.0045	0.0345	0.0875

Table B.41 J-test (CEL) : Sample size n = 100 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.0045	0.0165	0.035	0	0.0195	0.0625
α_n = 1e-04	0.005	0.0195	0.038	0.001	0.0225	0.0725
α_n = 2e-04	0.006	0.024	0.044	0.001	0.0265	0.083
α_n = 5e-04	0.009	0.0315	0.0575	0.001	0.0325	0.091
α_n = 8e-04	0.011	0.034	0.065	0.0015	0.036	0.099
$\alpha_n = 0.001$	0.0125	0.037	0.0695	0.002	0.039	0.106
$\alpha_n = 0.002$	0.015	0.0485	0.08	0.003	0.05	0.122
$\alpha_n = 0.01$	0.0325	0.083	0.1215	0.007	0.0835	0.1885
$\alpha_n = 0.05$	0.0735	0.1455	0.2015	0.024	0.1445	0.2695
$\alpha_n = 0.1$	0.093	0.172	0.232	0.0355	0.168	0.2985

Table B.42 J-test (CEL) : Sample size n = 200 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.004	0.019	0.04	5e-04	0.0225	0.073
α_n = 1e-04	0.0055	0.0215	0.0495	5e-04	0.024	0.081
α_n = 2e-04	0.007	0.026	0.053	5e-04	0.03	0.095
α_n = 5e-04	0.007	0.033	0.062	5e-04	0.035	0.1105
α_n = 8e-04	0.01	0.0385	0.0695	0.0015	0.0395	0.119
$\alpha_n = 0.001$	0.011	0.04	0.071	0.0015	0.043	0.123
$\alpha_n = 0.002$	0.0135	0.047	0.082	0.002	0.049	0.1305
$\alpha_n = 0.01$	0.017	0.049	0.0825	0.004	0.049	0.1195
$\alpha_n = 0.05$	0.019	0.037	0.059	0.004	0.0365	0.0965
$\alpha_n = 0.1$	0.018	0.0365	0.052	0.0035	0.0345	0.0855

Table B.43 J-test (CEL) : Sample size n = 300 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.036	0.0935	0.152	0.0155	0.103	0.2705
α_n = 1e-04	0.043	0.1065	0.177	0.019	0.115	0.3015
α_n = 2e-04	0.0495	0.1265	0.212	0.019	0.1315	0.351
α_n = 5e-04	0.0655	0.153	0.257	0.024	0.16	0.422
α_n = 8e-04	0.071	0.1665	0.265	0.0275	0.173	0.443
$\alpha_n = 0.001$	0.072	0.1715	0.2645	0.028	0.178	0.4455
$\alpha_n = 0.002$	0.0715	0.167	0.2485	0.0285	0.171	0.406
$\alpha_n = 0.01$	0.0425	0.098	0.1435	0.02	0.0975	0.2005
$\alpha_n = 0.05$	0.032	0.064	0.0845	0.0145	0.062	0.119
$\alpha_n = 0.1$	0.034	0.063	0.0835	0.014	0.0615	0.1145

Table B.44 J-test (CEL) : Sample size n = 100 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.0065	0.031	0.053	0.0025	0.0335	0.0915
α_n = 1e-04	0.0095	0.033	0.0555	0.0025	0.036	0.0945
α_n = 2e-04	0.01	0.0305	0.056	5e-04	0.034	0.093
α_n = 5e-04	0.008	0.032	0.0585	5e-04	0.035	0.0995
α_n = 8e-04	0.0075	0.036	0.0605	5e-04	0.0385	0.1065
$\alpha_n = 0.001$	0.007	0.0385	0.064	5e-04	0.041	0.11
$\alpha_n = 0.002$	0.0075	0.0435	0.0725	5e-04	0.044	0.1185
$\alpha_n = 0.01$	0.0235	0.0685	0.1075	0.0035	0.0685	0.1555
$\alpha_n = 0.05$	0.055	0.114	0.156	0.011	0.113	0.2185
$\alpha_n = 0.1$	0.0775	0.1385	0.1895	0.023	0.138	0.246

Table B.45 J-test (CEL) : Sample size n = 200 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.014	0.0305	0.0385	0.006	0.0315	0.0545
α_n = 1e-04	0.011	0.024	0.033	0.004	0.0255	0.0485
α_n = 2e-04	0.01	0.0235	0.033	0.0035	0.0255	0.0435
α_n = 5e-04	0.011	0.0225	0.031	0.0045	0.0235	0.0425
α_n = 8e-04	0.011	0.0215	0.0295	0.0045	0.0235	0.0425
$\alpha_n = 0.001$	0.0115	0.021	0.03	0.005	0.022	0.042
$\alpha_n = 0.002$	0.0115	0.0215	0.03	0.004	0.0215	0.047
$\alpha_n = 0.01$	0.0135	0.0245	0.037	0.005	0.0245	0.0575
$\alpha_n = 0.05$	0.018	0.036	0.0515	0.0055	0.0335	0.0675
$\alpha_n = 0.1$	0.02	0.039	0.0545	0.0065	0.038	0.072

Table B.46 J-test (CEL) : Sample size n = 300 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.003	0.0045	0.0095	0.002	0.0055	0.015
α_n = 1e-04	0.0025	0.0045	0.0095	0.0015	0.0045	0.0155
α_n = 2e-04	0.0025	0.0055	0.0085	0.0015	0.0055	0.0175
α_n = 5e-04	0.0025	0.007	0.0105	0.001	0.007	0.019
α_n = 8e-04	0.003	0.0065	0.0105	0.001	0.0065	0.0245
$\alpha_n = 0.001$	0.003	0.0065	0.0115	0.001	0.0065	0.026
$\alpha_n = 0.002$	0.0035	0.008	0.0165	0.001	0.0085	0.03
$\alpha_n = 0.01$	0.007	0.021	0.0305	0.002	0.02	0.046
$\alpha_n = 0.05$	0.0145	0.033	0.0475	0.0045	0.032	0.074
$\alpha_n = 0.1$	0.0185	0.0385	0.0595	0.0055	0.0385	0.0875

Table B.47 LR-test (CEL) : Sample size n = 100 (Normal errors)

	Asymptotic distribution			Gamma approximation		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.031	0.059	0.076	0.0165	0.061	0.1085
α_n = 1e-04	0.026	0.0475	0.0635	0.014	0.048	0.092
α_n = 2e-04	0.025	0.0415	0.0545	0.0125	0.0435	0.082
α_n = 5e-04	0.0195	0.0355	0.0485	0.009	0.037	0.0665
α_n = 8e-04	0.019	0.0325	0.0435	0.0085	0.033	0.0615
$\alpha_n = 0.001$	0.0175	0.032	0.042	0.008	0.0325	0.0575
$\alpha_n = 0.002$	0.0145	0.029	0.0355	0.005	0.03	0.052
$\alpha_n = 0.01$	0.0105	0.021	0.031	0.004	0.0215	0.0405
$\alpha_n = 0.05$	0.0065	0.017	0.0205	0.002	0.017	0.0285
$\alpha_n = 0.1$	0.0045	0.01	0.015	0.0015	0.0095	0.02

Table B.48 LR-test (CEL) : Sample size n = 200 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.0255	0.051	0.079	0.015	0.0555	0.106
α_n = 1e-04	0.021	0.045	0.0665	0.0125	0.048	0.0955
α_n = 2e-04	0.0185	0.0395	0.06	0.011	0.042	0.083
α_n = 5e-04	0.0165	0.033	0.053	0.009	0.035	0.073
α_n = 8e-04	0.0165	0.0315	0.051	0.009	0.0335	0.073
$\alpha_n = 0.001$	0.0145	0.032	0.0515	0.0095	0.033	0.072
$\alpha_n = 0.002$	0.0175	0.0315	0.0495	0.007	0.0315	0.067
$\alpha_n = 0.01$	0.0155	0.033	0.0445	0.006	0.033	0.065
$\alpha_n = 0.05$	0.0095	0.023	0.031	0.0025	0.0225	0.043
$\alpha_n = 0.1$	0.006	0.0135	0.0235	0.0015	0.012	0.033

Table B.49 LR-test (CEL) : Sample size n = 300 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.0315	0.0535	0.0775	0.02	0.057	0.1045
α_n = 1e-04	0.0315	0.054	0.074	0.019	0.0565	0.1005
α_n = 2e-04	0.0265	0.0505	0.07	0.0165	0.053	0.0955
α_n = 5e-04	0.0245	0.049	0.062	0.015	0.0495	0.087
α_n = 8e-04	0.025	0.0455	0.0595	0.015	0.0455	0.083
$\alpha_n = 0.001$	0.0235	0.044	0.059	0.0145	0.0445	0.0795
$\alpha_n = 0.002$	0.022	0.038	0.0535	0.0125	0.039	0.0725
$\alpha_n = 0.01$	0.0155	0.0315	0.042	0.0105	0.0315	0.058
$\alpha_n = 0.05$	0.0135	0.0205	0.0265	0.0075	0.0205	0.039
$\alpha_n = 0.1$	0.01	0.0165	0.0215	0.004	0.016	0.029

Table B.50 LR-test (CEL) : Sample size n = 100 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.0755	0.1185	0.1545	0.0485	0.119	0.1895
α_n = 1e-04	0.0665	0.0965	0.131	0.0395	0.0965	0.1635
α_n = 2e-04	0.053	0.0805	0.1075	0.0325	0.0815	0.142
α_n = 5e-04	0.0335	0.058	0.081	0.0215	0.0575	0.109
α_n = 8e-04	0.031	0.0495	0.068	0.019	0.05	0.0945
$\alpha_n = 0.001$	0.0295	0.0465	0.0615	0.0165	0.0465	0.088
$\alpha_n = 0.002$	0.0185	0.036	0.0495	0.01	0.036	0.066
$\alpha_n = 0.01$	0.009	0.0145	0.0225	0.0025	0.0145	0.0305
$\alpha_n = 0.05$	5e-04	0.0035	0.007	5e-04	0.0035	0.01
$\alpha_n = 0.1$	5e-04	0.001	0.002	5e-04	0.001	0.006

Table B.51 LR-test (CEL) : Sample size n = 200 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.072	0.1045	0.1275	0.053	0.1035	0.157
α_n = 1e-04	0.0665	0.089	0.1165	0.049	0.0895	0.1395
α_n = 2e-04	0.0585	0.0785	0.103	0.047	0.082	0.128
α_n = 5e-04	0.044	0.0685	0.084	0.0335	0.069	0.109
α_n = 8e-04	0.0405	0.0625	0.079	0.0295	0.063	0.104
$\alpha_n = 0.001$	0.037	0.058	0.0735	0.027	0.0595	0.1
$\alpha_n = 0.002$	0.031	0.0495	0.0645	0.02	0.0495	0.0925
$\alpha_n = 0.01$	0.015	0.032	0.043	0.0055	0.032	0.0585
$\alpha_n = 0.05$	0.0065	0.015	0.021	0.0015	0.0135	0.033
$\alpha_n = 0.1$	0.005	0.007	0.0125	5e-04	0.007	0.0205

Table B.52 LR-test (CEL) : Sample size n = 300 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.0825	0.1225	0.148	0.058	0.1225	0.181
α_n = 1e-04	0.072	0.103	0.133	0.05	0.1045	0.164
α_n = 2e-04	0.0645	0.091	0.122	0.0405	0.0925	0.147
α_n = 5e-04	0.0495	0.078	0.106	0.029	0.079	0.1305
α_n = 8e-04	0.042	0.067	0.097	0.025	0.066	0.1195
$\alpha_n = 0.001$	0.038	0.0645	0.0875	0.0205	0.0655	0.115
$\alpha_n = 0.002$	0.029	0.0505	0.0695	0.0135	0.0505	0.0925
$\alpha_n = 0.01$	0.0125	0.0305	0.044	0.005	0.0295	0.0605
$\alpha_n = 0.05$	0.0045	0.0135	0.0165	0.001	0.0135	0.0275
$\alpha_n = 0.1$	0.002	0.0075	0.013	0.001	0.007	0.017

Table B.53 LM-test (CEL) : Sample size n = 100 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.179	0.2755	0.3425	0.0505	0.102	0.1475
α_n = 1e-04	0.1415	0.232	0.2855	0.0325	0.077	0.117
α_n = 2e-04	0.108	0.186	0.239	0.02	0.053	0.0865
α_n = 5e-04	0.067	0.128	0.18	0.012	0.0335	0.056
α_n = 8e-04	0.0535	0.1015	0.155	0.007	0.0285	0.042
$\alpha_n = 0.001$	0.046	0.093	0.1415	0.0065	0.026	0.0395
$\alpha_n = 0.002$	0.033	0.0605	0.0955	0.0055	0.017	0.028
$\alpha_n = 0.01$	0.0125	0.026	0.036	0.001	0.0045	0.0125
$\alpha_n = 0.05$	0.0025	0.0045	0.007	0	0.001	0.0015
$\alpha_n = 0.1$	5e-04	0.0015	0.002	0	0	0.001

Table B.54 LM-test (CEL) : Sample size n = 200 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.1305	0.2175	0.3025	0.023	0.0615	0.0975
α_n = 1e-04	0.1005	0.176	0.2525	0.0155	0.039	0.072
α_n = 2e-04	0.076	0.1455	0.2055	0.01	0.0315	0.0545
α_n = 5e-04	0.0505	0.1045	0.152	0.0065	0.0185	0.041
α_n = 8e-04	0.043	0.0895	0.1305	0.004	0.0165	0.037
$\alpha_n = 0.001$	0.041	0.083	0.122	0.0035	0.0165	0.036
$\alpha_n = 0.002$	0.031	0.068	0.0975	0.0025	0.0125	0.0295
$\alpha_n = 0.01$	0.011	0.031	0.047	0.0015	0.004	0.014
$\alpha_n = 0.05$	0.001	0.002	0.0045	0	0.001	0.001
$\alpha_n = 0.1$	0	5e-04	0.001	0	0	0

Table B.55 LM-test (CEL) : Sample size n = 300 (Normal errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.119	0.2105	0.2805	0.017	0.0485	0.08
α_n = 1e-04	0.094	0.182	0.246	0.015	0.039	0.068
α_n = 2e-04	0.0725	0.1515	0.2085	0.0115	0.033	0.056
α_n = 5e-04	0.055	0.1105	0.1615	0.008	0.0245	0.0435
α_n = 8e-04	0.0425	0.0955	0.14	0.0065	0.0195	0.038
$\alpha_n = 0.001$	0.039	0.0865	0.1335	0.0065	0.017	0.037
$\alpha_n = 0.002$	0.0295	0.0625	0.1055	0.0055	0.0145	0.0285
$\alpha_n = 0.01$	0.014	0.025	0.042	0.003	0.01	0.015
$\alpha_n = 0.05$	0.001	0.0035	0.007	0	5e-04	0.0015
$\alpha_n = 0.1$	0	0	0	0	0	0

Table B.56 LM-test (CEL) : Sample size n = 100 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.252	0.3445	0.406	0.1125	0.171	0.2195
α_n = 1e-04	0.2115	0.2935	0.349	0.0955	0.149	0.186
α_n = 2e-04	0.1745	0.2465	0.302	0.078	0.1225	0.157
α_n = 5e-04	0.1305	0.1875	0.234	0.0515	0.0875	0.12
α_n = 8e-04	0.1135	0.16	0.1985	0.043	0.0785	0.105
$\alpha_n = 0.001$	0.1045	0.1495	0.1925	0.0385	0.0715	0.099
$\alpha_n = 0.002$	0.085	0.117	0.146	0.0255	0.058	0.082
$\alpha_n = 0.01$	0.028	0.0495	0.068	0.0065	0.015	0.0285
$\alpha_n = 0.05$	0.001	0.0045	0.0085	0.001	0.001	0.0025
$\alpha_n = 0.1$	0.001	0.0015	0.002	0.001	0.001	0.0015

Table B.57 LM-test (CEL) : Sample size n = 200 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.204	0.285	0.3475	0.0755	0.1285	0.1735
α_n = 1e-04	0.177	0.256	0.3045	0.0635	0.1115	0.144
α_n = 2e-04	0.143	0.2235	0.271	0.055	0.0965	0.125
α_n = 5e-04	0.1095	0.169	0.21	0.0375	0.07	0.0985
α_n = 8e-04	0.0905	0.145	0.1835	0.0315	0.055	0.082
$\alpha_n = 0.001$	0.081	0.1345	0.1735	0.0285	0.0495	0.0745
$\alpha_n = 0.002$	0.061	0.102	0.1375	0.018	0.0375	0.06
$\alpha_n = 0.01$	0.023	0.039	0.0525	0.0025	0.0125	0.026
$\alpha_n = 0.05$	0.0015	0.005	0.008	5e-04	0.001	0.002
$\alpha_n = 0.1$	5e-04	0.001	0.002	0	5e-04	5e-04

Table B.58 LM-test (CEL) : Sample size n = 300 (Skewed errors)

	Asymptotic distribution			**Gamma approximation**		
	Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
α_n = 5e-05	0.207	0.288	0.351	0.081	0.1375	0.1775
α_n = 1e-04	0.169	0.2415	0.3015	0.0655	0.11	0.1495
α_n = 2e-04	0.1385	0.211	0.266	0.0535	0.097	0.127
α_n = 5e-04	0.1015	0.153	0.201	0.034	0.068	0.098
α_n = 8e-04	0.086	0.133	0.1735	0.0275	0.052	0.084
$\alpha_n = 0.001$	0.076	0.122	0.162	0.024	0.0445	0.072
$\alpha_n = 0.002$	0.05	0.09	0.124	0.015	0.0315	0.049
$\alpha_n = 0.01$	0.0155	0.0325	0.047	0.003	0.0105	0.016
$\alpha_n = 0.05$	0.001	0.003	0.0055	5e-04	5e-04	0.002
$\alpha_n = 0.1$	5e-04	5e-04	5e-04	5e-04	5e-04	5e-04

Table B.59 J, LM and LR tests for CEL, using optimal α_n (Normal errors)

		Asymptotic distribution			**Gamma approximation**		
		Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
LR	n = 100	0.0105	0.0210	0.0310	0.0040	0.0215	0.0405
	n = 200	0.0175	0.0315	0.0495	0.0070	0.0315	0.0670
	n = 300	0.0235	0.0440	0.0590	0.0145	0.0445	0.0795
LM	n = 100	0.0125	0.0260	0.0360	0.0010	0.0045	0.0125
	n = 200	0.0310	0.0680	0.0975	0.0025	0.0125	0.0295
	n = 300	0.0390	0.0865	0.1335	0.0065	0.0170	0.0370
J	n = 100	0.0325	0.0830	0.1215	0.0070	0.0835	0.1885
	n = 200	0.0135	0.0470	0.0820	0.0020	0.0490	0.1305
	n = 300	0.0720	0.1715	0.2645	0.0280	0.1780	0.4455

Table B.60 J, LM and LR tests for CEL, using optimal α_n (Skewed errors)

		Asymptotic distribution			**Gamma approximation**		
		Size = 0.01	Size = 0.05	Size = 0.10	Size = 0.01	Size = 0.05	Size = 0.10
LR	n = 100	0.0005	0.0010	0.0020	0.0005	0.0010	0.0060
	n = 200	0.0310	0.0495	0.0645	0.0200	0.0495	0.0925
	n = 300	0.0420	0.0670	0.0970	0.0250	0.0660	0.1195
LM	n = 100	0.0010	0.0015	0.0020	0.0010	0.0010	0.0015
	n = 200	0.0610	0.1020	0.1375	0.0180	0.0375	0.060
	n = 300	0.0860	0.1330	0.1735	0.0275	0.0520	0.0840
J	n = 100	0.0775	0.1385	0.1895	0.0230	0.1380	0.2460
	n = 200	0.0115	0.0215	0.0300	0.0040	0.0215	0.0470
	n = 300	0.0030	0.0065	0.0105	0.0010	0.0065	0.0245

Table B.61 Probability of rejecting $H0 : d = 0.1$, using the asymptotic distribution (CEL, n=300, Normal errors).

	0.01	0.05	0.1
$\alpha_n =$ 5e-05	1e-03	0.0095	0.0310
$\alpha_n =$ 1e-04	1e-03	0.0095	0.0295
$\alpha_n =$ 2e-04	1e-03	0.0100	0.0295
$\alpha_n =$ 5e-04	1e-03	0.0095	0.0285
$\alpha_n =$ 8e-04	1e-03	0.0105	0.0285
$\alpha_n = 0.001$	1e-03	0.0110	0.0280
$\alpha_n = 0.002$	1e-03	0.0095	0.0275
$\alpha_n = 0.01$	5e-04	0.0095	0.0260
$\alpha_n = 0.05$	5e-04	0.0075	0.0185
$\alpha_n = 0.1$	0e+00	0.0050	0.0130

Figure B.1 Bias and RMSE, n=100, Normal errors

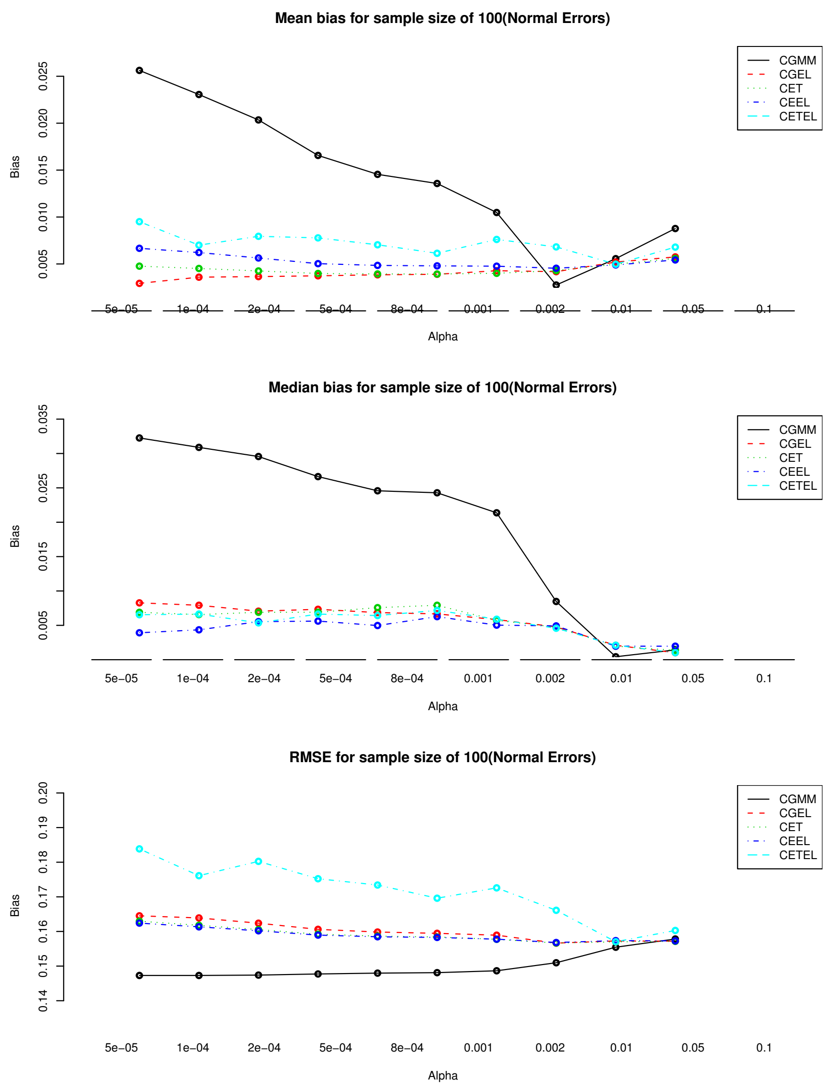


Figure B.2 Bias and RMSE, n=200, Normal errors

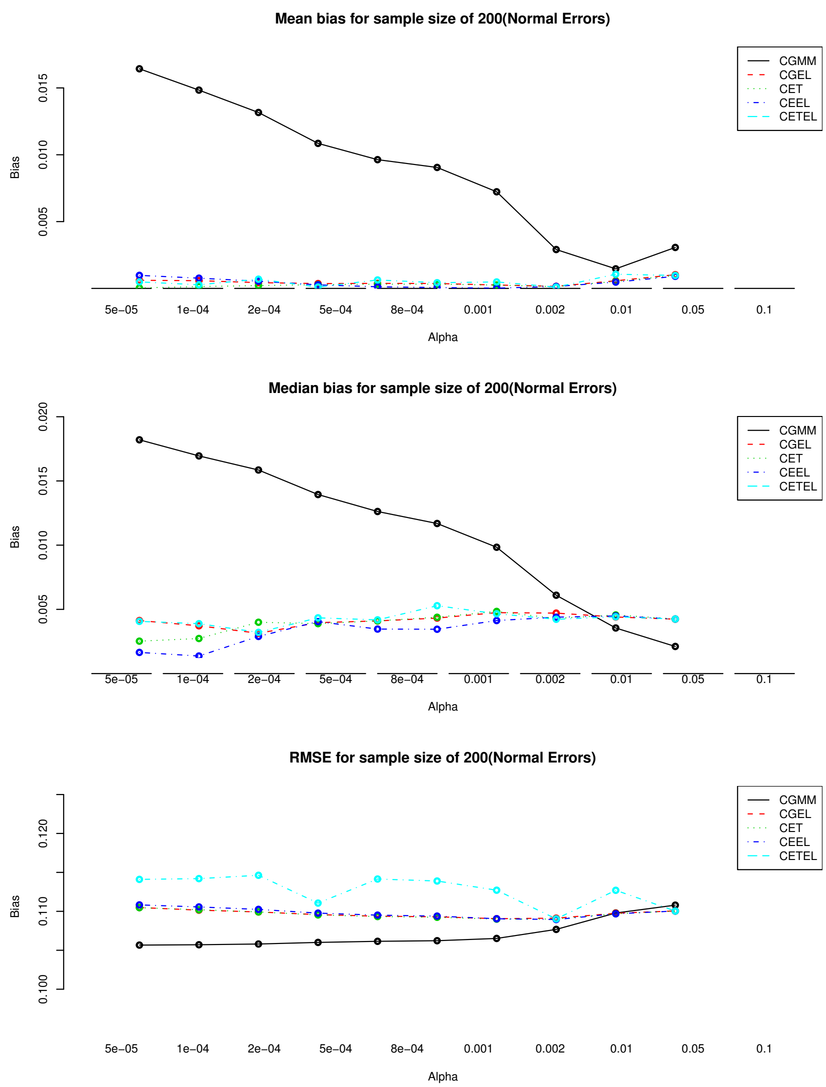


Figure B.3 Bias and RMSE, n=300, Normal errors

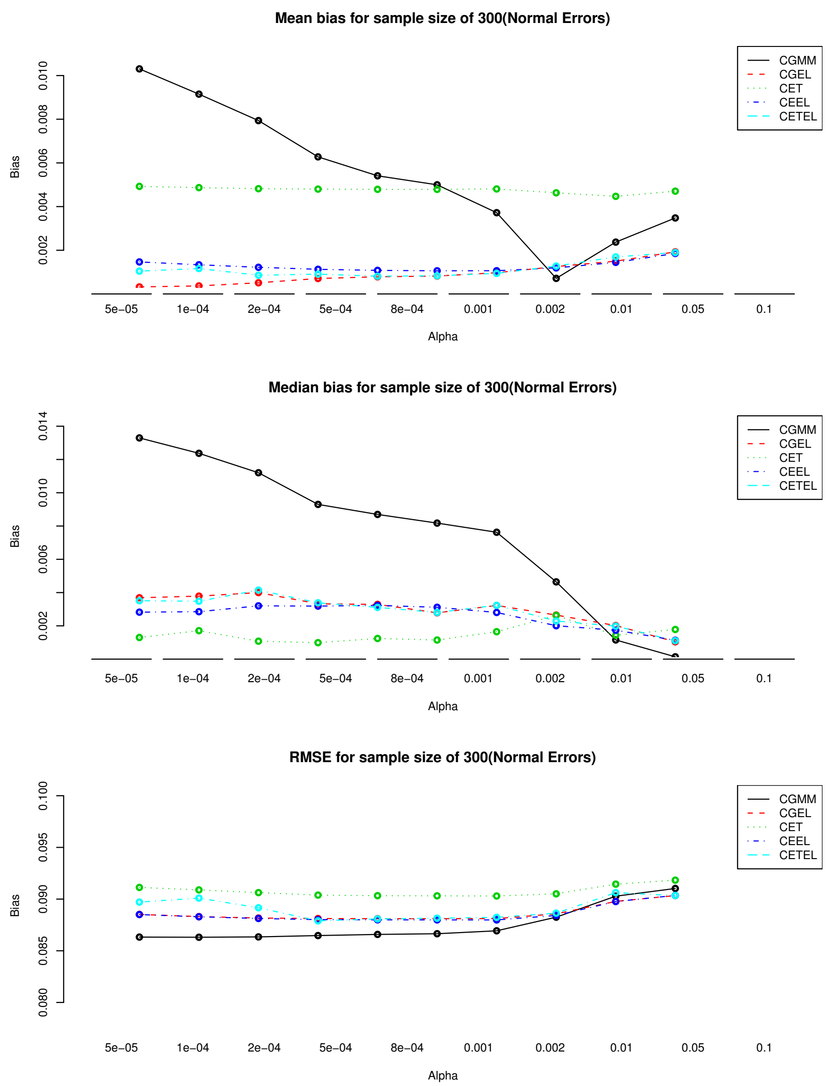


Figure B.4 Bias and RMSE, n=100, Skewed errors

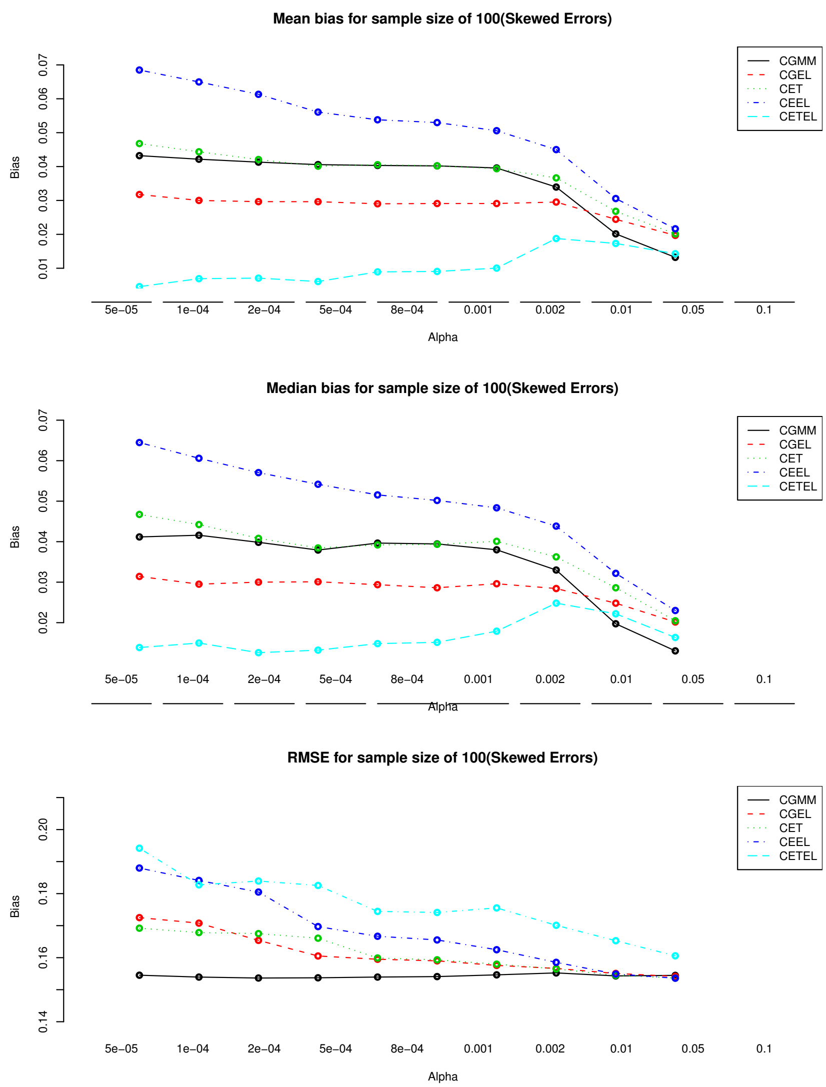


Figure B.5 Bias and RMSE, n=200, Skewed errors

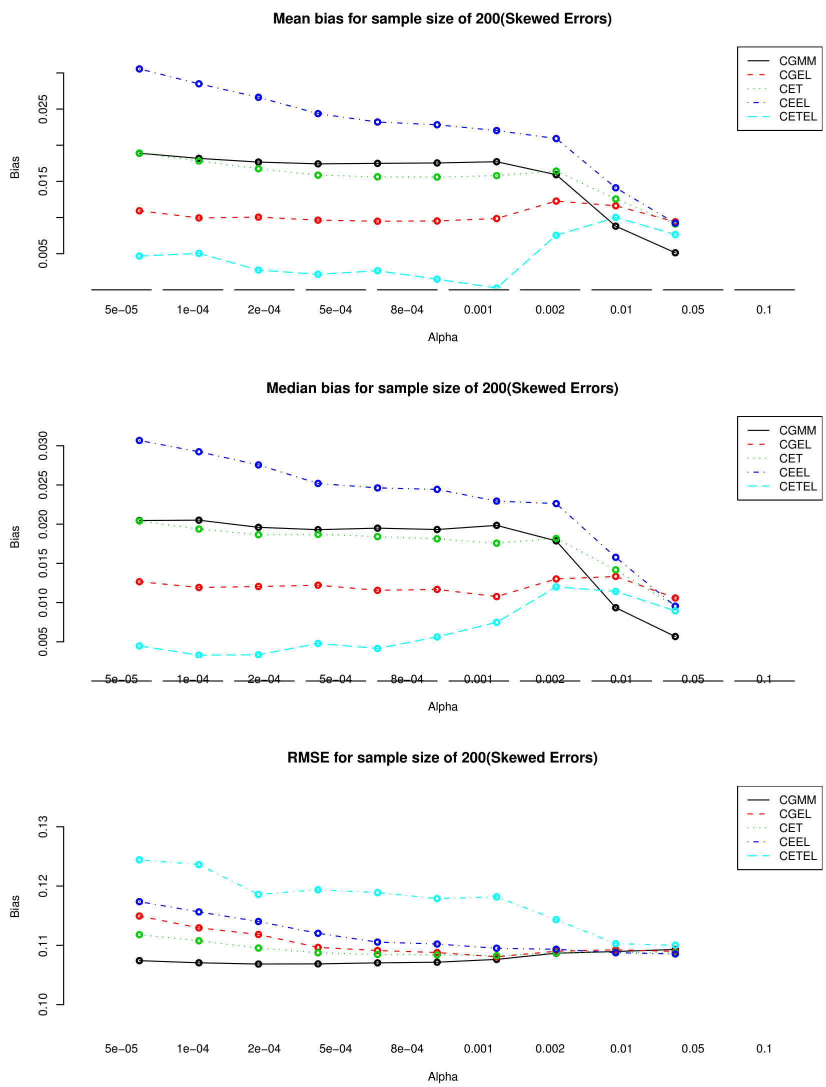
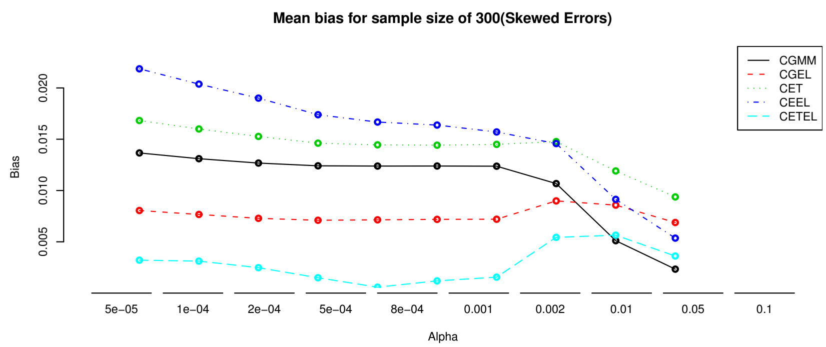
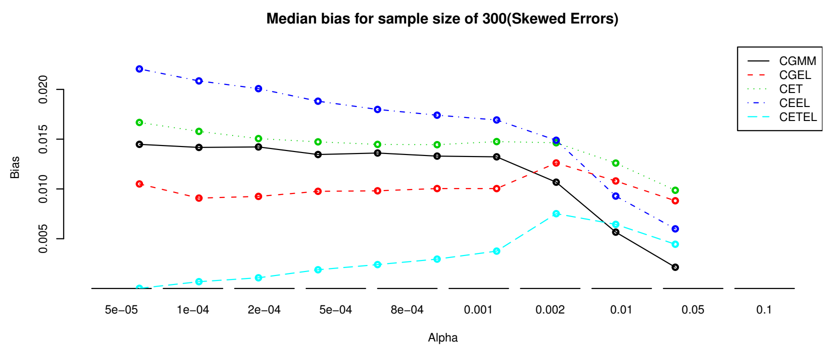


Figure B.6 Bias and RMSE, n=300, Skewed errors





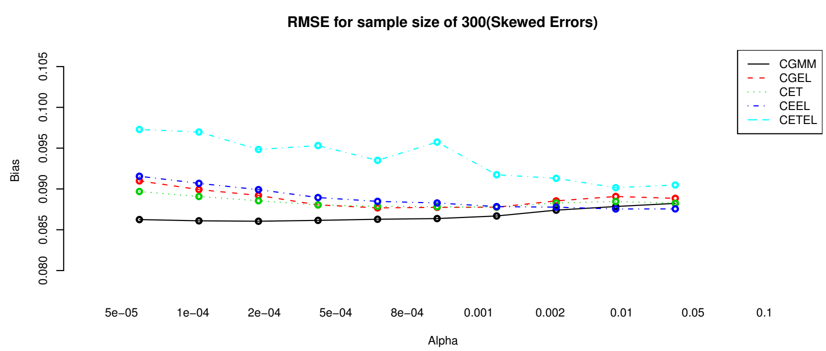


Figure B.7 Estimated λ for CEL, n=100, and $\alpha_n = 0.0001$.

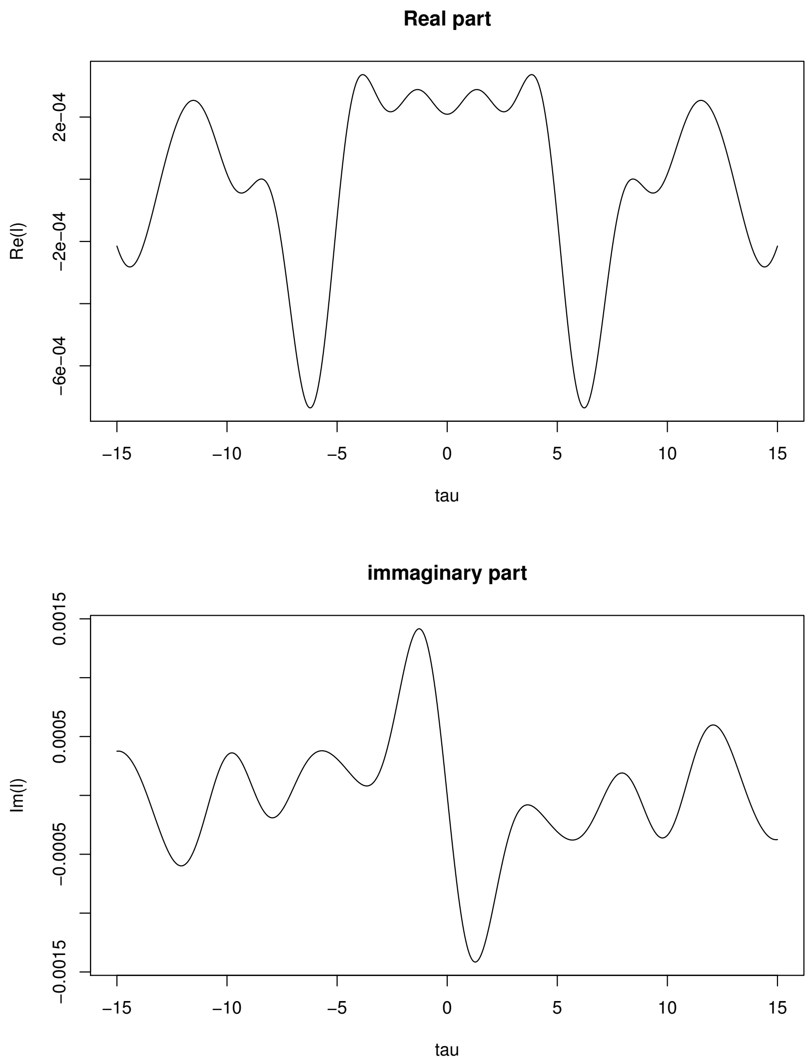
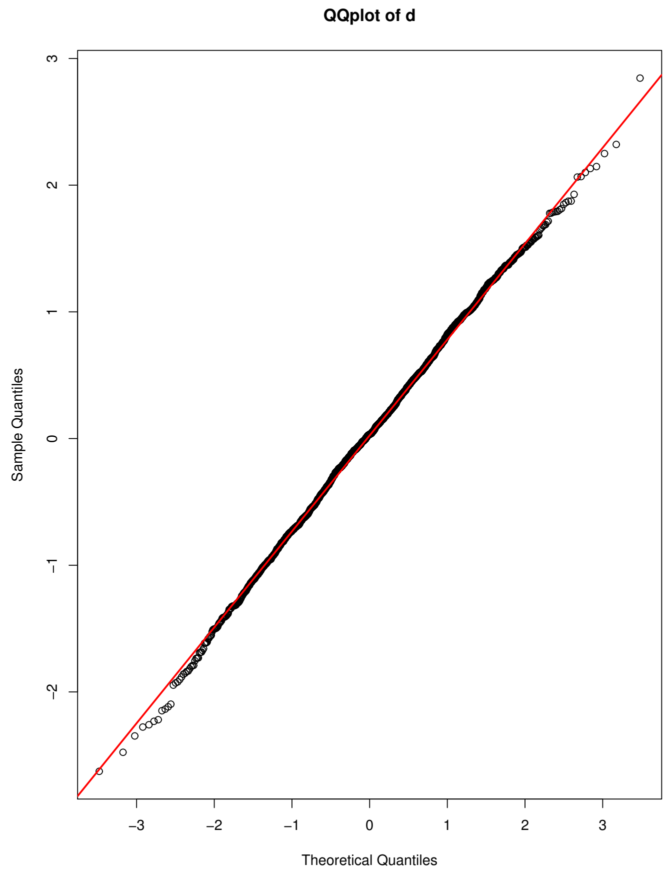


Figure B.8 QQ-plot for $\hat{d}$ (CEL, n = 300, normal errors, and $\alpha_n = 0.00005$).



B.4 Results : CIR process

Table B.62 Estimation of the CIR with sample size of 500

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

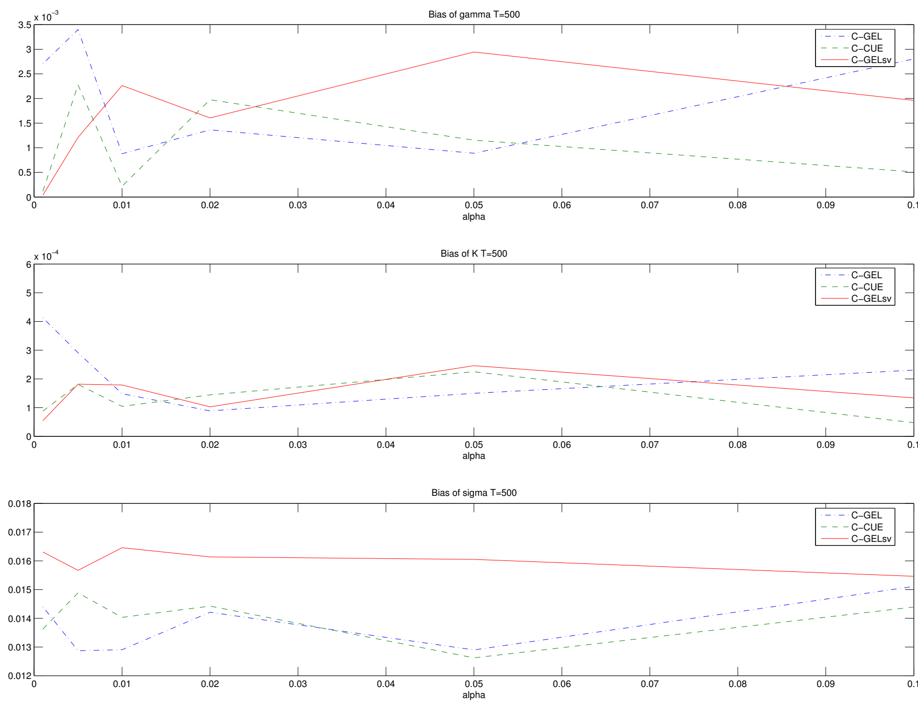


Figure B.9 Bias of the different estimators for T=500

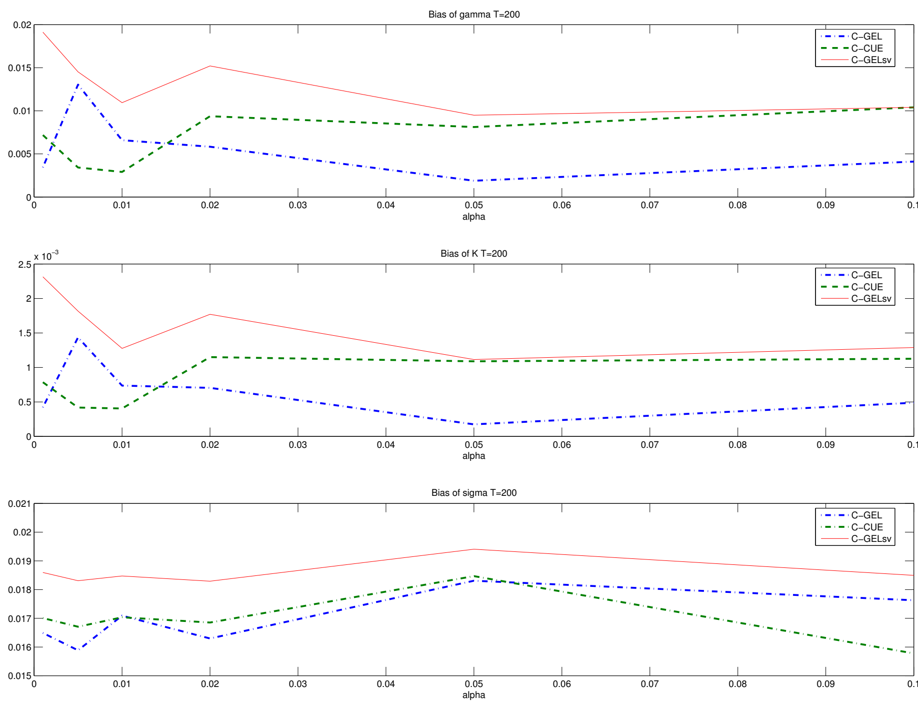


Figure B.10 Bias of the different estimators for T=200

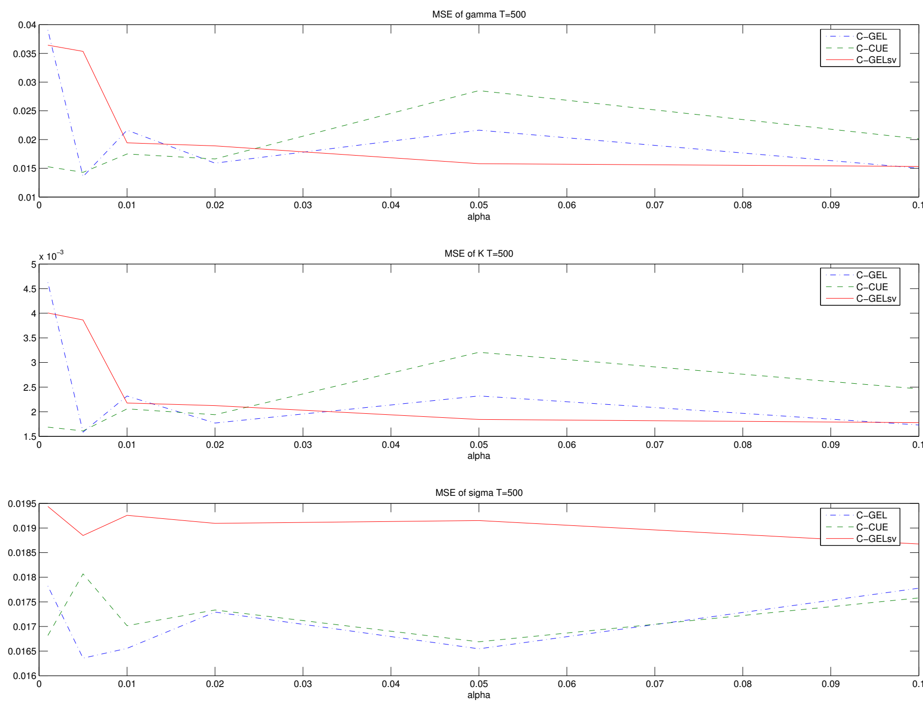


Figure B.11 MSE of the different estimators for T=500

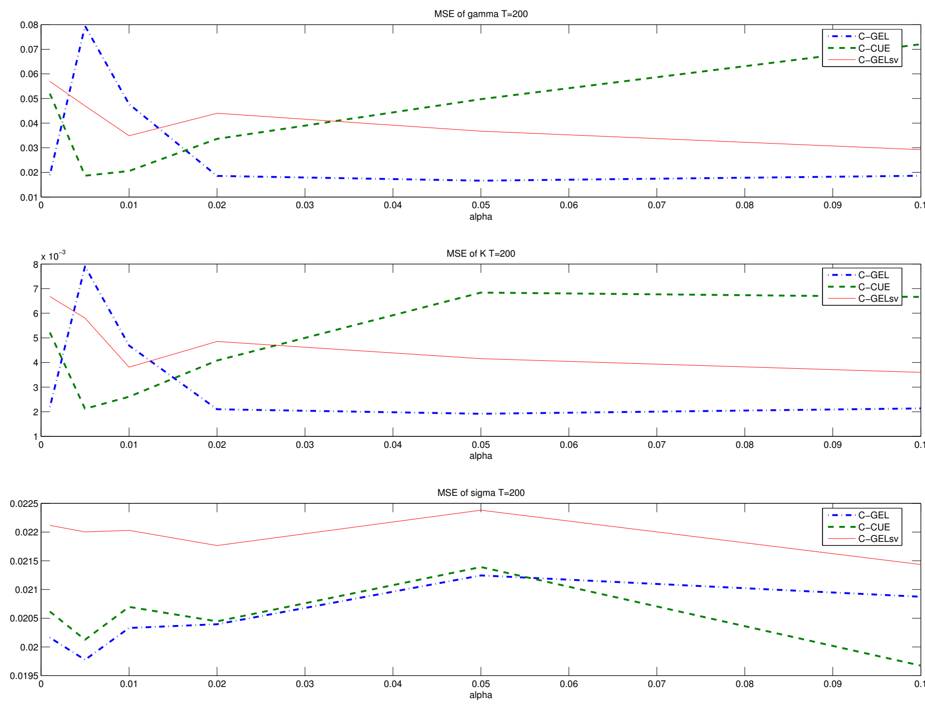


Figure B.12 MSE of the different estimators for T=200

ANNEXE C

APPENDIX : CHAPTER 3

C.1 Higher order expansion

C.1.1 Notation and setup

The regularized first order conditions of CGEL can be written as :

$$\frac{1}{n}\sum_{t=1}^{n} F_{1t}(\theta,\lambda) = 0 \tag{C.1}$$

and

$$\frac{1}{n}\sum_{t=1}^{n} F_{2t}(\theta,\lambda) = 0, \tag{C.2}$$

where

$$F_{1t} = \left[\frac{1}{n}\sum_{s=1}^{n} \rho''(<\lambda, g_s(\theta)>)g_s(\theta)g_s(\theta)\right]\rho'\left(<\lambda, g_t(\theta)>\right)g_t(\theta) + \alpha\lambda, \tag{C.3}$$

and

$$F_{2t} = \rho'\left(<\lambda, g_t(\theta)>\right)\left\langle \lambda, G_t(\theta)\right\rangle. \tag{C.4}$$

In what follows, we use for simplicity the linear operator notation instead of the explicit $<,>$ notation. Therefore, the inner product in $L^2(\pi)$, $< f, g >$, will be written as fg. Without loss of generality, we suppose that $\theta_0 = 0$. Since $\hat{\lambda} \xrightarrow{P} 0$, we expand the solution around $\theta = 0$ and $\lambda = 0$. We restrict our expansion to the case in which $\theta \in \mathbb{R}$. The extension to a more general case is straightforward. The following are linear operators, which are defined by their kernels. In each case, $g_t = g_t(0; .)$, and $G_t = [\partial g_t/\partial\theta](0; .)$.

$$\hat{g} = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau), \quad \hat{G} = \frac{1}{n}\sum_{t=1}^{n} G_t(\tau), \quad \hat{K} = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)g_t(\tau_2),$$

$$\hat{S} = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)g_t(\tau_2)g_t(\tau_3), \quad \hat{V} = \frac{1}{n}\sum_{t=1}^{n} G_t(\tau_1)g_t(\tau_2), \quad \hat{V}^* = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)G_t(\tau_2),$$

$$\hat{Q} = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)g_t(\tau_2)g_t(\tau_3)g_t(\tau_4), \quad \hat{B}\frac{1}{n}\sum_{t=1}^{n} G_t(\tau_1)G_t(\tau_2),$$

$$\hat{\Gamma} = \frac{1}{n}\sum_{t=1}^{n} G_t(\tau_1)g_t(\tau_2)g_t(\tau_3), \quad \hat{\Gamma}^* = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)g_t(\tau_2)G_t(\tau_3),$$

$$\hat{\Gamma}^{**} = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)G_t(\tau_2)g_t(\tau_3), \quad \hat{F} = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)g_t(\tau_2)g_t(\tau_3)g_t(\tau_4)g_t(\tau_5),$$

$$\hat{H} = \frac{1}{n}\sum_{t=1}^{n} G_t(\tau_1)G_t(\tau_2)g_t(\tau_3), \quad \hat{H}^* = \frac{1}{n}\sum_{t=1}^{n} G_t(\tau_1)g_t(\tau_2)G_t(\tau_3),$$

$$\hat{H}^{**} = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)G_t(\tau_2)G_t(\tau_3), \quad \hat{L} = \frac{1}{n}\sum_{t=1}^{n} G_t(\tau_1)g_t(\tau_2)g_t(\tau_3)g_t(\tau_4),$$

$$\hat{L}^* = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)G_t(\tau_2)g_t(\tau_3)g_t(\tau_4), \quad \hat{L}^{**} = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)g_t(\tau_2)G_t(\tau_3)g_t(\tau_4),$$

$$\hat{L}^{***} = \frac{1}{n}\sum_{t=1}^{n} g_t(\tau_1)g_t(\tau_2)g_t(\tau_3)G_t(\tau_4).$$

When applied to $f \in L^2(\pi)$, these operators integrate f with the first argument of the kernel. For example,

$$\hat{H}f = \int \frac{1}{n}\sum_{t=1}^{n} G_t(\tau_1)G_t(\tau_2)g_t(\tau_3)f(\tau_3)d\pi(\tau_3),$$

but

$$\hat{H}^*f = \int \frac{1}{n}\sum_{t=1}^{n} G_t(\tau_1)g_t(\tau_2)G_t(\tau_3)f(\tau_3)d\pi(\tau_3),$$

However, this is for the general case in which $g_t(\tau)$ is not the product of residuals and instruments. In the case we are interested in, $g_t(\tau) = (y_t - \theta W_t)Z(x_t, \tau)$, where the instruments $Z(x_t, \tau)$ are $e^{ix_t\tau}$. In that case, it is easy to see that $\hat{H} = \hat{H}^*$:

$$\hat{H}f = \int \frac{1}{n}\sum_{t=1}^{n} \left(-W_t e^{ix_t\tau_1}\right)\left(-W_t e^{ix_t\tau_2}\right)\left((y_t - \theta W_t)e^{ix_t\tau_3}\right) f(\tau_3)d\pi(\tau_3),$$

but

$$\hat{H}^*f = \int \frac{1}{n}\sum_{t=1}^n \left(-W_t e^{ix_t\tau_1}\right)\left((y_t-\theta W_t)e^{ix_t\tau_2}\right)\left(-W_t e^{ix_t\tau_3}\right) f(\tau_3)d\pi(\tau_3)$$
$$= \int \frac{1}{n}\sum_{t=1}^n \left(-W_t e^{ix_t\tau_1}\right)\left(-W_t e^{ix_t\tau_2}\right)\left((y_t-\theta W_t)e^{ix_t\tau_3}\right) f(\tau_3)d\pi(\tau_3) = \hat{H}f.$$

The expansion follows (Donald, Imbens and Newey, 2010) who derive the second order MSE of GEL in order to select the optimal number of instruments. We can write the first order conditions as a just identified model if we define $m_t(\lambda,\theta)\in L^2(\pi)\otimes\mathbb{R}$ as $\{F_{1t},F_{2t}\}'$. In the following, θ and λ refers to the solution of the first order conditions. For any function or operator, f, when we write $f(\theta,\lambda)$, the function is evaluated at the solution, and when we simply write f, it is evaluated a the true value $(0,0)$. Also, $g_t=g_t(\theta;.)$, $G_t=G_t(\theta;.)$, $g_{0t}=g_t(0;.)$, and $G_{0t}=G_t(0;.)=G_t$ because the model is linear. The estimators (λ,θ) are then defined as the solution to :

$$\hat{m}(\lambda,\theta)\equiv\frac{1}{n}\sum_{t=1}^n m_t(\lambda,\theta)=0$$

The expansion is :

$$0=\hat{m}(\lambda,\theta) = \hat{m}+\hat{M}\begin{pmatrix}\lambda\\ \theta\end{pmatrix}+(1/2)\left[\hat{A}_\theta\theta+\hat{A}_\lambda\lambda\right]\begin{pmatrix}\lambda\\ \theta\end{pmatrix}$$
$$+ (1/6)\left[\hat{E}_{\lambda\lambda}\lambda^2+\hat{E}_{\theta\theta}\theta^2+2\hat{E}_{\theta\lambda}\theta\lambda\right]\begin{pmatrix}\lambda\\ \theta\end{pmatrix}+R_{n,\alpha}$$

$\hat{M}$, $\hat{A}_i$, and $\hat{E}_{ij}$, for $i,j=\theta,\lambda$, are operators from $L^2(\pi)\otimes\mathbb{R}$ to $L^2(\pi)\otimes\mathbb{R}$ defined as :

$$\hat{M}=\frac{1}{n}\sum_{t=1}^n\frac{\partial m_t}{\partial\beta},\quad \hat{A}_\theta=\frac{1}{n}\sum_{t=1}^n\frac{\partial\hat{M}_t}{\partial\theta},\quad \hat{A}_\lambda=\frac{1}{n}\sum_{t=1}^n\frac{\partial\hat{M}_t}{\partial\lambda},$$

$$\hat{E}_{ij}=\frac{1}{n}\sum_{t=1}^n\frac{\partial^2\hat{M}_t}{\partial i\partial j}\quad i,j=\theta,\lambda,$$

where $\beta=\{\lambda,\theta\}'$. If we define M, E, and A as the population values of $\hat{M}$, $\hat{E}$, and $\hat{A}$, we can rewrite the expansion as :

$$\beta = -M^{-1}\hat{m}-M^{-1}(\hat{M}-M)\beta-(1/2)M^{-1}\left[A_\theta\theta+A_\lambda\lambda\right]\beta$$

$$- \quad (1/2)M^{-1}\left[(\hat{A}_\theta - A_\theta)\theta + (\hat{A}_\lambda - A_\lambda)\lambda\right]\beta$$
$$- \quad (1/6)M^{-1}\left[E_{\lambda\lambda}\lambda^2 + E_{\theta\theta}\theta^2 + 2E_{\theta\lambda}\theta\lambda\right]\beta + R_{n,\alpha}$$

C.1.2 Derivation of each term

$$\hat{M} = \begin{pmatrix} \frac{\partial \hat{m}[1]}{\partial \lambda} & \frac{\partial \hat{m}[1]}{\partial \theta} \\ & \\ \frac{\partial \hat{m}[2]}{\partial \lambda} & \frac{\partial \hat{m}[2]}{\partial \theta} \end{pmatrix}$$

$$\begin{aligned}\hat{M}(\lambda,\theta)[1,1] &= \left[\frac{1}{n}\sum_{t=1}^{n}\rho'''(\lambda g_t)g_t g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho'(\lambda g_t)g_t\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho''(\lambda g_t)g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho''(\lambda g_t)g_t g_t\right] + \alpha I\end{aligned}$$

$$\begin{aligned}\hat{M}(\lambda,\theta)[1,2] &= \left[\frac{1}{n}\sum_{t=1}^{n}\left\{\rho'''(\lambda g_t)(\lambda G_t)g_t g_t + \rho''(\lambda g_t)G_t g_t + \rho''(\lambda g_t)g_t G_t\right\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho'(\lambda g_t)g_t\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho''(\lambda g_t)g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\left\{\rho''(\lambda g_t)(\lambda G_t)g_t + \rho'(\lambda g_t)G_t\right\}\right]\end{aligned}$$

$$\hat{M}(\lambda,\theta)[2,1] = \frac{1}{n}\sum_{t=1}^{n}\left\{\rho''(\lambda g_t)(\lambda G_t)g_t + \rho'(\lambda g_t)G_t\right\}$$

$$\hat{M}(\lambda,\theta)[2,2] = \frac{1}{n}\sum_{t=1}^{n}\rho''(\lambda g_t)(\lambda G_t)^2$$

The last term comes from the fact that $\partial G_t/\partial\theta = 0$. It implies that

$$\hat{M} = \begin{pmatrix} [-\rho_3\hat{S}\hat{g} + \hat{K}^2 + \alpha I] & [(\hat{V} + \hat{V}^*)\hat{g} + \hat{K}\hat{G}] \\ -\hat{G} & 0 \end{pmatrix},$$

which implies in our specific case :

$$\hat{M} = \begin{pmatrix} [-\rho_3\hat{S}\hat{g} + \hat{K}^2 + \alpha I] & [2\hat{V}\hat{g} + \hat{K}\hat{G}] \\ -\hat{G} & 0 \end{pmatrix}, \tag{C.5}$$

$$M = \begin{pmatrix} [K^2 + \alpha I] & KG \\ -G & 0 \end{pmatrix}, \tag{C.6}$$

and

$$M^{-1} = \begin{pmatrix} \left[I - K_\alpha^{-1} G \Omega_\alpha G\right] (K^2 + \alpha I)^{-1} & -K_\alpha^{-1} G \Omega_\alpha \\ \Omega_\alpha G (K^2 + \alpha I)^{-1} & \Omega_\alpha \end{pmatrix}, \tag{C.7}$$

where $K_\alpha^{-1} = (K^2 + \alpha I)^{-1} K$, and $\Omega_\alpha = (G K_\alpha^{-1} G)^{-1}$.

The second term of the expansion requires the matrices $\hat{A}_\lambda$ and $\hat{A}_\theta$. We derive them term by term.

$$\begin{aligned} \hat{A}_\lambda(\lambda, \theta)[1,1] &= \left[\frac{1}{n}\sum_{t=1}^{n} \rho^{iv}(\lambda g_t) g_t g_t g_t g_t\right] \left[\frac{1}{n}\sum_{t=1}^{n} \rho'(\lambda g_t) g_t\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n} \rho'''(\lambda g_t) g_t g_t g_t\right] \left[\frac{1}{n}\sum_{t=1}^{n} \rho''(\lambda g_t) g_t g_t\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n} \rho'''(\lambda g_t) g_t g_t g_t\right] \left[\frac{1}{n}\sum_{t=1}^{n} \rho''(\lambda g_t) g_t g_t\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n} \rho''(\lambda g_t) g_t g_t\right] \left[\frac{1}{n}\sum_{t=1}^{n} \rho'''(\lambda g_t) g_t g_t g_t\right] \end{aligned}$$

$$\begin{aligned} \hat{A}_\lambda(\lambda, \theta)[1,2] &= \left[\frac{1}{n}\sum_{t=1}^{n} \{\rho^{iv}(\lambda g_t)(\lambda G_t) g_t g_t g_t + \rho'''(\lambda g_t) g_t g_t G_t + \rho'''(\lambda g_t) G_t g_t g_t + \rho'''(\lambda g_t) g_t G_t g_t\}\right] \\ &\quad \left[\frac{1}{n}\sum_{t=1}^{n} \rho'(\lambda g_t)\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n} \{\rho'''(\lambda g_t)(\lambda G_t) g_t g_t + \rho''(\lambda g_t) G_t g_t + \rho''(\lambda g_t) g_t G_t\}\right] \left[\frac{1}{n}\sum_{t=1}^{n} \rho''(\lambda g_t) g_t g_t\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n} \rho'''(\lambda g_t) g_t g_t g_t\right] \left[\frac{1}{n}\sum_{t=1}^{n} \{\rho''(\lambda g_t)(\lambda G_t) g_t + \rho'(\lambda g_t) G_t\}\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n} \rho''(\lambda g_t) g_t g_t\right] \left[\frac{1}{n}\sum_{t=1}^{n} \{\rho'''(\lambda g_t)(\lambda G_t) g_t g_t + \rho''(\lambda g_t) g_t G_t + \rho''(\lambda g_t) G_t g_t\}\right] \end{aligned}$$

$$\hat{A}_\lambda(\lambda, \theta)[2,1] = \frac{1}{n}\sum_{t=1}^{n} \left\{\rho'''(\lambda g_t)(\lambda G_t) g_t g_t + \rho''(\lambda g_t) g_t G_t + \rho''(\lambda g_t) G_t g_t\right\}$$

$$\hat{A}_\lambda(\lambda,\theta)[2,2] = \frac{1}{n}\sum_{t=1}^{n}\left\{\rho'''(\lambda g_t)(\lambda G_t)^2 g_t + 2\rho''(\lambda g_t)(\lambda G_t)G_t\right\}$$

It follows that

$$\hat{A}_\lambda = \begin{pmatrix} [-\rho_4\hat{Q}\hat{g} - \rho_3\hat{S}\hat{K} - 2\rho_3\hat{K}\hat{S}] & [-\rho_3(\hat{\Gamma}+\hat{\Gamma}^*+\hat{\Gamma}^{**})\hat{g} + (\hat{V}+\hat{V}^*)\hat{K} + \hat{K}(\hat{V}+\hat{V}^*) + \rho_3\hat{S}\hat{G}] \\ -(\hat{V}+\hat{V}^*) & 0 \end{pmatrix},$$

which implies in our specified case :

$$\hat{A}_\lambda = \begin{pmatrix} [-\rho_4\hat{Q}\hat{g} - 3\rho_3\hat{S}\hat{K}] & [-3\rho_3\hat{\Gamma}\hat{g} + 4\hat{V}\hat{K} - \rho_3\hat{S}\hat{G}] \\ -2\hat{V} & 0 \end{pmatrix}, \tag{C.8}$$

and

$$A_\lambda = \begin{pmatrix} [-3\rho_3 SK] & [4VK - \rho_3 SG] \\ -2V & 0 \end{pmatrix}. \tag{C.9}$$

From here, $\rho_t = \rho(\lambda g_t)$.

$$\begin{aligned}
\hat{A}_\theta(\lambda,\theta)[1,1] &= \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)g_t g_t g_t + \rho_t''' G_t g_t g_t + \rho_t''' g_t G_t g_t + \rho_t''' g_t g_t G_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t' g_t\right] \\
&+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t''' g_t g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t''(\lambda G_t)g_t + \rho_t' G_t\}\right] \\
&+ \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)g_t g_t + \rho_t'' G_t g_t + \rho_t'' g_t G_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right] \\
&+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)g_t g_t + \rho_t'' G_t g_t + \rho_t'' g_t G_t\}\right]
\end{aligned}$$

$$\begin{aligned}
\hat{A}_\theta(\lambda,\theta)[1,2] &= \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^2 g_t g_t + 2\rho_t'''(\lambda G_t)G_t g_t + 2\rho_t'''(\lambda G_t)g_t G_t + 2\rho_t'' G_t G_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t' g_t\right] \\
&+ 2\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)g_t g_t + \rho_t'' G_t g_t + \rho_t'' g_t G_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t''(\lambda G_t)g_t + \rho_t' G_t\}\right] \\
&+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)^2 g_t + 2\rho_t''(\lambda G_t)G_t\}\right]
\end{aligned}$$

$$\hat{A}_\theta(\lambda,\theta)[2,1] = \frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)^2 g_t + 2\rho_t''(\lambda G_t)G_t\}$$

$$\hat{A}_\theta(\lambda,\theta)[2,2] = \frac{1}{n}\sum_{t=1}^{n} \rho_t'''(\lambda G_t)^3$$

It follows that

$$\hat{A}_\theta = \begin{pmatrix} [-\rho_3(\hat{\Gamma}+\hat{\Gamma}^*+\hat{\Gamma}^{**})\hat{g} - \rho_3\hat{S}\hat{G} + (\hat{V}+\hat{V}^*)\hat{K} + \hat{K}(\hat{V}+\hat{V}^*)] & [2\hat{B}\hat{g} + (\hat{V}+\hat{V}^*)\hat{G} + \hat{G}(\hat{V}+\hat{V}^*)] \\ 0 & 0 \end{pmatrix}$$

which implies in our specific case :

$$\hat{A}_\theta = \begin{pmatrix} [-3\rho_3\hat{\Gamma}\hat{g} - \rho_3\hat{S}\hat{G} + 4\hat{V}\hat{K}] & [2\hat{B}\hat{g} + 4\hat{V}\hat{G}] \\ 0 & 0 \end{pmatrix}, \tag{C.10}$$

and

$$A_\theta = \begin{pmatrix} [-\rho_3 SG + 4VK] & [4VG] \\ 0 & 0 \end{pmatrix}. \tag{C.11}$$

The third term in the expansion includes $\hat{E}_{\lambda\lambda}$, $\hat{E}_{\lambda\theta}$, and $\hat{E}_{\theta\theta}$.

$$\begin{aligned} \hat{E}_{\lambda\lambda}(\lambda,\theta)[1,1] &= \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{v} g_t g_t g_t g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t' g_t\right] \\ &+ 3\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv} g_t g_t g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right] \\ &+ 3\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t''' g_t g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t''' g_t g_t g_t\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv} g_t g_t g_t g_t\right] \end{aligned}$$

$$\begin{aligned} \hat{E}_{\lambda\lambda}(\lambda,\theta)[1,2] &= \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{v}(\lambda G_t) g_t g_t g_t g_t + \rho_t^{iv} g_t g_t g_t G_t + \rho_t^{iv} g_t g_t G_t g_t + \rho_t^{iv} g_t G_t g_t g_t + \rho_t^{iv} G_t g_t g_t g_t\}\right] \\ &\quad \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t' g_t\right] \\ &+ 2\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t) g_t g_t g_t + \rho_t''' g_t g_t G_t + \rho_t''' G_t g_t g_t + \rho_t''' g_t G_t g_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t) g_t g_t + \rho_t'' G_t g_t + \rho_t'' g_t G_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t''' g_t g_t g_t\right] \\ &+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t^{iv} g_t g_t g_t g_t\right]\left[\sum_{t=1}^{n}\{\rho_t''(\lambda G_t) g_t + \rho_t' G_t\}\right] \end{aligned}$$

$$+ \quad 2\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t''' g_t g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)g_t g_t + \rho_t'' g_t G_t + \rho_t'' G_t g_t\}\right]$$

$$+ \quad \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)g_t g_t g_t + \rho_t''' g_t g_t G_t + \rho_t''' G_t g_t g_t + \rho_t''' g_t G_t g_t\}\right]$$

$$\hat{E}_{\lambda\lambda}(\lambda,\theta)[2,1] = \frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)g_t g_t g_t + \rho_t''' g_t g_t G_t + \rho_t''' g_t G_t g_t + \rho_t''' G_t g_t g_t\}$$

$$\hat{E}_{\lambda\lambda}(\lambda,\theta)[2,2] = \frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^2 g_t g_t + 2\rho_t'''(\lambda G_t)g_t G_t + 2\rho_t'''(\lambda G_t)G_t g_t + 2\rho_t'' G_t G_t\}$$

Il follows that :

$$\hat{E}_{\lambda\lambda}[1,1] = -\rho_5\hat{F}\hat{g} - 3\rho_4\hat{Q}\hat{K} - \rho_4\hat{K}\hat{Q} + 3\rho_3^2\hat{S}^2]$$

$$\hat{E}_{\lambda\lambda}[1,2] = -\rho_4(\hat{L}+\hat{L}^*+\hat{L}^{**}+\hat{L}^{***})\hat{g} - 2\rho_3(\hat{\Gamma}+\hat{\Gamma}^*+\hat{\Gamma}^{**})\hat{K} - 3\rho_3\hat{S}(\hat{V}+\hat{V}^*) - \rho_4\hat{Q}\hat{G} - \rho_3\hat{K}(\hat{\Gamma}+\hat{\Gamma}^*+\hat{\Gamma}^{**})$$

$$\hat{E}_{\lambda\lambda}[2,1] = \rho_3(\hat{\Gamma}+\hat{\Gamma}^*+\hat{\Gamma}^{**})$$

$$\hat{E}_{\lambda\lambda}[2,2] = -2\hat{B}$$

which implies in our specific case :

$$\hat{E}_{\lambda\lambda} = \begin{pmatrix} [-\rho_5\hat{F}\hat{g} - 4\rho_4\hat{Q}\hat{K} + 3\rho_3^2\hat{S}^2] & [-4\rho_4\hat{L}\hat{g} - 9\rho_3\hat{\Gamma}\hat{K} - 6\rho_3\hat{S}\hat{V} - \rho_4\hat{Q}\hat{G}] \\ 3\rho_3\hat{\Gamma} & -2\hat{B} \end{pmatrix}, \tag{C.12}$$

and

$$E_{\lambda\lambda} = \begin{pmatrix} [-4\rho_4 QK + 3\rho_3^2 S^2] & [-9\rho_3\Gamma K - 6\rho_3 SV - \rho_4 QG] \\ 3\rho_3\Gamma & -2B \end{pmatrix}, \tag{C.13}$$

$$\hat{E}_{\theta\theta}(\lambda,\theta)[1,1] = \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{v}(\lambda G_t)^2 g_t g_t g_t + 2\rho_t^{iv}(\lambda G_t) g_t g_t G_t + 2\rho_t^{iv}(\lambda G_t) g_t G_t g_t + 2\rho_t^{iv}(\lambda G_t) G_t g_t g_t\right.$$

$$\left. + 2\rho_t''' G_t G_t g_t + 2\rho_t''' G_t g_t G_t + 2\rho_t''' g_t G_t G_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t' g_t\right]$$

$$+ 2\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t) g_t g_t g_t + \rho_t''' G_t g_t g_t + \rho_t''' g_t G_t g_t + \rho_t''' g_t g_t G_t\}\right]$$

$$\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t''(\lambda G_t) g_t + \rho_t' G_t\}\right]$$

$$+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t''' g_t g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)^2 g_t + 2\rho_t''(\lambda G_t) G_t\}\right]$$

$$+ \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^2 g_t g_t + 2\rho_t'''(\lambda G_t) G_t g_t + 2\rho_t'''(\lambda G_t) g_t G_t + 2\rho_t'' G_t G_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right]$$

$$+ 2\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t) g_t g_t + \rho_t'' G_t g_t + \rho_t'' g_t G_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t) g_t g_t + \rho_t'' G_t g_t + \rho_t'' g_t G_t\}\right]$$

$$+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^2 g_t g_t + 2\rho_t'''(\lambda G_t) G_t g_t + 2\rho_t'''(\lambda G_t) g_t G_t + 2\rho_t'' G_t G_t\}\right]$$

$$\hat{E}_{\theta\theta}(\lambda,\theta)[1,2] = \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{v}(\lambda G_t)^3 g_t g_t + 3\rho_t^{iv}(\lambda G_t)^2 G_t g_t + 3\rho_t^{iv}(\lambda G_t)^2 g_t G_t + \right.$$

$$\left. + 6\rho_t'''(\lambda G_t) G_t G_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t' g_t\right]$$

$$+ 3\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^2 g_t g_t + 2\rho_t'''(\lambda G_t) G_t g_t + 2\rho_t'''(\lambda G_t) g_t G_t + 2\rho_t'' G_t G_t\}\right]$$

$$\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t''(\lambda G_t) g_t + \rho_t' G_t\}\right]$$

$$+ 3\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t) g_t g_t + \rho_t'' G_t g_t + \rho_t'' g_t G_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)^2 g_t + 2\rho_t''(\lambda G_t) G_t\}\right]$$

$$+ \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^3 g_t + 3\rho_t'''(\lambda G_t)^2 G_t\}\right]$$

$$\hat{E}_{\theta\theta}(\lambda,\theta)[2,1] = \frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^3 g_t + 3\rho_t'''(\lambda G_t)^2 G_t\}$$

$$\hat{E}_{\theta\theta}(\lambda,\theta)[2,2] = \frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^4\}$$

Evaluating them at (0,0) yields :

$$\hat{E}_{\theta\theta}[1,1] = -2\rho_3(\hat{H}+\hat{H}^*+\hat{H}^{**})\hat{g} - 2\rho_3(\hat{\Gamma}+\hat{\Gamma}^*+\hat{\Gamma}^{**})\hat{G} + 2\hat{B}\hat{K} - 2(\hat{V}+\hat{V}^*)^2 + 2\hat{K}\hat{B},$$

$$\hat{E}_{\theta\theta}[1,2] = 6\hat{B}\hat{G},$$

and $\hat{E}_{\theta\theta}[2,1] = \hat{E}_{\theta\theta}[2,2] = 0$. It follows for our case :

$$\hat{E}_{\theta\theta} = \begin{pmatrix} [-6\rho_3(\hat{H}\hat{g}+\hat{\Gamma}\hat{G}) + 4\hat{B}\hat{K} + 8\hat{V}^2] & [6\hat{B}\hat{G}] \\ 0 & 0 \end{pmatrix}, \tag{C.14}$$

and

$$E_{\theta\theta} = \begin{pmatrix} [-6\rho_3\Gamma G + 4BK + 8V^2] & [6BG] \\ 0 & 0 \end{pmatrix}. \tag{C.15}$$

$$\begin{aligned}
\hat{E}_{\theta\lambda}(\lambda,\theta)[1,1] = & \Bigg[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{v}(\lambda G_t)g_tg_tg_tg_t + \rho_t^{iv}g_tg_tg_tG_t + \rho_t^{iv}G_tg_tg_tg_t + \\
& + \rho_t^{iv}g_tG_tg_tg_t + \rho_t^{iv}g_tg_tG_tg_t\}\Bigg]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'g_t\right] \\
& + 2\Bigg[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)g_tg_tg_t + \rho_t'''G_tg_tg_t + \rho_t'''g_tG_tg_t \\
& + \rho_t'''g_tg_tG_t\}\Bigg]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t''g_tg_t\right] \\
& + \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t^{iv}g_tg_tg_tg_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t''(\lambda G_t)g_t + \rho_t'G_t\}\right] \\
& + \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'''g_tg_tg_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)g_tg_t + \rho_t''g_tG_t + \rho_t''G_tg_t\}\right] \\
& + \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)g_tg_t + \rho_t''G_tg_t + \rho_t''g_tG_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'''g_tg_tg_t\right] \\
& + \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'''g_tg_tg_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)g_tg_t + \rho_t''G_tg_t + \rho_t''g_tG_t\}\right] \\
& + \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t''g_tg_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)g_tg_tg_t + \rho_t'''G_tg_tg_t + \rho_t'''g_tG_tg_t + \rho_t'''g_tg_tG_t\}\right]
\end{aligned}$$

$$\hat{E}_{\theta\lambda}(\lambda,\theta)[2,1] = \frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^2g_tg_t + 2\rho_t'''(\lambda G_t)g_tG_t + 2\rho_t'''(\lambda G_t)G_tg_t + 2\rho_t''G_tG_t\}$$

$$\hat{E}_{\theta\lambda}(\lambda,\theta)[2,2] = \frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^3 g_t + 3\rho_t'''(\lambda G_t)^2 G_t\}$$

$$\begin{aligned}
\hat{E}_{\theta\lambda}(\lambda,\theta)[1,2] = & \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{v}(\lambda G_t)^2 g_t g_t g_t + 2\rho_t^{iv}(\lambda G_t) g_t g_t G_t + 2\rho_t^{iv}(\lambda G_t) G_t g_t g_t + \right.\\
& \left. + 2\rho_t^{iv}(\lambda G_t) g_t G_t g_t + 2\rho_t''' G_t g_t G_t + 2\rho_t''' g_t G_t G_t + 2\rho_t''' G_t G_t g_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t' g_t\right]\\
& + \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^2 g_t g_t + 2\rho_t'''(\lambda G_t) g_t G_t + 2\rho_t'''(\lambda G_t) G_t g_t + 2\rho_t'' G_t G_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right]\\
& + 2\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t) g_t g_t g_t + \rho_t''' g_t g_t G_t + \rho_t''' G_t g_t g_t + \rho_t''' g_t G_t g_t\}\right]\\
& \left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t''(\lambda G_t) g_t + \rho_t' G_t\}\right]\\
& + 2\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t) g_t g_t + \rho_t'' g_t G_t + \rho_t'' G_t g_t\}\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t) g_t g_t + \rho_t'' g_t G_t + \rho_t'' G_t g_t\}\right]\\
& + \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t''' g_t g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t'''(\lambda G_t)^2 g_t + 2\rho_t''(\lambda G_t) G_t\}\right]\\
& + \left[\frac{1}{n}\sum_{t=1}^{n}\rho_t'' g_t g_t\right]\left[\frac{1}{n}\sum_{t=1}^{n}\{\rho_t^{iv}(\lambda G_t)^2 g_t g_t + 2\rho_t'''(\lambda G_t) g_t G_t + 2\rho_t'''(\lambda G_t) G_t g_t + 2\rho_t'' G_t G_t\}\right]
\end{aligned}$$

We evaluate them at $(0,0)$:

$$\begin{aligned}
\hat{E}_{\theta\lambda}[1,1] = & -\rho_4(\hat{L} + \hat{L}^* + \hat{L}^{**} + \hat{L}^{***})\hat{g} - 2\rho_3(\hat{\Gamma} + \hat{\Gamma}^* + \hat{\Gamma}^{**})\hat{K} - \rho_4\hat{Q}\hat{G}\\
& -\rho_3(\hat{V} + \hat{V}^*)\hat{S} - 2\rho_3\hat{S}(\hat{V} + \hat{V}^*) - \rho_3\hat{K}(\hat{\Gamma} + \hat{\Gamma}^* + \hat{\Gamma}^{**})
\end{aligned},$$

$$\hat{E}_{\theta\lambda}[1,2] = -2\rho_3(\hat{H} + \hat{H}^* + \hat{H}^{**})\hat{g} - 2\rho_3(\hat{\Gamma} + \hat{\Gamma}^* + \hat{\Gamma}^{**})\hat{G} + 2(\hat{V} + \hat{V})^2 + 2\hat{B}\hat{K} + 2\hat{K}\hat{B},$$

and $\hat{E}_{\theta\lambda}[2,1] = -2\hat{B}$, $\hat{E}_{\theta\lambda}[2,2] = 0$. It follows for our specific case :

$$\hat{E}_{\theta\lambda} = \begin{pmatrix} [-4\rho_4\hat{L}\hat{g} - 9\rho_3\hat{\Gamma}\hat{K} - 6\rho_3\hat{S}\hat{V} - \rho_4\hat{Q}\hat{G}] & [-6\rho_3\hat{H}\hat{g} + 4\hat{B}\hat{K} - 6\rho_3\hat{\Gamma}\hat{G} + 8\hat{V}^2] \\ -2\hat{B} & 0 \end{pmatrix}, \tag{C.16}$$

and

$$E_{\theta\lambda} = \begin{pmatrix} [-9\rho_3\Gamma K - 6\rho_3 SV - \rho_4 QG] & [4BK - 6\rho_3\Gamma G + 8V^2] \\ -2B & 0 \end{pmatrix}, \tag{C.17}$$

C.1.3 Derivation of the MSE

The first term of the expansion is :

$$T_1 = \begin{pmatrix} T_1^\lambda \\ T_1^\theta \end{pmatrix} = -M^{-1}\hat{m} = \begin{pmatrix} -[I - K_\alpha^{-1}G\Omega_\alpha G]\tilde{K}_\alpha^{-1}\hat{g} \\ -\Omega_\alpha G\tilde{K}_\alpha^{-1}\hat{g} \end{pmatrix} = \begin{pmatrix} O_p\left(\frac{1}{\sqrt{\alpha n}}\right) \\ O_p\left(\frac{1}{\sqrt{n}}\right) \end{pmatrix},$$

where $\tilde{K}_\alpha^{-1} = (K^2+\alpha I)^{-1}\hat{K}$. The rate of convergence of λ depends on α through $\tilde{K}_\alpha^{-1}$, which is $O(1/\sqrt{\alpha})$ according to (Kress, 1999). Indeed, using the inequality $\sqrt{ab} \leq (a+b)/2$ and the fact that the eigenvalues of K are finite, then $\|K_\alpha^{-1}\| \leq \sup_i \lambda_i/(\lambda_i^2+\alpha) \leq C/\sqrt{\alpha}$, for some finite constant C. Notice that the result does not contradict Lemma 2 of chapter 1 in which $\lambda = O_p(n^{-1/2})$. In fact $\sup_i \lambda_i/(\lambda_i^2+\alpha)$ converges to $\sup_i 1/\lambda_i$ when α goes to zero. Since K is a strictly positive definite operator, $\lambda_i > 0\ \forall i$. The result holds if the speed of convergence of α is such that $\sup_i \hat{\lambda}_i/(\hat{\lambda}_i^2+\alpha) < \infty$ as n goes to infinity.

The second term is $-M^{-1}(\hat{M}-M)\beta \equiv T_2 = \{T_2^\lambda, T_2^\theta\}'$. Let $\Sigma = (I - K_\alpha^{-1}G\Omega_\alpha G)$ which is $O(1)$, then

$$\begin{aligned} T_2^\lambda = &- \{\Sigma(K^2+\alpha I)^{-1}[-\rho_3\hat{S}\hat{g} + (\hat{K}^2 - K^2)] + K_\alpha^{-1}G\Omega_\alpha(G-\hat{G})\}\lambda \\ &- \{\Sigma(K^2+\alpha I)^{-1}[2\hat{V}\hat{g} + (\hat{K}\hat{G} - KG)]\}\theta, \end{aligned}$$

and

$$\begin{aligned} T_2^\theta = &- \{\Omega_\alpha G(K^2+\alpha I)^{-1}[-\rho_3\hat{S}\hat{g} + (\hat{K}^2 - K^2)] + \Omega_\alpha(\hat{G}-G)\}\lambda \\ &- \{\Omega_\alpha G(K^2+\alpha I)^{-1}[2\hat{V}\hat{g} + (\hat{K}\hat{G} - KG)]\}\theta. \end{aligned}$$

We now have to analyze each term in order to determine their order. If we start with T_2^λ, its terms are :

$$(1):\ \rho_3\Sigma(K^2+\alpha I)^{-1}\hat{S}\hat{g}\lambda = O_p(1/\alpha)O_p(1/\sqrt{n})O_p(1/(\sqrt{\alpha n})) = O_p(1/(\alpha^{3/2}n)),$$

$$(2):\ -\Sigma(K^2+\alpha I)^{-1}(\hat{K}^2-K^2)\lambda = O_p(1/\alpha)O_p(1/\sqrt{n})O_p(1/(\sqrt{\alpha n})) = O_p(1/(\alpha^{3/2}n)),$$

$$(3):\ -K_\alpha^{-1}G\Omega_\alpha(G-\hat{G})\lambda = O_p(1/\sqrt{\alpha})O_p(\sqrt{\alpha})O_p(1/\sqrt{n})O_p(1/(\sqrt{\alpha n})) = O_p(1/(n\sqrt{\alpha})),$$

$$(4):\ -2\Sigma(K^2+\alpha I)^{-1}\hat{V}\hat{g}\theta = O_p(1/\alpha)O_p(1/\sqrt{n})O_p(1/\sqrt{n}) = O_p(1/(\alpha n)),$$

and

$$(5):\ -\Sigma(K^2+\alpha I)^{-1}(\hat{K}\hat{G}-KG)\theta = O_p(1/(\alpha n)),$$

while for θ, the terms are :

$$(1):\ \rho_3\Omega_\alpha G(K^2+\alpha I)^{-1}\hat{S}\hat{g}\lambda = O_p(1/(\alpha n)),$$

$$(2):\ -\Omega_\alpha G(K^2+\alpha I)^{-1}(\hat{K}^2-K^2)\lambda = O_p(1/(\alpha n)),$$

$$(3):\ -\Omega_\alpha(\hat{G}-G)\lambda = O_p(1/n),$$

$$(4):\ -2\Omega_\alpha G(K^2+\alpha I)^{-1}\hat{V}\hat{g}\theta = O_p(1/(n\sqrt{\alpha})),$$

$$(5):\ -\Omega_\alpha G(K^2+\alpha I)^{-1}(\hat{K}\hat{G}-KG)\theta = O_p(1/(n\sqrt{\alpha}))$$

The third term is $-(1/2)M^{-1}(A_\theta\theta + A_\lambda\lambda)\beta$ and can be written as follows

$$\begin{aligned}\begin{pmatrix} T_3^\lambda \\ T_3^\theta \end{pmatrix} &= -(1/2)M^{-1}\begin{pmatrix} A_\theta[1,1]\theta\lambda + A_\theta[1,2]\theta^2 + A_\lambda[1,1]\lambda^2 + A_\lambda[1,2]\theta\lambda \\ A_\theta[2,1]\theta\lambda + A_\theta[2,2]\theta^2 + A_\lambda[2,1]\lambda^2 + A_\lambda[2,2]\theta\lambda \end{pmatrix}, \\ &= -(1/2)M^{-1}\begin{pmatrix} A_\theta[1,1]\theta\lambda + A_\theta[1,2]\theta^2 + A_\lambda[1,1]\lambda^2 + A_\lambda[1,2]\theta\lambda \\ A_\lambda[2,1]\lambda^2 \end{pmatrix}\end{aligned}$$

Notice that $V(\tau_1,\tau_2) = E[G_t(\tau_1)\overline{g_t(\tau_2)}] = -E[W_t\epsilon_t e^{ix_t(\tau_1-\tau_2)}] = -\sigma_{\epsilon u}K/\sigma_\epsilon^2$. It implies that $\|(K^2+\alpha I)^{-1}V\| = |\sigma_{\epsilon u}|\|K_\alpha^{-1}\|/\sigma_\epsilon^2 = O(1/\sqrt{\alpha})$. Similarly, $\|(K^2+\alpha I)^{-1}VK\| = |\sigma_{\epsilon u}|\|(K^2+\alpha I)^{-1}K^2\|/\sigma_\epsilon^2 \leq C\sup_i \lambda_i^2/(\lambda_i^2+\alpha) = 1$. Therefore,

$$\begin{aligned}T_3^\lambda =& -(1/2)\Sigma(K^2+\alpha I)^{-1}[4VG\theta^2 + 2(4VK-\rho_3 SG)\theta\lambda - 3\rho_3 SK\lambda^2] \\ &- K_\alpha^{-1}G\Omega_\alpha V\lambda^2 \\ =& O_p(1/(n\sqrt{\alpha})) + O_p(1/(n\sqrt{\alpha})) + O_p(1/(\alpha^{3/2}n)) + O_p(1/(\alpha^2 n))\end{aligned}$$

$$\begin{aligned}T_3^\theta =& (-1/2)\Omega_\alpha G(K^2+\alpha I)^{-1}[4VG\theta^2 + 2(4VK-\rho_3 SG)\theta\lambda - 3\rho_3 SK\lambda^2] \\ &+ \Omega_\alpha V\lambda^2 \\ =& O_p(1/n) + O_p(1/n) + O_p(1/(\alpha n)) + O_p(1/(\alpha^{3/2}n)) + O_p(1/(n\sqrt{\alpha}))\end{aligned}$$

The fourth term is $(T_4^\lambda T_4^\theta)' = -(1/2)M^{-1}[(\hat{A}_\theta - A_\theta)\theta + (\hat{A}_\lambda - A_\lambda)\lambda]\beta$, where :

$$(\hat{A}_\theta - A_\theta) = \begin{pmatrix} -3\rho_3\hat{\Gamma}\hat{g} + \rho_3(SG-\hat{S}\hat{G}) + 4(\hat{V}\hat{K}-VK) & 2\hat{B}\hat{g} + 4(\hat{V}\hat{G}-VG) \\ 0 & 0 \end{pmatrix}$$

and

$$(\hat{A}_\lambda - A_\lambda) = \begin{pmatrix} -\rho_4\hat{Q}\hat{g} + 3\rho_3(SK - \hat{S}\hat{K}) & -3\rho_3\hat{\Gamma}\hat{g} + 4(\hat{V}\hat{K} - VK) + \rho_3(SG - \hat{s}\hat{G} \\ 2(V - \hat{V}) & 0 \end{pmatrix}$$

The terms T_4^λ and T_4^θ can be written as :

$$\begin{aligned}
-2T_4^\lambda =& \Big[\Sigma(K^2+\alpha I)^{-1}\Big\{-3\rho_3\hat{\Gamma}\hat{g} + \rho_3(SG - \hat{S}\hat{G}) + 4(\hat{V}\hat{K} - VK)\Big\}\theta \\
&+ \Big\{\Sigma(K^2+\alpha I)^{-1}[-\rho_4\hat{Q}\hat{g} + 3\rho_3(SK - \hat{S}\hat{K})] + 2K_\alpha^{-1}G\Omega_\alpha(V - \hat{V})\Big\}\lambda\Big]\lambda \\
&+ \Big[\Sigma(K^2+\alpha I)^{-1}\Big\{2\hat{B}\hat{g} + 4(\hat{V}\hat{G} - VG)\Big\}\theta \\
&+ \Sigma(K^2+\alpha I)^{-1}\Big\{-3\rho_3\hat{\Gamma}\hat{g} + \rho_3(SG - \hat{S}\hat{G}) + 4(\hat{V}\hat{K} - VK)\Big\}\lambda\Big]\theta \\
=& \Big[\Sigma(K^2+\alpha I)^{-1}[-\rho_4\hat{Q}\hat{g} + 3\rho_3(SK - \hat{S}\hat{K})] + 2K_\alpha^{-1}G\Omega_\alpha(V - \hat{V})\Big]\lambda^2 \\
&+ 2\Big[\Sigma(K^2+\alpha I)^{-1}\Big\{-3\rho_3\hat{\Gamma}\hat{g} + \rho_3(SG - \hat{S}\hat{G}) + 4(\hat{V}\hat{K} - VK)\Big\}\Big]\theta\lambda \\
&+ \Big[\Sigma(K^2+\alpha I)^{-1}\Big\{2\hat{B}\hat{g} + 4(\hat{V}\hat{G} - VG)\Big\}\Big]\theta^2 \\
=& \left[O_p\left(\frac{1}{\alpha\sqrt{n}}\right) + O_p\left(\frac{1}{\alpha\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\right]O_p\left(\frac{1}{\alpha n}\right) \\
&+ \left[O_p\left(\frac{1}{\alpha\sqrt{n}}\right) + O_p\left(\frac{1}{\alpha\sqrt{n}}\right) + O_p\left(\frac{1}{\alpha\sqrt{n}}\right)\right]O_p\left(\frac{1}{n\sqrt{\alpha}}\right) \\
&+ \left[O_p\left(\frac{1}{\alpha\sqrt{n}}\right) + O_p\left(\frac{1}{\alpha\sqrt{n}}\right)\right]O_p\left(\frac{1}{n}\right) \\
=& O_p\left(\frac{1}{\alpha^2 n^{3/2}}\right) + O_p\left(\frac{1}{(\alpha n)^{3/2}}\right) + O_p\left(\frac{1}{\alpha n^{3/2}}\right)
\end{aligned}$$

$$\begin{aligned}
-2T_4^\theta =& \Big[\Omega_\alpha G(K^2+\alpha I)^{-1}\Big\{-3\rho_3\hat{\Gamma}\hat{g} + \rho_3(SG - \hat{S}\hat{G}) + 4(\hat{V}\hat{K} - VK)\Big\}\theta \\
&+ \Big\{\Omega_\alpha G(K^2+\alpha I)^{-1}[-\rho_4\hat{Q}\hat{g} + 3\rho_3(SK - \hat{S}\hat{K})] - 2\Omega_\alpha(V - \hat{V})\Big\}\lambda\Big]\lambda \\
&+ \Big[\Omega_\alpha G(K^2+\alpha I)^{-1}\Big\{2\hat{B}\hat{g} + 4(\hat{V}\hat{G} - VG)\Big\}\theta \\
&+ \Omega_\alpha G(K^2+\alpha I)^{-1}\Big\{-3\rho_3\hat{\Gamma}\hat{g} + \rho_3(SG - \hat{S}\hat{G}) + 4(\hat{V}\hat{K} - VK)\Big\}\lambda\Big]\theta \\
=& \Big[\Omega_\alpha G(K^2+\alpha I)^{-1}[-\rho_4\hat{Q}\hat{g} + 3\rho_3(SK - \hat{S}\hat{K})] - 2\Omega_\alpha(V - \hat{V})\Big]\lambda^2 \\
&+ 2\Big[\Omega_\alpha G(K^2+\alpha I)^{-1}\Big\{-3\rho_3\hat{\Gamma}\hat{g} + \rho_3(SG - \hat{S}\hat{G}) + 4(\hat{V}\hat{K} - VK)\Big\}\Big]\theta\lambda \\
&+ \Big[\Omega_\alpha G(K^2+\alpha I)^{-1}\Big\{2\hat{B}\hat{g} + 4(\hat{V}\hat{G} - VG)\Big\}\Big]\theta^2 \\
=& \left[O_p\left(\frac{1}{\sqrt{\alpha n}}\right) + O_p\left(\frac{1}{\sqrt{\alpha n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\right]O_p\left(\frac{1}{\alpha n}\right)
\end{aligned}$$

$$
\begin{aligned}
&+\left[O_p\left(\frac{1}{\sqrt{\alpha n}}\right)+O_p\left(\frac{1}{\sqrt{\alpha n}}\right)+O_p\left(\frac{1}{\sqrt{\alpha n}}\right)\right]O_p\left(\frac{1}{n\sqrt{\alpha}}\right)\\
&+\left[O_p\left(\frac{1}{\sqrt{\alpha n}}\right)+O_p\left(\frac{1}{\sqrt{\alpha n}}\right)\right]O_p\left(\frac{1}{n}\right)\\
=&O_p\left(\frac{1}{(\alpha n)^{3/2}}\right)+O_p\left(\frac{1}{\alpha n^{3/2}}\right)+O_p\left(\frac{1}{\alpha^{1/2}n^{3/2}}\right)
\end{aligned}
$$

The fifth term is $\{T_5^{\lambda}T_5^{\theta}\}'=-(1/6)M^{-1}[E_{\lambda\lambda}\lambda^2+E_{\theta\theta}\theta^2+2E_{\theta\lambda}\theta\lambda]\beta$. The terms T_5^{λ} and T_5^{θ} are :

$$
\begin{aligned}
-6T_5^{\lambda}=&\left[\left\{\Sigma(K^2+\alpha I)^{-1}\left(-4\rho_4QK+3\rho_3^2S^2\right)+3\rho_3K_{\alpha}^{-1}G\Omega_{\alpha}\Gamma\right\}\lambda^2\right.\\
&+\left\{\Sigma(K^2+\alpha I)^{-1}(-6\rho_3\Gamma G+4BK+8V^2)\right\}\theta^2\\
&\left.+2\left\{\Sigma(K^2+\alpha I)^{-1}\left(-9\rho_3\Gamma K-6\rho_3SV-\rho_4QG\right)-2K_{\alpha}^{-1}G\Omega_{\alpha}B\right\}\theta\lambda\right]\lambda\\
&+\left[\left\{\Sigma(K^2+\alpha I)^{-1}\left(-9\rho_3\Gamma K-6\rho_3SV-\rho_4QG\right)-2K_{\alpha}^{-1}G\Omega_{\alpha}B\right\}\lambda^2\right.\\
&\left.+\left\{6\Sigma(K^2+\alpha I)^{-1}BG\right\}\theta^2+2\left\{\Sigma(K^2+\alpha I)^{-1}\left(4BK-6\rho_3\Gamma G+8V^2\right)\right\}\theta\lambda\right]\theta\\
=&\left[\Sigma(K^2+\alpha I)^{-1}\left(-4\rho_4QK+3\rho_3^2S^2\right)+3\rho_3K_{\alpha}^{-1}G\Omega_{\alpha}\Gamma\right]\lambda^3\\
&+3\left[\Sigma(K^2+\alpha I)^{-1}(-6\rho_3\Gamma G+4BK+8V^2)\right]\theta^2\lambda\\
&+3\left[\Sigma(K^2+\alpha I)^{-1}\left(-9\rho_3\Gamma K-6\rho_3SV-\rho_4QG\right)-2K_{\alpha}^{-1}G\Omega_{\alpha}B\right]\theta\lambda^2\\
&+\left[6\Sigma(K^2+\alpha I)^{-1}BG\right]\theta^3\\
=&\left[O_p\left(\frac{1}{\alpha}\right)+O_p(1)\right]O_p\left(\frac{1}{\alpha^{3/2}n^{3/2}}\right)+\left[O_p\left(\frac{1}{\alpha}\right)\right]O_p\left(\frac{1}{\alpha^{1/2}n^{3/2}}\right)\\
&+\left[O_p\left(\frac{1}{\alpha}\right)+O_p(1)\right]O_p\left(\frac{1}{\alpha n^{3/2}}\right)+\left[O_p\left(\frac{1}{\alpha}\right)\right]O_p\left(\frac{1}{n^{3/2}}\right)\\
=&O_p\left(\frac{1}{\alpha^{5/2}n^{3/2}}\right)+O_p\left(\frac{1}{\alpha^{3/2}n^{3/2}}\right)+O_p\left(\frac{1}{\alpha n^{3/2}}\right)+O_p\left(\frac{1}{\alpha^2n^{3/2}}\right)
\end{aligned}
$$

$$
\begin{aligned}
-6T_5^{\theta}=&\left[\left\{\Omega_{\alpha}G(K^2+\alpha I)^{-1}\left(-4\rho_4QK+3\rho_3^2S^2\right)-3\rho_3\Omega_{\alpha}\Gamma\right\}\lambda^2\right.\\
&+\left\{\Omega_{\alpha}G(K^2+\alpha I)^{-1}(-6\rho_3\Gamma G+4BK+8V^2)\right\}\theta^2\\
&\left.+2\left\{\Omega_{\alpha}G(K^2+\alpha I)^{-1}\left(-9\rho_3\Gamma K-6\rho_3SV-\rho_4QG\right)+2\Omega_{\alpha}B\right\}\theta\lambda\right]\lambda\\
&+\left[\left\{\Omega_{\alpha}G(K^2+\alpha I)^{-1}\left(-9\rho_3\Gamma K-6\rho_3SV-\rho_4QG\right)+2\Omega_{\alpha}B\right\}\lambda^2\right.\\
&\left.+\left\{6\Omega_{\alpha}G(K^2+\alpha I)^{-1}BG\right\}\theta^2+2\left\{\Omega_{\alpha}G(K^2+\alpha I)^{-1}\left(4BK-6\rho_3\Gamma G+8V^2\right)\right\}\theta\lambda\right]\theta\\
=&\left[\Omega_{\alpha}G(K^2+\alpha I)^{-1}\left(-4\rho_4QK+3\rho_3^2S^2\right)-3\rho_3\Omega_{\alpha}\Gamma\right]\lambda^3
\end{aligned}
$$

$$+ 3\left[\Omega_\alpha G(K^2+\alpha I)^{-1}(-6\rho_3\Gamma G + 4BK + 8V^2)\right]\theta^2\lambda$$
$$+ 3\left[\Omega_\alpha G(K^2+\alpha I)^{-1}(-9\rho_3\Gamma K - 6\rho_3 SV - \rho_4 QG) + 2\Omega_\alpha B\right]\theta\lambda^2$$
$$+ \left[6\Omega_\alpha G(K^2+\alpha I)^{-1}BG\right]\theta^3$$
$$= \left[O_p\left(\frac{1}{\sqrt{\alpha}}\right) + O_p\left(\sqrt{\alpha}\right)\right] O_p\left(\frac{1}{\alpha^{3/2}n^{3/2}}\right) + \left[O_p\left(\frac{1}{\sqrt{\alpha}}\right)\right] O_p\left(\frac{1}{\alpha^{1/2}n^{3/2}}\right)$$
$$+ \left[O_p\left(\frac{1}{\sqrt{\alpha}}\right) + O_p\left(\alpha\right)\right] O_p\left(\frac{1}{\alpha n^{3/2}}\right) + \left[O_p\left(\frac{1}{\sqrt{\alpha}}\right)\right] O_p\left(\frac{1}{n^{3/2}}\right)$$
$$= O_p\left(\frac{1}{\alpha^2 n^{3/2}}\right) + O_p\left(\frac{1}{\alpha n^{3/2}}\right) + O_p\left(\frac{1}{\alpha^{3/2}n^{3/2}}\right) + O_p\left(\frac{1}{\alpha^{1/2}n^{3/2}}\right)$$

Therefore, the result implies that :

$$\begin{aligned}\theta = &- \Omega_\alpha G(K^2+\alpha I)^{-1}\hat{K}\hat{g} \\ &+ \rho_3\Omega_\alpha G(K^2+\alpha I)^{-1}\hat{S}\hat{g}\lambda \\ &- \Omega_\alpha G(K^2+\alpha I)^{-1}(K^2-\hat{K}^2)\lambda \\ &- 4\Omega_\alpha G(K^2+\alpha I)^{-1}VK\theta\lambda \\ &+ \rho_3\Omega_\alpha G(K^2+\alpha I)^{-1}SG\theta\lambda \\ &+ \frac{3\rho_3}{2}\Omega_\alpha G(K^2+\alpha I)^{-1}SK\lambda^2 \\ &+ o_p\left(\frac{1}{\alpha n}\right) \\ \equiv &\sum_{i=1}^{6}\mathcal{T}_i + o_p\left(\frac{1}{\alpha n}\right),\end{aligned} \tag{C.18}$$

where λ and θ can be replaced by their respective first term. Indeed, adding more terms would only result in more $o_p(1/(\alpha n))$ elements.

C.1.4 The approximated MSE of θ

We want to derive :

$$MSE(\alpha) = E(n\theta^2) \approx E\left[\sum_{i=1}^{6}\sqrt{n}\mathcal{T}_i\right]^2 = \sum_{i=1}^{6}E(n\mathcal{T}_i^2) + 2\sum_{j<i}E(n\mathcal{T}_i\mathcal{T}_j) \tag{C.19}$$

For deriving the above expression, we'll be using the following notation when the operators are defined by kernels with dimension higher than 2. For example, let Q be the operator from $L^2(\pi)$ to $L^2(\pi)\times L^2(\pi)\times L^2(\pi)$ defined by the four dimensional kernel $q(\tau_1,\tau_2,\tau_3,\tau_4) = E(g_t(\tau_1)\ \overline{g_t(\tau_2)}\ \overline{g_t(\tau_3)}\ \overline{g_t(\tau_4)})$. Then

$$\int\int\int\int\int\int\int Q(\tau_1,\tau_2,\tau_3,\tau_4)k(\tau_4,\tau_5)\hat{g}(\tau_5)G(\tau_3)v(\tau_2,\tau_6)\hat{g}(\tau_6)\Sigma(\tau_1,\tau_7)\hat{g}(\tau_7)d\pi(\tau_1)\cdots d\pi(\tau_7)$$

will be written as $Q_{1234}K_{45}V_{26}\Sigma_{17}\hat{g}_5\hat{g}_6\hat{g}_7G_3$. Therefore, $\hat{S}\hat{g}\lambda$, in the second term $\mathcal{T}_2$ will be written as :

$$\begin{aligned}\hat{S}\hat{g}\lambda =& \int\int \hat{s}(\tau_1,\tau_2,\tau_3)\hat{g}(\tau_3)\lambda(\tau_2)d\pi(\tau_2)d\pi(\tau_3)\\ =&\hat{S}_{123}\hat{g}_3\lambda_2\end{aligned}$$

Each term has a the common factor $\Omega_\alpha G(K^2+\alpha I)^{-1}$. We therefore write them as :

$$\mathcal{T}_i = \Omega_\alpha G(K^2+\alpha I)^{-1}\Upsilon_i = \Omega_\alpha \sum_{j=1}^{\infty}\frac{1}{\mu_j^2+\alpha} < \phi_j, G >< \phi_j, \Upsilon_i >,$$

where μ_j and ϕ_j are the eigenvalues and eigenfunctions of K. Let the covariance operator Ξ_{ij} be defined by the kernel :

$$\Xi_{ij}(\tau_1,\tau_2) = nE[\Upsilon_i(\tau_1)\overline{\Upsilon_j(\tau_2)}],$$

then

$$\begin{aligned}E(n\mathcal{T}_i\mathcal{T}_j) =&\Omega_\alpha^2\overline{G}(K^2+\alpha I)^{-1}\Xi_{ij}(K^2+\alpha I)^{-1}G\\ =&\Omega_\alpha^2\int\int\left[\overline{G}(K^2+\alpha I)^{-1}\right](\tau_1)\left[(K^2+\alpha I)^{-1}G\right](\tau_2)\Xi_{ij}(\tau_1,\tau_2)d\pi(\tau_1)d\pi(\tau_2)\end{aligned}$$

for which the exact expression depends on the operator Ξ_{ij} and will be analyzed further in Section C.1.5. We consider now each term $E(n\mathcal{T}_i\mathcal{T}_j)$.

– $E(n\mathcal{T}_1^2)$

For the first term, we can derive the operator directly :

$$\begin{aligned}\Xi_{11} =&nE[K\hat{g}\overline{K\hat{g}}]\\ =&nKE(\hat{g}\overline{\hat{g}})K\\ =&nK\left(\frac{K}{n}\right)K\\ =&K^3 = O(1)\end{aligned}$$

Notice that $\hat{K}$ was replaced by K because

$$\begin{aligned}-\Omega_\alpha G(K^2+\alpha I)^{-1}\hat{K}\hat{g} =& -\Omega_\alpha G(K^2+\alpha I)^{-1}K\hat{g} + -\Omega_\alpha G(K^2+\alpha I)^{-1}(\hat{K}-K)\hat{g}\\ =& -\Omega_\alpha G(K^2+\alpha I)^{-1}K\hat{g} + O_p\left(\frac{1}{n\sqrt{\alpha}}\right)\\ =& -\Omega_\alpha G(K^2+\alpha I)^{-1}K\hat{g} + o_p\left(\frac{1}{n\alpha}\right).\end{aligned}$$

It follows that :

$$\begin{aligned}E(n\mathcal{T}_1^2) =&\Omega_\alpha^2 G(K^2+\alpha I)^{-1}K^3(K^2+\alpha I)^{-1}\overline{G}\\ =&\Omega_\alpha^2 GK_\alpha^{-1}KK_\alpha^{-1}\overline{G}\\ =&\Omega_\alpha^2\sum_{j=1}^{\infty}\frac{\mu_j^3}{(\mu_j^2+\alpha)^2} < \phi_j, G >^2\end{aligned} \tag{C.20}$$

Since

$$\Omega_\alpha^2 = \frac{1}{\left(\sum_{j=1}^{\infty}\frac{\mu_j}{(\mu_j^2+\alpha)^2} < \phi_j, G >^2\right)^2},$$

we can show that the derivative of $E(n\mathcal{T}_1^2)$ with respect to α is negative, the minimum being reached at $\alpha = 0$ in which case $E(n\mathcal{T}_1^2) = (GK^{-1}G)^{-1}$.

– $E(n\mathcal{T}_2^2)$

In the following, we redefine Σ as $(I - K_\alpha^{-1}G\Omega_\alpha\overline{G})K_\alpha^{-1}$. The first term of λ is therefore $T_1^\lambda = -\Sigma\hat{g} + (I - K_\alpha^{-1}G\Omega_\alpha\overline{G})(K^2-\alpha I)^{-1}(\hat{K}-K)\hat{g} = -\Sigma\hat{g} + O_p(1/(\alpha n))$. The second term can therefore be eliminated because it would create lower order elements. For the same reason, S can be used instead of $\hat{S}$.

$$\begin{aligned}\Xi_{22}(\tau_1,\tau_5) =&n\rho_3^2 E[S\hat{g}\Sigma\hat{g}]\overline{[S\hat{g}\Sigma\hat{g}]}(\tau_1,\tau_5)\\ =&n\rho_3^2 E[S_{123}\hat{g}_3\Sigma_{24}\hat{g}_4]\overline{[S_{123}\hat{g}_3\Sigma_{24}\hat{g}_4]}(\tau_1,\tau_5)\\ =&n\rho_3^2 E[S_{123}\hat{g}_3\Sigma_{24}\hat{g}_4 S_{567}\hat{g}_7\Sigma_{68}\hat{g}_8]\\ =&n\rho_3^2 S_{123}\Sigma_{24}S_{567}\Sigma_{68}E[\hat{g}_3\hat{g}_4\hat{g}_7\hat{g}_8]\end{aligned}$$

By the iid assumption,

$$
\begin{aligned}
E[\hat{g}_3\hat{g}_4\hat{g}_7\hat{g}_8] =& \frac{1}{n^4}\sum_{t_1}^{n}\sum_{t_2}^{n}\sum_{t_3}^{n}\sum_{t_4}^{n} E[g_{t_1}(\tau_3)g_{t_2}(\tau_4)g_{t_3}(\tau_7)g_{t_4}(\tau_8)] \\
=& \frac{1}{n^4}\{n(n-1)[k(\tau_3,\tau_4)k(\tau_7,\tau_8)+k(\tau_3,\tau_7)k(\tau_4,\tau_8)+k(\tau_3,\tau_8)k(\tau_4,\tau_7)]+nQ(\tau_3,\tau_4,\tau_7 \\
=& \frac{n-1}{n^3}[[k(\tau_3,\tau_4)k(\tau_7,\tau_8)+k(\tau_3,\tau_7)k(\tau_4,\tau_8)+k(\tau_3,\tau_8)k(\tau_4,\tau_7)]+\frac{Q(\tau_3,\tau_4,\tau_7,\tau_8)}{n^3} \\
=& \frac{1}{n^2}[[k(\tau_3,\tau_4)k(\tau_7,\tau_8)+k(\tau_3,\tau_7)k(\tau_4,\tau_8)+k(\tau_3,\tau_8)k(\tau_4,\tau_7)]+O\left(\frac{1}{n^3}\right)
\end{aligned}
$$

We can drop the $O(1/n^3)$ term which implies

$$
\begin{aligned}
\Xi_{22}(\tau_1,\tau_5) =& \frac{1}{n}\rho_3^2 S_{123}\Sigma_{24}S_{567}\Sigma_{68}[K_{34}K_{78}+K_{37}K_{48}+K_{38}K_{47}] \\
=& \frac{1}{n}\rho_3^2 S_{123}\Sigma_{24}S_{567}\Sigma_{68}\Delta_{3478} \\
=& \frac{1}{n}\rho_3^2[S\Sigma\Delta\Sigma S](\tau_1,\tau_5) \\
=& O\left(\frac{1}{\alpha n}\right),
\end{aligned}
\tag{C.21}
$$

where $\Delta_{3478} = [K_{34}K_{78}+K_{37}K_{48}+K_{38}K_{47}]$.

– $E(n\mathcal{T}_3^2)$

In order to analyze that term, we need an expression for $(K^2-\hat{K}^2)$, Since $k(\tau_1,\tau_2) = E[\epsilon_t^2 \exp(ix_t(\tau_1-\tau_2)] = \sigma_\epsilon^2 E[\exp(ix_t(\tau_1-\tau_2)]$, if we condition on x_t, we can write $\hat{K} = \hat{\sigma}_\epsilon^2/\sigma_\epsilon^2 K$, which implies that $(K^2-\hat{K}^2) = (1-\hat{\sigma}_\epsilon^4/\sigma_\epsilon^4)K^2$. It follows that :

$$
\begin{aligned}
\Xi_{33}(\tau_1,\tau_2) =& \frac{n}{\sigma_\epsilon^8}E\left[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)^2(K^2\Sigma\hat{g})\overline{(K^2\Sigma\hat{g})}\right](\tau_1,\tau_2) \\
=& \frac{n}{\sigma_\epsilon^8}\left[K^2\Sigma E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)^2\hat{g}\overline{\hat{g}}]\Sigma K^2\right](\tau_1,\tau_2),
\end{aligned}
$$

where

$$
\begin{aligned}
E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)^2\hat{g}\overline{\hat{g}}](\tau_1,\tau_2) &= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)^2 g_t(\tau_1)g_s(\tau_2)] \\
&= \frac{1}{n^2}\sum_{t=1}^{n} E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)^2 g_t(\tau_1)g_t(\tau_2)] + o(n^{-2}),
\end{aligned}
$$

with

$$\begin{aligned}
E[(\sigma_\epsilon^4-\hat\sigma_\epsilon^4)^2 g_t(\tau_1)g_t(\tau_2)] =& \frac{1}{n^4}E[e^{ix_t(\tau_1-\tau_2)}]\sum_{t_1=1}^n\sum_{t_2=1}^n\sum_{t_3=1}^n\sum_{t_4=1}^n E[(\sigma_\epsilon^4-\epsilon_{t_1}^2\epsilon_{t_2}^2)(\sigma_\epsilon^4-\epsilon_{t_3}^2\epsilon_{t_4}^2)\epsilon_t^2]\\
=& \frac{1}{\sigma_\epsilon^2 n^4}k(\tau_1,\tau_2)\sum_{t_1=1}^n\sum_{t_2=1}^n\sum_{t_3=1}^n\sum_{t_4=1}^n E[(\sigma_\epsilon^4-\epsilon_{t_1}^2\epsilon_{t_2}^2)(\sigma_\epsilon^4-\epsilon_{t_3}^2\epsilon_{t_4}^2)\epsilon_t^2]\\
=& \frac{1}{\sigma_\epsilon^2}k(\tau_1,\tau_2)\left(\frac{4\sigma_\epsilon^{10}}{n}(k_\epsilon-1)+O\left(\frac{1}{n^2}\right)\right)\\
=& \frac{4\sigma_\epsilon^8(k_\epsilon-1)}{n}k(\tau_1,\tau_2)+O\left(\frac{1}{n^2}\right),
\end{aligned}$$

where $k_\epsilon = E(\epsilon_t^4)/\sigma_\epsilon^4$. It follows that

$$E[(\sigma_\epsilon^4-\hat\sigma_\epsilon^4)^2\hat g\overline{\hat g}](\tau_1,\tau_2) = \frac{4\sigma_\epsilon^8(k_\epsilon-1)}{n^2}k(\tau_1,\tau_2)+o\left(\frac{1}{n^2}\right),$$

and, if we drop the $o(1/n^2)$ term,

$$\Xi_{33} = \frac{4(k_\epsilon-1)}{n}K^2\Sigma K\Sigma K^2 = O\left(\frac{1}{\alpha n}\right).$$

The third term can then be written explicitly as :

$$E(n\mathcal{T}_3^2) = \frac{4\Omega_\alpha^2(k_\epsilon-1)}{n}\overline{G}K_\alpha^{-1}K\Sigma K\Sigma K K_\alpha^{-1}G \tag{C.22}$$

– $E(n\mathcal{T}_4^2)$

We show above that the operator V is $-\sigma_{\epsilon u}K/\sigma_\epsilon^2$, which implies

$$\begin{aligned}
\Xi_{44}(\tau_1,\tau_2) =& \frac{16\Omega_\alpha^2\sigma_{\epsilon u}^2 n}{\sigma_\epsilon^4}E[K^2GK_\alpha^{-1}\hat g\Sigma\hat g]\overline{[K^2GK_\alpha^{-1}\hat g\Sigma\hat g]}(\tau_1,\tau_2)\\
=& \frac{16\Omega_\alpha^2\sigma_{\epsilon u}^2 n}{\sigma_\epsilon^4}K_{13}K_{34}E[G_5K_{\alpha 56}^{-1}\hat g_6 G_7 K_{\alpha 78}^{-1}\hat g_8\Sigma_{49}\hat g_9\hat g_{10}\Sigma_{10,11}]K_{11,12}K_{12,2}\\
=& \frac{16\Omega_\alpha^2\sigma_{\epsilon u}^2 n}{\sigma_\epsilon^4}K_{13}K_{34}G_5K_{\alpha 56}^{-1}G_7K_{\alpha 78}^{-1}\Sigma_{49}\Sigma_{10,11}K_{11,12}K_{12,2}E[\hat g_6\hat g_8\hat g_9\hat g_{10}]\\
=& \frac{16\Omega_\alpha^2\sigma_{\epsilon u}^2}{\sigma_\epsilon^4 n}K_{13}K_{34}G_5K_{\alpha 56}^{-1}G_7K_{\alpha 78}^{-1}\Sigma_{49}\Sigma_{10,11}K_{11,12}K_{12,2}\Delta_{689,10}\\
=& \frac{16\Omega_\alpha^2\sigma_{\epsilon u}^2}{\sigma_\epsilon^4 n}[K^2\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma K^2](\tau_1,\tau_2)\\
=& O\left(\frac{1}{\alpha n}\right)
\end{aligned}$$

It follows that :

$$E(n\mathcal{T}_4^2) = \frac{16\Omega_\alpha^4\sigma_{\epsilon u}^2}{\sigma_\epsilon^4 n}\overline{G}K_\alpha^{-1}K[\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma]KK_\alpha^{-1}G \tag{C.23}$$

Notice that we must keep in mind the version with the subscripts in order to respect the order of integration.

– $E(n\mathcal{T}_5^2)$

Because of the similarity with $E(nT_4^2)$, we can easily show that the term can be written as :

It follows that :

$$E(n\mathcal{T}_5^2) = \frac{\rho_3^2\Omega_\alpha^4}{n}\overline{G}(K^2+\alpha I)^{-1}SG[\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma]\overline{G}S(K^2+\alpha I)^{-1}G \tag{C.24}$$

with Ξ_{55} being $O\left(\frac{1}{\alpha^2 n}\right)$

– $E(n\mathcal{T}_6^2)$

$$\begin{aligned}
\Xi_{66}(\tau_1,\tau_2) =& \frac{9\rho_3^2 n}{4}E[S_{134}K_{45}\Sigma_{56}\hat{g}_6\Sigma_{37}\hat{g}_7]\overline{[S_{289}K_{9,10}\Sigma_{10,11}\hat{g}_{11}\Sigma_{8,12}\hat{g}_{12}]}\\
=& \frac{9\rho_3^2 n}{4}S_{134}K_{45}\Sigma_{56}\Sigma_{37}S_{289}K_{9,10}\Sigma_{10,11}\Sigma_{8,12}E[\hat{g}_6\hat{g}_7\hat{g}_{11}\hat{g}_{12}]\\
=& \frac{9\rho_3^2}{4n}S_{134}K_{45}\Sigma_{56}\Sigma_{37}S_{289}K_{9,10}\Sigma_{10,11}\Sigma_{8,12}\Delta_{67,11,12}\\
=& \frac{9\rho_3^2}{4n}SK\Sigma^2\Delta\Sigma^2KS\\
=& O\left(\frac{1}{\alpha^2 n}\right)
\end{aligned}$$

$$E(n\mathcal{T}_6^2) = \frac{9\rho_3^2\Omega_\alpha^2}{2n}\overline{G}(K^2+\alpha I)^{-1}SK\Sigma^2\Delta\Sigma^2KS(K^2+\alpha I)^{-1}G \tag{C.25}$$

– $E(n\mathcal{T}_1\mathcal{T}_2)$

From now on, only the Ξ_{ij} terms are derived.

$$\begin{aligned}\Xi_{12}(\tau_1,\tau_2) =&\rho_3 n E[K\hat{g}\overline{S\hat{g}\Sigma\hat{g}}]\\ =&\rho_3 n E[K_{13}\hat{g}_3\overline{\hat{g}_4}\Sigma_{45}\overline{\hat{g}_6}S_{652}]\\ =&\rho_3 n K_{13}\Sigma_{45}S_{652}E[\hat{g}_3\overline{\hat{g}_4\hat{g}_6}]\\ =&\frac{\rho_3}{n}K_{13}\Sigma_{45}S_{652}S_{346}\\ =&\frac{\rho_3}{n}KS\Sigma S\\ =&O\left(\frac{1}{n\sqrt{\alpha}}\right)\end{aligned}$$

– $E(n\mathcal{T}_1\mathcal{T}_3)$

$$\begin{aligned}\Xi_{13}(\tau_1,\tau_2) =& -\frac{n}{\sigma_\epsilon^4}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)K\hat{g}\overline{K^2\Sigma\hat{g}}]\\ =& -\frac{n}{\sigma_\epsilon^4}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)K_{13}\hat{g}_3\overline{\hat{g}_4}\Sigma_{45}K_{56}K_{62}]\\ =& -\frac{n}{\sigma_\epsilon^4}K_{13}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)\hat{g}_3\overline{\hat{g}_4}]\Sigma_{45}K_{56}K_{62}\\ =& -\frac{n}{\sigma_\epsilon^4}K_{13}\left[\frac{1}{n^2}\sum_{t=1}^n\sum_{s=1}^n E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)g_{t3}\overline{g_{s4}}]\right]\Sigma_{45}K_{56}K_{62}\\ =& -\frac{n}{\sigma_\epsilon^4}K_{13}\left[\frac{1}{n^2}\sum_{t=1}^n E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)g_{t3}\overline{g_{t4}}]\right]\Sigma_{45}K_{56}K_{62}+O\left(\frac{1}{n^2}\right)\end{aligned}$$

with

$$\begin{aligned}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)g_t(\tau_3)\overline{g_t(\tau_4)}] =&\frac{k(\tau_3,\tau_4)}{\sigma_\epsilon^2 n^2}\sum_{t_1=1}^n\sum_{t_2=1}^n E[(\sigma_\epsilon^4-\epsilon_{t_1}\epsilon_{t_2})\epsilon_t^2]\\ =&\frac{k(\tau_3,\tau_4)}{\sigma_\epsilon^2 n}\left(3\sigma_\epsilon^6(1-k_\epsilon)\right)+O\left(\frac{1}{n^2}\right),\end{aligned}$$

which implies

$$\begin{aligned}\Xi_{13}(\tau_1,\tau_2) =& -\frac{n}{\sigma_\epsilon^4}K_{13}\left[K_{34}\frac{3\sigma_\epsilon^4(1-k_\epsilon)}{n^2}\right]\Sigma_{45}K_{56}K_{62}+O\left(\frac{1}{n^2}\right)\\ =&\frac{3(k_\epsilon-1)}{n}K_{13}K_{34}\Sigma_{45}K_{56}K_{62}+O\left(\frac{1}{n^2}\right)\\ =&\frac{3(k_\epsilon-1)}{n}[K^2\Sigma K^2](\tau_1,\tau_2)\\ =&O\left(\frac{1}{n\sqrt{\alpha}}\right),\end{aligned}\tag{C.26}$$

where the smaller order term as been removed in the last equality.

- $E(n\mathcal{T}_1\mathcal{T}_4)$

$$\begin{aligned}\Xi_{14}(\tau_1,\tau_2) =& -\frac{4n\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^2}E[K\hat{g}\overline{K^2\bar{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}}]\\ =& -\frac{4n\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^2}E[K_{13}\hat{g}_3\overline{\hat{g}_4}\Sigma_{45}\overline{\hat{g}_6}K_{\alpha 67}^{-1}G_7K_{58}K_{82}]\\ =& -\frac{4\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^2 n}K_{13}\Sigma_{45}K_{\alpha 67}^{-1}G_7K_{58}K_{82}S_{346}\\ =& -\frac{4\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^2 n}K_{13}S_{346}K_{\alpha 67}^{-1}G_7\Sigma_{45}K_{58}K_{82}\\ =& -\frac{4\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^2 n}[KSK_\alpha^{-1}G\Sigma K^2](\tau_1,\tau_2)\\ =& O\left(\frac{1}{n\sqrt{\alpha}}\right)\end{aligned}$$

- $E(n\mathcal{T}_1\mathcal{T}_5)$

$$\begin{aligned}\Xi_{15}(\tau_1,\tau_2) =& -\rho_3 n\Omega_\alpha E[K\hat{g}\overline{SG\bar{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}}](\tau_1,\tau_2)\\ =& -\rho_3 n\Omega_\alpha E[K_{13}\hat{g}_3\overline{\hat{g}_4}\Sigma_{45}\overline{\hat{g}_6}K_{\alpha 67}^{-1}G_7\overline{G_8}S_{852}]\\ =& -\frac{\rho_3\Omega_\alpha}{n}K_{13}\Sigma_{45}K_{\alpha 67}^{-1}G_7\overline{G_8}S_{852}S_{346}\\ =& -\frac{\rho_3\Omega_\alpha}{n}K_{13}S_{346}K_{\alpha 67}^{-1}G_7\Sigma_{45}\overline{G_8}S_{852}\\ =& -\frac{\rho_3\Omega_\alpha}{n}[KSK_\alpha^{-1}G\Sigma\overline{G}S](\tau_1,\tau_2)\\ =& O\left(\frac{1}{n\sqrt{\alpha}}\right)\end{aligned}$$

- $E(n\mathcal{T}_1\mathcal{T}_6)$

$$\begin{aligned}
\Xi_{16}(\tau_1,\tau_2) =& -\frac{3n\rho_3}{2}E[K\hat{g}\overline{SK\Sigma\hat{g}\Sigma\hat{g}}](\tau_1,\tau_2)\\
=& -\frac{3n\rho_3}{2}E[K_{13}\hat{g}_3\overline{\hat{g}_4}\Sigma_{45}\overline{\hat{g}_6}\Sigma_{67}K_{78}S_{852}]\\
=& -\frac{3n\rho_3}{2}K_{13}\Sigma_{45}\Sigma_{67}K_{78}S_{852}E[\hat{g}_3\overline{\hat{g}_4\hat{g}_6}]\\
=& -\frac{3\rho_3}{2n}K_{13}\Sigma_{45}\Sigma_{67}K_{78}S_{852}S_{346}\\
=& -\frac{3\rho_3}{2n}[KS\Sigma^2KS](\tau_1,\tau_2)\\
=&O\left(\frac{1}{\alpha n}\right)
\end{aligned}$$

– $E(n\mathcal{T}_2\mathcal{T}_3)$

$$\begin{aligned}
\Xi_{23}(\tau_1,\tau_2) =& -\frac{\rho_3 n}{\sigma_\epsilon^4}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)S\hat{g}\Sigma\hat{g}\overline{K^2\Sigma\hat{g}}]\\
=& -\frac{\rho_3 n}{\sigma_\epsilon^4}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)S_{134}\hat{g}_4\Sigma_{35}\hat{g}_5\overline{\hat{g}_6}\Sigma_{67}K_{78}K_{82}]\\
=& -\frac{\rho_3 n}{\sigma_\epsilon^4}S_{134}\Sigma_{35}\Sigma_{67}K_{78}K_{82}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)\hat{g}_4\hat{g}_5\overline{\hat{g}_6}]\\
=& -\frac{\rho_3 n}{\sigma_\epsilon^4}S_{134}\Sigma_{35}\Sigma_{67}K_{78}K_{82}\underbrace{\left[\frac{1}{n^3}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{t_3=1}^{n}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)g_{t_1}(\tau_4)g_{t_2}(\tau_5)\overline{g_{t_3}(\tau_6)}]\right]}_{[\bullet]}
\end{aligned}$$

with

$$\begin{aligned}
[\bullet] =&\frac{1}{n^5}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{t_3=1}^{n}\sum_{t_4=1}^{n}\sum_{t_5=1}^{n}E[(\sigma_\epsilon^4-\epsilon_{t_4}^2\epsilon_{t_5}^2)g_{t_1}(\tau_4)g_{t_2}(\tau_5)\overline{g_{t_3}(\tau_6)}]\\
=&\frac{2S_\epsilon\sigma_\epsilon^2}{n^2}\left[k(\tau_4,\tau_5)E\left(e^{ix_t\tau_6}\right)+k(\tau_4,\tau_6)E\left(e^{ix_t\tau_5}\right)+k(\tau_5,\tau_6)E\left(e^{ix_t\tau_4}\right)\right]\\
=&-\frac{2S_\epsilon\sigma_\epsilon^2}{n^2}\Psi_{456},
\end{aligned}$$

where $S_\epsilon = E(\epsilon_t^3)$. It follows that

$$\begin{aligned}
\Xi_{23}(\tau_1,\tau_2) =&\frac{2\rho_3 S_\epsilon}{\sigma_\epsilon^2 n}S_{134}\Sigma_{35}\Sigma_{67}K_{78}K_{82}\Psi_{456}\\
=&\frac{2\rho_3 S_\epsilon}{\sigma_\epsilon^2 n}[S\Sigma\Psi\Sigma K^2](\tau_1,\tau_2)\\
=&O\left(\frac{1}{\alpha n}\right)
\end{aligned}$$

– $E(n\mathcal{T}_2\mathcal{T}_4)$

$$\begin{aligned}
\Xi_{24}(\tau_1,\tau_2) =& -\frac{4n\rho_3\Omega_\alpha\sigma_{\epsilon u}}{\sigma_\epsilon^2}E[S\hat{g}\Sigma\hat{g}\overline{K^2\bar{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}}]\\
=& -\frac{4n\rho_3\Omega_\alpha\sigma_{\epsilon u}}{\sigma_\epsilon^2}E[S_{134}\hat{g}_4\Sigma_{35}\hat{g}_5\overline{\hat{g}_6}\Sigma_{67}\overline{\hat{g}_8}K_{\alpha 89}^{-1}G_9K_{7,10}K_{10,2}]\\
=& -\frac{4\rho_3\Omega_\alpha\sigma_{\epsilon u}}{n\sigma_\epsilon^2}S_{134}\Sigma_{35}\Sigma_{67}K_{\alpha 89}^{-1}G_9K_{7,10}K_{10,2}\Delta_{4568}\\
=& -\frac{4\rho_3\Omega_\alpha\sigma_{\epsilon u}}{n\sigma_\epsilon^2}[S\Sigma\Delta K_\alpha^{-1}G\Sigma K^2](\tau_1,\tau_2)\\
=& O\left(\frac{1}{\alpha n}\right)
\end{aligned}$$

– $E(n\mathcal{T}_2\mathcal{T}_5)$

$$\begin{aligned}
\Xi_{25}(\tau_1,\tau_2) =& -n\rho_3^2\Omega_\alpha E[S\hat{g}\Sigma\hat{g}\overline{SG\bar{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}}]\\
=& -n\rho_3^2\Omega_\alpha E[S_{134}\hat{g}_4\Sigma_{35}\hat{g}_5\overline{\hat{g}_6}\Sigma_{67}\overline{\hat{g}_8}K_{\alpha 89}^{-1}G_9\overline{G_{10}}S_{10,72}]\\
=& -\frac{\rho_3^2\Omega_\alpha}{n}S_{134}\Sigma_{35}\Sigma_{67}K_{\alpha 89}^{-1}G_9\overline{G_{10}}S_{10,72}\Delta_{4568}\\
=& -\frac{\rho_3^2\Omega_\alpha}{n}[S\Sigma\Delta K_\alpha^{-1}G\Sigma\overline{G}S](\tau_1,\tau_2)\\
=& O\left(\frac{1}{\alpha n}\right)
\end{aligned}$$

– $E(n\mathcal{T}_2\mathcal{T}_6)$

$$\begin{aligned}
\Xi_{26}(\tau_1,\tau_2) =& -\frac{3\rho_3^2 n}{2}E[S\hat{g}\Sigma\hat{g}\overline{SK\Sigma\hat{g}\Sigma\hat{g}}]\\
=& -\frac{3\rho_3^2 n}{2}E[S_{134}\hat{g}_4\Sigma_{35}\hat{g}_5\overline{\hat{g}_6}\Sigma_{67}\overline{\hat{g}_8}\Sigma_{89}K_{9,10}S_{10,72}]\\
=& -\frac{3\rho_3^2}{2n}S_{134}\Sigma_{35}\Sigma_{67}\Sigma_{89}K_{9,10}S_{10,72}\Delta_{4568}\\
=& -\frac{3\rho_3^2}{2n}[S\Sigma\Delta\Sigma\Sigma KS](\tau_1,\tau_2)\\
=& O\left(\frac{1}{\alpha^{3/2}n}\right)
\end{aligned}$$

– $E(n\mathcal{T}_3\mathcal{T}_4)$

$$
\begin{aligned}
\Xi_{34}(\tau_1,\tau_2) =& \frac{4n\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^6} E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)K^2\Sigma\hat{g}\overline{K^2\bar{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}}]\\
=& \frac{4n\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^6} E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)K_{13}K_{34}\Sigma_{45}\hat{g}_5\overline{\hat{g}_6}\Sigma_{67}\overline{\hat{g}_8}K_{\alpha 89}^{-1}G_9K_{7,10}K_{10,2}]\\
=& \frac{4n\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^6} K_{13}K_{34}\Sigma_{45}\Sigma_{67}K_{\alpha 89}^{-1}G_9K_{7,10}K_{10,2}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)\hat{g}_5\overline{\hat{g}_6\hat{g}_8}]\\
=& -\frac{8\sigma_{\epsilon u}\Omega_\alpha S_\epsilon}{n\sigma_\epsilon^4} K_{13}K_{34}\Sigma_{45}\Sigma_{67}K_{\alpha 89}^{-1}G_9K_{7,10}K_{10,2}\Psi_{568}\\
=& -\frac{8\sigma_{\epsilon u}\Omega_\alpha S_\epsilon}{n\sigma_\epsilon^4}[K^2\Sigma\Psi K_\alpha^{-1}G\Sigma K^2](\tau_1,\tau_2)\\
=& O\left(\frac{1}{\alpha n}\right)
\end{aligned}
$$

– $E(n\mathcal{T}_3\mathcal{T}_5)$

$$
\begin{aligned}
\Xi_{35}(\tau_1,\tau_2) =& \frac{\rho_3\Omega_\alpha n}{\sigma_\epsilon^4} E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)K^2\Sigma\hat{g}\overline{SG\bar{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}}]\\
=& \frac{\rho_3\Omega_\alpha n}{\sigma_\epsilon^4} E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)K_{13}K_{34}\Sigma_{45}\hat{g}_5\overline{\hat{g}_6}\Sigma_{67}\overline{\hat{g}_8}K_{\alpha 89}^{-1}G_9\overline{G_{10}}S_{10,72}]\\
=& \frac{\rho_3\Omega_\alpha n}{\sigma_\epsilon^4} K_{13}K_{34}\Sigma_{45}\Sigma_{67}K_{\alpha 89}^{-1}G_9\overline{G_{10}}S_{10,72}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)\hat{g}_5\overline{\hat{g}_6\hat{g}_8}]\\
=& -\frac{2\rho_3 S_\epsilon\Omega_\alpha}{\sigma_\epsilon^2 n} K_{13}K_{34}\Sigma_{45}\Sigma_{67}K_{\alpha 89}^{-1}G_9\overline{G_{10}}S_{10,72}\Psi_{568}\\
=& -\frac{2\rho_3 S_\epsilon\Omega_\alpha}{\sigma_\epsilon^2 n}[K^2\Sigma\Psi K_\alpha^{-1}G\Sigma\overline{G}S](\tau_1,\tau_2)\\
=& O\left(\frac{1}{\alpha n}\right)
\end{aligned}
$$

– $E(n\mathcal{T}_3\mathcal{T}_6)$

$$\begin{aligned}\Xi_{36}(\tau_1,\tau_2) =&\frac{3n\rho_3}{2\sigma_\epsilon^4}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)K^2\Sigma\hat{g}\overline{SK\Sigma\hat{g}\Sigma\hat{g}}]\\
=&\frac{3n\rho_3}{2\sigma_\epsilon^4}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)K_{13}K_{34}\Sigma_{45}\hat{g}_5\overline{\hat{g}_6}\Sigma_{67}\overline{\hat{g}_8}\Sigma_{89}K_{9,10}S_{10,72}]\\
=&\frac{3n\rho_3}{2\sigma_\epsilon^4}K_{13}K_{34}\Sigma_{45}\Sigma_{67}\Sigma_{89}K_{9,10}S_{10,72}E[(\sigma_\epsilon^4-\hat{\sigma}_\epsilon^4)\hat{g}_5\overline{\hat{g}_6\hat{g}_8}]\\
=&-\frac{3\rho_3S_\epsilon}{\sigma_\epsilon^2n}K_{13}K_{34}\Sigma_{45}\Sigma_{67}\Sigma_{89}K_{9,10}S_{10,72}\Psi_{568}\\
=&-\frac{3\rho_3S_\epsilon}{\sigma_\epsilon^2n}[K^2\Sigma\Psi\Sigma^2KS](\tau_1,\tau_2)\\
=&O\left(\frac{1}{\alpha^{3/2}n}\right)\end{aligned}$$

- $E(n\mathcal{T}_4\mathcal{T}_5)$

$$\begin{aligned}\Xi_{45}(\tau_1,\tau_2) =&\frac{4n\sigma_{\epsilon u}\rho_3\Omega_\alpha^2}{\sigma_\epsilon^2}E[K^2\overline{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}\overline{SG\bar{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}}]\\
=&\frac{4n\sigma_{\epsilon u}\rho_3\Omega_\alpha^2}{\sigma_\epsilon^2}E[K_{13}K_{34}\overline{G_5}K_{\alpha 56}^{-1}\hat{g}_6\Sigma_{47}\hat{g}_7\overline{\hat{g}_8}\Sigma_{89}\overline{\hat{g}_{10}}K_{\alpha 10,11}^{-1}G_{11}\overline{G_{12}}S_{12,92}]\\
=&\frac{4n\sigma_{\epsilon u}\rho_3\Omega_\alpha^2}{\sigma_\epsilon^2}K_{13}K_{34}\overline{G_5}K_{\alpha 56}^{-1}\Sigma_{47}\Sigma_{89}K_{\alpha 10,11}^{-1}G_{11}\overline{G_{12}}S_{12,92}E[\hat{g}_6\hat{g}_7\overline{\hat{g}_8\hat{g}_{10}}]\\
=&\frac{4\sigma_{\epsilon u}\rho_3\Omega_\alpha^2}{\sigma_\epsilon^2n}K_{13}K_{34}\overline{G_5}K_{\alpha 56}^{-1}\Sigma_{47}\Sigma_{89}K_{\alpha 10,11}^{-1}G_{11}\overline{G_{12}}S_{12,92}\Delta_{678,10}\\
=&\frac{4\sigma_{\epsilon u}\rho_3\Omega_\alpha^2}{\sigma_\epsilon^2n}[K^2\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma\overline{G}S](\tau_1,\tau_2)\\
=&O\left(\frac{1}{\alpha n}\right)\end{aligned}$$

- $E(n\mathcal{T}_4\mathcal{T}_6)$

$$\begin{aligned}\Xi_{46}(\tau_1,\tau_2) =&\frac{6n\sigma_{\epsilon u}\rho_3\Omega_\alpha}{\sigma_\epsilon^2}E[K^2\overline{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}\overline{SK\Sigma\hat{g}\Sigma\hat{g}}]\\
=&\frac{6n\sigma_{\epsilon u}\rho_3\Omega_\alpha}{\sigma_\epsilon^2}E[K_{13}K_{34}\overline{G_5}K_{\alpha 56}^{-1}\hat{g}_6\Sigma_{47}\hat{g}_7\overline{\hat{g}_8}\Sigma_{89}\overline{\hat{g}_{10}}\Sigma_{10,11}K_{11,12}S_{12,92}]\\
=&\frac{6n\sigma_{\epsilon u}\rho_3\Omega_\alpha}{\sigma_\epsilon^2}K_{13}K_{34}\overline{G_5}K_{\alpha 56}^{-1}\Sigma_{47}\Sigma_{89}\Sigma_{10,11}K_{11,12}S_{12,92}E[\hat{g}_6\hat{g}_7\overline{\hat{g}_8\hat{g}_{10}}]\\
=&\frac{6\sigma_{\epsilon u}\rho_3\Omega_\alpha}{\sigma_\epsilon^2n}[K^2\Sigma\overline{G}K_\alpha^{-1}\Delta\Sigma^2KS](\tau_1,\tau_2)\\
=&O\left(\frac{1}{\alpha^{3/2}n}\right)\end{aligned}$$

– $E(n\mathcal{T}_5\mathcal{T}_6)$

$$\begin{aligned}
\Xi_{56}(\tau_1,\tau_2) =& \frac{3n\rho_3^2\Omega_\alpha}{2} E[SG\overline{G}K_\alpha^{-1}\hat{g}\Sigma\hat{g}\overline{SK\Sigma\hat{g}\Sigma\hat{g}}] \\
=& \frac{3n\rho_3^2\Omega_\alpha}{2} E[S_{134}G_4\overline{G_5}K_{\alpha 56}^{-1}\hat{g}_6\Sigma_{37}\hat{g}_7\overline{\hat{g}_8}\Sigma_{89}\overline{\hat{g}_{10}}\Sigma_{10,11}K_{11,12}S_{12,92}] \\
=& \frac{3n\rho_3^2\Omega_\alpha}{2} S_{134}G_4\overline{G_5}K_{\alpha 56}^{-1}\Sigma_{37}\Sigma_{89}\Sigma_{10,11}K_{11,12}S_{12,92}E[\hat{g}_6\hat{g}_7\overline{\hat{g}_8\hat{g}_{10}}] \\
=& \frac{3\rho_3^2\Omega_\alpha}{2n}[SG\Sigma\overline{G}K_\alpha^{-1}\Delta\Sigma^2 KS](\tau_1,\tau_2) \\
=& O\left(\frac{1}{\alpha^{3/2}n}\right)
\end{aligned}$$

The approximate MSE can therefore be written as :

$$E(n\theta^2) = \Omega_\alpha^2\overline{G}(K^2+\alpha I)^{-1}\Xi(K^2+\alpha I)G \tag{C.27}$$

with

$$
\begin{aligned}
\Xi = K^3 &+ \frac{\rho_3^2}{n}[S\Sigma\Delta\Sigma S] + \frac{4\Omega_\alpha^2(k_\epsilon - 1)}{n}K^2\Sigma K\Sigma K^2 \\
&+ \frac{16\Omega_\alpha^2\sigma_{\epsilon u}^2}{\sigma_\epsilon^4 n}K^2\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma K^2 + \frac{\rho_3^2\Omega_\alpha^2}{n}SG\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma\overline{G}S \\
&+ \frac{9\rho_3^2}{4n}SK\Sigma^2\Delta\Sigma^2 KS \\
&+ \frac{\rho_3}{n}KS\Sigma S \\
&+ \frac{3(k_\epsilon - 1)}{n}K^2\Sigma K^2 \\
&- \frac{4\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^2 n}KSK_\alpha^{-1}G\Sigma K^2 \\
&- \frac{\rho_3\Omega_\alpha}{n}[KSK_\alpha^{-1}G\Sigma\overline{G}S] \\
&- \frac{3\rho_3}{2n}KS\Sigma^2 KS \\
&+ \frac{2\rho_3 S_\epsilon}{\sigma_\epsilon^2 n}S\Sigma\Psi\Sigma K^2 \\
&- \frac{4\rho_3\Omega_\alpha\sigma_{\epsilon u}}{n\sigma_\epsilon^2}S\Sigma\Delta K_\alpha^{-1}G\Sigma K^2 \\
&- \frac{\rho_3^2\Omega_\alpha}{n}S\Sigma\Delta K_\alpha^{-1}G\Sigma\overline{G}S \\
&- \frac{3\rho_3^2}{2n}S\Sigma\Delta\Sigma\Sigma KS \\
&- \frac{8\sigma_{\epsilon u}\Omega_\alpha S_\epsilon}{n\sigma_\epsilon^4}K^2\Sigma\Psi K_\alpha^{-1}G\Sigma K^2 \\
&- \frac{2\rho_3 S_\epsilon\Omega_\alpha}{\sigma_\epsilon^2 n}K^2\Sigma\Psi K_\alpha^{-1}G\Sigma\overline{G}S \\
&- \frac{3\rho_3 S_\epsilon}{\sigma_\epsilon^2 n}K^2\Sigma\Psi\Sigma^2 KS \\
&+ \frac{4\sigma_{\epsilon u}\rho_3\Omega_\alpha^2}{\sigma_\epsilon^2 n}K^2\Sigma\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G\Sigma\overline{G}S \\
&+ \frac{6\sigma_{\epsilon u}\rho_3\Omega_\alpha}{\sigma_\epsilon^2 n}K^2\Sigma\overline{G}K_\alpha^{-1}\Delta\Sigma^2 KS \\
&+ \frac{3\rho_3^2\Omega_\alpha}{2n}SG\Sigma\overline{G}K_\alpha^{-1}\Delta\Sigma^2 KS
\end{aligned} \tag{C.28}
$$

C.1.5 Computing the approximate MSE

The approximate MSE requires a first step estimate of δ which can be the GMM estimate for a continuum using the identity operator since it does not depend on any regularization parameter. We derive here the estimate of some MSE terms. The other terms can be derived similarly. Each operator are replaced by its estimate using $\tilde{\epsilon}_i = y_i - \tilde{\delta} W_i$, where $\tilde{\delta}$ is the first step estimate. In the following, we write $\hat{K}(\tilde{\delta})$ as K. That notation holds also for the other operators. We also omit the hat over the variables for simplicity.

Notice that for each term, we need to compute Ω_α which is just the asymptotic covariance matrix of $\hat{\delta}$ which is given by equation (B.1) in Section (B.1).

– $E(n\mathcal{T}_1^2)$

For the first term, we need the eigenvalues, μ_i, and eigen-functions, ϕ_i, of K which are defined in Section (A)

$$\begin{aligned} E(n\mathcal{T}_1^2) =& \Omega_\alpha^2 \bar{G}(K^2+\alpha I)^{-1} K^3 (K^2+\alpha I)^{-1} G \\ =& \Omega_\alpha^2 \sum_{i=1}^n \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < G, \phi_i >^2 \\ =& \Omega_\alpha^2 \frac{w' H B \beta D_{32} \beta' B H w}{n^2}, \end{aligned}$$

where β, H, B and w are defined in Section (B.1). We also define D_{ab} as an $n \times n$ diagonal matrix with the i^{th} diagonal equals to $\mu_i^a/(\mu_i^2+\alpha)^b$. We also define the three following expressions :

$$P_{ab} = w' H B \beta D_{ab} \beta' B H w / n^2,$$

$$P_{tab} = -H_{t\bullet} B \beta D_{ab} \beta' B H w / n,$$

and

$$P_{tsab} = H_{t\bullet} B \beta D_{ab} \beta' B H_{\bullet s}$$

which allows us to simplify the first term :

$$E(n\mathcal{T}_1^2) = \Omega_\alpha^2 P_{32}, \tag{C.29}$$

– $E(n\mathcal{T}_2^2)$

The second term can be written as :

$$E(n\mathcal{T}_2^2) = \frac{\rho_3^2\Omega_\alpha^2}{n}\overline{G}(K^2+\alpha I)^{-1}\{S_{123}\Sigma_{24}S_{567}\Sigma_{68}\Delta_{3478}\}(K^2+\alpha I)^{-1}G$$

We first need to compute the operator Σ for which $(\Sigma f)(\tau)$, $f \in L^2(\pi)$, is :

$$\begin{aligned}(\Sigma f)(\tau) =&(K_\alpha^{-1}f)(\tau) - [K_\alpha^{-1}G\Omega_\alpha\bar{G}K_\alpha^{-1}f](\tau)\\
=&(K_\alpha^{-1}f)(\tau) - [\Omega_\alpha\bar{G}K_\alpha^{-1}f][K_\alpha^{-1}G](\tau)\\
=&\sum_{i=1}^n \frac{\mu_i}{\mu_i^2+\alpha} < \phi_i, f > \phi_i(\tau)\\
&-\Omega_\alpha\left[\sum_{i=1}^n \frac{\mu_i}{\mu_i^2+\alpha} < \phi_i, f >< G, \phi_i >\right]\sum_{i=1}^n \frac{\mu_i}{\mu_i^2+\alpha} < \phi_i, G > \phi_i(\tau)\\
=&\sum_{i=1}^n \frac{\mu_i}{\mu_i^2+\alpha} < \phi_i, f > \phi_i(\tau)\\
&-\Omega_\alpha\left[\frac{1}{n}\sum_{i=1}^n \frac{\mu_i}{\mu_i^2+\alpha}[\beta_i'BHw] < \phi_i, f >\right]\frac{1}{n}\sum_{i=1}^n \frac{\mu_i}{\mu_i^2+\alpha}[\beta_i'BHw]\phi_i(\tau),\end{aligned}$$

where $\phi_i(\tau) = \beta_i' g(\tau)/n$ (see Appendix (A)) [1]. The exact expression depends on f. We proceed step by step in order to see clearly the required expression. Let $S_{123} = S_\epsilon \frac{1}{n}\sum_{t=1}^n Z_t(\tau_1)\overline{Z_t(\tau_2)Z_t(\tau_3)}$, where $S_\epsilon = \sum_{i=1}^n \epsilon_i^3/n$, then :

$$\begin{aligned}\{S_{123}\Sigma_{24}S_{567}\Sigma_{68}\Delta_{3478}\} =&\frac{S_\epsilon}{n^2}\sum_{t,s=1}^n Z_t(\tau_1)\overline{Z_s(\tau_5)}[\overline{Z_t}\Sigma]_4][\overline{Z_s}\Sigma]_8\overline{Z_t(\tau_3)}Z_s(\tau_7)\Delta_{3478}\\
=&\frac{S_\epsilon}{n^2}\sum_{t,s=1}^n Z_t(\tau_1)\overline{Z_s(\tau_5)}[\overline{Z_t}\Sigma]_4][\overline{Z_s}\Sigma]_8\overline{Z_t(\tau_3)}Z_s(\tau_7)\Big[K_{34}K_{78}\\
&+K_{37}K_{48}+K_{38}K_{47}\Big]\\
=&\frac{S_\epsilon}{n^2}\sum_{t,s=1}^n Z_t(\tau_1)\overline{Z_s(\tau_5)}\Big[[\overline{Z_t}K\Sigma Z_t][\overline{Z_s}K\Sigma Z_s]\\
&+[\overline{Z_t}KZ_s][\overline{Z_t}\Sigma K\Sigma Z_s]+[\overline{Z_t}K\Sigma Z_s][\overline{Z_t}\Sigma KZ_s]\Big]\end{aligned}$$

1. Review the discussion at the end of Section (B.1) to compute the above terms. The eigenfunctions defined as $\phi_i(\tau) = \beta_i' g(\tau)/n$ are not normalized. We need to transform the β_i's to make the norm of ϕ_i equal to one. Once correctly transformed we can apply the formulas of the section.

It implies that

$$E(n\mathcal{T}_2^2) = \frac{\rho_3^2\Omega_\alpha^2 S_\epsilon^2}{n^3}\sum_{t,s=1}^{n}[\bar{G}(K^2+\alpha I)^{-1}Z_t][\bar{G}(K^2+\alpha I)^{-1}Z_s]\Big[+ [\overline{Z_t}K\Sigma Z_t][\overline{Z_s}K\Sigma Z_s] + [\overline{Z_t}KZ_s][\overline{Z_t}\Sigma K\Sigma Z_s] + [\overline{Z_t}K\Sigma Z_s][\overline{Z_t}\Sigma KZ_s]\Big]$$

with

$$\bar{G}(K^2+\alpha I)^{-1}Z_t = \sum_{i=1}^{n}\frac{1}{(\mu_i^2+\alpha)} < G,\phi_i >< Z_t,\phi_i > = -w'HB\beta D_{01}\beta' BH_{\bullet t}/n = P_{t01}$$

$$\overline{Z_t}K\Sigma Z_t = \sum_{i=1}^{n}\frac{\mu_i^2}{\mu_i^2+\alpha} < \phi_i, Z_t >^2 - \Omega_\alpha\left[\sum_{i=1}^{n}\frac{\mu_i^2}{\mu_i^2+\alpha} < \phi_i, Z_t >< G,\phi_i >\right]\sum_{i=1}^{n}\frac{\mu_i}{\mu_i^2+\alpha} < \phi_i, G >< \phi_i, Z_t >$$
$$= H_{t\bullet}B\beta D_{21}\beta' BH_{\bullet t} - \Omega_\alpha[w'HB\beta D_{21}\beta' BH_{\bullet t}/n][w'HB\beta D_{11}\beta' BH_{\bullet t}/n]$$
$$= P_{t21} - \Omega_\alpha P_{t21}P_{t11}$$

Similarly, we have :

$$[\overline{Z_t}KZ_s] = H_{t\bullet}B\beta D_{10}\beta' BH_{\bullet s} = P_{ts10}$$

$$[\overline{Z_t}K\Sigma Z_s] = P_{ts21} - \Omega_\alpha P_{t21}P_{s11}$$

$$[\overline{Z_t}\Sigma KZ_s] = P_{ts21} - \Omega_\alpha P_{t11}P_{s21}$$

For $[\overline{Z_t}\Sigma K\Sigma Z_s]$ we need the operator :

$$
\begin{aligned}
(\Sigma K\Sigma f)(\tau) = &\sum_{i=1}^{n} \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < \phi_i, f > \phi_i(\tau) \\
&- \Omega_\alpha \left[\sum_{i=1}^{n} \frac{\mu_i}{\mu_i^2+\alpha} < \phi_i, f >< G, \phi_i >\right] \sum_{i=1}^{n} \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < \phi_i, G > \phi_i(\tau) \\
&- \Omega_\alpha \left[\sum_{i=1}^{n} \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < \phi_i, f >< G, \phi_i >\right] \sum_{i=1}^{n} \frac{\mu_i}{\mu_i^2+\alpha} < \phi_i, G > \phi_i(\tau) \\
&+ \Omega_\alpha^2 \left[\sum_{i=1}^{n} \frac{\mu_i}{\mu_i^2+\alpha} < \phi_i, f >< G, \phi_i >\right] \left[\sum_{i=1}^{n} \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < \phi_i, G >^2\right] \\
&\qquad \sum_{i=1}^{n} \frac{\mu_i}{\mu_i^2+\alpha} < \phi_i, G > \phi_i(\tau),
\end{aligned}
$$

which implies

$$
[\overline{Z_t}\Sigma K\Sigma Z_s] = P_{ts32} - \Omega_\alpha P_{s11} P_{t32} - \Omega P_{s32} P_{t11} + \Omega^2 P_{s11} P_{32} P_{t11},
$$

and

$$
\begin{aligned}
E(n\mathcal{T}_2^2) = &\frac{\rho_3^2 \Omega_\alpha^2 S_\epsilon^2}{n^3} \sum_{t,s=1}^{n} P_{t01} P_{s01} \Big[(P_{t21} - \Omega_\alpha P_{t21} P_{t11})(P_{s21} - \Omega_\alpha P_{s21} P_{s11}) \\
&+ P_{ts10}(P_{ts32} - \Omega_\alpha P_{s11} P_{t32} - \Omega P_{s32} P_{t11} + \Omega_\alpha^2 P_{s11} P_{32} P_{t11}) \\
&+ (P_{ts21} - \Omega_\alpha P_{t21} P_{s11})(P_{ts21} - \Omega_\alpha P_{t11} P_{s21}) \Big]
\end{aligned}
$$

– $E(n\mathcal{T}_3^2)$

The term can be written as $\frac{4\Omega_\alpha^4(k_\epsilon - 1)}{n}\overline{G} K_\alpha^{-1} K\Sigma K\Sigma K K_\alpha^{-1} G$, where the operator $K_\alpha^{-1} K\Sigma K\Sigma K K_\alpha^{-1}$ is defined as :

$$(K_\alpha^{-1}K\Sigma K\Sigma KK_\alpha^{-1}f)(\tau) = \sum_{i=1}^n \frac{\mu_i^7}{(\mu_i^2+\alpha)^4} < \phi_i, f > \phi_i(\tau)$$
$$- \Omega_\alpha \left[\sum_{i=1}^n \frac{\mu_i^5}{(\mu_i^2+\alpha)^3} < \phi_i, f >< G, \phi_i >\right] \sum_{i=1}^n \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < \phi_i, G > \phi_i($$
$$- \Omega_\alpha \left[\sum_{i=1}^n \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < \phi_i, f >< G, \phi_i >\right] \sum_{i=1}^n \frac{\mu_i^5}{(\mu_i^2+\alpha)^3} < \phi_i, G > \phi_i($$
$$+ \Omega_\alpha^2 \left[\sum_{i=1}^n \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < \phi_i, f >< G, \phi_i >\right] \left[\sum_{i=1}^n \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < \phi_i, G >^2\right]$$
$$\sum_{i=1}^n \frac{\mu_i^3}{(\mu_i^2+\alpha)^2} < \phi_i, G > \phi_i(\tau).$$

which implies

$$E(n\mathcal{T}_3^2) = \frac{4\Omega_\alpha^4(k_\epsilon-1)}{n}\Big[w'HB\beta D_{74}\beta'BHw/n^2$$
$$- 2\Omega_\alpha[w'HB\beta D_{32}\beta'BHw][w'HB\beta D_{53}\beta'BHw]/n^4$$
$$+ \Omega_\alpha^2[w'HB\beta D_{32}\beta'BHw]^3/n^6\Big]$$
$$= \frac{4\Omega_\alpha^4(k_\epsilon-1)}{n}(P_{74} - 2\Omega_\alpha P_{32}P_{53} + \Omega_\alpha^2 P_{32}^3).$$

– $E(n\mathcal{T}_4^2)$

Since we don't observe $f(x_t)$ we cannot directly estimate $\sigma_{\epsilon u}$. However, $\sigma_{\epsilon u} = E(w_t\epsilon_t)$. We can therefore use the sample mean $\sum_t(\epsilon_t w_t)/n$. We first need to compute

$$\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G = \overline{G_5}K_{\alpha 56}^{-1}(K_{68}K_{9,10} + K_{69}K_{8,10} + K_{6,10}K_{89})K_{\alpha 87}^{-1}G_7$$
$$= [\overline{G}KK_\alpha^{-1}KG]K_{9,10} + [\overline{G}K_\alpha^{-1}K]_9[KK_\alpha^{-1}G]_{10}$$
$$+ [\overline{G}KK_\alpha^{-1}KG]K_{10,9}$$

We can therefore write $E(n\mathcal{T}_4^2)$ as :

$$
\begin{aligned}
E(n\mathcal{T}_4^2) =& \frac{16\Omega_\alpha^4\sigma_{\epsilon u}^2}{\sigma_\epsilon^4 n}[\overline{G}K_\alpha^{-1}][K\Sigma][\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G][\Sigma K][K_\alpha^{-1}G] \\
=& \frac{16\Omega_\alpha^4\sigma_{\epsilon u}^2}{\sigma_\epsilon^4 n}\Big[2[\overline{G}KK_\alpha^{-1}KG][\overline{G}K_\alpha^{-1}K\Sigma K\Sigma KK_\alpha^{-1}G] + [\overline{G}K_\alpha^{-1}K\Sigma KK_\alpha^{-1}G]^2\Big] \\
=& \frac{16\Omega_\alpha^4\sigma_{\epsilon u}^2}{\sigma_\epsilon^4 n}\Big[2w'HB\beta D_{31}\beta'BHw/n^2\Big\{w'HB\beta D_{74}\beta'BHw/n^2 \\
& - 2\Omega_\alpha[w'HB\beta D_{32}\beta'BHw][w'HB\beta D_{53}\beta'BHw]/n^4 \\
& + \Omega_\alpha^2[w'HB\beta D_{32}\beta'BHw]^3/n^6\Big\} \\
& + [w'HB\beta D_{53}\beta'BHw/n^2 - \Omega_\alpha(w'HB\beta D_{32}\beta'BHw)^2/n^4]^2\Big] \\
=& \frac{16\Omega_\alpha^4\sigma_{\epsilon u}^2}{\sigma_\epsilon^4 n}\Big[2P_{31}(P_{74} - 2\Omega_\alpha P_{32}P_{53} + \Omega_\alpha^2 P_{32}^3) + (P_{53} - \Omega_\alpha P_{32}^2)^2\Big]
\end{aligned}
$$

– $E(n\mathcal{T}_5^2)$

$$
\begin{aligned}
E(n\mathcal{T}_5^2) =& \frac{\rho_3^2\Omega_\alpha^4}{n}\overline{G}(K^2+\alpha I)^{-1}SG\Sigma[\overline{G}K_\alpha^{-1}\Delta K_\alpha^{-1}G]\Sigma\overline{G}S(K^2+\alpha I)^{-1}G \\
=& \frac{\rho_3^2\Omega_\alpha^4 S_\epsilon^2}{n^3}\sum_{t,s=1}^{n} P_{t01}P_{s01}[H_{t\bullet}w][H_{s\bullet}w][\overline{G}K_\alpha^{-1}]_6[\Sigma Z_s]_8\Big[K_{67}K_{8,10} \\
& + K_{68}K_{7,10} + K_{6,10}K_{78}\Big][\overline{Z_t}\Sigma]_7[K_\alpha^{-1}G]_{10} \\
=& \frac{\rho_3^2\Omega_\alpha^4 S_\epsilon^2}{n^3}\sum_{t,s=1}^{n} P_{t01}P_{s01}[H_{t\bullet}w][H_{s\bullet}w]\Big[2[\overline{G}K_\alpha^{-1}K\Sigma Z_t][\overline{G}K_\alpha^{-1}K\Sigma Z_s] \\
& + [\overline{G}K_\alpha^{-1}KK_\alpha^{-1}G][\overline{Z_t}\Sigma K\Sigma Z_s]\Big]
\end{aligned}
$$

with

$$[\overline{G}K_\alpha^{-1}K\Sigma Z_t] = (P_{t32} - \Omega_\alpha P_{32}P_{t11})$$

$$[\overline{G}K_\alpha^{-1}KK_\alpha^{-1}G] = P_{32},$$

and

$$\begin{aligned}\overline{Z_t}\Sigma K \Sigma Z_s =& H_{t\bullet} B\beta D_{32}\beta' B H_{s\bullet} \\ & - \Omega_\alpha [H_{t\bullet} B\beta D_{32}\beta' BHw][H_{s\bullet} B\beta D_{11}\beta' BHw]/n^2 \\ & - \Omega_\alpha [H_{t\bullet} B\beta D_{11}\beta' BHw][H_{s\bullet} B\beta D_{32}\beta' BHw]/n^2 \\ & + \Omega_\alpha^2 [H_{t\bullet} B\beta D_{11}\beta' BHw][w' HB\beta D_{32}\beta' BHw] \\ & [H_{s\bullet} B\beta D_{11}\beta' BHw]/n^4 \\ =& P_{ts32} - \Omega_\alpha (P_{t32}P_{s11} + P_{t11}P_{s32}) + \Omega_\alpha^2 P_{t11}P_{32}P_{s11}.\end{aligned} \tag{C.30}$$

It follows that

$$\begin{aligned}E(n\mathcal{T}_5^2) =& \frac{\rho_3^2 \Omega_\alpha^4 S_\epsilon^2}{n^3} \sum_{t,s=1}^n P_{t01}P_{s01}[H_{t\bullet}w][H_{s\bullet}w]\Big[2(P_{t32} - \Omega_\alpha P_{32}P_{t11})(P_{s32} - \Omega_\alpha P_{32}P_{s11}) \\ & + P_{32}(P_{ts32} - \Omega_\alpha (P_{t32}P_{s11} + P_{t11}P_{s32}) + \Omega_\alpha^2 P_{t11}P_{32}P_{s11})\Big].\end{aligned}$$

– $E(n\mathcal{T}_6^2)$

$$\begin{aligned}E(n\mathcal{T}_6^2) =& \frac{9\rho_3^2\Omega_\alpha^2 S_\epsilon}{4n^3} \sum_{t,s=1}^n P_{t01}P_{t02}[\overline{Z_t}K\Sigma]_6[\overline{Z_t}\Sigma]_7 \Delta_{678,10}[\Sigma Z_s]_8[\Sigma K Z_s]_{10} \\ =& \frac{9\rho_3^2\Omega_\alpha^2 S_\epsilon}{4n^3} \sum_{t,s=1}^n P_{t01}P_{t02}[\overline{Z_t}K\Sigma]_6[\overline{Z_t}\Sigma]_7 \Big[K_{67}K_{8,10} \\ & + K_{68}K_{7,10} + K_{6,10}K_{78}\Big][\Sigma Z_s]_8[\Sigma K Z_s]_{10} \\ =& \frac{9\rho_3^2\Omega_\alpha^2 S_\epsilon}{4n^3} \sum_{t,s=1}^n P_{t01}P_{t02}\Big[2[\overline{Z_t}K\Sigma K\Sigma Z_t][\overline{Z_s}K\Sigma K\Sigma Z_s] \\ & + [\overline{Z_t}K\Sigma K\Sigma K Z_s][\overline{Z_t}\Sigma K\Sigma Z_s]\Big],\end{aligned}$$

with

$$[\overline{Z_t}K\Sigma K\Sigma Z_s] = P_{ts42} - \Omega_\alpha (P_{t42}P_{s11} + P_{t21}P_{s32}) + \Omega_\alpha^2 P_{t21}P_{32}P_{s11},$$

$$\begin{aligned}[\overline{Z_t}K\Sigma K\Sigma KZ_s] =&H_{t\bullet}B\beta D_{52}\beta' BH_{s\bullet}\\ &- \Omega_\alpha[H_{t\bullet}B\beta D_{42}\beta' BHw][H_{s\bullet}B\beta D_{21}\beta' BHw]/n^2\\ &- \Omega_\alpha[H_{t\bullet}B\beta D_{21}\beta' BHw][H_{s\bullet}B\beta D_{42}\beta' BHw]/n^2\\ &+ \Omega_\alpha^2[H_{t\bullet}B\beta D_{21}\beta' BHw][w' HB\beta D_{32}\beta' BHw]\\ &[H_{s\bullet}B\beta D_{21}\beta' BHw]/n^4\\ =&P_{ts52} - \Omega_\alpha(P_{t42}P_{s21} + P_{t21}P_{s42}) + \Omega_\alpha^2 P_{t21}P_{s21}P_{32},\end{aligned}$$

and $[\overline{Z_t}\Sigma K\Sigma Z_s]$ given by the equation (C.30).

– $E(n\mathcal{T}_1\mathcal{T}_2)$

$$\begin{aligned}K_{13}\Sigma_{45}S_{652}S_{346} =&\frac{S_\epsilon^2}{n^2}\sum_{t,s=1}^n Z_{t6}Z_{t5}Z_{t2}Z_{s3}Z_{s4}Z_{s6}K_{13}\Sigma_{45}\\ =&\frac{S_\epsilon^2}{n^2}\sum_{t,s=1}^n [KZ_s]_1[\overline{Z_s}\Sigma Z_t][nH_{st}]\overline{Z_{t2}}\end{aligned}$$

which implies

$$\begin{aligned}E(n\mathcal{T}_1\mathcal{T}_2) =&\frac{\rho_3\Omega_\alpha^2 S_\epsilon^2}{n^2}\sum_{t,s=1}^n [\overline{G}K_\alpha^{-1}Z_s][\overline{Z_s}\Sigma Z_t][H_{st}][\overline{Z_t}(K^2+\alpha I)^{-1}G]\\ =& \frac{\rho_3\Omega_\alpha^2 S_\epsilon^2}{n^2}\sum_{t,s=1}^n P_{s11}H_{st}P_{t01}(P_{ts11} - \Omega_\alpha P_{t11}P_{s11})\end{aligned}$$

– $E(n\mathcal{T}_1\mathcal{T}_3)$

$$\begin{aligned}E(n\mathcal{T}_1\mathcal{T}_3) =&\frac{3(k_\epsilon - 1)\Omega_\alpha^2}{n}\overline{G}K_\alpha^{-1}K\Sigma KK_\alpha^{-1}G\\ =&\frac{3(k_\epsilon - 1)\Omega_\alpha^2}{n}(P_{53} - \Omega_\alpha P_{32}^2)\end{aligned}$$

– $E(n\mathcal{T}_1\mathcal{T}_4)$

$$\begin{aligned}\Xi_{14} =& -\frac{4\sigma_{\epsilon u}\Omega_\alpha}{\sigma_\epsilon^2 n}K_{13}S_{346}K_{\alpha 67}^{-1}G_7\Sigma_{45}K_{58}K_{82}\\ =& -\frac{4\sigma_{\epsilon u}\Omega_\alpha S_\epsilon}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n [KZ_t]_1[\overline{Z_t}K_\alpha^{-1}G][\overline{Z_t}\Sigma KK]_2\end{aligned}$$

$$E(n\mathcal{T}_1\mathcal{T}_4) = -\frac{4\sigma_{\epsilon u}\Omega_\alpha^3 S_\epsilon}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n [\overline{G}K_\alpha^{-1}Z_t]^2[\overline{Z_t}\Sigma KK_\alpha^{-1}G]$$
$$= -\frac{4\sigma_{\epsilon u}\Omega_\alpha^3 S_\epsilon}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n P_{t11}^2(P_{t32} - \Omega_\alpha P_{t11}P_{32})$$

– $E(n\mathcal{T}_1\mathcal{T}_5)$

$$\Xi_{15} = -\frac{\rho_3\Omega_\alpha}{n}K_{13}S_{346}K_{\alpha 67}^{-1}G_7\Sigma_{45}\overline{G_8}S_{852}$$
$$= -\frac{\rho_3\Omega_\alpha S_\epsilon^2}{n^3}\sum_{t,s=1}^n [KZ_t]_1[\overline{Z_t}K_\alpha^{-1}G][\overline{Z_t}\Sigma Z_s][\overline{G}Z_s]Z_{s2}$$

$$E(n\mathcal{T}_1\mathcal{T}_5) = -\frac{\rho_3\Omega_\alpha^3 S_\epsilon^2}{n^3}\sum_{t,s=1}^n [\overline{G}K_\alpha^{-1}Z_t][\overline{Z_t}K_\alpha^{-1}G][\overline{Z_t}\Sigma Z_s][\overline{G}Z_s][Z_s(K^2+\alpha I)^{-1}G]$$
$$= \frac{\rho_3\Omega_\alpha^3 S_\epsilon^2}{n^3}\sum_{t,s=1}^n P_{t11}^2 P_{s01}(P_{ts11} - \Omega_\alpha P_{t11}P_{s11})[w'H_{\bullet s}]$$

– $E(n\mathcal{T}_1\mathcal{T}_6)$

$$\Xi_{16} = -\frac{3\rho_3}{2n}K_{13}\Sigma_{45}\Sigma_{67}K_{78}S_{852}S_{346}$$
$$= -\frac{3\rho_3 S_\epsilon^2}{2n^3}\sum_{t,s=1}^n [KZ_t]_1[\overline{Z_t}\Sigma Z_s][\overline{Z_t}\Sigma KZ_s]Z_{s2}$$

$$E(n\mathcal{T}_1\mathcal{T}_6) = -\frac{3\rho_3\Omega_\alpha^2}{2n^3}\sum_{t,s=1}^n [\overline{G}K_\alpha^{-1}Z_t][\overline{Z_t}\Sigma Z_s][\overline{Z_t}\Sigma KZ_s][\overline{Z_s}(K^2+\alpha I)^{-1}G]$$
$$= -\frac{3\rho_3\Omega_\alpha^2}{2n^3}\sum_{t,s=1}^n P_{t11}P_{s01}(P_{ts11} - \Omega_\alpha P_{t11}P_{s11})(P_{ts21} - \Omega_\alpha P_{t11}P_{s21})$$

– $E(n\mathcal{T}_2\mathcal{T}_3)$

$$\Xi_{23}(\tau_1,\tau_2) = \frac{2\rho_3 S_\epsilon}{\sigma_\epsilon^2 n}S_{134}\Sigma_{35}\Sigma_{67}K_{78}K_{82}\Psi_{456}$$
$$= \frac{2\rho_3 S_\epsilon^2}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n Z_t(\tau_1)[\overline{Z_t}\Sigma]_5\overline{Z_{t4}}[K_{45}\bar{Z}_6 + K_{46}\bar{Z}_5 + K_{56}\bar{Z}_4][\Sigma KK]_{62}$$
$$= \frac{2\rho_3 S_\epsilon^2}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n Z_t(\tau_1)\Big[[\overline{Z_t}K\Sigma Z_t][\overline{Z}\Sigma KK]_2 + [\overline{Z_t}\Sigma\bar{Z}][\overline{Z_t}K\Sigma KK]_2 + [\iota' H_{\bullet t}][\overline{Z_t}\Sigma K\Sigma KK]_2\Big],$$

where $\bar{Z}$ without subscript t is the sample mean $\sum_{t=1}^n Z_t/n$ and ι is a vector of ones. Therefore,

$$E(n\mathcal{T}_2\mathcal{T}_3) = \frac{2\rho_3 S_\epsilon^2 \Omega_\alpha^2}{\sigma_\epsilon^2 n^2} \sum_{t=1}^{n} P_{t01}\Big[[\overline{Z_t}K\Sigma Z_t][\overline{Z}\Sigma K K_\alpha^{-1}G] + [\overline{Z_t}\Sigma\bar{Z}][\overline{Z_t}K\Sigma K K_\alpha^{-1}G] + [\iota' H_{\bullet t}][\overline{Z_t}\Sigma K\Sigma K K_\alpha^{-1}G]\Big],$$

with

$$[\overline{Z_t}K\Sigma Z_t] = P_{tt21} - \Omega_\alpha P_{t21}P_{t11},$$

$$[\overline{Z}\Sigma K K_\alpha^{-1}G] = P_{\iota 22} - \Omega_\alpha P_{\iota 11}P_{22},$$

$$[\overline{Z_t}K\Sigma K K_\alpha^{-1}G] = P_{t42} - \Omega_\alpha P_{t21}P_{32},$$

where $P_{\iota ab} = -\iota' H B\beta D_{ab}\beta' B H w/n^2$,

$$\overline{Z_t}\Sigma\bar{Z} = P_{\iota t11} - \Omega_\alpha P_{t11}P_{\iota 11}$$

where $P_{\iota tab} = \iota' H B\beta D_{ab}\beta' B H_{\bullet t}/n$, and

$$[\overline{Z_t}\Sigma K\Sigma K K_\alpha^{-1}G] = P_{t53} - \Omega_\alpha(P_{t32}P_{32} + P_{t11}P_{53}) + \Omega^2 P_{t11}P_{32}^2.$$

– $E(n\mathcal{T}_2\mathcal{T}_4)$

$$\begin{aligned}
\Xi_{24} =& -\frac{4\rho_3\Omega_\alpha\sigma_{\epsilon u}}{n\sigma_\epsilon^2} S_{134}\Sigma_{35}\Sigma_{67}K_{\alpha 89}^{-1}G_9 K_{7,10}K_{10,2}\Delta_{4568} \\
=& -\frac{4\rho_3\Omega_\alpha\sigma_{\epsilon u}S_\epsilon}{n^2\sigma_\epsilon^2}\sum_{t=1}^{n} Z_t(\tau_1)Z_{t4}[\overline{Z_t}\Sigma]_5[K_\alpha^{-1}G]_8[K_{45}K_{68} + K_{46}K_{58} + K_{48}K_{56}][\Sigma K K]_{62} \\
=& -\frac{4\rho_3\Omega_\alpha\sigma_{\epsilon u}S_\epsilon}{n^2\sigma_\epsilon^2}\sum_{t=1}^{n}\Big[[\overline{Z_t}K\Sigma Z_t][\overline{G}K_\alpha^{-1}K\Sigma K K]_2 + [\overline{Z_t}\Sigma K K_\alpha^{-1}G][\overline{Z_t}K\Sigma K K]_2 \\
& + [\overline{Z_t}K K_\alpha^{-1}G][\overline{Z_t}\Sigma K\Sigma K K]_2\Big]
\end{aligned}$$

$$E(n\mathcal{T}_2\mathcal{T}_4) = -\frac{4\rho_3\Omega_\alpha^3\sigma_{\epsilon u}S_\epsilon}{n^2\sigma_\epsilon^2}\sum_{t=1}^{n}\Big[(P_{tt21}-\Omega_\alpha P_{t21}P_{t11})[\overline{G}K_\alpha^{-1}K\Sigma K_\alpha^{-1}G]$$

$$+ (P_{t32}-\Omega_\alpha P_{t11}P_{32})[\overline{Z_t}K\Sigma K_\alpha^{-1}G]$$

$$+ P_{t21}[\overline{Z_t}\Sigma K\Sigma K_\alpha^{-1}G]\Big]$$

$$= -\frac{4\rho_3\Omega_\alpha^3\sigma_{\epsilon u}S_\epsilon}{n^2\sigma_\epsilon^2}\sum_{t=1}^{n}\Big[(P_{tt21}-\Omega_\alpha P_{t21}P_{t11})(P_{43}-\Omega_\alpha P_{32}P_{22})$$

$$+ (P_{t32}-\Omega_\alpha P_{t11}P_{32})(P_{t32}-\Omega_\alpha P_{t21}P_{22})$$

$$+ P_{t21}(P_{t43}-\Omega_\alpha(P_{t32}P_{22}+P_{t11}P_{43})+\Omega^2 P_{t11}P_{32}P_{22})\Big]$$

– $E(n\mathcal{T}_2\mathcal{T}_5)$

$$\Xi_{25} = -\frac{\rho_3^2\Omega_\alpha}{n}S_{134}\Sigma_{35}\Sigma_{67}K_{\alpha 89}^{-1}G_9\overline{G_{10}}S_{10,72}\Delta_{4568}$$

$$= -\frac{\rho_3^2\Omega_\alpha S_\epsilon^2}{n^3}\sum_{t,s=1}^{n}Z_t(\tau_1)\overline{Z_s(\tau_2)}[\overline{G}Z_s][\Sigma Z_s]_6[K_\alpha^{-1}G]_8[\overline{Z_t}\Sigma]_5 Z_{t4}[K_{45}K_{68}+K_{46}K_{58}+K_{48}K_{56}]$$

$$= -\frac{\rho_3^2\Omega_\alpha S_\epsilon^2}{n^3}\sum_{t,s=1}^{n}Z_t(\tau_1)\overline{Z_s(\tau_2)}[-w'H_{\bullet s}]\Big[[\overline{Z_t}K\Sigma Z_t][\overline{Z_s}\Sigma KK_\alpha^{-1}G]$$

$$+ [\overline{Z_t}K\Sigma Z_s][\overline{Z_t}\Sigma KK_\alpha^{-1}G]+[\overline{Z_t}KK_\alpha^{-1}G][\overline{Z_t}\Sigma K\Sigma Z_s]\Big]$$

$$E(n\mathcal{T}_2\mathcal{T}_5) = \frac{\rho_3^2\Omega_\alpha^3 S_\epsilon^2}{n^3}\sum_{t,s=1}^{n}P_{t01}P_{s01}[w'H_{\bullet s}]\Big[[\overline{Z_t}K\Sigma Z_t][\overline{Z_s}\Sigma KK_\alpha^{-1}G]$$

$$+ [\overline{Z_t}K\Sigma Z_s][\overline{Z_t}\Sigma KK_\alpha^{-1}G]+[\overline{Z_t}KK_\alpha^{-1}G][\overline{Z_t}\Sigma K\Sigma Z_s]$$

$$= \frac{\rho_3^2\Omega_\alpha^3 S_\epsilon^2}{n^3}\sum_{t,s=1}^{n}P_{t01}P_{s01}[w'H_{\bullet s}]\Big[(P_{tt21}-\Omega_\alpha P_{t21}P_{t11})(P_{s32}-\Omega_\alpha P_{s11}P_{32})$$

$$+ (P_{ts21}-\Omega_\alpha P_{t21}P_{s11})(P_{t32}-\Omega_\alpha P_{t11}P_{32})+$$

$$+ P_{t21}(P_{ts32}-\Omega_\alpha(P_{t32}P_{s11}+P_{t11}P_{s32})+\Omega^2 P_{t11}P_{32}P_{s11})\Big]$$

– $E(n\mathcal{T}_2\mathcal{T}_6)$

$$\begin{aligned}
\Xi_{26} &= -\frac{3\rho_3^2}{2n} S_{134}\Sigma_{35}\Sigma_{67}\Sigma_{89}K_{9,10}S_{10,72}\Delta_{4568} \\
&= -\frac{3\rho_3^2 S_\epsilon^2}{2n^3}\sum_{t,s=1}^{n} Z_t(\tau_1)\overline{Z_s(\tau_2)}Z_{t4}[\Sigma K Z_s]_8[\Sigma Z_s]_6[\overline{Z_t}\Sigma]_5[K_{45}K_{68}+K_{46}K_{58}+K_{48}K_{56}] \\
&= -\frac{3\rho_3^2 S_\epsilon^2}{2n^3}\sum_{t,s=1}^{n} Z_t(\tau_1)\overline{Z_s(\tau_2)}\Big[[\overline{Z_t}\Sigma K Z_t][\overline{Z_s}\Sigma K\Sigma K Z_s] \\
&\qquad + [\overline{Z_t}K\Sigma Z_s][\overline{Z_t}\Sigma K\Sigma K Z_s] + [\overline{Z_t}K\Sigma K Z_s][\overline{Z_t}\Sigma K\Sigma Z_s]
\end{aligned}$$

$$\begin{aligned}
E(n\mathcal{T}_2\mathcal{T}_6) = -\frac{3\rho_3^2\Omega_\alpha^2 S_\epsilon^2}{2n^3}\sum_{t,s=1}^{n} P_{t01}P_{s01}\Big[[\overline{Z_t}\Sigma K Z_t][\overline{Z_s}\Sigma K\Sigma K Z_s] \\
+ [\overline{Z_t}K\Sigma Z_s][\overline{Z_t}\Sigma K\Sigma K Z_s] + [\overline{Z_t}K\Sigma K Z_s][\overline{Z_t}\Sigma K\Sigma Z_s]
\end{aligned}$$

with

$$[\overline{Z_t}\Sigma K Z_t] = (P_{tt21} - \Omega_\alpha P_{t11}P_{t21})$$

$$[\overline{Z_t}K\Sigma Z_s] = (P_{ts21} - \Omega_\alpha P_{t21}P_{s11})$$

$$[\overline{Z_t}\Sigma K\Sigma K Z_s] = (P_{ts42} - \Omega_\alpha(P_{t32}P_{s21} + P_{t11}P_{s42}) + \Omega^2 P_{t11}P_{32}P_{s21})$$

$$[\overline{Z_t}K\Sigma K Z_s] = (P_{ts31} - \Omega_\alpha P_{t21}P_{s21})$$

$$[\overline{Z_t}\Sigma K\Sigma Z_s] = (P_{ts32} - \Omega_\alpha(P_{t32}P_{s11} + P_{t11}P_{s32}) + \Omega^2 P_{t11}P_{32}P_{s11})$$

– $E(n\mathcal{T}_3\mathcal{T}_4)$

$$\begin{aligned}
\Xi_{34} &= -\frac{8\sigma_{\epsilon u}\Omega_\alpha S_\epsilon}{n\sigma_\epsilon^4} K_{13}K_{34}\Sigma_{45}\Sigma_{67}K^{-1}_{\alpha 89}G_9K_{7,10}K_{10,2}\Psi_{568} \\
&= -\frac{8\sigma_{\epsilon u}\Omega_\alpha S_\epsilon}{n\sigma_\epsilon^4}[KK\Sigma]_{15}[K_\alpha^{-1}G]_8[\Sigma KK]_{6,2}[K_{56}\bar{Z}_8 + K_{58}\bar{Z}_6 + K_{68}\bar{Z}_5] \\
&= -\frac{8\sigma_{\epsilon u}\Omega_\alpha S_\epsilon}{n\sigma_\epsilon^4}\Big[[\overline{Z}K_\alpha^{-1}G][KK\Sigma K\Sigma KK]_{12} + [KK\Sigma KK_\alpha^{-1}G]_1[\overline{Z}\Sigma KK]_2 \\
&\qquad + [KK\Sigma\bar{Z}]_1[\overline{G}K_\alpha^{-1}K\Sigma KK]_2\Big]
\end{aligned}$$

$$E(n\mathcal{T}_3\mathcal{T}_4) = -\frac{8\sigma_{\epsilon u}\Omega_\alpha^3 S_\epsilon}{n\sigma_\epsilon^4}\Big[[\overline{Z}K_\alpha^{-1}G][\overline{G}K_\alpha^{-1}K\Sigma K\Sigma KK_\alpha^{-1}G] + [\overline{G}K_\alpha^{-1}K\Sigma KK_\alpha^{-1}G][\overline{Z}\Sigma KK_\alpha^{-1}G] + [\overline{G}K_\alpha^{-1}K\Sigma\bar{Z}][\overline{G}K_\alpha^{-1}K\Sigma KK_\alpha^{-1}G]\Big]$$

$$= -\frac{8\sigma_{\epsilon u}\Omega_\alpha^3 S_\epsilon}{n\sigma_\epsilon^4}\Big[P_{\iota 11}(P_{74} - \Omega_\alpha(P_{53}P_{32} + P_{32}P_{53}) + \Omega_\alpha^2 P_{32}^3) + 2(P_{53} - \Omega_\alpha P_{32}^2)(P_{\iota 32} - \Omega_\alpha P_{\iota 11}P_{32})\Big]$$

– $E(n\mathcal{T}_3\mathcal{T}_5)$

$$\Xi_{35} = -\frac{2\rho_3 S_\epsilon \Omega_\alpha}{\sigma_\epsilon^2 n} K_{13}K_{34}\Sigma_{45}\Sigma_{67}K_{\alpha 89}^{-1}G_9\overline{G_{10}}S_{10,72}\Psi_{568}$$

$$= -\frac{2\rho_3 S_\epsilon^2 \Omega_\alpha}{\sigma_\epsilon^2 n^2}\sum_{t=1}^{n}[KK\Sigma]_{15}[\overline{G}Z_t][\Sigma Z_t]_6[K_\alpha^{-1}G]_8\overline{Z_t(\tau_2)}[K_{56}\bar{Z}_8 + K_{58}\bar{Z}_6 + K_{68}\bar{Z}_5]$$

$$= \frac{2\rho_3 S_\epsilon^2 \Omega_\alpha}{\sigma_\epsilon^2 n^2}\sum_{t=1}^{n}[w'H_{\bullet t}]\overline{Z_t(\tau_2)}\Big[[\bar{Z}K_\alpha^{-1}G][KK\Sigma K\Sigma Z_t]_1 + [\bar{Z}\Sigma Z_t][KK\Sigma KK_\alpha^{-1}G]_1 + [\overline{Z_t}\Sigma KK_\alpha^{-1}G][KK\Sigma\bar{Z}]_1\Big]$$

$$E(n\mathcal{T}_3\mathcal{T}_5) = \frac{2\rho_3 S_\epsilon^2 \Omega_\alpha^3}{\sigma_\epsilon^2 n^2}\sum_{t=1}^{n}[w'H_{\bullet t}]P_{t01}\Big[P_{\iota 11}[\overline{G}K_\alpha^{-1}K\Sigma K\Sigma Z_t] + (P_{\iota t11} - \Omega_\alpha P_{\iota 11}P_{t11})[\overline{G}K_\alpha^{-1}K\Sigma KK_\alpha^{-1}G] + [\overline{Z_t}\Sigma KK_\alpha^{-1}G][\overline{G}K_\alpha^{-1}K\Sigma\bar{Z}]\Big]$$

$$= \frac{2\rho_3 S_\epsilon^2 \Omega_\alpha^3}{\sigma_\epsilon^2 n^2}\sum_{t=1}^{n}[w'H_{\bullet t}]P_{t01}\Big[P_{\iota 11}(P_{t53} - \Omega_\alpha(P_{53}P_{t11} + P_{32}P_{t32}) + \Omega_\alpha^2 P_{32}^2 P_{t11}) + (P_{\iota t11} - \Omega_\alpha P_{\iota 11}P_{t11})(P_{53} - \Omega_\alpha P_{32}^2) + (P_{t32} - \Omega_\alpha P_{t11}P_{32})(P_{\iota 32} - \Omega_\alpha P_{\iota 11}P_{32})\Big]$$

– $E(n\mathcal{T}_3\mathcal{T}_6)$

$$\Xi_{36} = -\frac{3\rho_3 S_\epsilon}{\sigma_\epsilon^2 n} K_{13} K_{34} \Sigma_{45} \Sigma_{67} \Sigma_{89} K_{9,10} S_{10,72} \Psi_{568}$$
$$= -\frac{3\rho_3 S_\epsilon^2}{\sigma_\epsilon^2 n^2} \sum_{t=1}^{n} \overline{Z_t(\tau_2)} [KK\Sigma]_{15} [\Sigma Z_t]_6 [\Sigma K Z_t]_8 [K_{56} \bar{Z}_8 + K_{58} \bar{Z}_6 + K_{68} \bar{Z}_5]$$
$$= -\frac{3\rho_3 S_\epsilon^2}{\sigma_\epsilon^2 n^2} \sum_{t=1}^{n} \overline{Z_t(\tau_2)} \Big[[\bar{Z} \Sigma K Z_t][KK\Sigma K \Sigma Z_t]_1$$
$$+ [\bar{Z}\Sigma Z_t][KK\Sigma K\Sigma K Z_t]_1 + [\overline{Z_t}\Sigma K\Sigma K Z_t][KK\Sigma \bar{Z}]_1 \Big]$$

$$E(n\mathcal{T}_3\mathcal{T}_6) = -\frac{3\rho_3 S_\epsilon^2 \Omega_\alpha^2}{\sigma_\epsilon^2 n^2} \sum_{t=1}^{n} P_{t01} \Big[(P_{\iota t21} - \Omega_\alpha P_{\iota 11} P_{t21}) [\overline{G} K_\alpha^{-1} K\Sigma K \Sigma Z_t]$$
$$+ (P_{\iota t11} - \Omega_\alpha P_{\iota 11} P_{t11}) [\overline{G} K_\alpha^{-1} K \Sigma K \Sigma K Z_t]$$
$$+ (P_{\iota 32} - \Omega_\alpha P_{\iota 11} P_{32}) [\overline{Z_t} \Sigma K \Sigma K Z_t] \Big]$$

with

$$[\overline{G} K_\alpha^{-1} K\Sigma K\Sigma Z_t] = P_{t53} - \Omega_\alpha (P_{53} P_{t11} + P_{32} P_{t32}) + \Omega_\alpha^2 P_{32}^2 P_{t11}$$

$$[\overline{G} K_\alpha^{-1} K\Sigma K\Sigma K Z_t] = P_{t63} - \Omega_\alpha (P_{53} P_{t21} + P_{32} P_{t42}) + \Omega_\alpha^2 P_{32}^2 P_{t21}$$

$$[\overline{Z_t} \Sigma K \Sigma K Z_t] = P_{tt42} - \Omega_\alpha (P_{t32} P_{t21} + P_{t11} P_{t42}) + \Omega_\alpha^2 P_{t11} P_{32} P_{t21}$$

– $E(n\mathcal{T}_4\mathcal{T}_5)$

$$\Xi_{45} = \frac{4\sigma_{\epsilon u}\rho_3 \Omega_\alpha^2}{\sigma_\epsilon^2 n} K_{13} K_{34} \overline{G_5} K_{\alpha 56}^{-1} \Sigma_{47} \Sigma_{89} K_{\alpha 10,11}^{-1} G_{11} \overline{G_{12}} S_{12,92} \Delta_{678,10}$$
$$= \frac{4\sigma_{\epsilon u}\rho_3 \Omega_\alpha^2 S_\epsilon}{\sigma_\epsilon^2 n^2} \sum_{t=1}^{n} \overline{Z_t(\tau_2)} [\overline{G} Z_t][KK\Sigma]_{17} [\overline{G} K_\alpha^{-1}]_6 [\Sigma Z_t]_8 [K_\alpha^{-1} G]_{10} [K_{67} K_{8,10}$$
$$+ K_{68} K_{7,10} + K_{6,10} K_{78}]$$
$$= -\frac{4\sigma_{\epsilon u}\rho_3 \Omega_\alpha^2 S_\epsilon}{\sigma_\epsilon^2 n^2} \sum_{t=1}^{n} \overline{Z_t(\tau_2)} [w' H_{\bullet t}] \Big[2[\overline{Z_t} \Sigma K K_\alpha^{-1} G][KK\Sigma K K_\alpha^{-1} G]_1$$
$$+ [\overline{G} K_\alpha^{-1} K K_\alpha^{-1} G][KKK\Sigma K \Sigma Z_t]_1 \Big]$$

$$E(n\mathcal{T}_4\mathcal{T}_5) = -\frac{4\sigma_{\epsilon u}\rho_3\Omega_\alpha^4 S_\epsilon}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n P_{t01}[w'H_{\bullet t}]\Big[2[\overline{Z_t}\Sigma KK_\alpha^{-1}G][\overline{G}K_\alpha^{-1}K\Sigma KK_\alpha^{-1}G]$$
$$+ [\overline{G}K_\alpha^{-1}KK_\alpha^{-1}G][\overline{G}K_\alpha^{-1}KK\Sigma K\Sigma Z_t]\Big]$$
$$= -\frac{4\sigma_{\epsilon u}\rho_3\Omega_\alpha^4 S_\epsilon}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n P_{t01}[w'H_{\bullet t}]\Big[2(P_{t32}-\Omega_\alpha P_{t11}P_{32})(P_{53}-\Omega_\alpha P_{32}^2)$$
$$+ P_{32}(P_{t63}-\Omega_\alpha(P_{63}P_{t11}+P_{42}P_{t32})+\Omega_\alpha^2 P_{42}P_{32}P_{t11})\Big]$$

– $E(n\mathcal{T}_4\mathcal{T}_6)$

$$\Xi_{46} = \frac{6\sigma_{\epsilon u}\rho_3\Omega_\alpha}{\sigma_\epsilon^2 n}K_{13}K_{34}\overline{G_5}K_{\alpha 56}^{-1}\Sigma_{47}\Sigma_{89}\Sigma_{10,11}K_{11,12}S_{12,92}\Delta_{678,10}$$
$$= \frac{6\sigma_{\epsilon u}\rho_3\Omega_\alpha S_\epsilon}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n \overline{Z_t(\tau_2)}[KK\Sigma]_{17}[\overline{G}K_\alpha^{-1}]_6[\Sigma KZ_t]_{10}[\Sigma Z_t]_8[K_{6,7}K_{8,10}$$
$$+ K_{68}K_{7,10} + K_{6,10}K_{78}]$$
$$= \frac{6\sigma_{\epsilon u}\rho_3\Omega_\alpha S_\epsilon}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n \overline{Z_t(\tau_2)}\Big[[\overline{Z}_t K\Sigma K\Sigma Z_t][KK\Sigma KK_\alpha^{-1}G]_1$$
$$+ [\overline{G}K_\alpha^{-1}K\Sigma Z_t][KK\Sigma K\Sigma KZ_t]_1$$
$$+ [\overline{G}K_\alpha^{-1}K\Sigma KZ_t][KK\Sigma K\Sigma Z_t]_1\Big]$$

$$E(n\mathcal{T}_4\mathcal{T}_6) = \frac{6\sigma_{\epsilon u}\rho_3\Omega_\alpha^3 S_\epsilon}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n P_{t01}\Big[[\overline{Z}_t K\Sigma K\Sigma Z_t][\overline{G}K_\alpha^{-1}K\Sigma KK_\alpha^{-1}G]$$
$$+ [\overline{G}K_\alpha^{-1}K\Sigma Z_t][\overline{G}K_\alpha^{-1}K\Sigma K\Sigma KZ_t]$$
$$+ [\overline{G}K_\alpha^{-1}K\Sigma KZ_t][\overline{G}K_\alpha^{-1}K\Sigma K\Sigma Z_t]\Big]$$
$$= \frac{6\sigma_{\epsilon u}\rho_3\Omega_\alpha^3 S_\epsilon}{\sigma_\epsilon^2 n^2}\sum_{t=1}^n P_{t01}\Big[(P_{53}-\Omega_\alpha P_{32}^2)[\overline{Z}_t K\Sigma K\Sigma Z_t]$$
$$+ (P_{t32}-\Omega_\alpha P_{32}P_{t11})[\overline{G}K_\alpha^{-1}K\Sigma K\Sigma KZ_t]$$
$$+ (P_{t42}-\Omega_\alpha P_{32}P_{t21})[\overline{G}K_\alpha^{-1}K\Sigma K\Sigma Z_t]\Big]$$

with

$$[\overline{Z}_t K\Sigma K\Sigma Z_t] = P_{tt42} - \Omega_\alpha(P_{t32}P_{t21}+P_{t11}P_{t42}) + \Omega_\alpha^2 P_{t11}P_{32}P_{t21}$$

$$[\overline{G}K_\alpha^{-1}K\Sigma K\Sigma KZ_t] = P_{t63} - \Omega_\alpha(P_{53}P_{t21} + P_{32}P_{t42}) + \Omega_\alpha^2 P_{32}^2 P_{t21}$$

$$[\overline{G}K_\alpha^{-1}K\Sigma K\Sigma Z_t] = P_{t53} - \Omega_\alpha(P_{53}P_{t11} + P_{32}P_{t32}) + \Omega_\alpha^2 P_{32}^2 P_{t11}$$

– $E(n\mathcal{T}_5\mathcal{T}_6)$

$$\begin{aligned}
\Xi_{56} =& \frac{3\rho_3^2\Omega_\alpha}{2n} S_{134}G_4\overline{G_5}K_{\alpha 56}^{-1}\Sigma_{37}\Sigma_{89}\Sigma_{10,11}K_{11,12}S_{12,92}\Delta_{678,10} \\
=& \frac{3\rho_3^2\Omega_\alpha S_\epsilon^2}{2n^3}\sum_{t,s=1}^n Z_t(\tau_1)\overline{Z_s(\tau_2)}[\overline{G}Z_t][\overline{G}K_\alpha^{-1}]_6[\Sigma KZ_s]_{10}[\overline{Z_t}\Sigma]_7[\Sigma Z_s]_8[K_{6,7}K_{8,10} \\
& + K_{68}K_{7,10} + K_{6,10}K_{78}] \\
=& \frac{3\rho_3^2\Omega_\alpha S_\epsilon^2}{2n^3}\sum_{t,s=1}^n Z_t(\tau_1)\overline{Z_s(\tau_2)}\Big[[\overline{G}K_\alpha^{-1}K\Sigma Z_t][\overline{Z_s}\Sigma K\Sigma KZ_s] \\
& + [\overline{G}K_\alpha^{-1}K\Sigma Z_s][\overline{Z_t}\Sigma K\Sigma KZ_s] \\
& + [\overline{G}K_\alpha^{-1}K\Sigma KZ_s][\overline{Z_t}\Sigma K\Sigma Z_s]\Big]
\end{aligned}$$

$$\begin{aligned}
E(n\mathcal{T}_5\mathcal{T}_6) =& \frac{3\rho_3^2\Omega_\alpha^3 S_\epsilon^2}{2n^3}\sum_{t,s=1}^n P_{t01}P_{s01}\Big[(P_{t32} - \Omega_\alpha P_{32}P_{t11})[\overline{Z_s}\Sigma K\Sigma KZ_s] \\
& + (P_{s32} - \Omega_\alpha P_{32}P_{s11})[\overline{Z_t}\Sigma K\Sigma KZ_s] \\
& + (P_{s42} - \Omega_\alpha P_{32}P_{s21})[\overline{Z_t}\Sigma K\Sigma Z_s]\Big]
\end{aligned}$$

with

$$[\overline{Z_t}\Sigma K\Sigma KZ_s] = P_{ts42} - \Omega_\alpha(P_{t32}P_{s21} + P_{t11}P_{s42}) + \Omega_\alpha^2 P_{t11}P_{32}P_{s21}$$

$$[\overline{Z_s}\Sigma K\Sigma KZ_s] = P_{ss42} - \Omega_\alpha(P_{s32}P_{s21} + P_{s11}P_{s42}) + \Omega_\alpha^2 P_{s11}P_{32}P_{s21}$$

$$[\overline{Z_t}\Sigma K\Sigma Z_s] = P_{ts32} - \Omega_\alpha(P_{t32}P_{s11} + P_{t11}P_{s32}) + \Omega_\alpha^2 P_{t11}P_{32}P_{s11}$$

CONCLUSION

In this thesis, we extended the generalized empirical likelihood method to allow the moment conditions to be defined on a continuum. We showed that the method is asymptotically equivalent to the generalized method of moment for a continuum. Using numerical simulations, we found that in finite samples, the properties of the estimator depend on a regularization parameter, α. In particular, we saw that CGEL dominates CGMM if we choose α properly. We therefore proposed a way of computing its optimal value based on the higher order asymptotic mean square error.

Although we covered many aspects of the CGEL method in this thesis, there is still much to learn about how to improve its finite sample properties. We showed that the method can be a good alternative for estimating models in several areas of economics. Therefore, it should be studied further. First, we should compare the properties of the CGEL estimator using the selection procedure of α proposed in chapter 3 with CGMM of (Carrasco, 2010). CGEL should be extended to allow weakly dependent data and a procedure for selecting α should be derived for the more general nonlinear case.

RÉFÉRENCES

Airayetpan, R. and A. Ramm. 2000. ≪ Dynamical systems and discrete methods for solving nonlinear ill-posed problem ≫, *Applied Mathematics Review*, vol. 1, p. 491–536.

Anatolyev, S. 2005. ≪ Gmm, gel, serial correlation, and asymptotic bias ≫, *Econometrica*, vol. 73, p. 983–1002.

Anatolyev, S. and N. Gospodinov. 2011. ≪ Specification testing in models with many instruments ≫, *Econometric Theory*, vol. 27, p. 427–441.

Antoine, B., H. Bonnal, and E. Renault. 2007. ≪ On the efficient use of the informational content of estimating equations : Implied probabilities and euclidean empirical likelihood ≫, *Journal od Econometrics*, vol. 138, p. 461–487.

Arellano, M., P. Hansen, and D. Sentana. 2011. ≪ Underidentification ? ≫, *forthcoming in Journal of Econometrics.*

Baggerly, K. 1998. ≪ Empirical likelihood as a good-of-fit measure ≫, *Biometrika*, vol. 3, no. 85, p. 535–547.

Berkowitz, J. 2001. ≪ Generalized spectral estimation of the consumption based pricing model ≫, *Journal of Econometrics*, vol. 104, p. 269–288.

Blaschke, B., A. Neubauer, and O. Scherzer. 1997. ≪ On the convergence rates for the iteratively regularized gauss-newton method ≫, *IMA Journal of Numerical Analysis*, vol. 17, p. 421–436.

Carrasco, M. 2010. ≪ A regularization approach to the many instruments problem ≫, *Working Paper : Université de Montréal.*

Carrasco, M., M. Chernov, J. Florens, and E. Ghysels. 2007. ≪ Efficient of general dynamic models with continuun of moment conditions ≫, *Journal of Econometrics*, vol. 140, p. 529–573.

Carrasco, M. and J. Florens. 2000. ≪ Generalization of gmm to a continuum of moment conditions ≫, *Econometric Theory*, vol. 16, p. 655–673.

———. 2002. ≪ Efficient gmm estimation using the empirical characteristic function ≫, *mimeo CREST, Paris.*

Carrasco, M., J. Florens, and E. Renault. 2007. ≪ Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization ≫, *Handbook of Econometrics*, vol. 6B, p. 5633–5751.

Carrasco, M. and R. Kotchoni. 2010. ≪ Efficient gmm estimation using the characteristic function ≫, *Mimeo, University of Montreal*.

Dagenais, M. 1983. ≪ Extension of the ridge regression technique to non-linear models with additive errors ≫, *Economics Letters*, vol. 12, no. 2, p. 169–174.

Donald, S., G. Imbens, and W. Newey. 2003. ≪ Empirical likelihood estimation and consistent tests with conditional moment restrictions ≫, *Journal of Econometrics*, vol. 117, p. 55–93.

Donald, S., G. Imbens, and W. Newey. 2010. ≪ Choosing the number of moments in conditional moment restrition models ≫, *Forthcoming in Econometrica*.

Donald, S. and W. Newey. 2001. ≪ Choosing the number of instruments ≫, *Econometrica*, vol. 69, p. 1161–1191.

Garcia, R., E. Renault, and D. Veredas. 2006. ≪ Estimation of stable distribution by indirect inference ≫, *Working Paper : UCL and CORE*.

Groetsch, C. 1993. ≪ Inverse problems in mathematical sciences ≫, *Wiesbaden : Vieweg*.

Guggenberger, P. 2008. ≪ Finite sample evidence suggesting a heavy tail problem of the generalized empirical likelihood estimator ≫, *Econometric Reviews*, vol. 26, p. 526–541.

Guggenberger, P. and J. Hahn. 2005. ≪ Finite sample properties of the two-step empirical likelihood estimator ≫, *Econometric Reviews*, vol. 24, no. 3, p. 247–263.

Hansen, L. 1982. ≪ Large sample properties of generalized method of moments estimators ≫, *Econometrica*, vol. 50, p. 1029–1054.

Hansen, L., J. Heaton, and A. Yaron. 1996. ≪ Finit-sample properties of some alternative gmm estimators ≫, *Journal of Business and Economic Statistics*, vol. 14, p. 262–280.

Imhof, J. 1961. ≪ Computing the distribution of quadratic forms in normal variables ≫, *Biometrika*, vol. 48, p. 419–426.

Jin, Q. 2000. ≪ On the iteratively regularized gauss-newton method for solving nonlinear ill-posed problems ≫, *Mathematics of Computation*, vol. 69, no. 232, p. 1603–1623.

Kitamura, Y. and M. Stutzer. 1997. ≪ An information-theoretic alternative to generalized method of moments estimation ≫, *Econometrica*, vol. 65, no. 5, p. 861–874.

Kitamura, Y., G. Tripathi, and H. Ahn. 2004. ≪ Empirical likelihood-based inference in conditional moment restriction models ≫, *Econometrica*, vol. 72, p. 1667–1714.

Kress, R. 1999. *Linear Integral Equations*. Springer.

Li, K. and Y. He. 2005. ≪ Taylor expansion algorithme for the branching solution of the navier-stokes equations ≫, *International Journal Numerical Analysis and Modeling*, vol. 2, no. 4, p. 459–478.

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

Newey, W. and R. Smith. 2004. ≪ Higher order properties of gmm and generalized empirical likelihood estimators ≫, *Econometrica*, vol. 72, p. 219–255.

Nolan, J. 2005. ≪ Modeling financial data with stable distributions ≫, *Math/Stat Department, American University*.

———. 2009. ≪ Stable distributions ≫, *Math/Stat Department, American University*.

Owen, A. 2001. *Empirical Likelihood*. Chapman and Hall.

Ramm, A. 2004a. ≪ Dynamical systems method for solving nonlinear operator equations ≫, *Int. J. Appl. Math. Sci.*, vol. 1, p. 97–110.

———. 2004b. ≪ Dynamical systems method for solving operator equations ≫, *Comm. Nonlinear Sci. and Numer. Simul.*, vol. 9, no. N2, p. 383–402.

Schennach, S. 2007. ≪ Point estimation with exponentially tilted empirical likelihood ≫, *Econometrica*, vol. 35, no. 2, p. 634–672.

Seidman, T. and C. Vogel. 1989. ≪ Well posedness and convergence of some regularisation methods for non-linear ill-posed problems ≫, *Inverse Problems*, vol. 5, p. 227–238.

Singleton, K. 2001. ≪ Estimation of affine pricing models using the empirical characteristic function ≫, *Journal of Econometrics*, no. 102, p. 111–141.

Smith, R. 1997. ≪ Alternative semi-parametric likelihood approaches to generalized method of moments estimation ≫, *The Economic Journal*, vol. 107, p. 503–519.

———. 2004. ≪ Gel criteria for moment condition models ≫, *CeMMAP working papers*.