Regularized Generalized Empirical Likelihood Estimators

Marine Carrasco
Université de Montréal

Rachidi Kotchoni*
Université Paris Nanterre

March 2, 2017

Abstract

Generalized Empirical Likelihood (GEL) estimators are solved by converting a high dimensional primal optimization problem into a dual Minimum Discrepancy problem with fewer parameters. When the GEL problem is subject to a continuum of restrictions, the duality relationships break down as the system of constraints becomes ill-posed. Duality is restored by solving a relaxed problem that leads to a family of Regularized GEL (RGEL) estimators. We show that the RGEL estimator is asymptotically normally distributed and efficient. An implementation strategy in one step inspired from the Three-Steps Empirical Likelihood of Antoine, Bonnal and Renault (2007) is proposed. Monte Carlo simulations based on a linear heteroskedastic model shows that the RGEL and the efficient two-steps CGMM of Carrasco and Florens (2000) have quite similar performances. However, the optimal regularization parameter of the RGEL converges to zero at a slower rate than the one of the CGMM estimator.

Keywords: Continuum of Moment Conditions - Duality - Generalized Empirical Likelihood - Regularization

JEL Classification:

1 Introduction

This paper studies a class of regularized generalized empirical likelihood (RGEL) estimators that are suitable for models that are specified as a large number or a continuum of moment restrictions. In fact, many economics theories lead to testable implications that are represented as conditional moment restrictions. Famous examples are given by the set of Euler equations stemming from a consumption capital asset pricing model (CCAPM) or dynamic stochastic general equilibrium (DSGE) models. Such conditional moment restrictions can be converted into a continuum of unconditional moment conditions with same information content (Bierens, 1982; Lavergne and Patilea, 2008).

Two econometric approaches have often been used to conduct parametric inference based on moment conditions. The first approach is the Generalized Method of Moments (GMM) developed by Hansen (1982), which consists of minimizing a quadratic form of an empirical counterpart of the moment conditions with respect to parameters of interest. Much less used than the GMM in practice, the second approach is the Generalized Empirical Likelihood (GEL) estimation, which consists of minimizing a discrepancy function of the unknown probability distribution function (PDF) that generated the data subject to the moment restrictions of interest. GEL methods have been developed as an attempt to improve the small sample properties of estimators that are derived from moment conditions (see Newey and Smith, 2004). The family of GEL estimators encompasses

*Corresponding Author. Address: CNRS - EconomiX (UMR 7235). Bâtiment G; 200, Avenue de la République; 92001 Nanterre cedex; FRANCE. Telephone: +33 1 40 97 59 47; Fax: +33 1 40 97 41 98; Email: rachidi.kotchoni@u-paris10.fr

the continuously updated estimator (CUE), the exponential tilting estimator (Kitamura and Stutzer, 1997; Schennack, 2007) and the standard empirical likelihood estimators (Owen, 1988, 1990; Qin and Lawless, 1994).

Newey and Smith (2004) derives a duality relationship between the GEL estimation based on Cressie and Read (1984)'s discrepancy function and the minimum discrepancy (MD) estimation formulated by Corcoran (1998). Versions of this duality relationship can be found in Borwein and Lewis (1991) and Borwein (1992). In the primal GEL problem, the set of unknown parameters includes the PDF of the data, which is potentially infinite dimensional. The dual MD problem replaces this PDF with a vector of dual parameter whose dimensionality equals the number of moment restrictions. This makes the dual optimization problem easier to handle when the number of moment restrictions is small relatively to the sample size.

When the number of moment restrictions is large or infinite, the system of equations represented by the moment restrictions becomes singular or ill-conditioned. Therefore, the GEL problem becomes ill-posed and the natural duality relationship between the GEL and the MD problems breaks down. This problem is studied by Borwein (1992) who showed that the maximum entropy solution to a system of an infinite number of constraints may exists while failing to satisfy the natural duality formula. In this case, Borwein shows that the solution to the dual problem GEL problem can be derived as a limit of a sequence of Tikhonov-type regularized solutions that are obtained from relaxed problems in which the constraints are allowed to be violated by a small margin.

There is a growing literature that tackles problems related to inference in linear models where the number of regressors exceeds the sample size (Candes and Tao, 2007; Gautier and Tsybakov, 2014) and in large nonlinear moment conditions models (Shi, 2016). For linear models, a sparseness condition on the slope coefficients is necessary in order to ensure identification. In moment condition models, emphasis is put on the selection of a small subset of moments restrictions that are most informative about the parameter of interest. In either case, some king of regularization or relaxation is need in order to ensure that the solution of the model exist and is well-behaved. The closest reference to our paper is Chaussé (2011), where a continuum of moment condition model is estimated using a GEL procedure. In that paper, Chaussé addresses the ill-poseness by modifying the first order conditions solved by the dual parameters. Our approach based on relaxation entails the addition of a penalty term to the Lagrangian so that the resulting first order conditions are regular.

We illustrate the failure of duality in contexts where the system of constraints consists of a large number of discrete moment conditions or a continuum of moment conditions. Duality is restored by employing the relaxation scheme proposed by Borwein which naturally leads to a class of Regularized Generalized Empirical Likelihood (RGEL) estimators. We show that the RGEL estimator is asymptotically normally distributed and efficient in the class of moment conditions-based estimators. An implementation strategy in one step inspired from the Three-Steps Empirical Likelihood of Antoine, Bonnal and Renault (2007) is proposed. Finally, a Monte Carlo simulation study based on a linear heteroskedastic model taken from Cragg (1983) and later used in Kitamura, Tripathi and Ahn (2004) is performed. The simulations results show that the inefficient first step CGMM estimator, the efficient two-steps CGMM and the RGEL have very similar performances for this model. One needs to use a magnifying glass before realizing that the RMSE of the RGEL estimator lies between those of the two CGMM estimators. This result confirms the consistency of the RGEL estimator and its asymptotic equivalence to the efficient CGMM estimator. However, the regularization parameter that permits to minimize the Root Mean Square Error of the RGEL converges to zero at a slower rate than the one needed to minimize the RMSE of the two-step CGMM estimator.

The remainder of the paper is organized as follows. Section 2 illustrates the failure of duality relationship between the MD and GEL problems when the number of moment restrictions is finitely

large. Section 3 illustrates the same problem with a continuum of moment condition. Section 4 presents the derivation of the RGEL estimators and Section 5 discusses their asymptotic properties. Section 6 describes an implementation strategy for the RGEL. Section 7 presents the simulation study and Section 8 concludes. An appendix collects the mathematical proofs.

2 Failure of Duality with a Large Number of Moment Conditions

This section presents the failure of duality in the presence of a large number of moment conditions. We have in mind a situation where a conditional moment restrictions is converted into a large number of moment conditions using instruments, or an asset pricing model where a stochastic discount factor is being estimated using a large number of assets.

Let $x_i \in \mathbb{R}^d, i = 1, ..., n$ be IID draws from a distribution indexed by a finite dimensional parameter θ.[1] Let $g_i(\theta) \equiv g(x_i, \theta) \in \mathbb{R}^m$ be a vector of m moment conditions satisfying:

$$E[g_i(\theta_0)] = (0, ..., 0)' \in \mathbb{R}^m, \tag{1}$$

where θ_0 is the true parameter value for the actual data generating process.

The GMM estimator of Hansen (1982) for θ_0 is given by:

$$\widehat{\theta}_{GMM} = \arg\min_{\theta \in \Theta} \widehat{g}_n(\theta)' \widehat{\Omega}^{-1} \widehat{g}_n(\theta), \tag{2}$$

where Θ is a compact parameter space, $\widehat{g}_n(\theta) = \frac{1}{n}\sum_{t=1}^n g_i(\theta)$ is the vector of sample averages of the moment conditions and $\widehat{\Omega}$ is a consistent estimator of $\Omega = \lim_{n\to\infty} Var(\sqrt{n}\widehat{g}_n(\theta))$, the asymptotic covariance matrix of the moment conditions. The GMM estimator is asymptotically normal and efficient within the family of estimators obtained by minimizing a quadratic form of $\widehat{g}_n(\theta)$. Carrasco and Florens (2003) showed that GMM estimators attain the maximum likelihood estimator efficiency if the score associated with the true likelihood function belongs to the closure of the linear space spanned by the moment conditions.

The GEL estimator of θ_0 is obtained as the solution of the following saddle point problem:

$$\begin{aligned} \widehat{\theta} &= \arg\min_{\theta} \min_{p} \varphi(p) \\ \text{s.t. } \sum_{i=1}^n p_i g_i(\theta) &= 0 \text{ and } \sum_{i=1}^n p_i = n, \end{aligned} \tag{3}$$

where $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ is a convex discrepancy function and $p = (p_1, ..., p_n)'$. Note that the number of parameters $(\theta, p) \in \mathbb{R}^{n+q}$ is larger than the sample size, which makes this problem intractable directly.

If we let $A(\theta)$ denote the $(m+1, n)$ matrix given by:

$$A(\theta) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ g_1(\theta) & g_2(\theta) & \cdots & g_n(\theta) \end{pmatrix},$$

then the system of constraints under (3) can be written as $A(\theta)p = b$, where:

$$b = (n, 0, ..., 0)' \in \mathbb{R}^{m+1}.$$

[1] For a discussion on empirical likelihood methods with dependent data, see Kitamura, Y. (1997), Kitamura, Y., Tripathi G. and Ahn H. (2004) and Kitamura, Y. (2006).

Let $A^*(\theta)$ denote the transpose (or adjoint) of $A(\theta)$. For all $\lambda = (\lambda_0, \lambda_1) \in \mathbb{R}^{m+1}$ with $\lambda_0 \in \mathbb{R}$ and $\lambda_1 \in \mathbb{R}^m$, we have:

$$A^*(\theta)\lambda = \lambda_0 \iota + g(\theta)\lambda_1, \tag{4}$$

where $\iota = (1, ..., 1)' \in \mathbb{R}^n$ and $g(\theta)$ is the (n, m) matrix whose i^{th} row is given by $g_i(\theta)'$.

The convex conjugate φ^* of φ satisfies:

$$\varphi^*(y) = \sup_p \left\{ y'p - \varphi(p) \right\}, \text{ for all } y \in \mathbb{R}^n.$$

Note that φ^* is also a mapping from $\mathbb{R}^n$ to $\mathbb{R}$. An explicit expression of φ^* can easily be derived for a wide range of specifications of φ. Under certain regularity conditions, Borwein and Lewis (1991) showed that the dual MD problem is given by:

$$\widehat{\theta}_{GEL} = \arg\min_{\theta \in \Theta} \sup_{\lambda \in \Lambda_n(\theta)} b'\lambda - \varphi^*(A^*(\theta)\lambda), \tag{5}$$

where $\Lambda_n(\theta) \subset \mathbb{R}^{m+1}$ is the set of all λ such that $\lambda_0 + g_i'\lambda_1$ belongs to the domain of φ^* for all $i = 1, ..., n$. The vector of parameters of interest in the dual problem is $(\theta, \lambda) \in \mathbb{R}^{q+m+1}$. The dual problem is tractable because the number of moment conditions m is generally much smaller than the sample size n.

Let the solution of the first part of the optimization problem above be given by:

$$\widehat{\lambda}(\theta) = \arg \sup_{\lambda \in \Lambda_n(\theta)} b'\lambda - \varphi^*(A^*(\theta)\lambda). \tag{6}$$

Substituting $\widehat{\lambda}(\theta)$ into (5) yields a concentrated dual objective function that depends on θ only. We have:

$$\widehat{\theta}_{GEL} = \arg\min_{\theta \in \Theta} b'\widehat{\lambda}(\theta) - \varphi^*\left(A^*(\theta)\widehat{\lambda}(\theta)\right). \tag{7}$$

The solution for p can then be deduced as:

$$\widehat{p} = \varphi^{*\prime}\left(A^*\left(\widehat{\theta}_{GEL}\right)\widehat{\lambda}\left(\widehat{\theta}_{GEL}\right)\right) \tag{8}$$

where $\varphi^{*\prime}$ is the gradient of φ^*.

A popular specification of φ is given by the Cressie-Read (1984)'s family of discrepancy functions i.e.:

$$\varphi(p) = \begin{cases} \sum_{i=1}^n \left(\frac{p_i^{1+\gamma} - 1}{\gamma(1+\gamma)} - \frac{p_i}{\gamma} \right), \gamma > -1, \gamma \neq 0 \\ \sum_{i=1}^n \left(-\ln p_i + p_i \right), \gamma = -1 \end{cases}, \tag{9}$$

The corresponding expression of $\varphi^*(p)$ is given by [2]:

$$\varphi^*(p) = \begin{cases} \sum_{i=1}^n \frac{(1+\gamma p_i)^{\frac{1+\gamma}{\gamma}} + 1}{(1+\gamma)}, \gamma > -1, \gamma \neq 0 \\ \sum_{i=1}^n \left(-\ln(1 - p_i) - 1 \right), \gamma = -1 \end{cases}. \tag{10}$$

The linear term in p included in the expression of $\varphi(p)$ is unusual in the literature. This linear term does not alter the degree of convexity of $\varphi(p)$ while it forces the domain of $\varphi^*(p)$ to admit zero as

[2] The logarithm discrepancy function is obtained by letting $\gamma \to -1$ and it yields the empirical likelihood (EL) estimator of Owen (1988) and Qin and Lawless (1994).

interior point. This property is desirable given that the true value of λ is zero when the moment conditions are valid.[3]

Using (6), it is straightforward to show that the optimal dual parameter (λ) solves:

$$\widehat{\lambda}(\theta) = \arg \sup_{\lambda \in \Lambda_n(\theta)} Q(\theta, \lambda), \tag{11}$$

where

$$Q(\theta, \lambda) = n\lambda_0 - \frac{1}{1+\gamma} \sum_{i=1}^{n} \left[\left(1 + \gamma\lambda_0 + \gamma g_i'(\theta)\lambda_1\right)^{\frac{1+\gamma}{\gamma}} + \frac{1}{\gamma} \right]. \tag{12}$$

is the dual objective function.

When $\gamma = 1$, $\varphi(p)$ and $\varphi^*(p)$ are quadratic functions. In this case, the expression of $\widehat{\lambda}(\theta)$ is available in closed form for a well-posed GEL problem. Indeed, let the empirical covariance matrix of the moment conditions be given by:

$$\widehat{\Omega} = \frac{1}{n} (g - \overline{g})' (g - \overline{g}),$$

where $\overline{g}$ is the (n, m) matrix whose rows are all equal to $\widehat{g}(\theta)'$. We have the following result.

Theorem 1 *Assume that φ is given by (9) with $\gamma = 1$ and that $g(\theta)$ is of full rank. The solution to the GEL problem based on the moment conditions $g(\theta)$ satisfies:*

$$p_i(\theta) = 1 - (g_i(\theta) - \widehat{g}(\theta))' \widehat{\Omega}^{-1} \widehat{g}(\theta), i = 1, ..., n, \tag{13}$$

$$\widehat{\lambda}_0(\theta) = \widehat{g}(\theta)' \widehat{\Omega}^{-1} \widehat{g}(\theta), \tag{14}$$

$$\widehat{\lambda}_1(\theta) = -\widehat{\Omega}^{-1} \widehat{g}(\theta) \text{ and} \tag{15}$$

$$\widehat{\theta} = \arg\min_{\theta \in \Phi} Q\left(\theta, \widehat{\lambda}(\theta)\right) \tag{16}$$

where

$$Q\left(\theta, \widehat{\lambda}(\theta)\right) = -n + \frac{n}{2} \widehat{g}(\theta)' \widehat{\Omega}^{-1} \widehat{g}(\theta).$$

As shown by (1), the quadratic discrepancy function leads to the continuously updated GMM estimator (CUE) of Hansen, Heaton and Yaron (1996). Note that unlike Back and Brown (1993) and Newey and Smith (2004), we do not obtain uniform empirical probabilities in the quadratic case[4]. This is due to the linear term included in the expression of φ. The centering of the moment conditions in the expressions of p_i and $\widehat{\Omega}$ stems from the restriction $\sum_{i=1}^{n} p_i = n$, which permits us to avoid normalizing the estimated empirical probabilities artificially. Interestingly, the expression of $\widehat{\lambda}_0(\theta)$ given by (14) coincides with the statistic used for Hansen (1982)'s overidentification test (i.e., J-test). Note that $\widehat{\lambda}_0(\theta)$ could be interpreted as a Hansen-Jagannathan distance in the context of an asset pricing model.[5]

The expression of $\widehat{\lambda}_1(\theta)$ given by Equation (15) exists if and only if the matrix g is of full rank so that $\widehat{\Omega}$ is invertible. Otherwise, the equation $\widehat{\Omega}\widehat{\lambda}_1(\theta) = -\widehat{g}(\theta)$ of which $\widehat{\lambda}_1(\theta)$ is solution is ill-posed and a closed form expression of $\widehat{\lambda}_1(\theta)$ does not exist. As shown by Borwein (1992), duality can be restored by regularizing the matrix $\widehat{\Omega}$.

[3] If we remove the linear term from the expression of φ, the convex conjugate fonction would be $\varphi^*(p) = \sum_{i=1}^{n} \frac{1}{1+\gamma} \left((\gamma p_i)^{\frac{1+\gamma}{\gamma}} + \frac{1}{\gamma}\right)$ for $(\gamma > -1, \gamma \neq 0)$. In this case, 0 would be at the boundary of the domain of $\varphi^*(p)$ since power functions are defined on $\mathbb{R}+$.

[4] Although, the weights $p_i(\theta)$ are asymptotically uniform due to $\widehat{g}(\theta) = o_p(1)$ as $n \to \infty$.

[5] The standard Hansen-Jaganathan distance uses the metrics $\frac{1}{n} g'g$ rather than $\widehat{\Omega} = \frac{1}{n} (g - \overline{g})' (g - \overline{g})$. Both metrics converge in probability to the same limit if the moment conditions are valid. If some moment conditions are invalid, $\widehat{\Omega}$ will remain consistent for the covariance matrix of the moment conditions. See for example Kan and Robotti (2009) on the use of the Hansen-Jaganathan distance for asset pricing model evaluation.

3 Failure of Duality with a Continuum of Moment Conditions

With the insight gained from the previous section, let us now consider a continuum of moment conditions $h(x_i, \tau, \theta)$ satisfying:

$$E[h(x_i, \tau, \theta)] = 0 \text{ for all } \tau \in \mathbb{R} \text{ and for all } i = 1, ..., n, \tag{17}$$

where $h_i(\tau, \theta) \equiv h(x_i, \tau, \theta)$ is real-valued. For instance, $h_i(\tau, \theta)$ could be the Laplace transform of a conditional PDF or unconditional moment restrictions that are obtained using a continuum of instruments, as in Bierens (1982).[6]

The GEL problem with a continuum of moment conditions is stated as:

$$\begin{aligned} \widehat{\theta} &= \arg\min_{\theta} \min_{p} \varphi(p) \\ \text{s.t. } \sum_{i=1}^{n} p_i h_i(\tau, \theta) &= 0, \tau \in \mathbb{R} \text{ and } \sum_{i=1}^{n} p_i = n. \end{aligned} \tag{18}$$

The constraints above can be written as $A(\tau, \theta) p = b$, for all $\tau \in \mathbb{R}$ where:

$$A(\tau, \theta) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ h_1(\tau, \theta) & h_2(\tau, \theta) & \cdots & h_n(\tau, \theta) \end{pmatrix}$$

and $b = (n, 0)'$.

The Lagrangian of the GEL problem above is given by:

$$\widetilde{\mathcal{L}}(p, \lambda, \theta) = \varphi(p) - \int \lambda(\tau)' [A(\tau, \theta) p - b] \pi(\tau) d\tau \tag{19}$$

where $\lambda(\tau) = (\lambda_0, \lambda_1(\tau))'$ is a continuum of Lagrange multipliers and $\pi(\tau)$ is a positive measure that sums to unity on $\mathbb{R}^{\dim(\tau)}$. This measure defines a scalar product $\langle ., . \rangle$ such that for every pair of functions f and g, we have:

$$\langle f, g \rangle = \int f(\tau) g(\tau) \pi(\tau) d\tau,$$

It is important to select π such that $h_i(\tau, \theta)$ and $E(h_i(\tau, \theta))$ are both included in $L^2(\pi)$, the set of all square integrable functions with respect to π. This is done by selecting a measure for which the moments exist at any order.

The following theorem characterizes the solution of the GEL problem with a continuum of moment conditions.

Theorem 2 *The solution to the GEL problem with a continuum of moment conditions satisfies:*

$$p_i\left(\theta, \widehat{\lambda}\right) = \varphi^{*\prime}\left(\widehat{\lambda}_0 + \int h_i(\tau, \theta) \widehat{\lambda}_1(\tau, \theta) \pi(\tau) d\tau\right), i = 1, ..., n \tag{20}$$

where

$$\widehat{\lambda}(\tau, \theta) = \arg\sup_{\lambda(\tau, \theta)} Q(\theta, \lambda), \tag{21}$$

and

$$Q(\theta, \lambda) = n\lambda_0 - \sum_{i=1}^{n} \varphi^*\left(\lambda_0 + \int h_i(\tau, \theta) \lambda_1(\tau) \pi(\tau) d\tau\right). \tag{22}$$

[6] The evaluation of the dual objective function associated with a Cressy-Read discrepancy does not necessarily return a real output. Therefore, the GEL with complex-valued moment functions raises additional issues that are not addressed here.

As stated by Theorem 2 above, the existence of $p\left(\theta,\widehat{\lambda}\right)$ depends on whether a solution to (21) exists or not. Unfortunately, this problem is always ill-posed. This is easily seen by considering a quadratic discrepancy function ($\gamma=1$) so that the dual objective function becomes:

$$Q(\theta,\lambda)=n\lambda_0-\frac{1}{2}\sum_{i=1}^{n}\left(1+\lambda_0+\int h_i(\tau,\theta)\,\lambda_1(\tau)\,\pi(\tau)\,d\tau\right)^2+\frac{n}{2} \tag{23}$$

The first order conditions for the maximization of $Q(\theta,\lambda)$ with respect to λ are:

$$\begin{aligned}\frac{\partial Q(\theta,\lambda)}{\partial\lambda_0} &= n-\sum_{i=1}^{n}\left(1+\lambda_0+\int h_i(\tau,\theta)\,\lambda_1(\tau)\,\pi(\tau)\,d\tau\right)=0 \text{ and}\\ \frac{\partial Q(\theta,\lambda)}{\partial\lambda_1(r)} &= -\sum_{i=1}^{n}\left(1+\lambda_0+\int h_i(\tau,\theta)\,\lambda_1(\tau)\,\pi(\tau)\,d\tau\right)h_i(r,\theta)\,\pi(r)=0,\end{aligned}$$

for all $r\in\mathbb{R}$.

The first equation immediately leads to:

$$\widehat{\lambda}_0(\theta)=-\int\widehat{h}(\tau,\theta)\,\lambda_1(\tau)\,\pi(\tau)\,d\tau, \tag{24}$$

where $\widehat{h}(\tau,\theta)=\frac{1}{n}\sum_{i=1}^{n}h_i(\tau,\theta)$. Substituting $\widehat{\lambda}_0(\theta)$ into the second equation yields:

$$\widehat{K}(\theta)\,\lambda_1(r)=-\widehat{h}(r,\theta), \tag{25}$$

where $\widehat{K}(\theta)$ is the empirical covariance operator associated with the continuum of moment conditions. The kernel of $\widehat{K}(\theta)$ is given by:

$$\widehat{k}(r,\tau)=\frac{1}{n}\sum_{i=1}^{n}\left(h_i(r,\theta)-\widehat{h}(r,\theta)\right)\left(h_i(\tau,\theta)-\widehat{h}(\tau,\theta)\right) \tag{26}$$

and we have:

$$\widehat{K}(\theta)\,f(r)\equiv\int\widehat{k}(r,\tau,\theta)\,f(\tau)\,\pi(\tau)\,d\tau.$$

Hence, the feasibility of the GEL estimator rests upon the existence of a solution to Equation (25). Unfortunately, the empirical operator $\widehat{K}(\theta)$ is degenerate and non invertible on $L^2(\pi)$.[7] Its theoretical counterpart $K(\theta)$ with kernel $k(r,\tau)=Cov(h_i(r,\theta),h_i(\tau,\theta))$ is invertible only on a dense subset of $L^2(\pi)$. However, $K(\theta)^{-1}$ is discontinuous so that small variations in f may result in large variations in $K(\theta)^{-1}f$ (See Carrasco and Florens, 2000; Carrasco, Florens and Renault, 2007).

4 Regularized GEL (RGEL) Estimators

As seen previously, the natural solution to the dual problem does not necessarily exist when the number of moment conditions is large or infinite. In the case of a finite number of moment restrictions, the first order condition to solve for λ_1 is given by:

$$\widehat{\Omega}\widehat{\lambda}_1(\theta)=-\widehat{g}(\theta).$$

[7] The inversion of $\widehat{K}(\theta)$ raises a problem that is similar to the Fourier inversion of an empirical characteristic function.

Here, the ill-posedness of the GEL problem translates into the matrix $\widehat{\Omega}$ being ill-conditioned or singular.

There are two practical approaches to deal with this issue. The first approach consists of solving a penalized version of the first order condition, as done for example in Chaussé (2011). In the second approach, one adds a penalty term to the objective function in the first place so that the resulting first order condition is naturally well-posed. Kotchoni (2014) termed the first approach *"approximate solution to the exact problem"* and the second approach *"exact solution to an approximate problem."* This paper follows the second approach, which is less ad hoc in the sense that it is justified by a relaxation argument on the moment restrictions.

Although we are primarily interested in the case with a continuum of moment conditions, we first present the case with discrete moment conditions for the sake of completeness.

4.1 RGEL with a Finitely Large Number of Moment Conditions

Following Borwein (1992), we relax the constraints of the GEL with discrete moment conditions problem as follows:

$$\begin{aligned} \widehat{\theta} &= \arg\min_{\theta}\min_{p}\varphi(p) \\ \text{s.t. } \sum_{i=1}^{n} p_i &= n \text{ and } \left|\sum_{i=1}^{n} p_i g_{i,k}\right| \leq \varepsilon,\ k=1,...,m, \end{aligned} \tag{27}$$

for some $\varepsilon > 0$. This problem is relaxed in the sense that the moment conditions are not requested to hold exactly. The margin of violation is controlled by the hyper-parameter ε. The Lagrangian for this problem is:

$$\mathcal{L}\left(p,\widetilde{\lambda},\theta\right) = \varphi\left(p\right) - \lambda_0\left(\sum_{i=1}^{n} p_i - n\right) + \sum_{k=1}^{m}\widetilde{\lambda}_{1,k}\left(\zeta_k\sum_{i=1}^{n} p_i g_{i,k} - \varepsilon\right).$$

where ζ_k is the sign of $\sum_{i=1}^{n} p_i g_{i,k}$.

The Kuhn-Tucker conditions for optimality are given by:

$$\begin{aligned} \widetilde{\lambda}_{1,k} &\geq 0, \text{ and} \\ \widetilde{\lambda}_{1,k}\left(\zeta_k\sum_{i=1}^{n} p_i g_{i,k} - \varepsilon\right) &= 0, \text{ for all } k. \end{aligned}$$

As $\widetilde{\lambda}_{1,k} \geq 0$, we have $\widetilde{\lambda}_{1,k} = \left|\widetilde{\lambda}_{1,k}\right| = \left|\widetilde{\lambda}_{1,k}\zeta_k\right|$. Therefore:

$$\mathcal{L}\left(p,\widetilde{\lambda},\theta\right) = \varphi\left(p\right) - \lambda_0\left(\sum_{i=1}^{n} p_i - n\right) + \sum_{k=1}^{m}\widetilde{\lambda}_{1,k}\zeta_k\left(\sum_{i=1}^{n} p_i g_{i,k}\right) - \varepsilon\sum_{k=1}^{m}\left|\widetilde{\lambda}_{1,k}\zeta_k\right|.$$

Letting $\lambda_{1,k} = -\widetilde{\lambda}_{1,k}\zeta_k$ leads to a Lagrangian with an L1 penalty term:

$$\mathcal{L}\left(p,\lambda,\theta\right) = \varphi\left(p\right) - \lambda_0\left(\sum_{i=1}^{n} p_i - n\right) - \sum_{k=1}^{m}\lambda_{1,k}\left(\sum_{i=1}^{n} p_i g_{i,k}\right) - \varepsilon\sum_{k=1}^{m}\left|\lambda_{1,k}\right|. \tag{28}$$

The resulting dual problem is given by:

$$\widehat{\theta} = \arg\min_{\theta\in\Theta}\sup_{\lambda\in\Lambda_n(\theta)} n\lambda_0 - \varphi^*\left(\lambda_0\iota + g\left(\theta\right)\lambda_1\right) - \varepsilon\sum_{k=1}^{m}\left|\lambda_{1,k}\right| \tag{29}$$

If we abstract from the presence of λ_0 and the L1 penalty term, the dual objective-function above is similar to the ones given in Newey and Smith (2004) and Dudik, Phillips and Schapire (2004).

The expressions of (28) and (29) suggests that other types of penalty may be considered. An L2 penalization leads to:

$$\begin{aligned}\mathcal{L}(p,\lambda,\theta) &= \varphi(p) - \lambda_0\left(\sum_{i=1}^{n} p_i - n\right) \\ &\quad - \sum_{k=1}^{m}\lambda_{1,k}\left(\sum_{i=1}^{n} p_i g_{i,k}\right) - \frac{\varepsilon}{2}\sum_{k=1}^{m}\|\lambda_{1,k}\|^2,\end{aligned} \tag{30}$$

with a corresponding dual problem given by:

$$\widehat{\theta} = \arg\min_{\theta\in\Theta}\sup_{\lambda\in\Lambda_n(\theta)} n\lambda_0 - \varphi^*(\lambda_0\iota + g(\theta)\lambda_1) - \frac{\varepsilon}{2}\|\lambda_1\|^2. \tag{31}$$

L1 penalizations (also known as LASSO in the linear regression framework) are known to produce sparse dual estimators by ignoring the constraints whose violation is not costly (see Tibshirani, 1996). L2 penalizations are appealing because of their smoothness properties.

The following theorem characterizes the solution of a GEL problem with quadratic discrepancy function and L2 penalty term. This case is worth examining separately as it lead to closed form expressions for the dual parameters.

Theorem 3 *Assume that φ is given by (9) with $\gamma = 1$ and that the penalty term is quadratic. Then, the solution to the RGEL problem based on the discrete set of moment conditions satisfies:*

$$p_{\varepsilon,i}\left(\theta,\widehat{\lambda}\right) = 1 + (\widetilde{g}_i - \widehat{g}(\theta))'\widehat{\lambda}_{\varepsilon,1}(\theta), i = 1,...,n, \tag{32}$$

$$\widehat{\lambda}_{\varepsilon,0}(\theta) = \widehat{g}(\theta)'\widehat{\Omega}_{\varepsilon}^{-1}\widehat{g}(\theta), \tag{33}$$

$$\widehat{\lambda}_{\varepsilon,1}(\theta) = -\widehat{\Omega}_{\varepsilon}^{-1}\widehat{g}(\theta), \text{ and} \tag{34}$$

$$\widehat{\theta}_{\varepsilon} = \arg\min_{\theta\in\Theta} Q\left(\theta,\widehat{\lambda}_{\varepsilon}(\theta)\right) \tag{35}$$

where $\widehat{\Omega}_{\varepsilon} = \widehat{\Omega} + \frac{\varepsilon}{n}I$ and:

$$Q\left(\theta,\widehat{\lambda}_{\varepsilon}(\theta)\right) = -n + \frac{n}{2}\widehat{g}(\theta)'\widehat{\Omega}_{\varepsilon}^{-1}\widehat{g}(\theta).$$

With a quadratic discrepancy function and L2 penalty term, the solution to ill-posedness boils down to a ridge regularization of $\widehat{\Omega}$. If the solution θ_0 to $E(\widehat{g}(\theta)) = 0$ is unique and the moment conditions are in fixed finite number and not redundant, then $\widehat{\Omega}$ should converge to a positive definite matrix Ω as $n \to \infty$. In this case, the regularization parameter (ε) can be eliminated from the asymptotic distribution of RGEL estimator by letting $\frac{\varepsilon}{n}$ go to zero at a certain rate as n diverges to infinity.

With the intuitions gained from above, let us consider a more general Cressie-Read discrepancy function. The following notation is used subsequently:

$$\begin{aligned}\widehat{G}_i &= \frac{\partial g_i(\theta)}{\partial\theta}; & \widehat{\Omega}_i &= (g_i - \widehat{g})(g_i - \widehat{g})', \\ v_{i,\varepsilon} &= \widehat{\lambda}_{0,\varepsilon}(\theta) + g_i'\lambda_{1,\varepsilon}(\theta); & \overline{v}_{\varepsilon} &= \widehat{\lambda}_{0,\varepsilon}(\theta) + \widehat{g}'\widehat{\lambda}_{1,\varepsilon}(\theta), \\ k(x) &= \frac{(1+\gamma x)^{\frac{1}{\gamma}} - 1}{x}; & k_{i,\varepsilon} &= \frac{nk(v_{i,\varepsilon})}{\sum_{j=1}^{n} k(v_{j,\varepsilon})} \text{ and} \\ w_{i,\varepsilon} &= 1 + \overline{v}_{\varepsilon}k(v_{i,\varepsilon}) + \frac{1}{n}\sum_{j=1}^{n} k(v_{j,\varepsilon})(g_j - \widehat{g})'\lambda_{1,\varepsilon}.\end{aligned}$$

The next theorem characterizes the solution of the GEL problem with a general Cressie-Read discrepancy and L2 penalization.

Theorem 4 *Assume that φ is given by (9) with $\gamma \geq -1$ $(\gamma \neq 0)$ and that the penalty term is quadratic. Then the solution to the RGEL problem based on the discrete set of moment conditions satisfies:*

$$p_{i,\varepsilon} = 1 + v_{i,\varepsilon} k\left(v_{i,\varepsilon}\right), i = 1, ..., n, \tag{36}$$

$$\lambda_{0,\varepsilon}\left(\theta\right) = \left(\frac{1}{n}\sum_{i=1}^{n} k_{i,\varepsilon} g_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} k\left(v_{i,\varepsilon}\right)\widehat{\Omega}_i + \frac{\varepsilon}{n}I\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} w_{i,\varepsilon} g_i\right), \tag{37}$$

$$\lambda_{1,\varepsilon}\left(\theta\right) = -\left(\frac{1}{n}\sum_{i=1}^{n} k\left(v_{i,\varepsilon}\right)\widehat{\Omega}_i + \frac{\varepsilon}{n}I\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} w_{i,\varepsilon} g_i\right), \tag{38}$$

and the regularized GEL estimator $\widehat{\theta}_\varepsilon$ solves:

$$\left(\frac{1}{n}\sum_{i=1}^{n} p_{i,\varepsilon}\widehat{G}_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n} k\left(v_{i,\varepsilon}\right)\widehat{\Omega}_i + \frac{\varepsilon}{n}I\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} w_{i,\varepsilon} g_i\right) = 0. \tag{39}$$

Equations analogue to the ones above are derived by Newey and Smith (2004) in the context of a well-posed GEL problem with a finite number of moment conditions. Note that (37) and (38) are not closed form solutions, as the right hand side expressions depends on the left hand side variables. The representations (37), (38) and (39) are simply intended to show that RGEL estimator based on a Cressie-Read discrepancy shares the same structure as the CUE and the GMM estimator.

Interestingly, our approach to specify the GEL problem leads to four different sets of empirical probabilities: $\{p_{i,\varepsilon}/n\}_{i=1}^n$, $\{k\left(v_{i,\varepsilon}\right)/n\}_{i=1}^n$, $\{w_{i,\varepsilon}/n\}_{i=1}^n$ and $\{k_{i,\varepsilon}/n\}_{i=1}^n$. Indeed, we have:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} p_{i,\varepsilon} &= \frac{1}{n}\sum_{i=1}^{n} k_{i,\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} w_{i,\varepsilon} = 1, \text{ and} \\ \lim_{v_{i,\varepsilon}\to 0}\frac{1}{n}\sum_{i=1}^{n} k\left(v_{i,\varepsilon}\right) &= 1. \end{aligned}$$

However, the quantity $\frac{1}{n}\sum_{i=1}^{n} k\left(v_{i,\varepsilon}\right)$ may deviate from unity in finite sample and it approaches unity as $n \to \infty$ only if the model is correctly specified. In Equation (39), $G \equiv E\left(\widehat{G}_i\right)$ is estimated using the set of empirical probabilities $\{p_{i,\varepsilon}/n\}_{i=1}^n$, $\Omega \equiv E\left(\widehat{\Omega}_i\right)$ is estimated using $\{k\left(v_{i,\varepsilon}\right)/n\}_{i=1}^n$ while $E\left(g_i\right)$ is estimated using $\{w_{i,\varepsilon}/n\}_{i=1}^n$. In Equation (37), $E\left(g_i\right)$ is estimated using two different set of probabilities, namely $\{k_{i,\varepsilon}/n\}_{i=1}^n$ and $\{w_{i,\varepsilon}/n\}_{i=1}^n$. This suggests that no single set of empirical probabilities is efficient in all context.

When the discrepancy function is quadratic ($\gamma = 1$), we retrieve the result of Theorem 3 by noting that:

$$k\left(v_{i,\varepsilon}\right) = k_{i,\varepsilon} = w_{i,\varepsilon} = 1 \text{ and } p_{i,\varepsilon} = 1 + v_{i,\varepsilon}.$$

With a logarithmic discrepancy function (i.e., $\gamma = -1$), the empirical probabilities are given:

$$\begin{aligned} p_{i,\varepsilon} &= k\left(v_{i,\varepsilon}\right) = k_{i,\varepsilon} = \frac{1}{1 - v_{i,\varepsilon}} \text{ and} \\ w_{i,\varepsilon} &= 1 - \frac{v_{i,\varepsilon}}{1 - v_{i,\varepsilon}}\sum_{j=1}^{n} k_{j,\varepsilon}\left(g_j - \widehat{g}\right)'\lambda_{1,\varepsilon}. \end{aligned}$$

For all admissible values of γ, all four sets of empirical probabilities converge to uniform weights. This suggests that the asymptotic distribution of the RGEL estimator is independent of γ. Said differently, all GEL estimators that are based on Cressie-Read discrepancy functions are asymptotically equivalent to a regularized CUE.

4.2 Regularized GEL with a Continuum of Moment Conditions

Here, we consider the relaxed GEL problem given by:

$$\begin{aligned} \widehat{\theta} &= \arg\min_{\theta} \min_{p} \varphi(p) \\ \text{s.t. } \left|\sum_{i=1}^{n} p_i h_i(\tau,\theta)\right| &\leq \varepsilon,\ \tau \in \mathbb{R} \text{ and } \sum_{i=1}^{n} p_i = n \end{aligned} \tag{40}$$

The Lagrangian for this problem is given by:

$$\begin{aligned} \widetilde{\mathcal{L}}\left(p,\widetilde{\lambda},\theta\right) &= \varphi(p) - \lambda_0\left(\sum_{i=1}^{n} p_i - n\right) \\ &\quad + \int \widetilde{\lambda}_1(\tau)\left(\zeta_\tau \sum_{i=1}^{n} p_i h_i(\tau,\theta) - \varepsilon\right)\pi(\tau)\,d\tau, \end{aligned}$$

where $\widetilde{\lambda} \equiv \left\{\widetilde{\lambda}(\tau) = \left(\lambda_0, \widetilde{\lambda}_1(\tau)\right), \tau \in \mathbb{R}\right\}$ and ζ_τ is the sign of $\sum_{i=1}^{n} p_i h_i(\tau,\theta)$.

The Kuhn-Tucker optimality conditions are:

$$\widetilde{\lambda}_1(\tau) \geq 0 \text{ and } \widetilde{\lambda}_1(\tau)\left(\zeta_\tau \sum_{i=1}^{n} p_i h_i(\tau,\theta) - \varepsilon\right) = 0, \text{ for all } \tau$$

If we let $\lambda_1(\tau) = -\widetilde{\lambda}_1(\tau)\zeta_k$, we obtain:

$$\begin{aligned} \widetilde{\mathcal{L}}(p,\lambda,\theta) &= \varphi(p) - \lambda_0\left(\sum_{i=1}^{n} p_i - n\right) \\ &\quad - \int \lambda_1(\tau)\sum_{i=1}^{n} p_i h_i(\tau,\theta)\,\pi(\tau)\,d\tau - \varepsilon\int |\lambda_1(\tau)|\,\pi(\tau)\,d\tau, \end{aligned} \tag{41}$$

This stems from the fact that $0 \leq \widetilde{\lambda}_1(\tau) = \left|\widetilde{\lambda}_1(\tau)\right| = \left|\widetilde{\lambda}_1(\tau)\zeta_\tau\right|$ at the optimum.

For the sake of differentiability, we adopt an L2 penalization subsequently:

$$\begin{aligned} \widetilde{\mathcal{L}}(p,\lambda,\theta) &= \varphi(p) - \lambda_0\left(\sum_{i=1}^{n} p_i - n\right) \\ &\quad - \int \lambda_1(\tau)\sum_{i=1}^{n} p_i h_i(\tau,\theta)\,\pi(\tau)\,d\tau - \frac{\varepsilon}{2}\int \lambda_1^2(\tau)\,\pi(\tau)\,d\tau. \end{aligned} \tag{42}$$

By following the same steps as in the proof of Theorem 2, it is straightforward to show that:

$$\widehat{p}_{i,\varepsilon} = \varphi^{*'}\left(\widehat{\lambda}_{0,\varepsilon} + \int h_i(\tau,\theta)\,\widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\,\pi(\tau)\,d\tau\right),$$

where $\widehat{\lambda}_{\varepsilon}(\tau,\theta) = \underset{\lambda(\tau,\theta),\tau\in\mathbb{R}}{\arg\sup}\, Q_{\varepsilon}(\theta,\lambda)$ and:

$$Q_{\varepsilon}(\theta,\lambda) = n\lambda_0 - \sum_{i=1}^{n}\varphi^*\left(\lambda_0 + \int h_i(\tau,\theta)\,\lambda_1(\tau)\,\pi(\tau)\,d\tau\right) - \frac{\varepsilon}{2}\int \lambda_1^2(\tau)\,\pi(\tau)\,d\tau \tag{43}$$

The next theorem characterizes the solution of the RGEL problem based on a continuum of moment condition, a quadratic discrepancy function and an L2 penalty term. As noted previously, the quadratic discrepancy case is worth examining separately because it leads to closed form expressions for the dual parameters.

Theorem 5 *Assume that φ is given by (9) with $\gamma = 1$ and that the penalty term is quadratic. Then the solution to the RGEL problem based on a continuum of moment condition satisfies:*

$$\begin{aligned}
p_{\varepsilon,i} &= 1 + \widehat{\lambda}_{\varepsilon,0}(\theta) + \int h_i(\tau,\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)\,\pi(\tau)\,d\tau && (44)\\
\widehat{\lambda}_{\varepsilon,0}(\theta) &= \int \widehat{h}(\tau,\theta)\left[\left(\widehat{K}(\theta) + \frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(\tau,\theta)\right]\pi(\tau)\,d\tau && (45)\\
\widehat{\lambda}_{\varepsilon,1}(\tau) &= -\left(\widehat{K}(\theta) + \frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(\tau,\theta) \;\; \textit{and} && (46)\\
\widehat{\theta}_{\varepsilon} &= \arg\min_{\theta\in\Theta} Q\left(\theta,\widehat{\lambda}_{\varepsilon}(\theta)\right) && (47)
\end{aligned}$$

where $\widehat{K}(\theta)$ is the empirical covariance operator associated with the moment conditions, I is the identity operator and:

$$Q\left(\theta,\widehat{\lambda}_{\varepsilon}(\theta)\right) = -n + \frac{n}{2}\int \widehat{h}(\tau,\theta)\left[\left(\widehat{K}(\theta) + \frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(\tau,\theta)\right]\pi(\tau)\,d\tau. \tag{48}$$

From the objective function given at Equation (48), we see that this RGEL problem is the continuously updated version of Carrasco and Florens (2000)'s Continuum-GMM estimator. This result is quite intuitive given what we found previously for the case with large number of discrete moment conditions.

The next step is to characterize the solution of the GEL problem bases on a continuum of moment conditions and a general Cressie-Read discrepancy. For that purpose, we need the following notation:

$$\begin{aligned}
v_{i,\varepsilon} &= \widehat{\lambda}_{0,\varepsilon} + \int h_i(\tau,\theta)\,\widehat{\lambda}_1(\tau,\theta)\,\pi(\tau)\,d\tau;\\
\overline{v}_{\varepsilon} &= \widehat{\lambda}_{0,\varepsilon} + \int \widehat{h}(\tau,\theta)\,\widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\,\pi(\tau)\,d\tau,\\
k(x) &= \frac{(1+\gamma x)^{\frac{1}{\gamma}} - 1}{x}; \qquad k_{i,\varepsilon} = \frac{k(v_{i,\varepsilon})}{\sum_{j=1}^{n} k(v_{j,\varepsilon})} \text{ and}\\
w_{i,\varepsilon} &= 1 + \overline{v}_{\varepsilon}k(v_{i,\varepsilon}) + \frac{1}{n}\sum_{j=1}^{n} k(v_{j,\varepsilon})\int\left(h_j(\tau,\theta) - \widehat{h}(\tau,\theta)\right)\widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\,\pi(\tau)\,d\tau.
\end{aligned}$$

We further let $\widetilde{K}_{\varepsilon}(\theta)$ denote the covariance operator with kernel given by:

$$\widetilde{k}(r,\tau) = \frac{1}{n}\sum_{i=1}^{n} k(v_{i,\varepsilon})\left(h_i(r,\theta) - \widehat{h}(r,\theta)\right)\left(h_i(\tau,\theta) - \widehat{h}(\tau,\theta)\right). \tag{49}$$

The following result is established.

Theorem 6 *Assume that φ is given by (9) with $\gamma \geq -1$ $(\gamma \neq 0)$ and that the penalty term is quadratic. Then the solution to the RGEL problem with a continuum of moment conditions satisfies:*

$$p_{i,\varepsilon} = (1+\gamma v_{i,\varepsilon}(\theta))^{\frac{1}{\gamma}}, \tag{50}$$

$$\widehat{\lambda}_{0,\varepsilon} = \int \widetilde{h}(\tau,\theta)\left(\widetilde{K}_{\varepsilon}(\theta)+\frac{\varepsilon}{n}I\right)^{-1}\widetilde{\widetilde{h}}(\tau,\theta)\pi(\tau)\,d\tau, \tag{51}$$

$$\widehat{\lambda}_{1,\varepsilon}(\tau,\theta) = -\left(\widetilde{K}_{\varepsilon}(\theta)+\frac{\varepsilon}{n}I\right)^{-1}\widetilde{\widetilde{h}}(\tau,\theta), \tag{52}$$

and $\widehat{\theta}_{\varepsilon}$ solves:

$$\int \widetilde{G}_{\varepsilon}(\tau,\theta)\left(\widetilde{K}_{\varepsilon}(\theta)+\frac{\varepsilon}{n}I\right)^{-1}\widetilde{\widetilde{h}}(\tau,\theta)\pi(\tau)\,d\tau = 0, \tag{53}$$

where

$$\widetilde{h}(\tau,\theta) = \sum_{i=1}^{n} k_{i,\varepsilon}h_i(\tau,\theta), \quad \widetilde{\widetilde{h}}(\tau,\theta) = \frac{1}{n}\sum_{i=1}^{n} w_{i,\varepsilon}h_i(\tau,\theta) \text{ and}$$

$$\widetilde{G}_{\varepsilon}(\tau,\theta) = \frac{1}{n}\sum_{i=1}^{n} p_{i,\varepsilon}\frac{\partial h_i(\tau,\theta)}{\partial\theta}.$$

Theorem 6 is simply an adaptation of Theorem 4 to the case of a continuum of moment restrictions. It must be recalled that the formulas shown in these theorems are not closed form expressions. For instance, $\widehat{\lambda}_{1,\varepsilon}$ depends on $w_{i,\varepsilon}$ via $\widetilde{\widetilde{h}}(\tau,\theta)$ and $w_{i,\varepsilon}$ is already a function of $\widehat{\lambda}_{1,\varepsilon}$. Hence, Equation (53) cannot be directly solved for θ.

5 Consistency and Asymptotic Normality

We now study the consistency and asymptotic normality of regularized GEL estimators with a focus on the case with a continuum of moment conditions because. The results presented subsequently do no directly extend - but can be adapted - to models that are specified in terms of a discrete set of moment conditions, including the cases where the number of moment conditions increases with the sample size. Subsequently, we further maintain that the discrepancy function is of Cressie-Read type and that the penalty term is quadratic. The following assumptions are needed in order to establish the consistency of the RGEL estimator.

Assumptions:

(A1) The equation $E[h_i(\tau,\theta)] = 0$ has a unique solution θ_0 which is an interior point of Θ, for π-almost all τ;

(A2) The parameter space Θ is compact;

(A3) For π-almost all τ, $\widehat{h}(\tau,\theta)$ is continuous at each θ with probability one;

(A4) For π-almost all τ, $E[\sup_{\theta\in\Theta}\|h_i(\tau,\theta)\|^{\alpha}] < \infty$ for some $\alpha > 2$.

These assumptions are quite standard in the literature. Assumption (A1) imposes the point-identification of the model. Assumption (A2) ensures that all continuous functions of θ attain their maximum and minimum in Θ and allows us to use standard convex optimization results. Assumption (A3) allows $\widehat{h}(\tau,\theta)$ to be discontinuous in θ in finite sample but imposes that $\widehat{h}(\tau,\theta)$ is increasingly continuous as $n \to \infty$. In particular, this assumption implies that $E[h_i(\tau,\theta)]$ is a continuous in θ. Assumption (A4) is rather technical but it is always satisfied when $h_i(\tau,\theta)$ is bounded, as in the case of moment conditions that are based on the empirical characteristic function. This assumption ensures that the asymptotic covariance operator of the moment function is bounded.

Under these assumptions, the following consistency result is obtained.

Theorem 7 *: As $n \to \infty$ and $\varepsilon = O\left(n^b\right)$ for $1/2 < b < 1$, we have:*

$$\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right) = O_p\left(n^{-b+1/2}\right), \tag{54}$$

$$\left\|\overline{\lambda}_{1,\varepsilon}\left(.,\widehat{\theta}_\varepsilon\right)\right\| = O_p\left(n^{-b+1/2}\right), \tag{55}$$

$$\left\|\overline{\lambda}_{0,\varepsilon}\left(\widehat{\theta}_\varepsilon\right)\right\| = O_p\left(n^{-b}\right) \text{ and} \tag{56}$$

$$\widehat{\theta}_\varepsilon \xrightarrow{p} \theta_0 \tag{57}$$

where $\widehat{\theta}_\varepsilon$ denotes the RGEL estimator.

The proof of this theorem is largely inspired from Newey and Smith (2004). The nonparametric flavor of the GEL with a continuum of moment condition is seen by noting that the rate of convergence of the moment function $\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)$ and those of the dual parameters $\overline{\lambda}_{1,\varepsilon}\left(.,\widehat{\theta}_\varepsilon\right)$ and $\overline{\lambda}_{0,\varepsilon}\left(\widehat{\theta}_\varepsilon\right)$ are less than parametric. One might consider setting the rate b at which ε diverges as large as possible in order to increase the rate of convergence of the dual parameters. However, it is shown in a preliminary result that $\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|$ is bounded in probability only if $b > 1/2$. Moreover, $\left\|\widehat{h}\left(.,\theta_0\right)\right\| = O_p\left(n^{-1/2}\right)$ is the best rate that can be attained. For an arbitrary $b > 1/2$, the rate of convergence of the moment function therefore is of the form:

$$\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\| = O_p\left(n^{-\min(b-1/2,1/2)}\right).$$

Hence, the rate of convergence of $\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)$ does not improve as soon as $b \geq 1$. Consequently, the optimal rate of divergence for ε lies in the range $1/2 < b < 1$, as assumed by Theorem 7. In fact, the theorem essentially warns against choosing ε such that $\frac{\varepsilon}{n}$ converges to zero too fast, since choosing $b > 1/2$ does not jeopardize the consistency of $\widehat{\theta}_\varepsilon$.

The consistency of $\widehat{\theta}_\varepsilon$ follows naturally from the convergence of $\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)$ and Assumption (1c). The next theorem gives the rate of convergence of $\widehat{\theta}_\varepsilon$.

Theorem 8 *: As $n \to \infty$ and $\varepsilon = O\left(n^b\right)$ for $1/2 < b < 1$, we have:*

$$\begin{aligned}\widehat{\theta}_\varepsilon - \theta_0 \simeq & \left(\int \widetilde{G}\left(\tau,\theta_0\right)\left(K + \frac{\varepsilon}{n}I\right)^{-1}\widehat{G}\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1} \\ & \times \int \widetilde{G}\left(\tau,\theta_0\right)\left(K + \frac{\varepsilon}{n}I\right)^{-1}\widehat{h}\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau = O_p\left(n^{-1/2}\right)\end{aligned}$$

where $\widehat{G}\left(\tau,\theta\right) = \frac{\partial\widehat{h}(\tau,\theta)}{\partial\theta'}$ and $\widetilde{G}\left(\tau,\theta\right) = \frac{1}{n}\sum_{i=1}^{n}\left(1+\gamma v_{i,\varepsilon}\right)^{\frac{1}{\gamma}}\frac{\partial h_i(\tau,\theta)}{\partial\theta}$

Theorem 8 gives a large sample approximation of the RGEL estimator which shows how it differs from the Continuum-GMM (CGMM) estimator of Carrasco and Florens (2000) in finite sample. This difference is summarized in the expression of $\widetilde{G}\left(\tau,\theta\right)$. Indeed, the Continuum-GMM estimator (denoted $\widehat{\theta}_{CGMM}$) satisfies:

$$\begin{aligned}\widehat{\theta}_{CGMM} - \theta_0 = & \left(\int \widehat{G}\left(\tau,\theta_0\right)\left(K + \frac{\varepsilon}{n}I\right)^{-1}\widehat{G}\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1} \\ & \times \int \widehat{G}\left(\tau,\theta_0\right)\left(K + \frac{\varepsilon}{n}I\right)^{-1}\widehat{h}\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau.\end{aligned}$$

However, $\widetilde{G}(\tau,\theta_0)$ converges to $G(\tau,\theta_0)$ as n goes to infinity. This suggests that the RGEL and Continuum-GMM estimators enjoy the same asymptotic distribution, as stipulated by the next theorem.

Theorem 9 *As $n\to\infty$ and $\varepsilon = O\left(n^b\right)$ for $1/2<b<1$ we have:*

$$\sqrt{n}\left(\widehat{\theta}_\varepsilon-\theta_0\right)\to N(0,\Sigma)$$

where $\Sigma = \left(\int G(\tau,\theta)K^{-1}G(\tau,\theta)\pi(\tau)d\tau\right)^{-1}$.

The proof of Theorem 9 is rather straightforward once we recall that K is the asymptotic covariance operator of $h_i(\tau,\theta)$. While the upper bound $b<1$ was imposed only for convenience in Theorems 7 and 8, here this upper bound is necessary in order to obtain an asymptotic distribution that is regularization parameter free.

6 Implementation Strategy

In the context of a finite number of moment conditions, Newey and Smith (2004) show that GEL estimators have smaller higher order bias than the two-step GMM estimator of Hansen (1982). However, the computation of GEL estimators entails a saddle point problem taking the form of a double optimization program. When the dimensionality of the parameter space is large, saddle point problems are difficult to solve numerically (see Kitamura, 2006; Dovonon, 2016). This explains why so few applied researchers dare use empirical likelihood methods in their work.

In an effort to reduce the computational burden associated with GEL estimators, Antoine, Bonnal and Renault (2007) proposed an implementation strategy in three steps. The initial two steps are exactly those required by the solution of the efficient GMM estimator. In the third step, a first order condition similar to Equation (39) is solved with respect to θ where the implied probabilities, the gradient of the moment conditions and covariance matrix of the moment conditions are pre-evaluated at the two-step GMM estimator. The overall procedure is termed *Three-Steps Euclidean Likelihood*. In the context of a continuum of moment conditions, the computation of the Three-Steps Euclidean Likelihood estimator would involve the following steps:

Step 1: Compute the first step CGMM estimator:

$$\widehat{\theta}_{CGMM1} = \arg\min_\theta \int h(\tau,\theta)^2\pi(\tau)d\tau.$$

Step 2: Compute the second step CGMM estimator:

$$\widehat{\theta}_{CGMM2} = \arg\min_\theta \int h(\tau,\theta)\left(\widehat{K}\left(\widehat{\theta}_{CGMM1}\right)+\frac{\varepsilon}{n}I\right)^{-1}h(\tau,\theta)\pi(\tau)d\tau.$$

where $\widehat{K}\left(\widehat{\theta}_{CGMM1}\right)$ is the empirical covariance operator $\widehat{K}$ evaluated at $\widehat{\theta}_{CGMM1}$ (See Equation 26).

Step 3: Solve the following first order condition for the RGEL estimator:

$$\int \widetilde{G}_\varepsilon\left(\tau,\widehat{\theta}_{CGMM2}\right)\left(\widetilde{K}_\varepsilon\left(\widehat{\theta}_{CGMM2}\right)+\frac{\varepsilon}{n}I\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n w_{i,\varepsilon}h_i(\tau,\theta)\right)\pi(\tau)d\tau = 0,$$

where $\widetilde{K}_\varepsilon$ and $w_{i,\varepsilon}$ are defined as previously and:

$$\begin{aligned}
\widetilde{G}_\varepsilon(\tau,\theta) &= \frac{1}{n}\sum_{i=1}^{n}\left(1+\widehat{v}_{\varepsilon,i}(\theta)\right)\frac{\partial h_i(\tau,\theta)}{\partial\theta}, \\
\widehat{v}_{\varepsilon,i}(\theta) &= \widehat{\lambda}_{\varepsilon,0}(\theta)+\int h_i(\tau,\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)\,\pi(\tau)\,d\tau, \\
\widehat{\lambda}_{\varepsilon,0}(\theta) &= \int \widehat{h}(\tau,\theta)\left[\left(\widehat{K}(\theta)+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(\tau,\theta)\right]\pi(\tau)\,d\tau, \\
\widehat{\lambda}_{\varepsilon,1}(\tau,\theta) &= -\left(\widehat{K}(\theta)+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(\tau,\theta),
\end{aligned}$$

The idea behind the three-steps Euclidean Likelihood is to take advantage of the availability of closed form expressions for the dual parameters under the quadratic discrepancy to consistently estimate some of the terms that enter into the expression of the first order condition solved by the empirical likelihood estimator. Indeed, the computational burden associated with the GEL estimator is attributable to the fact that the expression of the dual parameters, as functions of θ, are not available in closed form.

It turns out that the three steps above can be shrunk into a single step by directly solving the equation:

$$\int \widetilde{G}_\varepsilon(\tau,\theta)\left(\widetilde{K}_\varepsilon(\theta)+\frac{\varepsilon}{n}I\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}w_{i,\varepsilon}h_i(\tau,\theta)\right)\pi(\tau)\,d\tau=0. \tag{58}$$

To evaluate the expression on the LHS, we approximate the dual parameters by their expressions under quadratic discrepancy. The empirical probabilities and all other quantities are computed using their formulas implied by the actual discrepancy function. In particular, we have:

$$\widetilde{G}_\varepsilon(\tau,\theta)=\frac{1}{n}\sum_{i=1}^{n}\left(1+\gamma\widehat{v}_{\varepsilon,i}(\theta)\right)^{1/\gamma}\frac{\partial h_i(\tau,\theta)}{\partial\theta},$$

where $\widehat{v}_{\varepsilon,i}(\theta)=\gamma\widehat{\lambda}_{\varepsilon,0}(\theta)+\gamma\int h_i(\tau,\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)\,\pi(\tau)\,d\tau$. This is the approach advocated for the Monte Carlo simulations presented in the next section.

Upon assigning a value to θ, one evaluates $\widehat{\lambda}_{\varepsilon,0}(\theta)$ and $\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)$ using the formulas above. Next, one used θ, $\widehat{\lambda}_{\varepsilon,0}(\theta)$ and $\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)$ to evaluate the empirical probabilities $p_{i,\varepsilon}$, $k(v_{i,\varepsilon})$, $k_{i,\varepsilon}$ and $w_{i,\varepsilon}$. Finally, the empirical probabilities are used to evaluate the RHS of (58). A numerical solver can therefore start from an initial value of θ (for instance, $\widehat{\theta}_{CGMM2}$) and iterate until convergence.

7 Monte Carlo Simulations

This simulation study is based on a linear heteroscedastic model used in Cragg (1983), Newey (1993), and Kitamura, Tripathi and Ahn (2004) and Lavergne and Patilea (2008). The first subsection presents the simulation design and the second subsection presents the results

7.1 Simulation Design

We consider the linear model:

$$y_t=X_t\theta_0+\varepsilon_t,\ t=1,...,T, \tag{59}$$

where:

$$\begin{aligned}\theta_0 &= (\theta_{01},\theta_{02},\theta_{03})' = (1,2,3)';\\ X_t &= (1, x_{1,t}, x_{2,t}),\ \ln x_{i,t} \sim N(0,1) \text{ for } i=1,2;\\ \varepsilon_t|X_t &\sim N(0,\sigma_t^2) \text{ with } \sigma_t^2 = 0.2 + 0.1\,(X_t\iota) + 0.05\,(X_t\iota)^2 \text{ and } \iota = (1,1,1)'.\end{aligned}$$

All observations are independent across t. Our goal is to estimate the parameter θ_0.

This model implies the following conditional moment restriction:

$$E\left(\varepsilon_t|X\right) \equiv E\left(y_t - X_t\theta|X\right) = 0.$$

Hence, one may consider estimating θ_0 using the continuum of unconditional moment conditions given by:

$$h_t\left(r,\theta\right) = \left(y_t - X_t\theta\right)\exp(-\tau_1 x_{1,t} - \tau_2 x_{2,t}),\ \text{for all } \tau = (\tau_1,\tau_2) \in \mathbb{R}^2_+. \tag{60}$$

Note that the regressors are strictly positive so that $\exp(-\tau_1 x_{1,t} - \tau_2 x_{2,t})$ is bounded for all $\tau \in \mathbb{R}^2_+$.

First, we compute the OLS estimator as:

$$\widehat{\theta}_{OLS} = \left(X'X\right)^{-1}X'Y,$$

where X is a $T \times 3$ matrix whose t^{th} row is X_t and Y is a $T \times 1$ vector whose t^{th} element is y_t.

Second, we compute a CGMM estimator using the identity operator as metrics along with exponential weights:

$$\widehat{\theta}_{CGMM1} = \arg\min_{\theta} \int_0^\infty \widehat{h}\left(\tau,\theta\right)^2 \pi\left(\tau\right) d\tau. \tag{61}$$

where $\pi\left(\tau\right) = e^{-\tau_1-\tau_2}$. The objective function is approximated using Gauss-Laguerre rule:

$$\int_0^\infty \widehat{h}\left(\tau,\theta\right)^2 \pi\left(\tau\right) d\tau \simeq \sum_{j=1}^{N^2} \omega_j \widehat{h}\left(\tau^{(j)},\theta\right)^2$$

where ω_j, $j = 1,...,N^2$ are quadrature weights associated with the quadrature points $\tau^{(j)}$ in $\mathbb{R}^2_+$. Indeed, the set $\tau^{(j)}, j = 1,...,N^2$ consists of pairs $(\tau_{1,k},\tau_{2,l})$ in a lexicographic order.

Third, we compute the efficient two-steps CGMM estimator of Carrasco and Florens (2000) as follows:

$$\widehat{\theta}_{CGMM2} = \arg\min_{\theta} \int_0^\infty h\left(\tau,\theta\right)\left(\widehat{K}\left(\widehat{\theta}_{CGMM1}\right) + \frac{\varepsilon}{n}I\right)^{-1} h\left(\tau,\theta\right)\pi\left(\tau\right) d\tau.$$

To approximate the inverse of $\widehat{K}\left(\theta\right) + \frac{\varepsilon}{n}I$, we first note that the quadrature rule implies:

$$\widehat{K}\widehat{h}\left(\tau,\theta\right) = \int_0^\infty \widehat{k}\left(r,\tau\right) h\left(s,\theta\right)\pi\left(s\right) d\left(s\right) \simeq \sum_{k=1}^{N^2} \omega_k \widehat{k}\left(s_k,\tau\right)\widehat{h}\left(s_k,\theta\right),$$

Let $\widehat{h}(\theta)$ be the vector $\left(\widehat{h}\left(s_1,\theta\right),\widehat{h}\left(s_2,\theta\right),...,\widehat{h}\left(s_{N^2},\theta\right)\right)'$ and $\widehat{W}$ the matrix with elements $W_{j,k} = \omega_k \widehat{k}\left(s_k,s_j\right)$ where ω_k is the weight associated with s_k. The previous equation can be rewritten as: $\widehat{K}\widehat{h}\left(\theta\right) \simeq \widehat{W}\widehat{h}\left(\theta\right)$, which indicates that $\widehat{W}$ is the matrix approximation of $\widehat{K}$ implied by the Gauss-Laguerre rule. Therefore:

$$\left(\widehat{K}\left(\theta\right) + \frac{\varepsilon}{n}I\right)^{-1}\widehat{h}_T\left(\theta\right) \simeq \left(\widehat{W} + \frac{\varepsilon}{n}I\right)^{-1}\widehat{h}\left(\theta\right).$$

Let $\widetilde{h}(\theta)$ denote the vector with k^{th} element given by $\widetilde{h}(s_k,\theta) = \omega_k \widehat{h}(s_k,\theta)$. The objective function of the two-steps CGMM is evaluated as:

$$\int_0^\infty h(\tau,\theta)\left(\widehat{K}\left(\widehat{\theta}_{CGMM1}\right)+\frac{\varepsilon}{n}I\right)^{-1} h(\tau,\theta)\,\pi(\tau)\,d\tau \simeq \widetilde{h}(\theta)'\left(\widehat{W}+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(\theta)$$

Finally, we compute the (*approximate*) RGEL estimator obtained by solving Equation (58) in one-step as explained in the previous section. The gradient of the moment function needed to evaluate this RHS of this equation is given by:

$$\frac{\partial h_t(r,\theta)}{\partial\theta} = -\exp(-\tau_1 x_{1,t} - \tau_2 x_{2,t})X_t.$$

The integrals involved in the RHS are evaluated using the same quadrature rule as previously.

We consider three different sample sizes ($T = 50$, $T = 250$ and $T = 1000$), three different values of γ ($\gamma - 0.95$, $\gamma = 1$ and $\gamma = 2$) and eight different values for $\epsilon = \frac{\varepsilon}{n}$ on the following grid:

$$\epsilon \in \left\{10^{-8}, 10^{-7}, ...10^{-2}, 10^{-1}\right\}.$$

The results presented in the next section are based on $M = 1000$ replications.

7.2 Simulation Results

Before comparing the performance of the estimators, we first examine the behavior of the empirical probabilities $p_{i,\varepsilon}$, $k(v_{i,\varepsilon})$, $k_{i,\varepsilon}$ and $w_{i,\varepsilon}$ as the sample size and the regularization parameter ($\epsilon = \frac{\varepsilon}{n}$) vary. Figure 1 shows the trajectory of the empirical probabilities computed using the samples simulated at the last replication and $\gamma = 0.95$. First, we note that $w_{i,\varepsilon}$ converges to uniform probabilities faster than $p_{i,\varepsilon}$, $k(v_{i,\varepsilon})$ and $k_{i,\varepsilon}$. Second, negative empirical probabilities are more likely to occur for small values of T and ϵ. All empirical probabilities eventually become positive as T increases for fixed ϵ or as ϵ increases for fixed T. This clearly indicates that the behavior of the empirical probabilities is closely related to the spectrum of the asymptotic covariance operator associated with the moment conditions. Indeed, the smallest eigenvalue of this covariance operator moves away from zero as T increases for fixed ϵ or as ϵ increases for fixed T. This behavior of the empirical probabilities implies that the optimal regularization parameter will taker larger values when the sample is smaller.

Figure 2 shows the root-mean square errors (RMSE) of the two-steps CGMM and RGEL estimators. For all values of T and γ, the RMSE of the two-steps CGMM estimator is minimized around $\epsilon = 10^{-8}$ (the smallest value of the grid) while the RMSE of the RGEL estimator is minimized between 10^{-3} and 10^{-1}.[8] This finding supports that the convergence rate of the optimal regularization parameter is slower for the RGEL than for the two-steps CGMM. This is not surprising given the semi-parametric nature of the GEL problem.[9] The RMSE of the RGEL estimator is minimized at relatively large values of the regularization parameter, in a region where all empirical probabilities are likely positive.

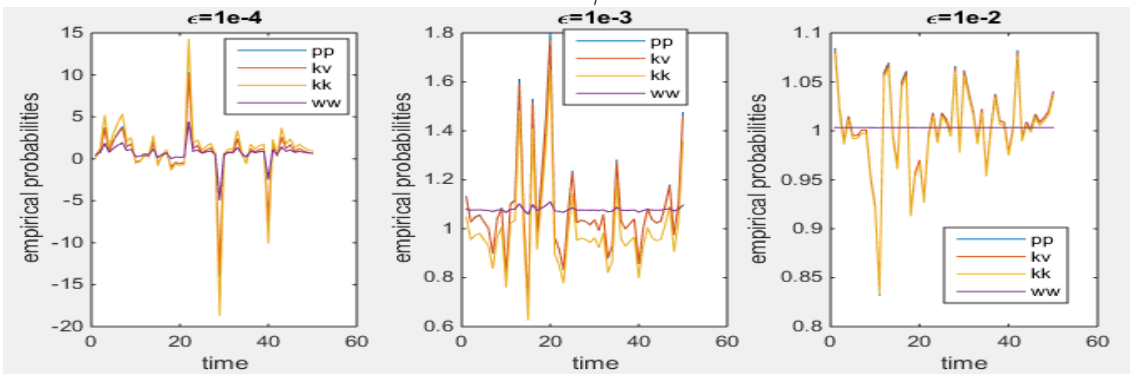
Figure 3 compares the RMSE of the RGEL estimator for different values of γ. It is seen that when ϵ is very small (e.g., 10^{-8}) so that negative empirical probabilities are highly prevalent, the RMSE of the RGEL estimator is decreasing in γ. This empirical finding is consistent with the theoretical results of Newey and Smith (2004) who show that the Empirical Likelihood estimator

[8] Note that the estimated value of the optimal regularization parameter depends on the precision of the computer used to perform the simulation (i.e., 32 bits, 64 bits, etc.).

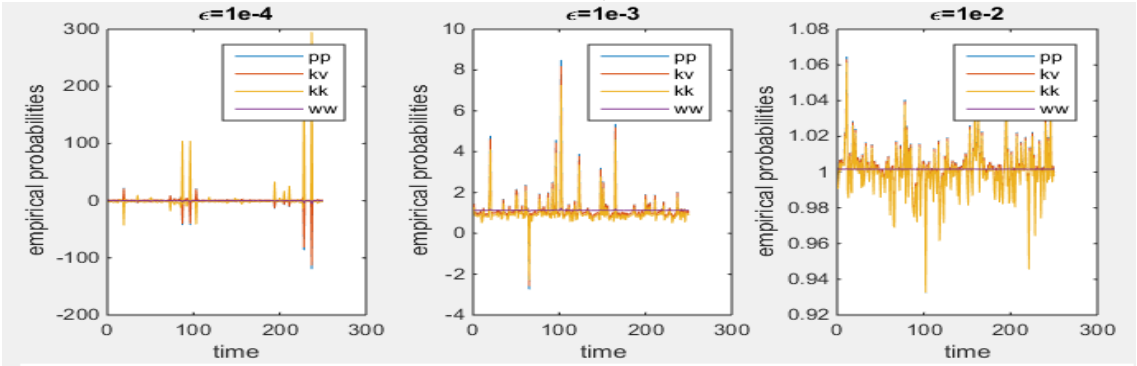
[9] Note that the dual parameter is infinite dimensional in the GEL with a continuum of moment conditions.

(i.e., the GEL estimator specialized to the logarithmic discrepancy) has the smallest higher order bias of all GEL estimators. This result may be interpreted by saying that the RGEL estimator is more and more resilient to negative empirical probabilities as γ decreases to -1. All RGEL estimators become equivalent in terms of RMSE as γ increases from -0.95 to 2 (the maximum value considered). However, the behavior of the RGEL estimator for small values of the regularization parameter prescribes to always use the logarithmic discrepancy, especially if the investigator does not consider selecting ϵ optimally.

Figure 1
Trajectories of empirical probabilities
$T = 50$ and $\gamma = 0.95$



$T = 250$ and $\gamma = 0.95$



$T = 1000$ and $\gamma = 0.95$

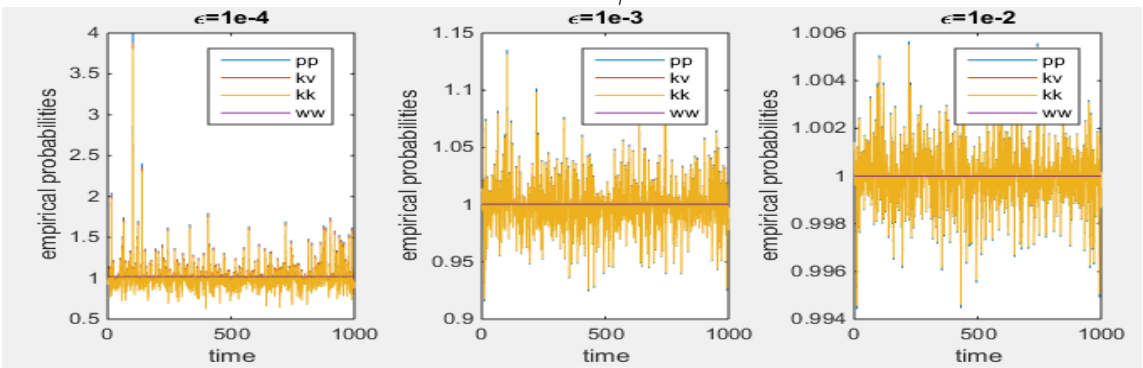
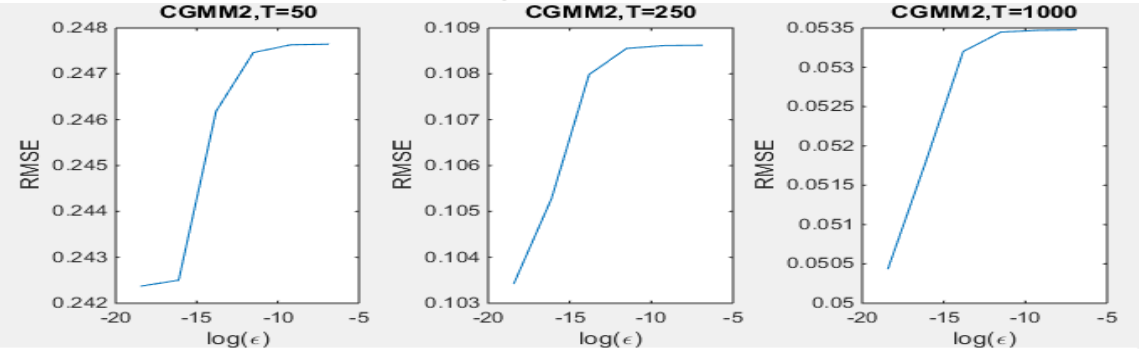


Figure 2
RMSE of the two-step CGMM and RGEL estimators
Two-step CGMM estimator



RGEL estimator

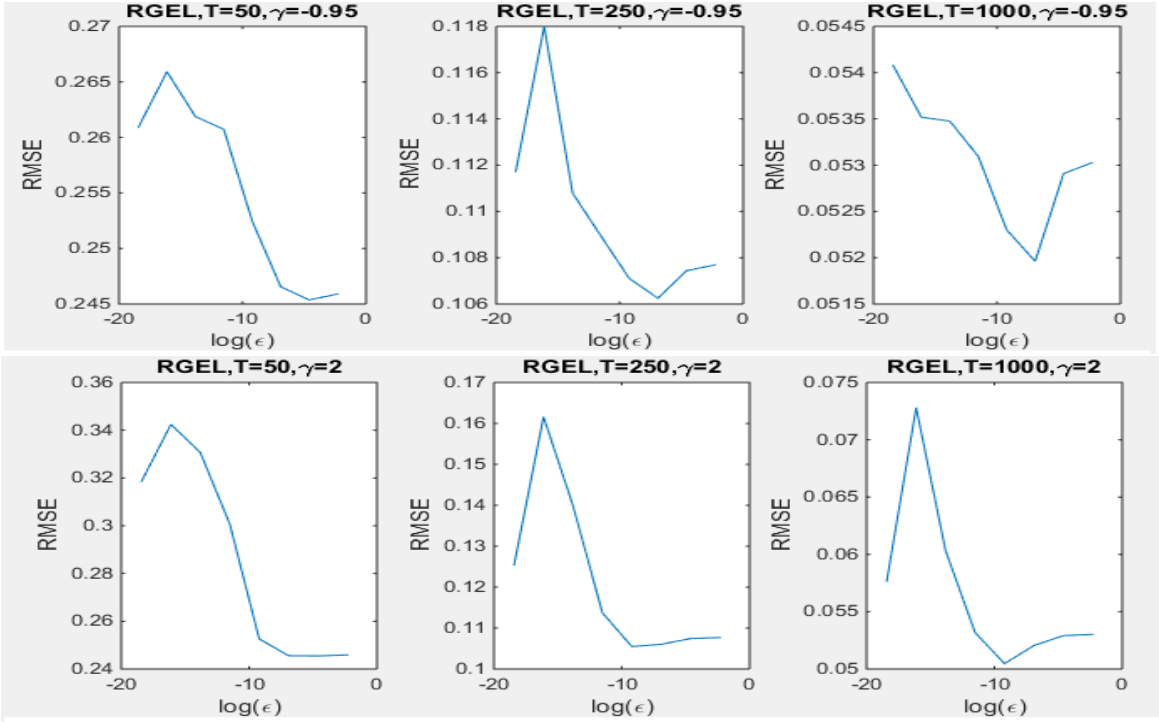


Figure 3
Sensitivity of the RGEL estimator to the choice of discrepancy function

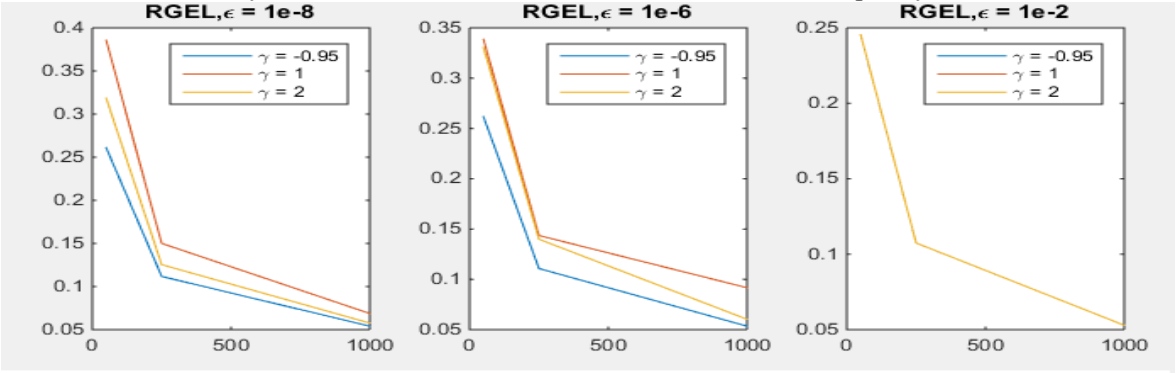
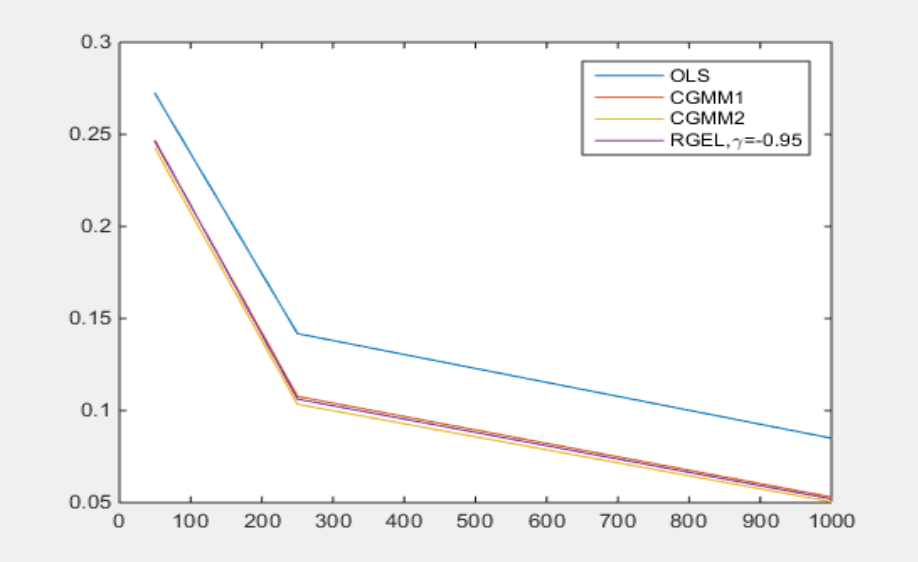


Figure 4 compares the RMSE of the OLS estimator, the first-step inefficient CGMM estimator, the second-step efficient CGMM estimator and the best RGEL estimators for different sample size. The ranking of the estimators remains the same for all sample sizes. The OLS estimator is the least efficient, which is not surprising given that the data generating process is quite heteroscedasticity. The performances of the other three estimators are roughly similar. If one looks through a magnifying glass, one discovers that the efficient two-step CGMM estimator slightly outperforms the RGEL estimator. On the other hand, the first-step CGMM slightly underperforms the RGEL. The main conclusion that emerges from these empirical results is that the RGEL estimator is $\sqrt{T}$ consistent and asymptotically equivalent to the efficient two-step CGMM estimator, as suggested by Theorem 9. The key point to note here is that efficiency is achieved by both estimators at regularization parameters that converge to zero at different rates.

Figure 4
Comparing the OLS, CGMM1, CGMM2 to the best RGEL estimator
(x-axis is the sample size)



8 Conclusion

This paper studies a class of regularized generalized empirical likelihood (RGEL) estimators that are suitable for models specified as a large number or a continuum of moment restrictions. In particular, we have in mind a situation where a conditional moment restriction is converted into a continuum of unconditional moment conditions. A duality relationship exists between the GEL estimation based on Cressie and Read (1984)'s discrepancy function and the minimum discrepancy estimation of Corcoran (1998). This duality relationship is a special case of a more general result established in Borwein and Lewis (1991) in the context of entropy-like minimization problems. When the number of moment restrictions is large or infinite as assumed in this paper, the system of equations represented by the moment restrictions becomes singular and the natural duality between the GEL and the MD problems breaks down. Indeed, Borwein (1992) showed that the maximum entropy solution to a system of an infinite number of constraints may exists while failing to satisfy the natural duality formula. Duality is restored by employing the relaxation scheme proposed by

Borwein. This approach leads to the class of Regularized Generalized Empirical Likelihood (RGEL) estimators studied in this paper.

We show that the RGEL estimator is asymptotically normally distributed and equivalent to the two-step CGMM estimator. An implementation strategy in one step inspired from the Three-Steps Empirical Likelihood of Antoine, Bonnal and Renault (2007) is proposed. Monte Carlo simulations based on a linear heteroskedastic model show that the RGEL estimator outperforms the first step CGMM estimator (in terms of RMSE) and slightly underperforms the efficient two-step CGMM estimator. We also find that the regularization parameter that permit to minimize the Root Mean Square Error of the RGEL converges to zero at a slower rate than the one needed to do the same for the two-step CGMM estimator.

References

[1] Almeida, C. and R. Garcia. 2012. Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators. *Journal of Econometrics* 170, 2, pp. 519–537.

[2] Antoine, B., Bonnal, H. and E. Renault. 2007. On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood. *Journal of Econometrics* 138, 461-487.

[3] Back, K. and D.P. Brown. 1993. Implied Probabilities in GMM Estimators. *Econometrica* 61, pp. 971-975.

[4] Bierens, H. 1982. Consistent model specification tests. *Journal of Econometrics* 20, pp. 105–134.

[5] Borwein, J.M. 1992. On the failure of maximum entropy reconstruction for Fredholm equations and other infinite systems. *Mathematical Programming*, 61, 1-3, pp. 251-261.

[6] Borwein, J.M. and A.S. Lewis. 1991. Duality Relationships for Entropy-Like Minimization Problems. SIAM *Journal of Control and Optimization*, 29, 2, pp. 325-338.

[7] Candes, E. and T. Tao. 2007. The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics* 35, 6, pp. 2313-2351.

[8] Carrasco, M. and J.P. Florens. 2014. On the Asymptotic Efficiency of GMM. *Econometric Theory* 30, 2, pp. 372-406.

[9] Carrasco, M., Florens, J.P. and E. Renault. 2007. Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization. In: *Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics*, vol. 6.

[10] Corcoran, S.A. 1998. Bartlett Adjustment of Empirical Discrepancy Statistics. *Biometrika* 85, pp 967-972.

[11] Cressie, N. and T. Read. 1984. Multinomial Goodness-of-Fit Tests. *Journal of the Royal Statistical Society, Series B.* 46. pp. 440-464.

[12] Dovonon, P. 2016. Large Sample Properties of the Three-Step Euclidean Likelihood Estimators under Model Misspecification. *Econometric Reviews* 35, 4, pp. 465-514.

[13] Dudik, M., Phillips, S. J. and R.E. Schapire. 2007. Maximum Entropy Density Estimation with Generalized Regularization and an Application to Species Distribution Modeling. *Journal of Machine Learning Research* 8, pp. 1217-1260.

[14] Gautier, E. and A.B. Tsybakov. 2014. High-dimensional instrumental variables regression and confidence sets. Available at: https://arxiv.org/abs/1105.2454

[15] Hansen, L.P. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50, pp. 1029-1054.

[16] Hansen, L.P., Heaton P. and A. Yaron. 1996. Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business & Economic Statistics* 14, pp. 362-280.

[17] Hansen L.P. and R. Jagannathan. 1997. Assessing Specification Errors in Stochastic Discount Factor Models. *Journal of Finance*, 52, 2, pp. 557-589.

[18] Imbens, W.G., Spady R. H. and P. Johnson. 1998. Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica* 66, pp. 333-357.

[19] Jagannathan, R. and Z. Wang. 1996. The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, pp. 3–53.

[20] Kan, R. and C. Robotti. 2009. Model Comparison Using the Hansen-Jagannathan Distance. *Review of Financial Studies*, Oxford University Press for Society for Financial Studies, 22(9), pp. 3449-3490.

[21] Kitamura, Y. 1997. Empirical Likelihood Methods with Weakly Dependent Processes. Annals of Statistics 25, pp. 2084-2102.

[22] Kitamura, Y. and M. Stutzer. 1997. An Information-Theorethic Alternative to Generalized Method of Moments Estimation. *Econometrica* 65, pp. 861-874.

[23] Kitamura, Y., Tripathi G. and H. Ahn. 2004. Empirical Likelihood-based Inference in Conditional Moment Restriction Models. *Econometrica* 72, pp. 1667-1714.

[24] Kitamura, Y. 2006. Empirical Likelihood Methods in Econometrics: Theory and Practice. *Cowles Foundation for Research in Economics*, Yale University, Discussion paper No. 1569.

[25] Kotchoni, R. 2014. The indirect continuous-GMM estimation. *Computational Statistics and Data Analysis* 76, pp. 464–488

[26] Lavergne, P., and V. Patilea. 2013. Smooth minimum distance estimation and testing with conditional estimating equations: uniform in bandwidth theory. *Journal of Econometrics* 177, 1, pp. 47–59

[27] Newey, K.W. and R.J. Smith. 2004. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* 72, pp. 219-255.

[28] Owen, A.B. 1988. Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika*, 75: 2, pp. 237-249.

[29] Owen, A.B. 1990. Empirical Likelihood Ratio Confident Regions. *Annals of Statistics* 18, pp. 90-120.

[30] Qin, J. and J.Lawless. 1994. Empirical Likelihood and General Estimating Equations. *Annals of Statistics* 22, pp. 300-325.

[31] Schennach, S.M. 2007. Point Estimation with Exponentially Tilted Empirical Likelihood. *Annals of Statistics* 35, pp. 634-672.

[32] Shi, Z. 2016. Econometric estimation with high-dimensional moment equalities. *Journal of Econometrics* 195, pp. 104–119

[33] Smith, R.J. 1997. Alternative Semi-parametric Likelihood Approaches to Generalized Method of Moments Estimation. *Economic Journal*, 107, pp. 503-519.

[34] Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B,* 58(1), pp. 267–288.

Appendix

Proof of Theorem 1. When $\gamma = 1$, the dual objective function is:

$$Q(\theta, \lambda) = n\lambda_0 - \frac{1}{2}\sum_{i=1}^{n}\left[\left(1 + \lambda_0 + g_i'\lambda_1\right)^2 + 1\right]$$

The result for the empirical probabilities stems from Equation (6):

$$\widehat{p}_i(\theta) = 1 + \lambda_0(\theta) + g_i'\lambda_1(\theta)$$

Next, consider the first order conditions of the maximization of $Q(\theta, \lambda)$ w.r.t λ_0 and λ_1:

$$n - \sum_{i=1}^{n}\left(1 + \lambda_0(\theta) + g_i'\lambda_1(\theta)\right) = 0 \text{ and } \sum_{i=1}^{n}\left(1 + \lambda_0(\theta) + g_i'\lambda_1(\theta)\right)g_i = 0$$

The first FOC gives immediately $\lambda_0(\theta) = -\widehat{g}(\theta)'\lambda_1(\theta)$, where $\widehat{g}(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_i$. Hence $\widehat{p}_i = 1 + (g_i - \widehat{g}(\theta))'\lambda_1(\theta)$. Replacing this expression of $\widehat{\lambda}_0(\theta)$ into the second FOC yields:

$$\sum_{i=1}^{n}\left(1 + \left(g_i' - \widehat{g}(\theta)'\right)\lambda_1(\theta)\right)g_i = \sum_{i=1}^{n} g_i + \sum_{i=1}^{n} g_i\left(g_i - \widehat{g}(\theta)\right)'\widehat{\lambda}_1(\theta) = 0.$$

Therefore:

$$\begin{aligned}\lambda_1(\theta) &= -\left[\sum_{i=1}^{n} g_i\left(g_i - \widehat{g}(\theta)\right)'\right]^{-1}\sum_{i=1}^{n} g_i = -\left[\frac{1}{n}g'(g - \overline{g})\right]^{-1}\widehat{g}(\theta) \\ &= -\left[\frac{1}{n}(g - \overline{g})'(g - \overline{g})\right]^{-1}\widehat{g}(\theta) = -\widehat{\Omega}^{-1}\widehat{g}(\theta)\end{aligned}$$

where $g - \overline{g}$ is the (n, m) matrix with i^{th} row given by $(g_i - \widehat{g}(\theta))'$. Consequently,

$$\widehat{p}_i(\theta) = 1 - (g_i - \widehat{g}(\theta))'\widehat{\Omega}^{-1}\widehat{g}(\theta) \text{ and } \lambda_0(\theta) = \widehat{g}(\theta)'\widehat{\Omega}^{-1}\widehat{g}(\theta)$$

Finally, replacing $\lambda_0(\theta)$ and $\lambda_1(\theta)$ in the objective function yields:

$$\begin{aligned}Q(\theta, \lambda) &= -n\widehat{g}(\theta)'\lambda_1(\theta) - \frac{1}{2}\sum_{i=1}^{n}\left[\left(1 + (g_i - \widehat{g}(\theta))'\lambda_1(\theta)\right)^2 + 1\right] \\ &= -n\widehat{g}(\theta)'\lambda_1(\theta) - n - \frac{1}{2}\lambda_1'(\theta)\left[\sum_{i=1}^{n}(g_i - \widehat{g}(\theta))(g_i - \widehat{g}(\theta))'\right]\lambda_1(\theta) \\ &= -n + \frac{n}{2}\widehat{g}(\theta)'\left[\frac{1}{n}(g - \overline{g})'(g - \overline{g})\right]^{-1}\widehat{g}(\theta) = -n + \frac{n}{2}\widehat{g}(\theta)'\widehat{\Omega}^{-1}\widehat{g}(\theta).\end{aligned}$$

Q.E.D. ■

Proof of Theorem 2. The first order condition for p:

$$\frac{\partial\mathcal{L}}{\partial p} = \varphi'(p) - \int A(\tau, \theta)'\lambda(\tau)\pi(\tau)\,d\tau = 0 \Rightarrow \widehat{p} = \varphi'^{-1}\left(\int A(\tau, \theta)'\lambda(\tau)\pi(\tau)\,d\tau\right),$$

where $\varphi'^{-1}(\varphi'(y)) = y$. By definition, we have $\varphi^*(y) = y'\widehat{p} - \varphi(\widehat{p})$ where $\widehat{p} = \varphi'^{-1}(y)$. Using the chain rule, it is straightforward to show that $\varphi^{*\prime}(y) = \varphi'^{-1}(y)$, so that:

$$\widehat{p} = \varphi^{*\prime}\left(\int A(\tau)'\lambda(\tau, \theta)\pi(\tau)\,d\tau\right),$$

where $\lambda(\tau,\theta)$ is the solution of:

$$\lambda(\tau,\theta) = \underset{\lambda(\tau,\theta),\tau\in\mathbb{R}}{\arg\sup} \left(\int \lambda(\tau)\pi(\tau)\,d\tau\right)' b + \varphi(\widehat{p}) - \left(\int A(\tau,\theta)'\lambda(\tau)\pi(\tau)\,d\tau\right)'\widehat{p}.$$

This problem is identical to:

$$\lambda(\tau,\theta) = \underset{\lambda(\tau,\theta),\tau\in\mathbb{R}}{\arg\sup}\, Q(\theta,\lambda),$$

and

$$Q(\theta,\lambda) = n\lambda_0 - \sum_{i=1}^{n}\varphi^*\left(\lambda_0 + \int h_i(\tau,\theta)\lambda_1(\tau)\pi(\tau)\,d\tau\right).$$

Q.E.D. ■

Proof of Theorem 3. The Lagrangian of the primal problem of interest is:

$$\widetilde{\mathcal{L}} = \sum_{i=1}^{n}\left(\frac{p_i^2-1}{2} - p_i\right) - \lambda_0\left(\sum_{i=1}^{n}p_i - n\right) - \sum_{k=1}^{m}\lambda_{1,k}\left(\sum_{i=1}^{n}p_i g_{i,k}\right) - \frac{\varepsilon}{2}\|\lambda_1\|^2$$

The first order condition for p_i is $p_i - 1 - \lambda_0 - g_i'\lambda_1 = 0$, which yields the first result: $\widehat{p}_i = 1 + \lambda_0(\theta) + g_i'\lambda_1(\theta)$. Substituting $\widehat{p}_i$ into the objective function yields:

$$\widetilde{\mathcal{L}} = n\lambda_0 - \frac{1}{2}\sum_{i=1}^{n}\left(1+\lambda_0+g_i'\lambda_1\right)^2 - \frac{n}{2} - \frac{\varepsilon}{2}\|\lambda_1\|^2 \equiv Q_\varepsilon(\theta,\lambda)$$

The first order conditions for the maximization of $Q_\varepsilon(\theta,\lambda)$ w.r.t λ are:

$$n - \sum_{i=1}^{n}\left(1+\lambda_0+g_i'\lambda_1\right) = 0 \text{ and } \sum_{i=1}^{n} g_i\left(1+\lambda_0+g_i'\lambda_1\right) + \varepsilon\lambda_1 = 0$$

The first equation yields the second result: $\lambda_0(\theta) = -\widehat{g}(\theta)'\lambda_1$. Substituting into the second FOC yields:

$$\sum_{i=1}^{n} g_i + \left(\sum_{i=1}^{n} g_i\left(g_i - \widehat{g}(\theta)\right)' + \varepsilon I\right)\lambda_1 = 0,$$

which yields the third result: $\lambda_{\varepsilon,1}(\theta) = -\left(\widehat{\Omega} + \frac{\varepsilon}{n}I\right)^{-1}\widehat{g}(\theta)$. This implies

$$\begin{aligned}\widehat{p}_i &\equiv \widehat{p}_{\varepsilon,i}(\theta) = 1 + \left(g_i - \widehat{g}(\theta)\right)'\lambda_{\varepsilon,1}(\theta) \text{ and}\\ \lambda_0(\theta) &\equiv \lambda_{\varepsilon,0}(\theta) = \widehat{g}(\theta)'\left(\widehat{\Omega} + \frac{\varepsilon}{n}I\right)^{-1}\widehat{g}(\theta).\end{aligned}$$

Finally, substituting $\lambda_{\varepsilon,0}$ and $\lambda_{\varepsilon,1}$ into the expression of $Q_\varepsilon(\theta,\lambda)$ yields:

$$\begin{aligned}Q_\varepsilon\left(\theta,\widehat{\lambda}_\varepsilon(\theta)\right) &= -n\widehat{g}(\theta)'\lambda_1 - n - \frac{1}{2}\lambda_1'\left[\sum_{i=1}^{n}\left(g_i - \widehat{g}(\theta)\right)\left(g_i - \widehat{g}(\theta)\right)' + \varepsilon I\right]\lambda_1\\ &= -n + \frac{n}{2}\widehat{g}(\theta)'\left(\widehat{\Omega} + \frac{\varepsilon}{n}I\right)^{-1}\widehat{g}(\theta).\end{aligned}$$

Q.E.D. ■

Proof of Theorem 4. The Lagrangian for the primal problem is:

$$\widetilde{\mathcal{L}} = \sum_{i=1}^{n}\left(\frac{p_i^{\gamma+1}-1}{(\gamma+1)\gamma} - \frac{p_i}{\gamma}\right) - \lambda_0\left(\sum_{i=1}^{n} p_i - n\right) - \sum_{k=1}^{m}\lambda_{1,k}\left(\sum_{i=1}^{n} p_i g_{i,k}(\theta)\right) - \frac{\varepsilon}{2}\|\lambda_1\|^2$$

The first order condition for p_i is: $\frac{p_i^{\gamma}}{\gamma} - \frac{1}{\gamma} - \lambda_0 - g_i'(\theta)\lambda_1 = 0$. Solving for p_i yields the first result:

$$p_{i,\varepsilon} = \left(1 + \gamma\lambda_{0,\varepsilon} + \gamma g_i'(\theta)\lambda_{1,\varepsilon}\right)^{\frac{1}{\gamma}} = \left(1+\gamma v_{i,\varepsilon}\right)^{\frac{1}{\gamma}} = 1 + v_{i,\varepsilon}k(v_{i,\varepsilon}).$$

where $v_{i,\varepsilon} \equiv \lambda_{0,\varepsilon} + g_i'\lambda_{1,\varepsilon}$ and where $k(v_{i,\varepsilon}) \equiv \left(\left(1+\gamma v_{i,\varepsilon}\right)^{\frac{1}{\gamma}} - 1\right)/v_{i,\varepsilon}$. The dual objective function (**??**) is obtained by substituting $p_{i,\varepsilon}$ into $\widetilde{\mathcal{L}}$. The first order conditions for λ_0 and λ_1 yields:

$$n - \sum_{i=1}^{n}\left(1+\gamma v_{i,\varepsilon}\right)^{\frac{1}{\gamma}} = 0 \text{ and } \sum_{i=1}^{n}\left(1+\gamma v_{i,\varepsilon}\right)^{\frac{1}{\gamma}} g_i'(\theta) + \varepsilon\lambda_{1,\varepsilon} = 0.$$

The first equation implies:

$$0 = \sum_{i=1}^{n}\left(1 - \left(1+\gamma v_{i,\varepsilon}\right)^{\frac{1}{\gamma}}\right) = \sum_{i=1}^{n} v_{i,\varepsilon}k(v_{i,\varepsilon}) = \lambda_{0,\varepsilon}\sum_{i=1}^{n} k(v_{i,\varepsilon}) + \sum_{i=1}^{n} k(v_{i,\varepsilon})g_i'\lambda_{1,\varepsilon}.$$

Solving for $\lambda_{0,\varepsilon}$ yields

$$\lambda_{0,\varepsilon} = -\left(\frac{1}{n}\sum_{i=1}^{n} k_{i,\varepsilon}g_i'\right)\lambda_{1,\varepsilon},$$

where $k_{i,\varepsilon} = nk(v_{i,\varepsilon})/\sum_{j=1}^{n}k(v_{j,\varepsilon})$. The second equation implies:

$$0 = \sum_{i=1}^{n} p_{i,\varepsilon}g_i(\theta) + \varepsilon\lambda_{1,\varepsilon} = \sum_{i=1}^{n} v_{i,\varepsilon}k(v_{i,\varepsilon})g_i(\theta) + \varepsilon\lambda_{1,\varepsilon} + n\widehat{g}(\theta),$$

Hence:

$$\begin{aligned}
0 &= \left(\sum_{i=1}^{n} k(v_{i,\varepsilon})g_i g_i' + \varepsilon I\right)\lambda_{1,\varepsilon} + \sum_{i=1}^{n}\left(1+\lambda_{0,\varepsilon}k(v_{i,\varepsilon})\right)g_i \\
&= \left(\sum_{i=1}^{n} k(v_{i,\varepsilon})\widehat{\Omega}_i + \varepsilon I\right)\lambda_{1,\varepsilon} + \sum_{i=1}^{n}\left(1+\widehat{\lambda}_{0,\varepsilon}k(v_{i,\varepsilon})\right)g_i \\
&\quad + \left(\sum_{i=1}^{n} k(v_{i,\varepsilon})\left(g_i\widehat{g}' + \widehat{g}g_i' - \widehat{g}\widehat{g}'\right)\right)\lambda_{1,\varepsilon} \\
&= \left(\sum_{i=1}^{n} k(v_{i,\varepsilon})\widehat{\Omega}_i + \varepsilon I\right)\lambda_{1,\varepsilon} \\
&\quad + \sum_{i=1}^{n}\left(1 + \overline{v}_{\varepsilon}k(v_{i,\varepsilon}) + \frac{1}{n}\sum_{j=1}^{n} k(v_{j,\varepsilon})\left(g_j - \widehat{g}\right)'\lambda_{1,\varepsilon}\right)g_i,
\end{aligned}$$

where $\widehat{\Omega}_i \equiv \left(g_i - \widehat{g}\right)\left(g_i - \widehat{g}\right)'$ and $\overline{v}_{\varepsilon} \equiv \widehat{\lambda}_{0,\varepsilon} + \widehat{g}'\lambda_{1,\varepsilon}$. Solving for $\lambda_{1,\varepsilon}$ yields:

$$\lambda_{1,\varepsilon}(\theta) = -\left(\frac{1}{n}\sum_{i=1}^{n} k(v_{i,\varepsilon})\widehat{\Omega}_i + \frac{\varepsilon}{n}I\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} w_{i,\varepsilon}g_i\right),$$

where $w_{i,\varepsilon} = 1+\overline{v}_\varepsilon k(v_{i,\varepsilon})+\frac{1}{n}\sum_{j=1}^n k(v_{j,\varepsilon})(g_j-\widehat{g})'\lambda_{1,\varepsilon}$. Replacing $\lambda_{1,\varepsilon}$ into $\lambda_{0,\varepsilon} = -\left(\frac{1}{n}\sum_{i=1}^n k_{i,\varepsilon}g_i'\right)\lambda_{1,\varepsilon}$ yields the expression provided in the theorem for $\lambda_{0,\varepsilon}$. Applying the Envelope Theorem to $Q_\varepsilon(\theta,\lambda_{1,\varepsilon}(\theta))$ yields $\sum_{i=1}^n \widehat{p}_i G_i'\widehat{\lambda}_{1,\varepsilon} = 0$. Substituting $\widehat{\lambda}_{1,\varepsilon}$ and normalizing by $\frac{1}{n}$ yields the last result. Q.E.D. ■

Proof of Theorem 5. When φ is quadratic, its convex conjugate is given by $\varphi^*(x) = \frac{(1+x)^2+1}{2}$. Therefore, $\varphi^{*\prime}(x) = 1+x$ so that:

$$p_i = \varphi^{*'}\left(\lambda_0 + \int h_i(\tau,\theta)\lambda_1(\tau,\theta)\pi(\tau)d\tau\right) = 1+\lambda_0+\int h_i(\tau,\theta)\lambda_1(\tau,\theta)\pi(\tau)d\tau$$

where $\lambda(\tau,\theta) = \underset{\lambda(\tau,\theta),\tau\in\mathbb{R}}{\arg\sup}\, Q(\theta,\lambda)$ and

$$\begin{aligned} Q(\theta,\lambda) &= n\lambda_0 - \frac{1}{2}\sum_{i=1}^n\left(1+\lambda_0+\int h_i(\tau,\theta)\lambda_1(\tau)\pi(\tau)d\tau\right)^2 \\ &\quad -\frac{n}{2}-\frac{\varepsilon}{2}\int\lambda_1^2(\tau)\pi(\tau)d\tau \end{aligned}$$

The first order conditions solved by λ_0 and $\lambda_1(\tau)$ are:

$$\begin{aligned} n-\sum_{i=1}^n\left(1+\lambda_0+\int h_i(\tau,\theta)\lambda_1(\tau)\pi(\tau)d\tau\right) &= 0 \\ \sum_{i=1}^n\left(1+\lambda_0+\int h_i(\tau,\theta)\lambda_1(\tau)\pi(\tau)d\tau\right)h_i(r,\theta)+\varepsilon\lambda_1(r) &= 0, \end{aligned}$$

for all $r\in\mathbb{R}$. The first equation yields

$$\widehat{\lambda}_0(\theta) = -\int\widehat{h}(\tau,\theta)\lambda_1(\tau)\pi(\tau)d\tau$$

where $\widehat{h}(\tau,\theta) = \frac{1}{n}\sum_{i=1}^n h_i(\tau,\theta)$. Substituting $\widehat{\lambda}_0(\theta)$ into the second equation yields:

$$\left(\widehat{K}(\theta)+\frac{\varepsilon}{n}I\right)\lambda_1(r) = -\widehat{h}(r,\theta),$$

where $\widehat{K}(\theta)$ is the empirical covariance operator associated with the moment conditions and I is the identity operator. Solving this equation yields:

$$\widehat{\lambda}_{\varepsilon,1}(r) = -\left(\widehat{K}(\theta)+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(r,\theta).$$

Substituting $\widehat{\lambda}_{\varepsilon,1}(\tau)$ into the expression of $\widehat{\lambda}_0(\theta)$ yields:

$$\widehat{\lambda}_{\varepsilon,0}(\theta) = \int\widehat{h}(\tau,\theta)\left[\left(\widehat{K}(\theta)+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(\tau,\theta)\right]\pi(\tau)d\tau$$

Substituting $\widehat{\lambda}_{\varepsilon,0}(\theta)$ and $\widehat{\lambda}_{\varepsilon,1}(r)$ into the expression of $Q\left(\theta,\widehat{\lambda}_{\varepsilon}(\theta)\right)$ yields.

$$\begin{aligned}
Q\left(\theta,\widehat{\lambda}_{\varepsilon}\right) &= n\widehat{\lambda}_{\varepsilon,0}-\frac{1}{2}\sum_{i=1}^{n}\left(1+\widehat{\lambda}_{\varepsilon,0}+\int h_i(\tau,\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)\,\pi(\tau)\,d\tau\right)^2 \\
&\quad -\frac{n}{2}-\frac{\varepsilon}{2}\int\widehat{\lambda}_{\varepsilon1}^2(\tau,\theta)\,\pi(\tau)\,d\tau \\
&= n\widehat{\lambda}_{\varepsilon,0}-\frac{n}{2}-\frac{n}{2}\widehat{\lambda}_{\varepsilon,0}^2-\frac{1}{2}\sum_{i=1}^{n}\left(\int h_i(\tau,\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)\,\pi(\tau)\,d\tau\right)^2 \\
&\quad -n\widehat{\lambda}_{\varepsilon,0}-\sum_{i=1}^{n}\int h_i(\tau,\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)\,\pi(\tau)\,d\tau \\
&\quad -\widehat{\lambda}_{\varepsilon,0}\sum_{i=1}^{n}\int h_i(\tau,\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)\,\pi(\tau)\,d\tau-\frac{n}{2}-\frac{\varepsilon}{2}\int\widehat{\lambda}_{\varepsilon1}^2(\tau,\theta)\,\pi(\tau)\,d\tau \\
&= -n+n\widehat{\lambda}_{\varepsilon,0}+\frac{n}{2}\widehat{\lambda}_{\varepsilon,0}^2-\frac{1}{2}\sum_{i=1}^{n}\left(\int h_i(\tau,\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau,\theta)\,\pi(\tau)\,d\tau\right)^2 \\
&\quad -\frac{\varepsilon}{2}\int\widehat{\lambda}_{\varepsilon1}^2(\tau,\theta)\,\pi(\tau)\,d\tau
\end{aligned}$$

■

But note that:

$$\begin{aligned}
&\sum_{i=1}^{n}\left(\int h_i(\tau,\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau)\,\pi(\tau)\,d\tau\right)^2 \\
&= n\int\widehat{K}(\theta)\,\widehat{\lambda}_{\varepsilon,1}(\tau)\,\widehat{\lambda}_{\varepsilon,1}(\tau)\,\pi(\tau)\,d\tau+n\widehat{\lambda}_{\varepsilon,0}^2 \\
&= n\int\left(\widehat{K}(\theta)+\frac{\varepsilon}{n}\right)\widehat{\lambda}_{\varepsilon,1}(\tau)\,\widehat{\lambda}_{\varepsilon,1}(\tau)\,\pi(\tau)\,d\tau+n\widehat{\lambda}_{\varepsilon,0}^2-\varepsilon\int\widehat{\lambda}_{\varepsilon1}^2(\tau,\theta)\,\pi(\tau)\,d\tau \\
&= n\widehat{\lambda}_{\varepsilon,0}+n\widehat{\lambda}_{\varepsilon,0}^2-\varepsilon\int\widehat{\lambda}_{\varepsilon1}^2(\tau,\theta)\,\pi(\tau)\,d\tau
\end{aligned}$$

Finally,

$$\begin{aligned}
Q\left(\theta,\widehat{\lambda}_{\varepsilon}(\theta)\right) &= -n+\frac{n}{2}\widehat{\lambda}_{\varepsilon,0}^2-\frac{1}{2}\left(n\widehat{\lambda}_{\varepsilon,0}+n\widehat{\lambda}_{\varepsilon,0}^2-\varepsilon\int\widehat{\lambda}_{\varepsilon1}^2(\tau,\theta)\,\pi(\tau)\,d\tau\right) \\
&\quad +n\widehat{\lambda}_{\varepsilon,0}-\frac{\varepsilon}{2}\int\widehat{\lambda}_{\varepsilon1}^2(\tau,\theta)\,\pi(\tau)\,d\tau \\
&= -n+\frac{n}{2}\int\widehat{h}(\tau,\theta)\left[\left(\widehat{K}(\theta)+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(\tau,\theta)\right]\pi(\tau)\,d\tau
\end{aligned}$$

Q.E.D.

Proof of Theorem 6. Note that:

$$\widehat{p}_i=\varphi^{*'}\left(\widehat{\lambda}_0(\theta)+\int h_i(\tau,\theta)\,\widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\,\pi(\tau)\,d\tau\right)=(1+\gamma v_{i,\varepsilon})^{\frac{1}{\gamma}},$$

where $\varphi^*(x) = \frac{(1+\gamma x)^{\frac{1+\gamma}{\gamma}}+1}{1+\gamma}$ is the convex conjugate of φ and:

$$\begin{aligned} v_{i,\varepsilon} &= \widehat{\lambda}_{0,\varepsilon} + \int h_i(\tau,\theta)\,\widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\,\pi(\tau)\,d\tau, \\ \widehat{\lambda}_\varepsilon(\tau,\theta) &= \underset{\lambda(\tau),\tau\in\mathbb{R}}{\arg\sup} Q_\varepsilon(\theta,\lambda) \text{ and} \\ Q_\varepsilon(\theta,\lambda) &= n\lambda_0 - \frac{1}{1+\gamma}\sum_{i=1}^n \left(1+\gamma\lambda_0+\gamma\int h_i(\tau,\theta)\,\lambda_1(\tau)\,\pi(\tau)\,d\tau\right)^{\frac{1+\gamma}{\gamma}} \\ &\quad -\frac{n}{1+\gamma} - \frac{\varepsilon}{2}\int \lambda_1^2(\tau)\,\pi(\tau)\,d\tau. \end{aligned}$$

The first order conditions solved by $\widehat{\lambda}_\varepsilon(\tau,\theta) = \left(\widehat{\lambda}_{0,\varepsilon}, \widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\right)$ are:

$$\begin{aligned} n - \sum_{i=1}^n (1+\gamma v_{i,\varepsilon})^{\frac{1}{\gamma}} &= 0 \text{ and} \\ \sum_{i=1}^n (1+\gamma v_{i,\varepsilon})^{\frac{1}{\gamma}} h_i(r,\theta) + \varepsilon\widehat{\lambda}_{1,\varepsilon}(r) &= 0. \end{aligned}$$

The first equation implies:

$$\begin{aligned} 0 &= n - \sum_{i=1}^n (v_{i,\varepsilon}k(v_{i,\varepsilon})+1) = -\sum_{i=1}^n v_{i,\varepsilon}k(v_{i,\varepsilon}) \\ &= -\widehat{\lambda}_{0,\varepsilon}\sum_{i=1}^n k(v_{i,\varepsilon}) - \sum_{i=1}^n k(v_{i,\varepsilon})\int h_i(\tau,\theta)\,\widehat{\lambda}_1(\tau,\theta)\,\pi(\tau)\,d\tau \end{aligned}$$

where $k(x) \equiv \left((1+\gamma x)^{\frac{1}{\gamma}} - 1\right)/x$. Solving for $\widehat{\lambda}_{0,\varepsilon}$ yields

$$\widehat{\lambda}_{0,\varepsilon} = -\int \widetilde{h}(\tau,\theta)\,\widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\,\pi(\tau)\,d\tau,$$

where $\widetilde{h}(\tau,\theta) = \frac{1}{n}\sum_{i=1}^n k_{i,\varepsilon}h_i(\tau,\theta)$ and $k_{i,\varepsilon} = nk(v_{i,\varepsilon})/\sum_{j=1}^n k(v_{j,\varepsilon})$. The first order conditions for $\widehat{\lambda}_1(r,\theta)$ implies:

$$\begin{aligned} 0 &= \sum_{i=1}^n (v_{i,\varepsilon}k(v_{i,\varepsilon})+1)\,h_i(r,\theta) + \varepsilon\widehat{\lambda}_{1,\varepsilon}(r,\theta) \\ &= \sum_{i=1}^n k(v_{i,\varepsilon})\int h_i(r,\theta)\,h_i(\tau,\theta)\,\widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\,\pi(\tau)\,d\tau + \varepsilon\widehat{\lambda}_{1,\varepsilon}(r,\theta) \\ &\quad + \sum_{i=1}^n \left(1+\widehat{\lambda}_{0,\varepsilon}k(v_{i,\varepsilon})\right)h_i(r,\theta). \end{aligned}$$

Note that

$$\begin{aligned}
&\sum_{i=1}^{n} k(v_{i,\varepsilon}) \int h_i(r,\theta)\, h_i(\tau,\theta)\, \widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\, \pi(\tau)\, d\tau + \varepsilon \widehat{\lambda}_{1,\varepsilon}(r,\theta) \\
= \;& \left(n\widetilde{K}_{\varepsilon}(\theta) + \varepsilon I\right) \widehat{\lambda}_{1,\varepsilon}(r,\theta) \\
&+ \sum_{i=1}^{n} k(v_{i,\varepsilon}) \left\{ \int \widehat{h}(\tau,\theta)\, \widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\, \pi(\tau)\, d\tau \right\} h_i(r,\theta) \\
&+ \sum_{j=1}^{n} k(v_{j,\varepsilon}) \left\{ \int \left(h_j(\tau,\theta) - \widehat{h}(\tau,\theta)\right) \widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\, \pi(\tau)\, d\tau \right\} \widehat{h}(r,\theta)
\end{aligned}$$

where $\widetilde{K}_{\varepsilon}(\theta)$ is the linear operator with kernel $\widetilde{k}(r,\tau)$ given by:

$$\widetilde{k}(r,\tau) = \frac{1}{n} \sum_{i=1}^{n} k(v_{i,\varepsilon}) \left(h_i(r,\theta) - \widehat{h}(r,\theta)\right) \left(h_i(\tau,\theta) - \widehat{h}(\tau,\theta)\right).$$

Substituting into the first order condition yields:

$$0 = \left(n\widetilde{K}_{\varepsilon}(\theta) + \varepsilon I\right) \widehat{\lambda}_{1,\varepsilon}(r,\theta) + \sum_{i=1}^{n} w_{i,\varepsilon} h_i(r,\theta)$$

where

$$w_{i,\varepsilon} = 1 + \overline{v}_{\varepsilon} k(v_{i,\varepsilon}) + \frac{1}{n} \sum_{j=1}^{n} k(v_{j,\varepsilon}) \int \left(h_j(\tau,\theta) - \widehat{h}(\tau,\theta)\right) \widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\, \pi(\tau)\, d\tau$$

and $\overline{v}_{\varepsilon} = \widehat{\lambda}_{0,\varepsilon} + \int \widehat{h}(\tau,\theta)\, \widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\, \pi(\tau)\, d\tau$. Solving for $\widehat{\lambda}_{1,\varepsilon}(r,\theta)$ leads to:

$$\begin{aligned}
\widehat{\lambda}_{1,\varepsilon}(r,\theta) &= -\frac{1}{n} \sum_{i=1}^{n} w_{i,\varepsilon} \left(\widetilde{K}_{\varepsilon}(\theta) + \frac{\varepsilon}{n} I\right)^{-1} h_i(r,\theta) \\
&= -\left(\widetilde{K}_{\varepsilon}(\theta) + \frac{\varepsilon}{n} I\right)^{-1} \widetilde{\widetilde{h}}(\tau,\theta),
\end{aligned}$$

where $\widetilde{\widetilde{h}}(\tau,\theta) = \frac{1}{n} \sum_{i=1}^{n} w_{i,\varepsilon} h_i(\tau,\theta)$. Replacing into $\widehat{\lambda}_{0,\varepsilon}$ yields:

$$\widehat{\lambda}_{0,\varepsilon} = \int \widetilde{h}(\tau,\theta) \left(\widetilde{K}_{\varepsilon}(\theta) + \frac{\varepsilon}{n} I\right)^{-1} \widetilde{\widetilde{h}}(\tau,\theta)\, \pi(\tau)\, d\tau.$$

To obtain the first order condition solved by θ, we apply the envelope theorem to $Q\left(\theta, \widehat{\lambda}_{\varepsilon}(\theta)\right)$ and normalize by $\frac{1}{n}$. This leads to:

$$\int \widetilde{G}(\tau,\theta)\, \widehat{\lambda}_{1,\varepsilon}(\tau)\, \pi(\tau)\, d\tau = 0,$$

where $\widetilde{G}(\tau,\theta) = \frac{1}{n} \sum_{i=1}^{n} \widehat{p}_i \frac{\partial h_i(\tau,\theta)}{\partial \theta}$. Finally, substituting $\widehat{\lambda}_{1,\varepsilon}(\tau)$ yields:

$$\int \widetilde{G}(\tau,\theta) \left(\widetilde{K}_{\varepsilon}(\theta) + \frac{\varepsilon}{n} I\right)^{-1} \widetilde{\widetilde{h}}(\tau,\theta)\, \pi(\tau)\, d\tau = 0.$$

Q.E.D. ■

Subsequently, we consider a shrinking neighborhood of zero define as:

$$\Lambda_n = \left\{ \lambda \in \mathbb{R} \times H : \left(\lambda_0^2 + \int \lambda_1^2 (\tau)\, \pi(\tau)\, d\tau \right)^{1/2} < n^{-\zeta} \right\}$$

for some $\zeta \in [1/\alpha, 1/2]$, $\zeta > 0$, where H is the Hilbert space to which $\lambda_1(\tau)$ belongs. Analogues of Lemmas A2 and A3 of NS (2004) are established below.

Lemma 10 (Lemma 1) *Suppose* $\overline{\theta} \in \Theta$, $\underset{n\to\infty}{P\lim} \overline{\theta} = \theta_0$, $\widehat{h}(\tau, \overline{\theta}) = O_p\left(n^{-1/2}\right)$ *for all* τ *and* $\varepsilon = o(1)$. *Then* $\overline{\lambda}_\varepsilon(\overline{\theta}) = \underset{\lambda \in \Lambda_n(\overline{\theta})}{\arg\max} Q_\varepsilon(\overline{\theta}, \lambda)$ *exists with probability approaching one,* $\left\|\overline{\lambda}_{1,\varepsilon}(\overline{\theta})\right\| = O_p\left(\varepsilon^{-1} n^{1/2}\right)$, $\overline{\lambda}_{0,\varepsilon}(\overline{\theta}) = O_p\left(\varepsilon^{-1}\right)$ *and:*

$$\sup_{\lambda \in \Lambda_n(\overline{\theta})} Q_\varepsilon(\overline{\theta}, \lambda) \leq Q_\varepsilon(\overline{\theta}, 0) + O_p\left(\varepsilon^{-2} n^2\right),$$

as $n \to \infty$, *with* $\left\|\overline{\lambda}_{1,\varepsilon}(\overline{\theta})\right\| = \left(\int \overline{\lambda}_{1,\varepsilon}^2(\tau, \overline{\theta})\, \pi(\tau)\, d\tau \right)^{1/2}$.

Proof of Lemma 1. According to Assumption 1(a), we have:

$$\begin{aligned}
& \sup_{\theta \in \Theta, \lambda \in \Lambda_n, 1 \leq i \leq n} \left| \int h_i(\tau, \theta)\, \lambda_1(\tau)\, \pi(\tau)\, d\tau \right| \\
\leq & \sup_{\theta \in \Theta, \lambda \in \Lambda_n} \left(\int \lambda_1^2(\tau)\, \pi(\tau)\, d\tau \right)^{1/2} \max_{1 \leq i \leq n} \left(\int h_i^2(\tau, \theta)\, \pi(\tau)\, d\tau \right)^{1/2} \\
\leq & n^{-\zeta} \max_{1 \leq i \leq n} \left(\int h_i^2(\tau, \theta)\, \pi(\tau)\, d\tau \right)^{1/2} = O_p\left(n^{-\zeta + 1/\alpha}\right)
\end{aligned}$$

By the triangular inequality, we have:

$$\begin{aligned}
\left| \lambda_0 + \int h_i(\tau, \theta)\, \lambda_1(\tau)\, \pi(\tau)\, d\tau \right| &\leq |\lambda_0| + \left| \int h_i(\tau, \theta)\, \lambda_1(\tau)\, \pi(\tau)\, d\tau \right| \\
&= n^{-\zeta} + O_p\left(n^{-\zeta + 1/\alpha}\right) = o_p(1),
\end{aligned}$$

As zero is an interior point of the domain φ_i^*, $\Lambda_n \subseteq \Lambda_n(\theta)$ with probability approaching one for all $\theta \in \Theta$ and $\lambda \in \Lambda_n$. Therefore, $\overline{\lambda}_\varepsilon = \underset{\lambda \in \Lambda_n}{\arg\max} Q_\varepsilon(\overline{\theta}, \lambda)$ exists with probability approaching one. To pursue, we need a second order expansion of $Q_\varepsilon(\theta, \lambda)$ around $\lambda = 0$. We have:

$$\begin{aligned}
Q_\varepsilon(\theta, \lambda) &= n\lambda_0 - \frac{1}{1+\gamma} \sum_{i=1}^{n} (1 + \gamma v_i)^{\frac{1+\gamma}{\gamma}} - \frac{n}{1+\gamma} - \frac{\varepsilon}{2} \int \lambda_1^2(\tau)\, \pi(\tau)\, d\tau, \\
v_i &= \lambda_0 + \int h_i(\tau, \theta)\, \lambda_1(\tau)\, \pi(\tau)\, d\tau,
\end{aligned}$$

The derivatives needed for the expansion are given by:

$$\begin{aligned}
\frac{\partial Q_\varepsilon(\theta,\lambda)}{\partial\lambda_0} &= n-\sum_{i=1}^{n}(1+\gamma v_i)^{\frac{1}{\gamma}}, \\
\frac{\partial Q_\varepsilon(\theta,\lambda)}{\partial\lambda_1(r)} &= -\sum_{i=1}^{n}(1+\gamma v_i)^{\frac{1}{\gamma}} h_i(r,\theta)\pi(r)\,dr-\varepsilon\lambda_1(r)\pi(r)\,dr, \\
\frac{\partial^2 Q_\varepsilon(\theta,\lambda)}{\partial\lambda_0^2} &= -\sum_{i=1}^{n}(1+\gamma v_i)^{\frac{1-\gamma}{\gamma}}, \\
\frac{\partial^2 Q_\varepsilon(\theta,\lambda)}{\partial\lambda_0\partial\lambda_1(r)} &= -\sum_{i=1}^{n}(1+\gamma v_i)^{\frac{1-\gamma}{\gamma}} h_i(r,\theta)\pi(r)\,dr, \\
\frac{\partial^2 Q_\varepsilon(\theta,\lambda)}{\partial(\lambda_1(r))^2} &= -\sum_{i=1}^{n}(1+\gamma v_i)^{\frac{1-\gamma}{\gamma}} (h_i(r,\theta)\pi(r)\,dr)^2-\varepsilon\pi(r)\,dr, \\
\frac{\partial^2 Q_\varepsilon(\theta,\lambda)}{\partial\lambda_1(r)\partial\lambda_1(\tau)} &= -\sum_{i=1}^{n}(1+\gamma v_i)^{\frac{1-\gamma}{\gamma}} h_i(r,\theta)h_i(\tau,\theta)\pi(r)\pi(\tau)\,dr d\tau,\ r\neq\tau.
\end{aligned}$$

Hence, the desired expansion is.

$$\begin{aligned}
Q_\varepsilon\left(\overline{\theta},\overline{\lambda}_\varepsilon\right) &= Q_\varepsilon\left(\overline{\theta},0\right)-\int\sum_{i=1}^{n}h_i\left(r,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}(r)\pi(r)\,dr-\frac{1}{2}\overline{\lambda}_{0,\varepsilon}^2\sum_{i=1}^{n}\varphi_{2,i}^* \\
&\quad -\overline{\lambda}_{0,\varepsilon}\int\sum_{i=1}^{n}\varphi_{2,i}^* h_i\left(r,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}(r)\pi(r)\,dr-\frac{1}{2}\varepsilon\int\overline{\lambda}_{1,\varepsilon}^2(r)\pi(r)\,dr \\
&\quad -\frac{1}{2}\int\int\sum_{i=1}^{n}\varphi_{2,i}^* h_i\left(r,\overline{\theta}\right)h_i\left(\tau,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}(r)\overline{\lambda}_{1,\varepsilon}(\tau)\pi(r)\pi(\tau)\,dr d\tau.
\end{aligned}$$

where $\varphi_{2,i}^*=(1+\gamma\widetilde{v}_i)^{\frac{1-\gamma}{\gamma}}$ and $\widetilde{v}_i$ is the function v_i evaluated at some λ lying between $\overline{\lambda}_\varepsilon$ and 0. Combining the last two terms yields:

$$\begin{aligned}
Q_\varepsilon\left(\overline{\theta},\overline{\lambda}_\varepsilon\right) &= Q_\varepsilon\left(\overline{\theta},0\right)-\int\sum_{i=1}^{n}h_i\left(r,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}(r)\pi(r)\,dr-\frac{1}{2}\overline{\lambda}_{0,\varepsilon}^2\sum_{i=1}^{n}\varphi_{2,i}^* \\
&\quad -\overline{\lambda}_{0,\varepsilon}\int\sum_{i=1}^{n}\varphi_{2,i}^* h_i\left(r,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}(r)\pi(r)\,dr \\
&\quad -\frac{n}{2}\int\overline{\lambda}_{1,\varepsilon}(r)\left(\widetilde{K}_1^*+\frac{\varepsilon}{n}I\right)\overline{\lambda}_{1,\varepsilon}(r)\pi(r)\,dr,
\end{aligned}$$

where

$$\widetilde{K}_1^*\overline{\lambda}_{1,\varepsilon}(r)=\int\frac{1}{n}\sum_{i=1}^{n}\varphi_{2,i}^* h_i\left(r,\overline{\theta}\right)h_i\left(\tau,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}(\tau)\pi(\tau)\,d\tau$$

Now, note that:

$$\begin{aligned} Q_\varepsilon\left(\overline{\theta},0\right) &\leq Q_\varepsilon\left(\overline{\theta},\overline{\lambda}_\varepsilon\right) \leq Q_\varepsilon\left(\overline{\theta},0\right) + n\left\|\widehat{h}\left(.,\overline{\theta}\right)\right\| \left\|\overline{\lambda}_{1,\varepsilon}\right\| - \frac{1}{2}\overline{\lambda}_{0,\varepsilon}^2 \sum_{i=1}^n \varphi_{2,i}^* \\ &\quad -\overline{\lambda}_{0,\varepsilon}\int \sum_{i=1}^n \varphi_{2,i}^* h_i\left(r,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}\left(r\right)\pi\left(r\right)dr \\ &\quad -\frac{n}{2}\int \overline{\lambda}_{1,\varepsilon}\left(r\right)\left(\widetilde{K}_1^* + \frac{\varepsilon}{n}I\right)\overline{\lambda}_{1,\varepsilon}\left(r\right)\pi\left(r\right)dr. \end{aligned}$$

As $\overline{\lambda}_{0,\varepsilon} = -\int \widetilde{h}\left(\tau,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}\left(\tau,\overline{\theta}\right)\pi\left(\tau\right)d\tau$, it is straightforward to note that:

$$\begin{aligned} \frac{1}{2}\overline{\lambda}_{0,\varepsilon}^2 \sum_{i=1}^n \varphi_{2,i}^* &= \frac{n}{2}\int \overline{\lambda}_{1,\varepsilon}\left(r\right)\widetilde{K}_2^*\overline{\lambda}_{1,\varepsilon}\left(r\right)\pi\left(r\right)dr, \\ \overline{\lambda}_{0,\varepsilon}\int \sum_{i=1}^n \varphi_{2,i}^* h_i\left(r,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}\left(r\right)\pi\left(r\right)dr &= -n\int \overline{\lambda}_{1,\varepsilon}\left(r\right)\widetilde{K}_{1,2}^*\overline{\lambda}_{1,\varepsilon}\left(r\right)\pi\left(r\right)dr \end{aligned}$$

where $\widetilde{K}_2^*$ and $\widetilde{K}_{1,2}^*$ are specific operators. Replacing the RHS of the expressions above into the previous inequality leads to:

$$\begin{aligned} 0 &\leq n\left\|\widehat{h}\left(.,\overline{\theta}\right)\right\| \left\|\overline{\lambda}_{1,\varepsilon}\right\| \\ &\quad -\frac{n}{2}\int \overline{\lambda}_{1,\varepsilon}\left(r\right)\left(\widetilde{K}_1^* - 2\widetilde{K}_{1,2}^* + \widetilde{K}_2^* + \frac{\varepsilon}{n}I\right)\overline{\lambda}_{1,\varepsilon}\left(r\right)\pi\left(r\right)dr \\ &\leq n\left\|\widehat{h}\left(.,\overline{\theta}\right)\right\| \left\|\overline{\lambda}_{1,\varepsilon}\right\| - nC_\varepsilon\left\|\overline{\lambda}_{1,\varepsilon}\right\|^2 \end{aligned}$$

where C_ε is the smalest eigenvalue of $\widetilde{K}_1^* - 2\widetilde{K}_{1,2}^* + \widetilde{K}_2^* + \frac{\varepsilon}{n}I$.

$$C_\varepsilon\left\|\overline{\lambda}_{1,\varepsilon}\right\| \leq \left\|\widehat{h}\left(.,\overline{\theta}\right)\right\|$$

As the empirical operators $\widetilde{K}_1^*$, $\widetilde{K}_{1,2}^*$ and $\widetilde{K}_2^*$ are degenerate, we have $C_\varepsilon = O\left(\frac{\varepsilon}{n}\right)$. Therfore,

$$\left\|\overline{\lambda}_{1,\varepsilon}\right\| \leq C_\varepsilon^{-1}\left\|\widehat{h}\left(.,\overline{\theta}\right)\right\| = O\left(\varepsilon^{-1}n\right) \times O\left(n^{-1/2}\right) = O\left(\varepsilon^{-1}n^{1/2}\right)$$

Next, we note that

$$\begin{aligned} \left|\overline{\lambda}_{0,\varepsilon}\right| &= \left|\int \widetilde{h}\left(\tau,\overline{\theta}\right)\overline{\lambda}_{1,\varepsilon}\left(\tau,\overline{\theta}\right)\pi\left(\tau\right)d\tau\right| \\ &\leq \left\|\widetilde{h}\left(.,\overline{\theta}\right)\right\| \left\|\overline{\lambda}_{1,\varepsilon}\left(.,\overline{\theta}\right)\right\| \\ &= \left\|\widehat{h}\left(.,\overline{\theta}\right) + \left(\widetilde{h}\left(.,\overline{\theta}\right) - \widehat{h}\left(.,\overline{\theta}\right)\right)\right\| \left\|\overline{\lambda}_{1,\varepsilon}\left(.,\overline{\theta}\right)\right\| \\ &\leq \left\|\widehat{h}\left(.,\overline{\theta}\right)\right\| \left\|\overline{\lambda}_{1,\varepsilon}\left(.,\overline{\theta}\right)\right\| + \left\|\widetilde{h}\left(.,\overline{\theta}\right) - \widehat{h}\left(.,\overline{\theta}\right)\right\| \left\|\overline{\lambda}_{1,\varepsilon}\left(.,\overline{\theta}\right)\right\| \end{aligned}$$

Recall that $\widetilde{h}\left(\tau,\overline{\theta}\right) = \frac{1}{n}\sum_{i=1}^n k_{i,\varepsilon}h_i\left(\tau,\overline{\theta}\right)$ with $k_{i,\varepsilon} = nk(v_{i,\varepsilon})/\sum_{j=1}^n k(v_{j,\varepsilon})$ and $v_{i,\varepsilon} = \overline{\lambda}_{0,\varepsilon} + \int h_i\left(\tau,\theta\right)\overline{\lambda}_{1,\varepsilon}\left(\tau\right)\pi\left(\tau\right)d\tau = O_p\left(n^{-\zeta+1/\alpha}\right)$. As $x \to 0$, we have $k(x) \equiv \left((1+\gamma x)^{\frac{1}{\gamma}} - 1\right)/x \simeq$

$1+(1-\gamma)x$. Hence,

$$\begin{aligned} k_{i,\varepsilon} &\simeq \frac{n\left(1+(1-\gamma)v_{i,\varepsilon}\right)}{n+(1-\gamma)\sum_{j=1}^{n}v_{j,\varepsilon}} = \frac{1+(1-\gamma)v_{i,\varepsilon}}{1+\frac{1-\gamma}{n}\sum_{j=1}^{n}v_{j,\varepsilon}} \\ &= \frac{\left(1+(1-\gamma)v_{i,\varepsilon}\right)\left(1-\frac{1-\gamma}{n}\sum_{j=1}^{n}v_{j,\varepsilon}\right)}{\left(1+\frac{1-\gamma}{n}\sum_{j=1}^{n}v_{j,\varepsilon}\right)\left(1-\frac{1-\gamma}{n}\sum_{j=1}^{n}v_{j,\varepsilon}\right)} \\ &\simeq 1+(1-\gamma)\left(v_{i,\varepsilon}-\frac{1}{n}\sum_{j=1}^{n}v_{j,\varepsilon}\right) = 1+O_p\left(n^{-\zeta+1/\alpha}\right). \end{aligned}$$

Replacing into $\widetilde{h}\left(\tau,\overline{\theta}\right)$ leads to:

$$\begin{aligned} \widetilde{h}\left(\tau,\overline{\theta}\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(1+(1-\gamma)\left(v_{i,\varepsilon}-\frac{1}{n}\sum_{j=1}^{n}v_{j,\varepsilon}\right)\right)h_i\left(\tau,\overline{\theta}\right) \\ &= \widehat{h}\left(\tau,\overline{\theta}\right)+\frac{1-\gamma}{n}\sum_{i=1}^{n}\left(v_{i,\varepsilon}-\frac{1}{n}\sum_{j=1}^{n}v_{j,\varepsilon}\right)h_i\left(\tau,\overline{\theta}\right). \end{aligned}$$

Hence

$$\left\|\widetilde{h}\left(.,\overline{\theta}\right)-\widehat{h}\left(.,\overline{\theta}\right)\right\|^2 = (1-\gamma)^2\int\left(\frac{1}{n}\sum_{i=1}^{n}\left(v_{i,\varepsilon}-\frac{1}{n}\sum_{j=1}^{n}v_{j,\varepsilon}\right)h_i\left(\tau,\overline{\theta}\right)\right)^2\pi\left(\tau\right)d\tau$$

and

$$\left\|\widetilde{h}\left(.,\overline{\theta}\right)-\widehat{h}\left(.,\overline{\theta}\right)\right\| \leq |1-\gamma|\left\|\widehat{h}\left(.,\overline{\theta}\right)\right\|^2\max_{1\leq i\leq n}\left|v_{i,\varepsilon}-\frac{1}{n}\sum_{j=1}^{n}v_{j,\varepsilon}\right| = O_p\left(n^{-1/2-\zeta+1/\alpha}\right)$$

This shows that $\left\|\widehat{h}\left(.,\overline{\theta}\right)\right\|$ dominates $\left\|\widetilde{h}\left(.,\overline{\theta}\right)-\widehat{h}\left(.,\overline{\theta}\right)\right\|$ so that:

$$\begin{aligned} \left|\overline{\lambda}_{0,\varepsilon}\right| &\leq \left\|\widehat{h}\left(.,\overline{\theta}\right)\right\|\left\|\overline{\lambda}_{1,\varepsilon}\left(.,\overline{\theta}\right)\right\|+\left\|\widetilde{h}\left(.,\overline{\theta}\right)-\widehat{h}\left(.,\overline{\theta}\right)\right\|\left\|\overline{\lambda}_{1,\varepsilon}\left(.,\overline{\theta}\right)\right\| \\ &= O_p\left(n^{-1/2}\right)\times O\left(\varepsilon^{-1}n^{1/2}\right) = O\left(\varepsilon^{-1}\right) \end{aligned}$$

Given what precedes, the dominant term of the second order expansion of $Q_\varepsilon\left(\overline{\theta},\overline{\lambda}_\varepsilon\right)$ is the quadratic term in $\overline{\lambda}_{1,\varepsilon}$. Therefore, we have:

$$Q_\varepsilon\left(\overline{\theta},\overline{\lambda}_\varepsilon\right) \leq Q_\varepsilon\left(\overline{\theta},0\right)+O_p\left(\varepsilon^{-2}n^2\right)$$

Q.E.D. ■

Lemma 11 (Lemma 2) *Suppose Assumption 1 is satisfied and let $\varepsilon = O\left(n^b\right)$ for $1/2 < b < 1$. Then we have $\left\|\widehat{h}\left(\tau,\widehat{\theta}_\varepsilon\right)\right\| = O_p\left(n^{-b+1/2}\right)$ as $n\to\infty$.*

Proof of Lemma 2. Let us define:

$$\widetilde{\lambda}_1\left(\tau,\widehat{\theta}_\varepsilon\right)=-n^{-\zeta}\frac{\widehat{h}\left(\tau,\widehat{\theta}_\varepsilon\right)}{\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|}\text{ and }\widetilde{\lambda}_0=-\int\widehat{h}\left(\tau,\widehat{\theta}_\varepsilon\right)\widetilde{\lambda}_1\left(\tau,\widehat{\theta}_\varepsilon\right)\pi\left(\tau\right)d\tau.$$

for some ζ such that $0<\zeta<2b-1$. A second order Taylor expansion of $Q_\varepsilon\left(\widehat{\theta}_\varepsilon,\widetilde{\lambda}\right)$ around $\lambda=0$ yields:

$$\begin{aligned}Q_\varepsilon\left(\widehat{\theta}_\varepsilon,\widetilde{\lambda}\right) &= Q_\varepsilon\left(\widehat{\theta}_\varepsilon,0\right)-\int\sum_{i=1}^{n}h_i\left(r,\widehat{\theta}_\varepsilon\right)\widetilde{\lambda}_1\left(r\right)\pi\left(r\right)dr-\frac{1}{2}\widetilde{\lambda}_0^2\sum_{i=1}^{n}\varphi_{2,i}^*\\ &\quad-\widetilde{\lambda}_0\int\sum_{i=1}^{n}\varphi_{2,i}^*h_i\left(r,\widehat{\theta}_\varepsilon\right)\widetilde{\lambda}_1\left(r\right)\pi\left(r\right)dr\\ &\quad-\frac{n}{2}\int\widetilde{\lambda}_1\left(r\right)\left(\widetilde{K}_1^*+\frac{\varepsilon}{n}I\right)\widetilde{\lambda}_1\left(r\right)\pi\left(r\right)dr,\end{aligned}$$

where $\widetilde{\varphi}_{2,i}^*$ and $\widetilde{K}^*$ are defined similarly as in the proof of Lemma 1. Replacing $\widetilde{\lambda}_0$ by its expression yields:

$$\begin{aligned}Q_\varepsilon\left(\widehat{\theta}_\varepsilon,\widetilde{\lambda}\right) &= Q_\varepsilon\left(\widehat{\theta}_\varepsilon,0\right)-n\int\widehat{h}\left(r,\widehat{\theta}_\varepsilon\right)\widetilde{\lambda}_1\left(r\right)\pi\left(r\right)dr\\ &\quad-\frac{n}{2}\int\widetilde{\lambda}_1\left(r\right)\left(\widetilde{K}_1^*-2\widetilde{K}_{1,2}^*+\widetilde{K}_2^*+\frac{\varepsilon}{n}I\right)\widetilde{\lambda}_1\left(r\right)\pi\left(r\right)dr\\ &\geq Q_\varepsilon\left(\widehat{\theta}_\varepsilon,0\right)-n\int\widehat{h}\left(r,\widehat{\theta}_\varepsilon\right)\widetilde{\lambda}_1\left(r\right)\pi\left(r\right)dr-nC\left\|\widetilde{\lambda}_1\right\|^2,\end{aligned}$$

where $\widetilde{K}_1^*$, $\widetilde{K}_{1,2}^*$ and $\widetilde{K}_2^*$ are defined similarly as in the proof of Lemma 1 and C satisfies:

$$\frac{1}{2}\int f\left(r\right)\left(\widetilde{K}_1^*-2\widetilde{K}_{1,2}^*+\widetilde{K}_2^*+\frac{\varepsilon}{n}I\right)f\left(r\right)\pi\left(r\right)dr\leq C\int f\left(r\right)^2\pi\left(r\right)dr,\ \forall\ f$$

Replacing $\widetilde{\lambda}_1\left(\tau,\widehat{\theta}_\varepsilon\right)$ by its expression leads to:

$$Q_\varepsilon\left(\widehat{\theta}_\varepsilon,\widetilde{\lambda}\right)\geq Q_\varepsilon\left(\widehat{\theta}_\varepsilon,0\right)+n^{1-\zeta}\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|-Cn^{1-2\zeta}.$$

Now, note that:

$$\begin{aligned}&Q_\varepsilon\left(\widehat{\theta}_\varepsilon,0\right)+n^{1-\zeta}\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|-Cn^{1-2\zeta}\\ \leq\ &Q_\varepsilon\left(\widehat{\theta}_\varepsilon,\widetilde{\lambda}\right)\leq Q_\varepsilon\left(\widehat{\theta}_\varepsilon,\lambda_\varepsilon\left(\widehat{\theta}_\varepsilon\right)\right)\leq Q_\varepsilon\left(\theta_0,\lambda_\varepsilon\left(\widehat{\theta}_\varepsilon\right)\right)\\ \leq\ &\sup_{\lambda\in\Lambda_n(\theta_0)}Q_\varepsilon\left(\theta_0,\lambda\right)\leq Q_\varepsilon\left(\theta_0,0\right)+O_p\left(\varepsilon^{-2}n^2\right),\end{aligned}$$

where the second inequality follows from a supremum over λ for given $\widehat{\theta}_\varepsilon$, the third one follows from the fact that $\widehat{\theta}_\varepsilon$ is a minimizer over θ, the fourth one is a logical implication of the supremum and and the last one follows from Lemma 1. By noting that $Q_\varepsilon\left(\theta_0,0\right)=Q_\varepsilon\left(\widehat{\theta}_\varepsilon,0\right)$, taking the first and last term yields:

$$\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|\leq Cn^{-\zeta}+O_p\left(\varepsilon^{-2}n^{1+\zeta}\right)=Cn^{-\zeta}+O_p\left(n^{1+\zeta-2b}\right).$$

Hence, $\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|$ is bounded in probability if $0<\zeta<2b-1$ as assumed. To complete the proof, consider $\widetilde{\lambda}_1=-\zeta_n\widehat{h}\left(\tau,\widehat{\theta}_\varepsilon\right)$ and $\widetilde{\lambda}_0=-\int\widehat{h}\left(\tau,\widehat{\theta}_\varepsilon\right)\widetilde{\lambda}_1\left(\tau,\widehat{\theta}_\varepsilon\right)\pi(\tau)\,d\tau$ for some $\zeta_n\to 0$. From what precedes, we have:

$$Q_\varepsilon\left(\widehat{\theta}_\varepsilon,0\right)+n\zeta_n\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|^2-nC\zeta_n^2\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|^2\leq Q_\varepsilon(\theta_0,0)+O_p\left(\varepsilon^{-2}n^2\right).$$

Therefore:

$$n\zeta_n(1-\zeta_nC)\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|^2\leq O_p\left(\varepsilon^{-2}n^2\right)=O_p\left(n^{-2b+2}\right).$$

As $n\to\infty$, we have $(1-\zeta_nC)=O(1)$. Hence $\zeta_n\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|^2\leq O_p\left(\varepsilon^{-2}n\right)$, for all possible sequence $\zeta_n\to 0$. It thus follows that $\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|=O_p\left(n^{-b+1/2}\right)$. Q.E.D. ■

Proof of Theorem 7. The fonction $h(\tau,\theta)=E[h_i(\tau,\theta)]$ is continuous in θ. By Lemma 2, $\widehat{h}\left(\tau,\widehat{\theta}_\varepsilon\right)\xrightarrow{p}0$, and the Uniform Weak Law of Large Numbers implies that $\sup_\theta\left\|\widehat{h}(\tau,\theta)-h(\tau,\theta)\right\|\xrightarrow{p}0$. Further note that:

$$\begin{aligned}\left\|\widehat{h}\left(\tau,\widehat{\theta}_\varepsilon\right)\right\| &= \left\|\widehat{h}\left(\tau,\widehat{\theta}_\varepsilon\right)-h\left(\tau,\widehat{\theta}_\varepsilon\right)+h\left(\tau,\widehat{\theta}_\varepsilon\right)\right\|\\ &\leq \left\|\widehat{h}\left(\tau,\widehat{\theta}_\varepsilon\right)-h\left(\tau,\widehat{\theta}_\varepsilon\right)\right\|+\left\|h\left(\tau,\widehat{\theta}_\varepsilon\right)\right\|\xrightarrow{p}0.\end{aligned}$$

It thus follows that $\left\|h\left(\tau,\widehat{\theta}_\varepsilon\right)\right\|\xrightarrow{p}0$. As $h(\tau,\theta)=0$ has a unique solution θ_0, it follows that $\widehat{\theta}_\varepsilon\xrightarrow{p}\theta_0$. By Lemma 2, we have $\left\|\widehat{h}\left(.,\widehat{\theta}_\varepsilon\right)\right\|=O_p\left(n^{-b+1/2}\right)$. By letting $\overline{\theta}=\widehat{\theta}_\varepsilon$ in Lemma 1, we obtain $\left\|\overline{\lambda}_{1,\varepsilon}\left(.,\widehat{\theta}_\varepsilon\right)\right\|=O_p\left(n^{-b+1/2}\right)$ and $\left\|\overline{\lambda}_{0,\varepsilon}\left(\widehat{\theta}_\varepsilon\right)\right\|=O_p\left(n^{-b}\right)$. Q.E.D. ■

Proof of Theorem 8. The optimal dual variable $\widehat{\lambda}_\varepsilon(\theta)=\left\{\widehat{\lambda}_\varepsilon(r,\theta)=\left(\widehat{\lambda}_{0,\varepsilon}(\theta),\widehat{\lambda}_{1,\varepsilon}(r,\theta)\right),r\in\mathbb{R}^d\right\}$ solves:

$$\begin{aligned}\frac{\partial Q_\varepsilon(\theta,\lambda)}{\partial\lambda_0} &= n-\sum_{i=1}^n(1+\gamma v_i)^{\frac{1}{\gamma}}=0,\\ \frac{\partial Q_\varepsilon(\theta,\lambda)}{\partial\lambda_1(r)} &= -\sum_{i=1}^n(1+\gamma v_i)^{\frac{1}{\gamma}}h_i(r,\theta)\pi(r)\,dr-\varepsilon\lambda_1(r)\pi(r)\,dr=0,\end{aligned}$$

where $v_i=\lambda_0+\int h_i(\tau,\theta)\lambda_1(\tau,\theta)\pi(\tau)\,d\tau$. We need the second order derivatives of $Q_\varepsilon(\theta,\lambda)$ to compute a first order expansion of the system above. We have:

$$\begin{aligned}\frac{\partial^2Q_\varepsilon(\theta,\lambda)}{\partial\lambda_0^2} &= -\sum_{i=1}^n(1+\gamma v_i)^{\frac{1-\gamma}{\gamma}},\\ \frac{\partial^2Q_\varepsilon(\theta,\lambda)}{\partial\lambda_0\partial\lambda_1(r)} &= -\sum_{i=1}^n(1+\gamma v_i)^{\frac{1-\gamma}{\gamma}}h_i(r,\theta)\pi(r)\,dr,\\ \frac{\partial^2Q_\varepsilon(\theta,\lambda)}{\partial(\lambda_1(r))^2} &= -\sum_{i=1}^n(1+\gamma v_i)^{\frac{1-\gamma}{\gamma}}(h_i(r,\theta)\pi(r)\,dr)^2-\varepsilon\pi(r)\,dr,\\ \frac{\partial^2Q_\varepsilon(\theta,\lambda)}{\partial\lambda_1(r)\partial\lambda_1(\tau)} &= -\sum_{i=1}^n(1+\gamma v_i)^{\frac{1-\gamma}{\gamma}}h_i(r,\theta)h_i(\tau,\theta)\pi(r)\pi(\tau)\,dr d\tau,\ r\neq\tau.\end{aligned}$$

Evaluating these derivatives at $\lambda = 0$ yields:

$$\begin{aligned}
\frac{\partial Q_\varepsilon(\theta,\lambda)}{\partial \lambda_0}|_{\lambda=0} &= 0; \quad \frac{\partial Q_\varepsilon(\theta,\lambda)}{\partial \lambda_1(r)}|_{\lambda=0} = -\sum_{i=1}^{n} h_i(r,\theta)\,\pi(r)\,dr \\
\frac{\partial^2 Q_\varepsilon(\theta,\lambda)}{\partial \lambda_0^2}|_{\lambda=0} &= -n; \quad \frac{\partial^2 Q_\varepsilon(\theta,\lambda)}{\partial \lambda_0 \partial \lambda_1(\tau)}|_{\lambda=0} = -\sum_{i=1}^{n} h_i(\tau,\theta)\,\pi(\tau)\,d\tau, \\
\frac{\partial^2 Q_\varepsilon(\theta,\lambda)}{\partial (\lambda_1(r))^2}|_{\lambda=0} &= -\sum_{i=1}^{n} (h_i(r,\theta)\,\pi(r)\,dr)^2 - \varepsilon\pi(r)\,dr, \\
\frac{\partial^2 Q_\varepsilon(\theta,\lambda)}{\partial \lambda_1(r)\,\partial \lambda_1(\tau)}|_{\lambda=0} &= -\sum_{i=1}^{n} h_i(r,\theta)\,h_i(\tau,\theta)\,\pi(r)\,\pi(\tau)\,dr d\tau,\ r \neq \tau.
\end{aligned}$$

Hence, a first order expansion of the system $\left(\frac{\partial Q_\varepsilon(\theta,\widehat{\lambda}_{0,\varepsilon})}{\partial \lambda_0}, \frac{\partial Q_\varepsilon(\theta,\widehat{\lambda}_{0,\varepsilon})}{\partial \lambda_1(r)}\right)$ around $\lambda = 0$ yields:

$$\begin{aligned}
\widehat{\lambda}_{0,\varepsilon} + \int \widehat{h}(\tau,\theta)\,\widehat{\lambda}_{1,\varepsilon}(\tau)\,\pi(\tau)\,d\tau &\simeq 0 \\
\widehat{\lambda}_{0,\varepsilon}\widehat{h}(r,\theta) + \int \frac{1}{n}\sum_{i=1}^{n} h_i(r,\theta)\,h_i(\tau,\theta)\,\widehat{\lambda}_{1,\varepsilon}(\tau)\,\pi(\tau)\,d\tau + \frac{\varepsilon}{n}\widehat{\lambda}_{1,\varepsilon}(r) &\simeq -\widehat{h}(r,\theta)
\end{aligned}$$

Multiplying the first equation by $\widehat{h}(r,\theta)$ and substracting from the second equation yields:

$$\int \left(\frac{1}{n}\sum_{i=1}^{n} h_i(r,\theta)\,h_i(\tau,\theta) - \widehat{h}(r,\theta)\,\widehat{h}(\tau,\theta)\right)\widehat{\lambda}_{1,\varepsilon}(\tau)\,\pi(\tau)\,d\tau + \frac{\varepsilon}{n}\widehat{\lambda}_{1,\varepsilon}(r) \simeq -\widehat{h}(r,\theta),$$

which is equivalent to

$$\left(K + \frac{\varepsilon}{n}I\right)\widehat{\lambda}_{1,\varepsilon}(r) \simeq -\widehat{h}(r,\theta)$$

where K is the linear operator with kernel

$$h(r,\tau) = \frac{1}{n}\sum_{i=1}^{n}\left(h_i(r,\theta) - \widehat{h}(r,\theta)\right)\left(h_i(\tau,\theta) - \widehat{h}(\tau,\theta)\right).$$

Finally, solving for $\widehat{\lambda}_{1,\varepsilon}$ yields:

$$\begin{aligned}
\widehat{\lambda}_{1,\varepsilon}(r,\theta) &\simeq -\left(K + \frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(r,\theta) \text{ and} \\
\widehat{\lambda}_{0,\varepsilon}(\theta) &\simeq -\int \widehat{h}(\tau,\theta)\left(K + \frac{\varepsilon}{n}I\right)^{-1}\widehat{h}(\tau,\theta)\,\pi(\tau)\,d\tau
\end{aligned}$$

Next, we consider the first order condition solved by $\widehat{\theta}_\varepsilon$:

$$0 \simeq \frac{\partial Q_\varepsilon\left(\widehat{\theta}_\varepsilon, \widehat{\lambda}_\varepsilon\left(\widehat{\theta}_\varepsilon\right)\right)}{\partial \theta} = -\int \sum_{i=1}^{n} (1 + \gamma v_{i,\varepsilon})^{\frac{1}{\gamma}} \frac{\partial h_i\left(\tau, \widehat{\theta}_\varepsilon\right)}{\partial \theta}\widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\,\pi(\tau)\,d\tau$$

where $v_{i,\varepsilon} = \widehat{\lambda}_{0,\varepsilon} + \int h_i(\tau,\theta)\,\widehat{\lambda}_{1,\varepsilon}(\tau,\theta)\,\pi(\tau)\,d\tau$. A first order expansion of $\widehat{h}\left(r,\widehat{\theta}_\varepsilon\right)$ at θ_0 yields:

$$\widehat{h}\left(r,\widehat{\theta}_\varepsilon\right) \simeq \widehat{h}(r,\theta_0) + \frac{\partial \widehat{h}(r,\theta_0)}{\partial \theta}\left(\widehat{\theta}_\varepsilon - \theta_0\right).$$

Therefore, $\widehat{\lambda}_{1,\varepsilon}\left(r,\widehat{\theta}_{\varepsilon}\right)$ has the following expansion:

$$\begin{aligned}\widehat{\lambda}_{1,\varepsilon}\left(r,\widehat{\theta}_{\varepsilon}\right) &\simeq -\left(K+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}\left(r,\widehat{\theta}_{\varepsilon}\right)\\ &= -\left(K+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}\left(r,\theta_0\right)-\left(K+\frac{\varepsilon}{n}I\right)^{-1}\frac{\partial\widehat{h}\left(r,\theta_0\right)}{\partial\theta'}\left(\widehat{\theta}_{\varepsilon}-\theta_0\right),\end{aligned}$$

where K is evaluated at $\widehat{\theta}_{\varepsilon}$. Replacing the latter expansion into the FOC solved by $\widehat{\theta}_{\varepsilon}$ and solving for $\widehat{\theta}_{\varepsilon}-\theta_0$ yields:

$$\begin{aligned}\widehat{\theta}_{\varepsilon}-\theta_0 \simeq &\left(\int\frac{1}{n}\sum_{i=1}^{n}\left(1+\gamma v_{i,\varepsilon}\right)^{\frac{1}{\gamma}}\frac{\partial h_i\left(\tau,\widehat{\theta}_{\varepsilon}\right)}{\partial\theta}\left(K+\frac{\varepsilon}{n}I\right)^{-1}\frac{\partial\widehat{h}\left(\tau,\theta_0\right)}{\partial\theta'}\pi\left(\tau\right)d\tau\right)^{-1}\\ &\times\int\frac{1}{n}\sum_{i=1}^{n}\left(1+\gamma v_{i,\varepsilon}\right)^{\frac{1}{\gamma}}\frac{\partial h_i\left(\tau,\widehat{\theta}_{\varepsilon}\right)}{\partial\theta}\left(K+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\end{aligned}$$

Finally, $\widehat{\theta}_{\varepsilon}$ is replaced by θ_0 in the RHS at a higher order cost. Q.E.D. ■

Proof of Theorem 9. Under $\varepsilon=O\left(n^b\right)$ for $1/2<b<1$, Theorem 8 establishes that:

$$\begin{aligned}\widehat{\theta}_{\varepsilon}-\theta_0 \simeq &\left(\int\widetilde{G}\left(\tau,\theta\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}\widehat{G}\left(\tau,\theta\right)\pi\left(\tau\right)d\tau\right)^{-1}\\ &\times\int\widetilde{G}\left(\tau,\theta\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}\left(\tau,\theta\right)\pi\left(\tau\right)d\tau\end{aligned}$$

where $\widehat{G}\left(\tau,\theta\right)=\frac{\partial\widehat{h}(\tau,\theta)}{\partial\theta'}$ and $\widetilde{G}\left(\tau,\theta\right)=\frac{1}{n}\sum_{i=1}^{n}\left(1+\gamma v_{i,\varepsilon}\right)^{\frac{1}{\gamma}}\frac{\partial h_i(\tau,\theta)}{\partial\theta}$. As $n\to\infty$, we have $v_{i,\varepsilon}\to 0$ and both $\widehat{G}\left(\tau,\theta\right)$ and $\widetilde{G}\left(\tau,\theta\right)$ converge to $G\left(\tau,\theta\right)=E\left(\frac{\partial h_i(\tau,\theta)}{\partial\theta'}\right)$. Hence, we have:

$$\begin{aligned}\widehat{\theta}_{\varepsilon}-\theta_0 \simeq &\left(\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1}\\ &\times\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}\widehat{h}\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\end{aligned}$$

By the Central Limit Theorem, $\sqrt{n}\widehat{h}\left(\tau,\theta_0\right)$ converges to a Gaussian element with mean zero and covariance operator K. Taking the variance yields:

$$\begin{aligned}&Var\left[\sqrt{n}\left(\widehat{\theta}_{\varepsilon}-\theta_0\right)\right]\\ =&\left(\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1}\\ &\times\int\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}Cov\left[\sqrt{n}\widehat{h}\left(\tau,\theta_0\right),\sqrt{n}\widehat{h}\left(r,\theta_0\right)\right]\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(r,\theta_0\right)\pi\left(\tau\right)\pi\left(r\right)d\tau dr\\ &\times\left(\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1}\end{aligned}$$

We have

$$\begin{aligned}&Cov\left[\sqrt{n}\widehat{h}\left(\tau,\theta_0\right),\sqrt{n}\widehat{h}\left(r,\theta_0\right)\right]\\ =&\frac{1}{n}Cov\left[\sum_{i=1}^{n}h_i\left(\tau,\theta_0\right),\sum_{k=1}^{n}h_k\left(r,\theta_0\right)\right]=\frac{1}{n}\sum_{i=1}^{n}Cov\left[h_i\left(\tau,\theta_0\right)h_i\left(r,\theta_0\right)\right]\\ =&Cov\left[h_i\left(\tau,\theta_0\right)h_i\left(r,\theta_0\right)\right]=k\left(\tau,r\right)\end{aligned}$$

Hence

$$\begin{aligned}
& Var\left[\sqrt{n}\left(\widehat{\theta}_{\varepsilon}-\theta_0\right)\right] \\
= & \left(\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1} \\
& \times\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}\int k\left(\tau,r\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(r,\theta_0\right)\pi\left(\tau\right)\pi\left(r\right)d\tau dr \\
& \times\left(\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1} \\
= & \left(\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1} \\
& \times\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}K\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau \\
& \left(\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1}
\end{aligned}$$

As $n\rightarrow\infty$, $\frac{\varepsilon}{n}\rightarrow 0$ as long as $\varepsilon=O\left(n^b\right)$ for $1/2<b<1$. Hence:

$$\begin{aligned}
\lim_{n\rightarrow\infty}Var\left[\sqrt{n}\left(\widehat{\theta}_{\varepsilon}-\theta_0\right)\right] &= \left(\int G\left(\tau,\theta_0\right)K^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1} \\
&\quad\times\left(\int G\left(\tau,\theta_0\right)K^{-1}KK^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right) \\
&\quad\left(\int G\left(\tau,\theta_0\right)\left(K+\frac{\varepsilon}{n}I\right)^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1} \\
&= \left(\int G\left(\tau,\theta_0\right)K^{-1}G\left(\tau,\theta_0\right)\pi\left(\tau\right)d\tau\right)^{-1}
\end{aligned}$$

Q.E.D. ■