UNIL | Université de Lausanne
Département de finance

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FINANCE RESEARCH SEMINAR SUPPORTED BY UNIGESTION

"On the Empirical Saddlepoint Approximation with Application to Asset Pricing"

Prof. Benjamin HOLCBLAT
University of Luxembourg

Abstract

Moment-based estimation often yields instable estimates, such as the RRA (relative risk aversion) estimate in consumption-based asset pricing. This paper establishes novel theoretical results for the ESP (empirical saddlepoint) approximation, and then use them to investigate this instability. We prove that there exists an intensity distribution of the solutions to empirical moment conditions, and approximate it with the integral of the ESP approximation, calling the result the ESP intensity. Global consistency and asymptotic normality of the ESP intensity are proved. The application provides an explanation for the instability of the RRA estimates reported in the literature (fat and long right tail of the ESP approximation), and it suggests that consumption-based asset-pricing theory is more consistent with data than standard inference approaches indicate.

Thursday, March 29, 2018, 10:30-12:00
Room 109, Extranef building at the University of Lausanne

ON THE EMPIRICAL SADDLEPOINT APPROXIMATION WITH APPLICATION TO ASSET PRICING

BENJAMIN HOLCBLAT

Abstract. Moment-based estimation often yields instable estimates, such as the RRA (relative risk aversion) estimate in consumption-based asset pricing. This paper establishes novel theoretical results for the ESP (empirical saddlepoint) approximation, and then use them to investigate this instability. We prove that there exists an intensity distribution of the solutions to empirical moment conditions, and approximate it with the integral of the ESP approximation, calling the result the ESP intensity. Global consistency and asymptotic normality of the ESP intensity are proved. The application provides an explanation for the instability of the RRA estimates reported in the literature (fat and long right tail of the ESP approximation), and it suggests that consumption-based asset-pricing theory is more consistent with data than existing inference approaches indicate.

Keywords: Saddlepoint approximation; Moment-based estimation; Multiple roots to estimating equations; Schmetterer-Jennrich lemma; Empirical consumption-based asset pricing.

JEL classification: C1, G12.

1. Introduction

When moment conditions are nonlinear, moment-based estimation is often found to provide unstable estimates. In particular, in empirical consumption-based asset pricing (Hansen and Singleton, 1982), the literature has found little common ground about the value of the relative risk aversion (RRA) of the representative agent. On the one hand, in a majority of studies, point estimates from economically similar moment conditions are generally outside of each other's confidence intervals. On the other hand, in a minority of studies, authors report or warn against "the trap of blowing up standard errors" (Cochrane, 2001, p. 210). One possible explanation is the inadequacy of consumption-based asset-pricing theories. But models are not always rejected (e.g., Vissing-Jørgensen and Attanasio, 2003; Savov, 2011), and simulations point to the insufficiency of existing moment-based methods for consumption-based asset pricing (e.g., Kocherlakota, 1990; Hansen, Heaton and Yaron, 1996 and other papers in that issue of JBES; Gregory, Lamarche and Smith, 2002).

This paper establishes novel theoretical results for the ESP (empirical saddlepoint) approximation, and then use them to investigate the instability of moment-based estimation in empirical

Date: October 11, 2015.

consumption-based asset pricing. On the *theoretical* side, we prove the existence of the distribution of the solutions to estimating equations. Such a result yields a generalization of the Schemetter-Jennrich lemma (Schmetterer, 1966, Ch. 5; Jennrich, 1969, Lemma 2). Then, when estimating equations are also empirical moment conditions,[1] we prove global consistency and asymptotic normality of the ESP approximation in the sense of the Prokhorov metric. On the *empirical* side, the paper sheds light on empirical consumption-based asset pricing. The ESP approximation of the distribution of the relative risk aversion (RRA) estimator suggests that the key equilibrium implication of consumption-based asset-pricing theory is more consistent with data than standard inference approaches indicate. Moreover, the fat and long right tail of the ESP approximation provides an explanation for the large variations and large values of the RRA often reported in the literature.

1.1. **Literature overview.** The contributions of the present paper relates to several strands of literature. We distinguish four of them: the SP (saddlepoint) and ESP literatures, the literature on estimating equations, and the literature on empirical consumption-based asset pricing. The ESP approximation is the empirical counterpart of the SP approximation. Following Esscher (1932) and Daniels (1954), the literature in statistics (e.g., Tingley and Field, 1990; Jensen, 1992; Robinson, Ronchetti and Young, 2003, Broda and Kan, 2015) and econometrics (e.g., Phillips, 1978; Holly and Phillips, 1979; Phillips, 1982; Kundhi and Rilstone, 2013) has long shown that ESP and SP approximations can provide accurate approximations of distributions (Field and Ronchetti, 1990, p. 130). More recently, Imbens (1997), Ronchetti and Trojani (2003), and Sowell (2007) propose to derive more accurate confidence intervals and tests for GMM. Czellar and Ronchetti (2010) proposes more accurate tests for indirect inference. Sowell (2009) proposes an ESP-based point estimator to automatically correct the higher-order bias of GEL (generalized empirical likelihood) estimators. Aït-Sahalia and Yu (2006) proposes a saddlepoint approximation of a transition density for likelihood-based inference of continuous-time Markov processes.

A first contribution of the present paper to the ESP literature is the global consistency and asymptotic normality of the measure that has the ESP approximation as a Radon-Nikodyn derivative w.r.t. (with respect to) the Lebesgue measure. We call the later the ESP intensity. When estimating equations are empirical moment conditions, we prove that the ESP intensity

[1]In the present paper, we distinguish between estimating equations and empirical moment conditions: estimating equations that have zero expectation are empirical moment conditions.

is consistent and asymptotically normal in the sense of the Prokhorov metric, which means that it converges to a point mass at the unknown parameter θ_0 like a Gaussian distribution with a standard deviation that goes to zero at the rate square root of the sample size T (Theorems 1 and 2 on p. 15). When there are several solutions to the moment conditions, we deduce that the ESP intensity converges to a sum of point masses each centered at one of the solutions to the moment conditions like a sum of Gaussian distributions (Remark 2 on p. 17). Such results are not available in the literature for nonlinear empirical moment conditions. Ronchetti and Welsh (1994) proves that the ESP approximation is equal to the corresponding saddlepoint approximation modulo $O_{\mathbb{P}}(T^{-\frac{1}{2}})$ in a neighborhood of the unknown parameter θ_0 shrinking at the rate square root of the sample size T. Almudevar, Field and Robinson (2000, Theorem 2) shows that the saddlepoint approximation is equal to the distribution of a solution to the empirical moment condition in a neighborhood of the unknown parameter θ_0, modulo $O(T^{-1})$ with a probability converging to one at the rate $1 - \mathrm{e}^{-cT}$, where $c > 0$. Jensen (1995, pp. 106–107; 114) explains the difficulty tackled by Almudevar, Field and Robinson (2000). The available results have the advantage to provide a characterization of the approximation error, but they are only local, so that our results complement them.

A contribution to both the SP and the ESP literature is the existence of the object approximated by the SP and ESP intensities (i.e., integrals of SP and ESP approximations) when applied to solutions of nonlinear estimating equations. When estimating equations are nonlinear, there can be several solutions to them, even if the asymptotic estimating equations have a unique solution: the implicit function theorem can only guarantee local uniqueness. Then, the distribution of the solutions is not a probability distribution, but an intensity distribution as pointed out in Skovgaard (1985; 1990), Jensen and Wood (1998), and Almudevar, Field and Robinson (2000). Under the assumption that the expected number of solutions is finite, the present paper shows that the solutions to the estimating equations defines a point random field with its corresponding intensity distribution (Propositions 1 and Corollary 2 on pp. 7–8).

A corollary of this existence result is the measurability of each of the solutions to the estimating equations (Corollary 1 on p. 7). Under the assumption that the expected number of solutions is finite, such corollary generalizes the Schmetterer-Jennrich lemma (Schmetterer, 1966, Ch. 5; Jennrich, 1969, Lemma 2), which guarantees the measurability of extremum estimators. Most estimators used in econometrics and statistics fall within the class of extremum estimators, which are also called optimization estimators (e.g., Gallant and White, 1988). The Schmetterer-Jennrich lemma guarantees the measurability of one of the potentially multiple global maximizers of the objective function (or, equivalently global minimizers). More recently, the literature (Ferger, 2004, Corollary 1) has established and used the measurability of the largest and smallest global maximizers of the objective function (e.g., Seijo and Sen, 2011). In extremum estimation, the first-order conditions of the objective function defines estimating equations, so that our result guarantees the measurability of each local extremum of the objective function. This result is practically relevant as it is often difficult to guarantee that the maximizer found is a global maximizer of the objective function (e.g., Amemiya, 1985, p. 110–111). Moreover, if there are several global maximizers, the found global maximizer may not be the largest or smallest, or even correspond to the one of the Schmetterer-Jennrich lemma. In some cases, an alternative to establishing measurability is the introduction of outer probability measures, but it forbids the use of standard probability theory, and it creates significant mathematical complications (e.g., van der Vaart and Wellner, 1996). Thus, several literatures, such as the literature on multiple roots (e.g., Perlman, 1983; Lehmann and Casella, 1998, sec. 6.4; Amemiya, 1985, theo. 4.1.2; Reeds, 1985; Small, Wang and Yang, 2000), implicitly assumes our result.

The present paper also contributes to the empirical consumption-based asset pricing literature. We estimate the RRA of the representative agent using GMM (Pearson, 1894; Hansen, 1982), CU (continuously updated) GMM (Hansen, Heaton and Yaron, 1996), which is an example of GEL (generalized empirical likelihood) estimators, CU GMM for weak identification (Stock and Wright, 2000). Following Julliard and Ghosh (2012), the estimation relies on standard data sets, and on a key moment condition that is as consistent with Lucas (1978) as with more recent consumption-based asset-pricing models (Barro, 2006; Gabaix, 2012). GMM and CU GMM provide almost the same results. They summarize the uncertainty about the RRA by a Gaussian distribution centered at the point estimate, which yield confidence regions that include values inconsistent with standard finance theory (negative RRA). In accordance with empirical observations in the literature (e.g., Hansen, Heaton and Yaron, 1996), CU GMM for weak identification provides incredibly large confidence regions for RRA so that they seem uninformative in practice. The ESP approximation indicates that the non-Gaussian structure of the uncertainty about the RRA estimator is behind the instability of the results previously reported in the literature. It also suggests that consumption-based asset-pricing theory is more consistent with data than other inference approaches suggest.

1.2. **Organization of the paper.** The paper is organized as follows. Section 2 presents the ESP estimand, which is the quantity approximated by integrals of the ESP approximation, and the ESP intensity, which is the measure defined by integrals of the ESP approximation. Section 3 establishes the consistency and asymptotic normality of the ESP intensity in the sense of the Prokhorov metric. Section 4 presents empirical evidence from consumption-based asset pricing. Short proofs of the results of sections 2–3 are in the appendix. Supplemental material contains detailed proofs and additional empirical evidence from asset pricing.

2. The ESP estimand and the ESP intensity

Subsection 2.1 defines and studies the ESP estimand, which is the quantity approximated by the integral of the ESP approximation, i.e., the intensity distribution of the solutions to the estimating equations. Subsection 2.2 studies finite-sample properties of the ESP approximation, and the ESP intensity, which is an integral of the former one.

2.1. **The ESP estimand.** We require the following Assumption 1 to define the estimand.

Assumption 1. **(a)** $(X_t)_{t=1}^{\infty}$ *is a sequence of random vectors of dimension* p *on a complete probability sample space* $(\mathbf{\Omega}, \mathcal{E}, \mathbb{P})$. **(b)** *Let the measurable space* $(\mathbf{\Theta}, \mathcal{B}(\mathbf{\Theta}))$ *be s.t. (such that)* $\mathbf{\Theta} \subset \mathbf{R}^m$ *is compact and* $\mathcal{B}(\mathbf{\Theta})$ *denotes the Borel* σ*-algebra on* $\mathbf{\Theta}$. **(c)** *The moment function* $\psi : \mathbf{R}^p \times \mathbf{\Theta} \to \mathbf{R}^m$ *is* $\mathcal{B}(\mathbf{R}^p) \otimes \mathcal{B}(\mathbf{\Theta}) / \mathcal{B}(\mathbf{R}^m)$*-measurable, where* $\mathcal{B}(\mathbf{R}^p) \otimes \mathcal{B}(\mathbf{\Theta})$ *denotes the product* σ*-algebra.* **(d)** *For the sample size at hand,* T*, the expectation of the number of solutions to the estimating equations is finite:* $\sum_{n=1}^{\infty} n p_{n,T} < \infty$ *where* $p_{n,T}$ *is the probability of having* n *solutions to the estimating equations.*

Assumptions 1(a) and (b) are weak and standard. Without loss of generality (e.g., Folland, 1984/1999, theo. 1.9), we assume completeness of the probability space to manipulate measure-zero sets. Compactness of the parameter space is a convenient mathematical assumption that is relevant in practice. A computer can only handle a bounded parameter space. Assumption 1(c) is the first departure from the GMM (generalized method of moments) literature. It requires equality between the dimension of the parameter space and the number of moment conditions. The reason is simple. In general, if the number of restrictions (estimating equations) exceeds the degrees of freedom (dimension of the parameter space), there is no solution to a system of equations. Thus, the probability weight that $\theta \in \mathbf{\Theta}$ solves the estimating equations is zero. Then, an approximation of the finite-sample distribution of the solutions to overrestricting estimating equations is generally not useful. However, following Newey and McFadden (1994, p. 2232), overrestricted estimating equations can be transformed into just-restricted estimating equations through an extension of the parameter space (Holcblat, 2012), so that Assumption 1(c) is mild in theory. Assumption 1(d), another mild departure from the GMM literature, means that the tails of the probability distribution of the number of solutions to the estimating equations are not too thick. Almudevar, Field and Robinson (2000) prove that Assumption 1(d) is implied by conditions in the spirit of the implicit function theorem combined with conditions on the distribution of the estimating equations normalized by the derivative of the latter. From a technical point of view, Assumption 1(d) allows us to use the standard point random-field theory, which is necessary to handle multiple solutions to nonlinear estimating equations. Skovgaard (1985; 1990) introduces this notion in the saddlepoint literature. However, the existing saddlepoint literature has usually attempted to narrow multiplicity to unicity, and thus evacuate point random-field theory in the end. To our knowledge, Sowell (2007) is the only paper that regards the ability of the ESP approximation to account for multiple solutions as an advantage, although he does not formalize it. His reliance on two-step GMM, a framework which requires a unique solution to the moment conditions, limits the possibility of such a theoretical development. In this paper, we take advantage of point random-field theory to to exploit the ability of the ESP approximation to account for multiple solutions to estimating equations. The following definition specializes the general definition of point random fields for our purpose.

Definition 1 (Point random field)**.** *Denote with $\mathcal{N}_{\mathbf{\Theta}}$ the space of finite simple counting measures on $\mathcal{B}(\mathbf{\Theta})$, i.e., the space consisting of finite integer-valued measures, N, s.t. for all $\theta \in \mathbf{\Theta}$, $N(\{\theta\}) \in \{0,1\}$. Denote with $\mathcal{B}(\mathcal{N}_{\mathbf{\Theta}})$ the Borel σ-algebra on $\mathcal{N}_{\mathbf{\Theta}}$ generated by the Prokhorov metric. A point random field (or point process) is a measurable mapping from $(\mathbf{\Omega}, \mathcal{E}, \mathbb{P})$ to $(\mathcal{N}_{\mathbf{\Theta}}, \mathcal{B}(\mathcal{N}_{\mathbf{\Theta}}))$.*[2]

In the present paper, a point random field is a mapping that maps each sample $(X_t(\omega))_{t=1}^{T}$ to the corresponding set of solutions to the estimating equations. More precisely, for a given sample size T, it maps each realization $\omega \in \mathbf{\Omega}$ to a counting measure, $N_T(\omega, .)$, which, in turn, maps any measurable subsets A of $\mathbf{\Theta}$ to the number of solutions to the estimating equations

[2]In the mathematical literature, the definition is typically more general. A point random field is defined as a measurable mapping to the space of integer-valued measures that are finite on bounded sets (e.g., Matthes, Kerstan and Mecke, 1974; Kallenberg, 1975; Daley and Vere-Jones, 2008).

contained in A, $N_T(\omega, A)$. The following proposition proves that it is actually the case $\mathbb{P}$-a.s. This is the core result of subsection 2.1.

Proposition 1. *Let $\#A$ denote the cardinality of the set A (i.e., the number of elements in A). Under Assumption 1, there exists a point random field, $N_T(.,.)$ and a $\mathbb{P}$-null set F, s.t. for all $\omega \in \mathbf{\Omega} \setminus F$ and $A \in \mathcal{B}(\mathbf{\Theta})$,*

$$N_T(\omega, A) = \# \left\{ \theta \in A : \frac{1}{T} \sum_{t=1}^{T} \psi\left(X_t(\omega), \theta\right) = 0 \right\}$$

Proof. See Appendix A.1 (p. 29). □

An implication of Proposition 1, which is of interest on its own and, which is used in the proofs of this paper, is the $\mathcal{E}/\mathcal{B}(\mathbf{R}^m)$-measurability of each of the solutions to the estimating equations.

Corollary 1 (Measurability of solutions to nonlinear estimating equations)**.** *Let $\dot{F}^c := \{\omega \in \mathbf{\Omega} \setminus F : 1 \leqslant N_T(\omega, \mathbf{\Theta})\}$. Assume that, for all $x \in \mathbf{R}^p$, $\theta \mapsto \psi(x, \theta)$ is continuous. Under Assumption 1, $\mathbb{P}$-a.s., each of the solutions to the estimating equations is $\mathcal{E}/\mathcal{B}(\mathbf{\Theta})$-measurable, i.e., for all $\omega \in \mathbf{\Omega} \setminus F$, if $\dot{\theta} \in \mathbf{\Theta}$ is such that $\frac{1}{T}\sum_{t=1}^{T} \psi(X_t(\omega), \dot{\theta}) = 0_{m\times 1}$, then there exits θ_T^* $\mathcal{E}/\mathcal{B}(\mathbf{\Theta})$-measurable s.t. $\theta_T^*(\omega) = \dot{\theta}$ and, for all $\tilde{\omega} \in \dot{F}^c$, $\frac{1}{T}\sum_{t=1}^{T} \psi(X_t(\tilde{\omega}), \theta_T^*(\tilde{\omega})) = 0_{m\times 1}$.*

Proof. See Proposition 9 (p. 33). □

Corollary 1 states that, $\mathbb{P}$-a.s., for each solution $\dot{\theta} \in \mathbf{\Theta}$ to the estimating equations, there exists a random element θ_T^* that is equal to $\dot{\theta}$, and that solves the estimating equations whenever they have a solution. As explained in the introduction, Corollary 1 generalizes Schmetterer-Jennrich's measurability result (Schmetterer, 1966, Ch. 5; Jennrich, 1969, Lemma 2), and Corollary 1 is of practical relevance beyond the saddlepoint literature.

Hereafter, for simplicity, we drop the dependence of the point random field on realizations $\omega \in \mathbf{\Omega}$. The distribution of the solutions to the estimating equations corresponds to the intensity measure associated with the point random field $N_T(.)$. If there can be only one solution to the estimating equations, the intensity measure is the probability distribution of the solution. But in the case of multiple solutions, we have to generalize probability measures into intensity measures.

Definition 2 (Intensity measure)**.** *Denote with $\mathcal{T} := (\mathcal{T}_n)_{n\geqslant 1}$ a dissecting system of $\mathbf{\Theta}$, i.e., a nested sequence of finite partitions $\mathcal{T}_n := \{A_{n,i} : i = 1, \ldots, k_n\}$ of Borel sets $A_{n,i}$ that separate* *all points of* $\mathbf{\Theta}$ *as* $n \to \infty$.[3] *The intensity measure of a finite point random field,* N_T*, is defined for all* $A \in \mathcal{B}(\mathbf{\Theta})$ *by*

$$\mathbb{F}_T(A) := \lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbb{P}\{N_T(A_{n,i}) = 1\}, \tag{1}$$

where $\mathcal{T}_n(A) := \{A_{n,i} \cap A : i = 1, \ldots, k_n \text{ and } A_{n,i} \in \mathcal{T}_n\}$.

Definition 2 defines the ESP estimand, i.e., the object that is approximated by the integral of the ESP approximation. The idea behind Definition 2 is the following. A singleton $\{\theta\}$ can contain at most one solution to the estimating equations, i.e., $\{\theta\}$ *is* or *is not* a solution. Thus, an intensity measure of a subset of $A \subset \mathbf{\Theta}$ can be defined as the sum of the probability weights that each of its elements contains a solution. There being an infinite number of elements, a sequence of increasingly thinner partitions has to be introduced to formalize the idea. Under Assumption 1, Definition 2 is equivalent to the general mathematical definition of intensity measures (e.g., Daley and Vere-Jones, 2008).[4]

The upcoming Corollary 2 ensures the existence of the intensity measure, i.e., the quantity approximated by the integral of the ESP approximation.

Corollary 2 (Existence of the ESP estimand)**.** *Under Assumptions 1, there exists an intensity measure* $\mathbb{F}_T$ *of the point random field* N_T*, i.e., there exists a finite measure* $\mathbb{F}_T : \mathcal{B}(\mathbf{\Theta}) \to \mathbf{R}_+$*, which satisfies equation (1) in Definition 2.*

Proof. Apply Lemma 2 (Appendix A.2 on p. 30) □

Corollary 2 follows from Proposition 1, and known results from point random-field theory. Namely, it follows from the existence of dissecting systems, stability of dissecting systems under restriction to subsets, finiteness and countable additivity of $\mathbb{F}_T$, and invariance of the intensity measure w.r.t. dissecting systems. The following proposition clarifies the relation between intensity measures and probability measures. It adapts a result from point random field theory.

---

[3]More precisely, a sequence $\mathcal{T} := (\mathcal{T}_n)_{n\geqslant 1}$ of sets $\mathcal{T}_n := \{A_{n,i} : 1 \leqslant i \leqslant k_n\}$ consisting of a finite number of Borel sets $A_{n,i}$ is a dissecting system of $\mathbf{\Theta}$ if

i) (partition properties) $A_{n,i} \cap A_{n,j} = \varnothing$ for $i \neq j$ and $A_{n,1} \cup \ldots \cup A_{n,k_n} = \mathbf{\Theta}$;
ii) (nesting property) $A_{n-1,j} \cap A_{n,j} = A_{n,j}$ or $\varnothing$; and
iii) (point-separating property) $\forall(\theta_1, \theta_2) \in \mathbf{\Theta}^2$ s.t. $\theta_1 \neq \theta_2$, $\exists n \in \mathbf{N}$ s.t. $\theta_1 \in A_{n,i}$ implies $\theta_2 \notin A_{n,i}$.

[4]It is also a generalization of the concepts of compensator or integrated intensity that are typically used to model defaults with point processes over the half-real line in continuous-time finance (e.g., Duffie, 1992/2001, chap. 11).

Proposition 2. *Under Assumptions 1, for $\mathbb{F}_T$-almost every $\theta \in \mathbf{\Theta}$,*

$$\mathbb{F}_T(A_n(\theta)) = \mathbb{P}\{N_T(A_n(\theta)) = 1\}(1 + \varepsilon_n(\theta))$$

where $\varepsilon_n(\theta) \downarrow 0$ as $n \to \infty$, and $A_n(\theta)$ denotes the element of $\mathcal{T}_n := \{A_{n,i}\}_{1 \leqslant i \leqslant k_n}$ that contains θ.

Proof. See Appendix A.3 (p. 30). □

In accordance with the idea behind Definition 2, the intensity measure of a sufficiently small set is approximately the probability that it contains one solution.

Theorem 1(iii) in Almudevar, Field and Robinson (2000) is a precursor of Proposition 2, which is its counterpart in our setup. Almudevar, Field and Robinson (2000) also formalize the point random field introduced by Skovgaard (1985, p. 95), and thus our subsection 2.1 is close to their section 2. The main differences between their setup and ours are the following. They implicitly assume the existence of the point random fields that they define, while we prove the existence of the point random field that we define (see Proposition 1). Because they construct a point process that discards continuums or accumulations of solutions to estimating equations, their setup does not need to rule them out, while we immediately rule them out $\mathbb{P}$-a.s. thanks to Assumption 1(d). They need additional assumptions (Assumption A2 in Almudevar, Field and Robinson, 2000) and results (Theorem 1 in Almudevar, Field and Robinson, 2000) to define their setup, while we can adapt point random-field theory without additional assumption. For example, if the support of the distribution of the vector of data, X, is discrete, in contrast to our setup, theirs does not hold.

2.2. **The ESP intensity.** The ESP intensity is the measure with the ESP approximation as its Radon-Nikodyn derivative w.r.t. the Lebesgue measure. In this subsection, firstly, we define the rough ESP approximation. Although the rough ESP approximation seems appropriate in practice, for mathematical reasons we need to introduce a smooth version of it. Thus, secondly, we show how we can define the (smooth) ESP approximation by arbitrarily slightly modifying the rough ESP approximation. Then, we define the ESP intensity measure of a subset as the integral of the ESP approximation over the subset. As in the previous subsection, T remains fixed to the size of the sample at hand. The definition of the rough ESP approximation requires the following assumption.

Assumption 2. *There exists* $\varepsilon > 0$ *s.t. for all* $x \in \mathbf{R}^p$, $\theta \mapsto \psi(x, \theta)$ *is continuously differentiable in* $\{\theta \in \mathbf{R}^m : \rho(\theta, \mathbf{\Theta}) < \varepsilon\}$ *where* $\rho(\theta, \mathbf{\Theta}) := \inf_{\dot{\theta} \in \mathbf{\Theta}} \|\theta - \dot{\theta}\|$.

Assumption 2 means that $\psi(., .)$ is continuously differentiable with respect to its second argument in an ε-neighborhood of $\mathbf{\Theta}$. This is a mild and convenient variant of the more standard assumption that requires continuous differentiability of $\psi(., .)$ in $\mathbf{\Theta}$.[5] Assumption 2 allows the application of the implicit function theorem on the boundary of $\mathbf{\Theta}$ when necessary. We define the rough ESP approximation as follows.

Definition 3 (Rough ESP approximation)**.** *The rough ESP approximation is*

$$\hat{f}_{\theta_T^*, sp}(\theta) := \exp\left\{T \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}\left(\frac{T}{2\pi}\right)^{m/2} |\Sigma_T(\theta)|_{\det}^{-\frac{1}{2}} \tag{2}$$

where $|.|_{\det}$ *denotes the determinant function,* $\psi_t(.) := \psi(X_t, .)$ *and*

$$\begin{aligned}
\Sigma_T(\theta) &:= \left[\sum_{t=1}^{T} \hat{w}_{t,\theta} \frac{\partial \psi_t(\theta)'}{\partial \theta}\right]^{-1} \left[\sum_{t=1}^{T} \hat{w}_{t,\theta} \psi_t(\theta)\psi_t(\theta)'\right] \left[\sum_{t=1}^{T} \hat{w}_{t,\theta} \frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1}, \\
\hat{w}_{t,\theta} &:= \frac{\exp\left[\tau_T(\theta)'\psi_t(\theta)\right]}{\sum_{i=1}^{T} \exp\left[\tau_T(\theta)'\psi_i(\theta)\right]}, \\
\tau_T(\theta) &\ s.t. \sum_{t=1}^{T} \psi_t(\theta) \exp\left[\tau_T(\theta)'\psi_t(\theta)\right] = 0_{m\times 1},
\end{aligned} \tag{3}$$

wherever it exists.

Approximation (2) was first studied by Ronchetti and Welsh (1994), who extended the work of Feuerverger (1989) for means to Z-estimators. Approximation (2) is constructed point-wise. For each $\theta \in \mathbf{\Theta}$, the ET (exponential tilting) term indicates the extent of the change of measure that is needed to set the estimating equations (3) to zero. The remaining terms discount the ET term according to the variance of the solution to the estimating equations. For each $\theta \in \mathbf{\Theta}$, the bigger the change of measure or the variance, the smaller the rough ESP approximation.

We call approximation (2) the rough ESP approximation to distinguish it from the (smooth) ESP approximation below. Despite its name, the rough ESP approximation is unique and continuous wherever it exists. Moreover, its domain of definition is $\mathcal{B}(\mathbf{\Theta})$-measurable.

[5]From a practical point of view, if ε is set to half of the smallest machine epsilon available, Assumption 2 is typically not different from the differentiability of $\psi(., .)$ in $\mathbf{\Theta}$.

Proposition 3. *Define the set $\hat{\mathbf{\Theta}}_T \subset \mathbf{\Theta}$ where the rough ESP approximation exists, that is, let*

$$\hat{\mathbf{\Theta}}_T := \left\{ \theta \in \mathbf{\Theta} : \exists \tau_T(\theta) \in \mathbf{R}^m \ s.t. \ \sum_{t=1}^{T} \psi_t(\theta) \mathrm{e}^{\tau_T(\theta)' \psi_t(\theta)} = 0_{m \times 1} \ and \ |\Sigma_T(\theta)|_{\det} \neq 0 \right\}.$$

Under Assumptions 1(a)–(c) and 2,

i) *$\hat{\mathbf{\Theta}}_T$ is an open subset of $\mathbf{\Theta}$;*

ii) *the rough ESP approximation, $\hat{f}_{\theta^*,sp}(.)$, is continuous and unique in $\hat{\mathbf{\Theta}}_T$.*

Proof. See Appendix A.4 (p. 30). □

The continuity of the rough ESP approximation is remarkable for a nonparametric estimate of a distribution obtained without smoothing. Nevertheless, the rough ESP approximation can have two undesirable properties. Primarily, it ignores the information provided by the absence of a solution to the tilting equation (3) because the rough ESP approximation does not exist when there is no solution. The following proposition clarifies the information provided by the absence of a solution to the tilting equation (3).

Proposition 4. *Denote with $[\![1, T]\!]$ the integers in $[1, T]$. Under Assumptions 1(a)–(c) and 2, for all $\theta \in \mathbf{\Theta}$, there exists $\tau \in \mathbf{R}^m$ s.t. $\sum_{t=1}^{T} \psi_t(\theta) \exp[\tau' \psi_t(\theta)] = 0_{m \times 1}$, if and only if there exists a probability distribution $(p_1, p_2, \ldots, p_T)$, with $\sum_{t=1}^{T} p_t = 1$ and $p_t > 0$ for all $t \in [\![1, T]\!]$, s.t. $\sum_{t=1}^{T} \psi_t(\theta) p_t = 0_{m \times 1}$.*

Proof. See Appendix A.5 (p. 30). □

Proposition 4 states a result implicitly used in Theorem 1 in Schennach (2005). It is a direct implication of the duality between the solution to the tilting equation (3) and maximization of entropy under moment conditions (e.g., Csiszár, 1975, sec.3(A)).

Proposition 4 indicates that the particular form of change of measure applied to the empirical distribution through ET does not restrict the set of parameter values that admits a solution to the tilting equation (3). In terms of parameter values that have a solution to the tilting equation (3), ET spans a class of probability measures as rich as the class of probability measures that have the data points as support (i.e., the class of probability measures equivalent to the empirical distribution). Thus, the absence of a solution to the tilting equation for a parameter value $\theta \in \mathbf{\Theta}$ means that the sample at hand does not provide support for this parameter value being a solution to the estimating equations. Consequently, we set the ESP approximation to zero for parameter values without a solution to the tilting equation.[6]

The second undesirable property of the ESP approximation is that it may not be defined for $\theta \in \mathbf{\Theta}$ s.t. $|\Sigma_T(\theta)|_{\det} = 0$, although there exists $\tau \in \mathbf{R}^m$ s.t. $\sum_{t=1}^{T} \psi_t(\theta) \mathrm{e}^{\tau' \psi_t(\theta)} = 0_{m \times 1}$. In this case, the ESP approximation would not provide an indication of the probability of being a solution to the estimating equations for a parameter value consistent with the tilted estimating equations. Assumption 3 rules out such a case.

Assumption 3. *Define* $\overline{\mathbf{\Theta}}_T := \left\{\theta \in \mathbf{\Theta} : \exists \tau \in \mathbf{R}^m \text{ s.t. } \sum_{t=1}^{T} \psi_t(\theta) \mathrm{e}^{\tau' \psi_t(\theta)} = 0\right\}$. *Assume* $\overline{\mathbf{\Theta}}_T = \hat{\mathbf{\Theta}}_T$.

Assumption 3, which can be numerically checked in practice, appears mild as a square matrix is generically non-singular. Moreover, from a mathematical point of view, Assumption 3 is not needed in the sense that no proofs in this paper require it. Assumption 3 combined with Proposition 4, allows us to meaningfully set the ESP approximation to zero on the complement of $\hat{\mathbf{\Theta}}_T$. Nevertheless, we also want to preserve continuity so that we introduce a continuous version of the rough ESP approximation. The integral of this continuous version corresponds to the ESP intensity.

Definition 4 (ESP intensity)**.** *For* $\eta > 0$, *and any set* $A \subset \mathbf{\Theta}$, *denote* $A^{-\eta} := \{a \in A : \rho(a, \partial A \cap \partial(A^c)) \geqslant \eta\}$ *where* A^c *denotes the complement of* A *in* $\mathbf{\Theta}$. *Under the notation of Proposition 3, for a small* $\eta > 0$, *the ESP intensity is a set function* $\tilde{\mathbb{F}}_T$ *s.t. for all* $A \in \mathcal{B}(\mathbf{\Theta})$,

$$\tilde{\mathbb{F}}_T(A) := \int_A \tilde{f}_{\theta_T^*, sp}(\theta) \mathrm{d}\theta$$

where for all $\omega \in \mathbf{\Omega}$, $\theta \mapsto \tilde{f}_{\theta_T^*, sp}(\theta)$ *is a positive continuous function, s.t., for all* $\theta \in \hat{\mathbf{\Theta}}_T^{-\eta}$, $\tilde{f}_{\theta_T^*, sp}(\theta) = \hat{f}_{\theta_T^*, sp}(\theta)$, *for all* $\theta \in \hat{\mathbf{\Theta}}_T^c$, $\tilde{f}_{\theta_T^*, sp}(\theta) = 0$, *and for all* $\theta \in \hat{\mathbf{\Theta}}_T \cap \left(\hat{\mathbf{\Theta}}_T^{-\eta}\right)^c$, $\tilde{f}_{\theta_T^*, sp}(\theta) \in [0, \bar{f}_T]$ *with* $\bar{f}_T := \sup_{\theta \in \left\{\partial \hat{\mathbf{\Theta}}_T^{-\eta}\right\}} \hat{f}_{\theta_T^*, sp}(\theta)$ *if* $\hat{\mathbf{\Theta}}_T^{-\eta} \neq \emptyset$, *or* 0 *otherwise. Moreover, for all* $\theta \in \mathbf{\Theta}$, $\omega \mapsto \tilde{f}_{\theta_T^*, sp}(\theta)$ *is* $\mathcal{E}/\mathcal{B}(\mathbf{R})$*-measurable.*

The following Proposition 6 ensures the existence of the ESP intensity, and guarantees that it is an intensity measure.

[6]In this way, we are in the spirit of Schennach (2005), in which a function of the parameter, which consists of the product of ET weights multiplied by a prior and which is interpreted as a Bayesian posterior, is extended to parameter values without solution to the tilting equation by setting the function to zero. Note, however, that the ESP approximation is an approximation of the distribution of the solutions to the estimating equations, and that we stick to its mathematical definition so that we do not interpret it as a Bayesian posterior (see supplemental material).

Proposition 5 (Existence and property of the ESP intensity)**.** *Under Assumptions 1–2,*

i) *the ESP intensity* $\tilde{\mathbb{F}}(.)$ *exists;*

ii) *the ESP intensity* $\tilde{\mathbb{F}}(.)$ *is an intensity measure, i.e., there exists a point random field,* $\tilde{N}_T(.)$ *on a probability space* $(\tilde{\mathbf{\Omega}}, \tilde{\mathcal{E}}, \tilde{\mathbb{P}})$*, s.t., for all* $A \in \mathcal{B}(\mathbf{\Theta})$*,* $\tilde{\mathbb{F}}_T(A) := \lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \tilde{\mathbb{P}}\{\tilde{N}_T(A_{n,i}) = 1\}$*, where* $\mathcal{T}_n(A)$ *is a dissecting system restricted to* A*.*

Proof. See Appendix A.7 (p. 31). □

3. Asymptotic limit of the ESP intensity

Whereas in the previous sections T remains fixed to the size of the sample at hand, in this section T goes to infinity. We here establish the consistency and asymptotic normality of the ESP intensity in the sense of the Prokhorov metric. By consistency, we mean convergence of the ESP intensity measure to a Dirac distribution at the true parameter. By asymptotic normality, we mean convergence of the standardized ESP intensity to a standard normal distribution. Both results are new to the literature. The key tools of our proof are the implicit function theorem, and a slight generalization of the Laplace approximation presented in Kass, Tierney and Kadane (1990). For simplicity, we adapt the basic assumptions of Kass, Tierney and Kadane (1990).

To study the asymptotic behavior of the ESP intensity, we should first be able to study the asymptotic behavior of the ESP estimand. Assumption 4, combined with Assumptions 6(a)–(b) below, allows us to study of the asymptotic limit of the ESP estimand.

Assumption 4. **(a)** $(X_t)_{t=1}^{\infty}$ *are i.i.d.* **(b)** *In the parameter space* $\mathbf{\Theta}$*, there exists a unique solution* $\theta_0 \in \operatorname{int}(\mathbf{\Theta})$ *to the moment conditions* $\mathbb{E}\left[\psi(X,\theta)\right] = 0_{m\times 1}$*.* **(c)** $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}} \|\psi(X,\theta)\|\right] < \infty$*.* **(d)** $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}} \left\|\frac{\partial\psi(X,\theta)}{\partial\theta'}\right\|\right] < \infty$*.* **(e)** $\left|\mathbb{E}\left[\frac{\partial\psi(X,\theta_0)}{\partial\theta'}\right]\right|_{\det} \neq 0$*.*

Assumption 4 is basic and standard. Assumption 4(a) can be relaxed to allow time dependence along the lines of Kitamura and Stutzer (1997). We require such an assumption for simplicity. Assumption 4(b) requires the estimating equations to also be empirical moment conditions (see footnote 1 on p. 2), and it ensures global identification. It can be relaxed as pointed out in the upcoming Remark 2 (p. 17). Assumption 4(c) ensures convergence of the solution to the empirical moment conditions to the true parameter. Assumptions 4(d) and (e) ensure the existence of solutions to the empirical moment conditions.

The remaining assumptions of this subsection allow us to study the asymptotic behavior of the ESP intensity. Assumption 5 ensures the asymptotic existence of the ESP intensity in a set that includes a neighborhood of the true parameter.

Assumption 5. *Define the set*

$$\hat{\mathbf{\Theta}} := \left\{ \theta \in \mathbf{\Theta} : \exists \tau(\theta) \in \mathbf{R}^m \text{ s.t. } \begin{array}{c} \exists r > 0,\, \forall \tau \in B_r(\tau(\theta)),\, \mathbb{E}\left[\mathrm{e}^{\tau' \psi(X,\theta)}\right] < \infty \\ \left\| \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)} \frac{\partial \psi(X,\theta)}{\partial \theta}'\right] \right\| < \infty \\ |\Sigma(\theta)|_{\det} \neq 0 \\ \mathbb{E}\left[\psi(X,\theta)\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] = 0_{m\times 1} \end{array} \right\},$$

where $\Sigma(\theta) := \left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X,\theta)} \frac{\partial \psi(X,\theta)}{\partial \theta}'\right]^{-1} \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)} \psi(X,\theta)\psi(X,\theta)'\right] \left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X,\theta)} \frac{\partial \psi(X,\theta)}{\partial \theta'}\right]^{-1}$ **(a)** *There exists* $\bar{r} > 0$ *s.t. there exists* $\dot{T} \in \mathbf{N}$, *so that for all* $T \geqslant \dot{T}$, $B_{\bar{r}}(\theta_0) \subset \left\{\hat{\mathbf{\Theta}}_T \cap \hat{\mathbf{\Theta}}\right\}$. *Define a fixed* $\eta \in]0, \bar{r}[$. **(b)** *For all* $\dot{\theta} \in \hat{\mathbf{\Theta}}^{-\eta}$, *there exist* $r_1, r_2 > 0$ *s.t.* $\mathbb{E}\left[\sup_{(\tau,\theta) \in B_{r_1}(\tau(\dot{\theta})) \times B_{r_2}(\dot{\theta})} \|\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\|\right] < \infty$.

The set $\hat{\mathbf{\Theta}}$ corresponds to the parameter values where the limit of the rough ESP intensity exists. In particular, the first two conditions ensure that $|\Sigma(\theta)|_{\det} < \infty$ by a standard result on Laplace transforms. Assumption 5(a) ensures that the rough ESP intensity is asymptotically well defined in a fixed neighborhood of the true parameter. Assumption 5(b) allows us to obtain the continuity of $\theta \mapsto \tau(\theta)$ by an implicit-function theorem.

Assumption 6 ensures the validity of the Laplace approximation in a fixed neighborhood of the true parameter, and thus in a fixed neighborhood of any solution to the empirical moment conditions for T big enough by consistency.

Assumption 6. **(a)** *For all* $x \in \mathbf{R}^p$, *the function* $\theta \mapsto \psi(x,\theta)$ *is three times continuously differentiable in a neighborhood of* θ_0. **(b)** *For all* $k \in [\![1, 2]\!]$, *there exists* $r > 0$, $\mathbb{E}\left[\sup_{\theta \in B_r(\theta_0)} \|D^k \psi(X,\theta)\|\right] < \infty$ *where* D^k *denotes the differential operator w.r.t.* θ *of order* k. **(c)** *There exist* $M \geqslant 0$, $\dot{T} \in \mathbf{N}$ *and* $r > 0$, *so that for all* $T \geqslant \dot{T}$ *and* $\theta \in B_r(\theta_0)$, $\left\| D\left\{|\Sigma_T(\theta)|_{\det}^{-\frac{1}{2}}\right\}\right\| < M$ $\mathbb{P}$*-a.s.* **(d)** *There exist* $\dot{T} \in \mathbf{N}$ *and* $r > 0$, *so that for all* $T \geqslant \dot{T}$ *and* $\theta \in B_r(\theta_0)$, $\left\| D^3 \left\{\ln\left[\frac{1}{T}\sum_{i=1}^T \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}\right\| < M$ $\mathbb{P}$*-a.s.* **(e)** *There exists* $r > 0$, $\left\|\mathbb{E}\left[\sup_{\theta \in B_r(\theta_0)} \psi(X,\theta)\psi(X,\theta)'\right]\right\| < \infty$.

Assumptions 6(a), (c) and (e), adapted from Kass, Tierney and Kadane (1990), essentially ensure the existence and boundedness of the derivatives of the ESP intensity terms in a neighborhood of the true parameter. Assumption 6(b), combined with Assumption 4, ensures the asymptotic normality of the solution to the empirical moment conditions. Assumption 6(e) ensures the validity of the implicit function theorem for the tilting parameter, $\tau_T(\theta)$, at any solution to the empirical moment conditions for T big enough.

Assumption 7 ensures the convergence of the ESP intensity to zero outside a neighborhood of the true parameter.

Assumption 7. *Let $\eta > 0$ be defined as in Assumption 5(a).* **(a)** *For all $\varepsilon > 0$, there exists $\dot{T} \in \mathbf{N}$ and $M \geqslant 0$ s.t. $T \geqslant \dot{T}$ implies that, for all $\theta \in \hat{\mathbf{\Theta}}^{-\eta}$, $\mathrm{e}^{-\varepsilon T} \left|\Sigma_T(\theta)\right|_{\det}^{-\frac{1}{2}} \leqslant M$ $\mathbb{P}$-a.s.* **(b)** *For all $\dot{\theta} \in \hat{\mathbf{\Theta}}^{-\eta}$, there exist $r_1, r_2 > 0$ s.t. $\mathbb{E}\left[\sup_{(\tau,\theta)\in B_{r_1}(\tau(\dot{\theta}))\times B_{r_2}(\dot{\theta})} \mathrm{e}^{\tau'\psi(X,\theta)}\right] < \infty$.*

Assumption 7 corresponds to assumption (iii) in Kass, Tierney and Kadane (1990). Assumption 7(a) rules out more than exponential divergence of the Jacobian of the ESP intensity. This is a mild assumption. Assumption 7(b) is a convenient variant of Assumption 4 in Kitamura and Stutzer (1997). It is not as strong as it may appear because observed quantities have typically finite support, which, in turn, implies that they have finite moments. Assumption 7(b) is a common type of assumption in entropy-based inference.

Under the above assumptions, we the consistency of the ESP intensity.

Theorem 1 (Consistency)**.** *Under Assumptions 1(a)–(c),2, 4, 5, 6(a)(c)(d)(e) and 7, as $T \to \infty$, the ESP smooth intensity, $\tilde{f}_{\theta_T^*,sp}(.)$, converges in distribution (or narrowly converges) to the Dirac distribution $\delta_{\theta_0}(.)$ $\mathbb{P}$-a.s.:*

$$\forall \varphi \in C_b, \qquad \int_{\mathbf{\Theta}} \varphi(\theta)\tilde{f}_{\theta_T^*,sp}(\theta)\mathrm{d}\theta \to \int_{\mathbf{\Theta}} \varphi(\theta)\delta_{\theta_0}(\theta)\mathrm{d}\theta \quad \mathbb{P}\text{-a.s.},$$

where C_b denotes the space of continuous bounded functions.

Proof. See Appendix A.8 (p. 31). □

Theorem 1 means that the ESP intensity measure converges to a point mass at the true parameter as the sample size increases. By the triangle inequality for the Prokhorov metric, Theorem 1 also implies that the ESP intensity and the ESP estimand (e.g., the intensity distribution of the solutions) converge towards each other as sample size increases.

The counterpart of Theorem 1 in Bayesian inference is the consistency of posterior distributions (e.g., Doob's theorem). However, Theorem 1 is stronger than theorems on the consistency of posterior distributions, in the sense that the ESP intensity integrates to one asymptotically, although it is not normalized by its integral (e.g., Chen, 1985; Kim, 2002; Ghosh and Ramamoorthi, 2003, sec. 1.4; Chernozhukov and Hong, 2003). This remarkable property is all the more surprising as ESP approximations are the result of a pointwise construction.

A second standard convergence result for Bayesian posterior distributions is asymptotic normality (or the Laplace–Bernstein–von Mises' theorem). We also provide its counterpart for the ESP intensity.

Theorem 2 (Asymptotic Normality)**.** *Let $a, b \in \boldsymbol{\Theta}$ s.t. $a \leqslant b$ where "$a \leqslant b$" means that every component of $b - a$ is nonnegative. Then, under Assumptions 1(a)–(c),2, 4, 5, 6(a)(c)(d)(e) and 7, as $T \to \infty$,*

$$\int_{D_T(a,\theta_T^*,b)} \tilde{f}_{\theta_T^*,sp}(\theta)\mathrm{d}\theta \to \frac{1}{(2\pi)^{\frac{m}{2}}} \int_{D(a,b)} \mathrm{e}^{-\frac{1}{2}s's}\mathrm{d}s \quad \mathbb{P}\text{-}a.s.,$$

where $D_T(a,\theta_T^*,b) := \left\{\theta : \theta_T^* + T^{-\frac{1}{2}}\left[\Sigma_T(\theta_T^*)\right]^{\frac{1}{2}} a \leqslant \theta \leqslant \theta_T^* + T^{-\frac{1}{2}}\left[\Sigma_T(\theta_T^*)\right]^{\frac{1}{2}} b\right\}$ with any solution, θ_T^*, to the empirical moment conditions, and $[\Sigma_T(\theta_T^*)]^{\frac{1}{2}}$ s.t. $\Sigma_T(\theta_T^*) = \left([\Sigma_T(\theta_T^*)]^{\frac{1}{2}}\right)'$ $\times [\Sigma_T(\theta_T^*)]^{\frac{1}{2}}$ and $[\Sigma_T(\theta_T^*)] := \left[\frac{1}{T}\sum_{t=1}^T \frac{\partial\psi_t(\theta_T^*)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T \psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]\left[\frac{1}{T}\sum_{t=1}^T \frac{\partial\psi_t(\theta_T^*)'}{\partial\theta}\right]^{-1}$, and $D(a,b) := \{s : a \leqslant s \leqslant b\}$.

Proof. See Appendix A.8 (p. 31). □

Theorem 2 indicates that the ESP intensity converges asymptotically to a point mass at the true parameter like a Gaussian distribution with a standard deviation that goes to zero at the rate $T^{-\frac{1}{2}}$. Theorem 2 is in line with the well-known asymptotic normality of a solution to empirical moment conditions. Theorem 2 is close to Theorem 5 in Sowell (2007), although the latter does not provide the asymptotic normality of the ESP intensity.

Remark 1. While in "$D_T(a,\theta_T^*,b)$," "θ_T^*" denotes a random variable that maps an $\omega \in \boldsymbol{\Omega}$ to *one* of the potentially multiple solutions to the empirical moment conditions, in some other places in the paper "θ_T^*" implicitly denotes the random correspondence that maps an $\omega \in \boldsymbol{\Omega}$ to the *set* of solutions to the empirical moment conditions (which has finite cardinality $\mathbb{P}$-a.s. by Assumption 1(d)). For example, a few lines above, in the subscript of $\tilde{f}_{\theta_T^*,sp}(.)$, "$\theta_T^*$" refers to the latter because an intensity is, by construction, about all the possible solutions (see Definition

2). For simplicity, we refrain from introducing two different notations. Moreover, the difference between the two meanings disappears when there can be only one solution to the empirical moment conditions. ◇

Remark 2. Theorems 1 and 2 can be extended to the case in which there are multiple solutions to the moment conditions $\mathbb{E}[\psi(X,\theta)] = 0_{m\times 1}$. Choose a partition of the parameter space such that each element of the partition contains only one solution to the moment conditions. Then, apply Theorems 1 and 2 to each element of the partition. ◇

4. Empirical evidence from asset pricing

4.1. **Setup.** In this section, we present empirical evidence from consumption-based asset pricing using the ESP approximation and the main existing moment-based methods. For brevity and clarity, we only estimate the RRA of the representative agent. We rely on a key moment condition of consumption-based asset-pricing theory,

$$\mathbb{E}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_t^m - R_t^f)\right] = 0, \tag{4}$$

where $\frac{C_t}{C_{t-1}}$ is the growth consumption and $(R_t^m - R_t^f)$ the market return in excess of the risk-free rate. The moment condition (4) is as consistent with Lucas (1978) as with more recent consumption-based asset-pricing models, such as Barro (2006) or Gabaix (2012). The moment condition and data are similar to Julliard and Ghosh (2012) corresponding to standard US data at yearly frequency from Shiller's website spanning from 1890 to 2009. Supplemental material includes empirical evidence from another data set. See Julliard and Ghosh (2012) for a more detailed data description.

We estimate the RRA using GMM (Hansen, 1982), CU GMM (Hansen, Heaton and Yaron, 1996) as an example of GEL estimators, CU GMM for lack of identification (Stock and Wright, 2000), which generalizes Anderson and Rubin (1949), and the ESP approximation. Although, for simplicity we restrained ourselves to the i.i.d. case in the previous sections, it does not matter here for implementation as there is no serial correlation theoretically (the moment condition (4) corresponds to a martingale difference) and empirically (e.g., Hall, 2005, pp. 86–87). In the case of the ESP approximation, we normalize the latter so that it integrates to one, and then we use the estimator and the confidence region defined in Holcblat and Grønneberg (2015). The estimator is the mode of the ESP approximation, and thus it corresponds to the estimator introduced in Sowell (2009). The 95% confidence region is the shortest closed set $R_{.95,T}$ s.t. $\frac{\tilde{\mathbb{F}}_T(R_{.95,T})}{\tilde{\mathbb{F}}_T(\boldsymbol{\Theta})} \geqslant .95$.

Remark 3. The popularity of moment-based estimation in consumption-based asset pricing, and more generally in economics is due to the fact that moment-based estimation does not necessarily require the specification of a family of distributions for the data (e.g., Hansen, 2001; Hall, 2005, pp. 1–2; Hansen, 2013, sec. 3). Typically, an economic model does not imply such family of distributions, except for tractability reasons. Imposing a family of distributions makes it difficult to disentangle the part of the inference results due to the empirical relevance of the economic model from the part due to these additional restrictions. Under regularity conditions, assuming a distribution corresponds to imposing an infinite number of extra moment restrictions: A characteristic function uniquely determines a probability distribution; and if the characteristic function of a random variable X is analytic in a neighborhood of zero, then it can be expanded at zero into an infinite Taylor series $\mathbb{E}\left(\mathrm{e}^{iuX}\right) = \sum_{j=0}^{\infty} \frac{(iu)^j}{j!} \mathbb{E}\left(X^j\right)$ where i denotes here the imaginary unit. ◇

4.2. **Empirical evidence.** Table 1 reports the GMM results. On Table 1(A) and (A zoom), the GMM objective function is relatively flat on a large area so that it does not have a well-separated global minimum. This is a common feature in empirical consumption-based asset pricing (e.g., Stock and Wright, 2000; Hall, 2005, pp. 62-64), which generates unstable point estimates. However, as shown on Table 1(B), standard GMM summarizes inference as if the uncertainty about the true parameter corresponded to a Gaussian distribution centered at the global minimum, and with a variance corresponding to the local curvature. Thus, there is a dichotomy between the information extracted from data through the GMM objective function and the asymptotic Gaussian template used to summarized it.

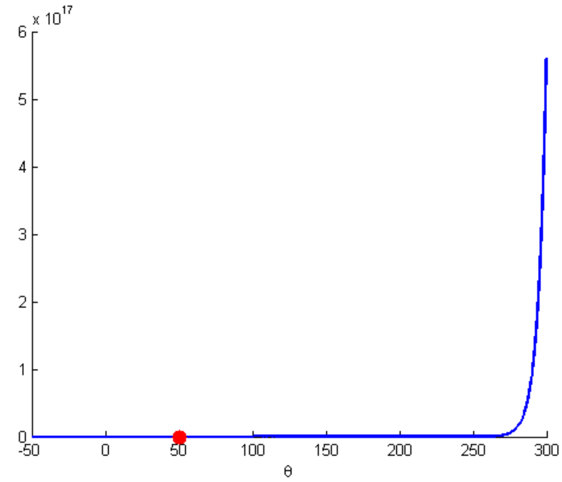
Tables 2(A) and (B) shows the results from CU GMM. The point estimate and the standard confidence region based are almost identical to the GMM ones. In fact, more generally, in the just-restricted case, which is the case considered in the paper, GMM, EL, ET and CU GMM should yield the same point estimate (the solution to the realized empirical moment conditions) by construction. Nevertheless, on Table 2(C), the confidence regions for weak identification, the S-sets, are quite different. By definition, in our case, an S-set is

$$\left\{\theta \in \boldsymbol{\Theta} : TQ_{T,\mathrm{CU}}(\theta) < c_\alpha\right\},$$

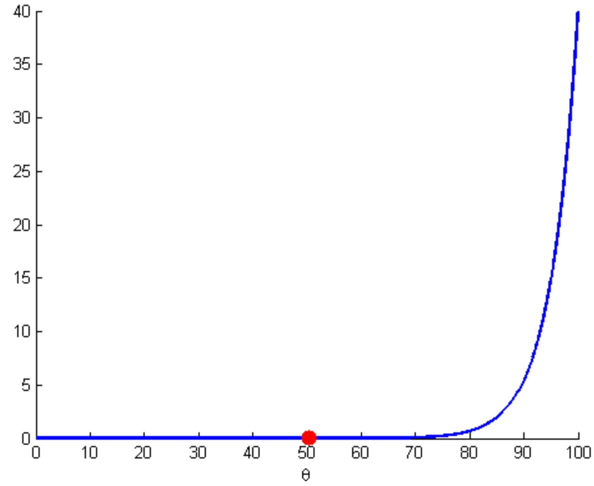
TABLE 1. **GMM inference (1890–2009)**

$$\frac{1}{2009-1889}\sum_{t=1890}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_t^m - R_t^f)\right] = 0,$$

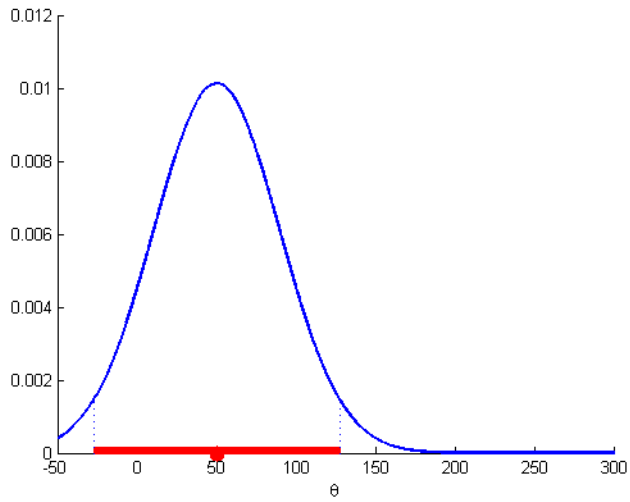
$R_t^m :=$ gross market return, R_t^f :=risk-free asset gross return, $C_t :=$ consumption,
θ :=relative risk aversion,
$\hat{\theta}_{\text{GMM}} = 50.3$, $\hat{I}_{.05}^{\text{GMM}} = [-26.9, 127.4]$.



(A) GMM objective function and point estimate.



(A zoom) GMM objective function and point estimate.



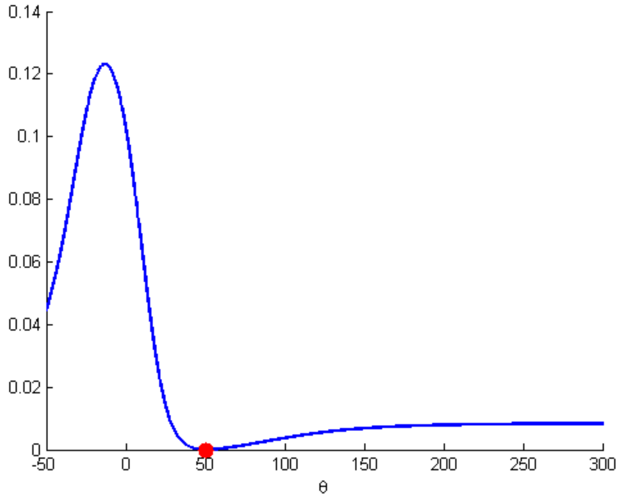
(B) Gaussian distribution, point estimate and confidence interval.

where c_α is the α quantile of a chi-square of degree one and $Q_{T,\text{CU}}(.)$ is the CU GMM objective function, i.e., $Q_{T,\text{CU}}(\theta) := \left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right]' \left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right]$. Now, as documented in the literature (e.g., Hansen, Heaton and Yaron, 1996), CU GMM objective functions tend to be flat and low in the tails. Thus, an S-set can be huge, as in Table 2(C), so that it is not very informative. In the less favourable case with overrestricting moment

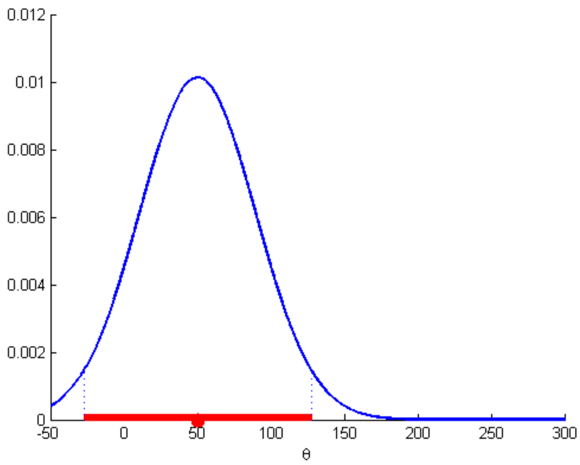
Table 2. **Continuously updated (CU) GMM inference (1890–2009)**

$$\frac{1}{2009-1890}\sum_{t=1890}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_t^m - R_t^f)\right] = 0,$$

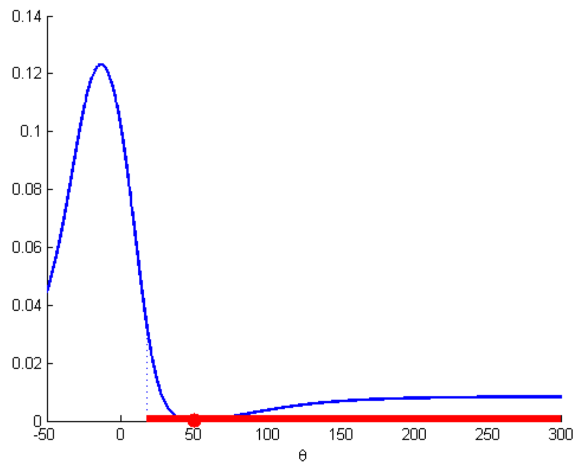
R_t^m := gross market return, R_t^f :=risk-free asset gross return, C_t := consumption,
θ :=relative risk aversion,
$\hat{\theta}_{\mathrm{CU}} = 50.3$ (bullet), $\hat{I}^{\mathrm{CU}}_{.05} = [-26.9, 127.4]$ (stripe in B),
$\hat{I}^{\mathrm{S}}_{.05} = [18.2, \mathit{3890}]$ (lower bound in italic; stripe in C)
Rk: We constrain the numerical search for point estimate to discard large values of θ.



(A) Objective function and point estimate.



(B) Gaussian dist., point estimate,confidence interval.



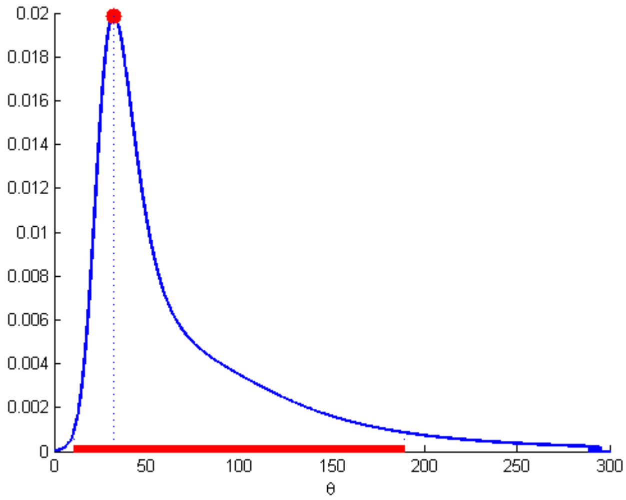
(C) Right-side truncated S-set.

conditions, S-sets are generally empty (e.g., Stock and Wright, 2000) so that they are not very informative.

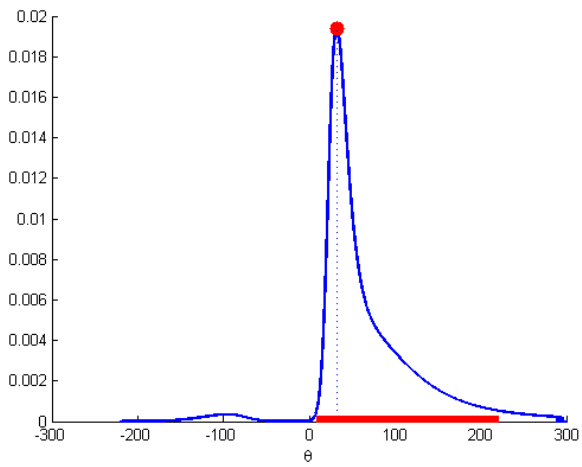
Table 3 displays the results for the ESP approximation. The ESP approximation has a fat and long right tail, which explains the large variations and large values of the RRA often reported in the literature. However, unlike for GMM and CU GMM, the point estimate is well-separated

TABLE 3. **ESP approximation (1890–2009)**

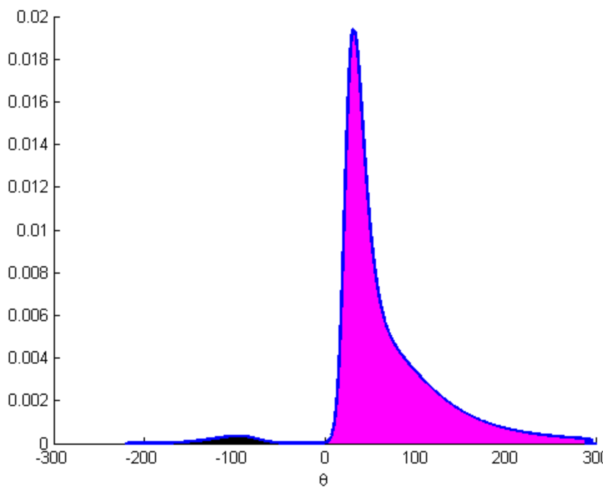
$\frac{1}{2009-1889}\sum_{t=1890}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_t^m - R_t^f)\right] = 0,$
$R_t^m :=$ gross market return, $R_t^f :=$risk-free asset gross return, $C_t :=$ consumption,
$\theta :=$relative risk aversion,
$\hat{\theta}_T^u = 32.21$;
Case with support restricted to $\mathbf{R}_+$: $\hat{I}_{.05}$=$[10.50, 188.85]$ (stripe on A), ESP support $=[0, 289.0]$
Case without restriction: $\hat{I}_{.05} = [9.0, 220.1]$ (stripe on B), ESP support $= [-218.2, 289.0]$



(A) ESP approximation, point estimate and confidence interval.



(B) ESP appr., point estimate and conf. interval.



(C) $\tilde{\mathbb{F}}_T(\mathbf{R}_-) < \tilde{\mathbb{F}}_T(\mathbf{R}_+)$.

as the ESP approximation is not flat around its mode. This is due to the additional information that is captured by the variance term of the ESP approximation,$|\Sigma_T(\theta)|^{-\frac{1}{2}}$ (see equation (2) on p.10). In the absence of the variance term, the ESP point estimate would be the same as GMM, CU GMM, EL, and ET point estimates.

The ESP results indicate that consumption-based asset pricing theory is more consistent with data than other inference approaches suggest. First, in line with financial theory, negative values for the RRA have almost no estimated weight (see Table 3(C)), while confidence intervals from other approaches often include negative values (e.g., Table 1 on p.19; p.93 in Hall, 2005). Second, the empirical key moment condition from consumption-based asset pricing theory has an estimated positive probability weight to hold. Proposition 4 on p.11 indicates if the moment condition was inconsistent with data, the ESP approximation would be zero everywhere. These findings are encouraging for consumption-based asset pricing theory because the moment condition (4) do not resort to Epstein-Zin-Weil preferences (Epstein and Zin, 1989) or other advanced preferences, which yield more flexible stochastic discount factors.

5. Conclusion

Several areas, such as empirical consumption-based asset pricing, have been a challenge for moment-based estimation: when moment conditions are nonlinear, estimates are often instable. The present paper establishes novel theoretical results for the ESP (empirical saddlepoint) approximation, and then use the ESP approximation to investigate the instability of RRA estimate in empirical consumption-based asset pricing. On the theoretical side, existence of the intensity distribution of the solutions to estimating equations is established. This distribution is the quantity estimated by SP and ESP approximations. An important corollary of this existence result is a generalization of the Schemetter-Jennrich lemma, which is extensively used in many areas. The present paper also establishes global consistency and asymptotic normality of the measure induced by the ESP approximation. On the empirical side, the paper sheds light on empirical consumption-based asset pricing. The ESP approximation of the RRA suggests that the key equilibrium implication of consumption-based asset-pricing theory is more consistent with data than standard inference approaches indicate. The fat and long right tail of the RRA ESP approximation provides an explanation for the large variations and large values of the RRA often reported in the literature.

References

Aït-Sahalia, Y. and Yu, J.: 2006, Saddlepoint approximations for continuous-time Markov processes, *Journal of Econometrics* **134**(2), 507–551.

Aliprantis, C. D. and Border, K.: 2006 (1994), *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd edn, Spinger.

Almudevar, A., Field, C. and Robinson, J.: 2000, The density of multivariate M-estimates, *The Annals of Statistics* **28**(1), 275–297.

Amemiya, T.: 1985, *Advanced Econometrics*, Harvard University Press.

Anderson, T. W. and Rubin, H.: 1949, Estimation of the parameters of a single equation in a complete system of stochastic equations, *The Annals of Mathematical Statistics* **20**(1), 46–63.

Barro, R. J.: 2006, Rare disasters and asset markets in the twentieth century, *The Quarterly Journal of Economics* **121**(3), 823–866.

Broda, S. and Kan, R.: 2015, On distributions of ratios, *Biometrika* . Forthcoming.

Chen, C.-F.: 1985, On asymptotic normality of limiting density functions with Bayesian implications, *Journal of the Royal Statistical Society. Series B (Methodological)* **47**(3), 540–546.

Chernozhukov, V. and Hong, H.: 2003, An MCMC approach to classical estimation, *Journal of Econometrics* **115**, 293–346.

Cochrane, J. H.: 2001, *Asset Pricing*, Princeton University Press.

Csiszár, I.: 1975, I-divergence geometry of probability distributions and minimization problems, *The Annals of Probability* **3**(1), 146–158.

Czellar, V. and Ronchetti, E.: 2010, Accurate and robust tests for indirect inference, *Biometrika* **97**(3), 621–630.

Daley, D. J. and Vere-Jones, D.: (1988) 2008, *An Introduction to the Theory of Point Processes. General Theory and Structure (vol. 2)*, Probability and Its Applications, second edn, Springer.

Daniels, H. E.: 1954, Saddlepoint approximations in statistics, *The Annals of Mathematical Statistics* **25**(4), 631–650.

Duffie, D.: (1992) 2001, *Dynamic Asset Pricing Theory*, third edn, Princeton University Press.

Epstein, L. G. and Zin, S. E.: 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* **57**(4), 937–969.

Esscher, F.: 1932, On the probability function in the collective theory of risk, *Scandinavian Actuarial Journal* pp. 175–195.

Ferger, D.: 2004, A continuous mapping theorem for the argmax-functional in the non-unique case, *Statistica Neerlandica* **58**(1), 83–96.

Feuerverger, A.: 1989, On the empirical saddlepoint approximation, *Biometrika* **76**(3), 457–464.

Field, C.: 1982, Small sample asymptotic expansions for multivariate M-estimates, *The Annals of Statistics,* **10**(3), 672–689.

Field, C. A. and Ronchetti, E.: 1990, *Small Sample Asymptotics*, Lecture notes-Monograph Series, Institute of Mathematical Statistics.

Folland, G. B.: (1984) 1999, *Real Analysis. Modern Techniques and Their Applications*, Pure & Applied Mathematics, Wiley-Interscience.

Fristedt, B. E. and Gray, L. F.: 1997, *A Modern Approach to Probability Theory*, Springer.

Gabaix, X.: 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *The Quarterly Journal of Economics* **127**(2), 645–700.

Gallant, R. and White, H.: 1988, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Blacklwell.

Ghosh, J. K. and Ramamoorthi, R. V.: 2003, *Bayesian Nonparametrics*, Statistics, Springer.

Gouriéroux, C. and Monfort, A.: (1989) 1996, *Statistique et Modèles Económétriques*, Col. "Economie et Statistiques Avancées", série ENSAE et CEPE, 2nd edn, Economica. Translated to English by Quang Vuong under the title "Statistics and Econometric Models".

Gregory, A. W., Lamarche, J.-F. and Smith, G. W.: 2002, Information-theoretic estimation of preference parameter: macroeconomic applications and simulation evidence, *Journal of Econometrics* **107**, 213–233.

Hall, A. R.: 2005, *Generalized Method of Moments*, Advanced Texts in Econometrics, Oxford University Press.

Hansen, L. P.: 1982, Large sample properties of generalized method of moments estimators, *Econometrica* **50**(4), 1029–1054.

Hansen, L. P.: 2001, *International Encyclopedia of the Social and Behavior Sciences*, Elsevier, chapter Method of Moments. Also available online under the title "Generalized Method of Moments Estimation: A Time Series Perspective".

Hansen, L. P.: 2013, Uncertainty outside and inside economic models, *Nobel Prize Lecture.*

Hansen, L. P., Heaton, J. and Yaron, A.: 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* **14**(3), 262–280.

Hansen, L. P. and Singleton, K. J.: 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* **50**(5), 1269–1286.

Hiriart-Urruty, J.-B. and Lemaréchal, C.: (1993) 1996, *Convex Analysis and Minimization Algorithms*, Comprehensive Studies in Mathematics.

Holcblat, B.: 2012, *A Classical Moment-Based Inference Framework with Bayesian Properties*, PhD thesis, Carnegie Mellon University.

Holcblat, B. and Grønneberg, S.: 2015, On econometric practice, multiple use of the same data and past-realized data. Working paper. Available at SSRN: http://ssrn.com/abstract=2639266.

Holly, A. and Phillips, P. C. B.: 1979, A saddlepoint approximation to the distribution of the k-class estimator of a coefficient in a simultaneous system, *Econometrica* **47**(6), 1527–1547.

Imbens, G. W.: 1997, One-step estimators for over-identified generalized method of moments models, *The Review of Economic Studies* **64**(3), 359–383.

Jennrich, R. I.: 1969, Asymptotic properties of non-linear least squares estimators, *The Annals of Mathematical Statistics* **40**(2), 633–643.

Jensen, J. L.: 1995, *Saddlepoint Approximations*, Oxford Statistical Science Series, Oxford University Press.

Jensen, J. L. and Wood, A. T.: 1998, Large deviation and other results for minimum contrast estimators, *Annals of the Institute of Statistical Mathematics* **50**(4), 673–695.

Julliard, C. and Ghosh, A.: 2012, Can rare events explain the equity premium puzzle?, *Review of Financial Studies* **25**(10).

Kallenberg, O.: (1997) 2002, *Foundation of Modern Probability*, Probability and Its Applications, second edn, Springer.

Kass, R. E., Tierney, L. and Kadane, J. B.: 1990, *Bayesian and Likelihood Methods In Statistics and Econometrics: Essays in Honor of George A. Barnard*, Amsterdam: North Holland, chapter "The Validity of posterior expansions based on Laplace's Method", pp. 473–498.

Kim, J.-Y.: 2002, Limited information likelihood and Bayesian analysis, *Journal of Econometrics* **107**, 175–193.

Kitamura, Y. and Stutzer, M.: 1997, An information-theoretic alternative to generalized method of moments estimation, *Econometrica* **65**(4), 861–874.

Kocherlakota, N. R.: 1990, On tests of representative consumer asset pricing models, *Journal of Monetary Economics* **26**(2), 285–304.

Kumagai, S.: 1980, An implicit function theorem: Comment, *Journal of Optimization Theory and Applications* **31**(2), 285–288.

Kundhi, G. and Rilstone, P.: 2013, Edgeworth and saddlepoint expansions for nonlinear estimators, *Econometric Theory* **29**, 1057–1078.

Laplace, P.-S.: (1774) 1891, *Œuvres complètes de Laplace*, Vol. 8, Gauthier-Villars, chapter "Mémoire sur la probabilité des causes par les événements", pp. 31–71.

Lehmann, E. L. and Casella, G.: 1998 (1983), *Theory of Point Estimation*, Texts in Statistics, second edn, Springer.

Lucas, R. E.: 1978, Asset prices in an exchange economy, *Econometrica* **46**(6), 1429–1445.

Matthes, K., Kerstan, J. and Mecke, J.: (1974) 1978, *Infinitely Divisible Point Processes*, Probability & Mathematical Statistics Monograph, Wiley.

Monfort, A.: (1980) 1996, *Cours de Probabilités*, "Economie et statistiques avancées", ENSAE et CEPE, third edn, Economica.

Newey, W. K. and McFadden, D. L.: 1994, *Handbook of Econometrics*, Vol. 4, Elsevier Science Publishers, chapter "Large Sample Estimation and Hypothesis Testing", pp. 2113–2247.

Pearson, K.: 1894, Contribution to the mathematical theory of evolution, *Philosophical Transactions of the Royal Society* pp. 71–110.

Perlman, M. D.: 1983, The limiting behavior of multiple roots of the likelihood equation, *Technical Report No. 31, Department of statistics, University of Washington* . Published also in *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday*, 339–370. (M. Rizvi, J. Rustagi, D. Siegmund, eds.) Academic, New York.

Phillips, P. C. B.: 1978, Edgeworth and saddlepoint approximations in the first-order non circular autoregression, *Biometrika* **65**(1), 91–98.

Phillips, P. C. B.: 1982, Exact small theory in the simultaneous equations model, *Cowles Foundation Discussion Paper NO. 621* .

Reeds, J. A.: 1985, Asymptotic number of roots of Cauchy location likelihood equations, *The Annals of Statistics* **13**(2), 775–784.

Roberts, A. W. and Varberg, D. E.: 1973, *Convex Functions*, Pure and Applied Mathematics, Academic Press.

Robinson, J., Ronchetti, E. and Young, G. A.: 2003, Saddlepoint approximations and tests based on multivariate M-estimates, *Annals of Statistics* **31**(4), 1154–1169.

Rogers, L. and Williams, D.: (1979) 2008, *Diffusions, Markov Processes and Martingales. Foundations (vol 1)*, Cambridge Mathematical Library, second edn, Cambridge University Press.

Ronchetti, E. and Trojani, F.: 2003, Saddlepoint approximations and test statistics for accurate inference in overidentified moment conditions models, *Working paper, National Centre of Competence in Research, Financial Valuation and Risk Management* .

Ronchetti, E. and Welsh, A. H.: 1994, Empirical saddlepoint approximations for multivariate M-estimators, *Journal of the Royal Statistical Society.Series B (Methodological),* **56**(2), 313–326.

Savov, A.: 2011, Asset pricing with garbage, *Journal of Finance* **66**(1), 177–201.

Schennach, S. M.: 2005, Bayesian exponentially tilted empirical likelihood, *Biometrika* **92**(1), 31–46.

Schervish, M. J.: (1995) 1997, *Theory of Statistics*, Statistics, Springer.

Seijo, E. and Sen, B.: 2011, A continuous mapping theorem for the smallest argmax functional, *Electronic Journal of Statistics* **5**, 421–439.

Skovgaard, I. M.: 1985, Large deviation approximations for maximum likelihood estimators, *Probability and Mathematical Statistics* **6**(2), 89–107.

Skovgaard, I. M.: 1990, On the density of minimum contrast estimators, *The Annals of Statistics* **18**(2), 779–789.

Small, C. G., Wang, J. and Yang, Z.: 2000, Eliminating multiple root problems in estimation, *Statistical Science* **15**(4), 313–341.

Smith, D. C.: 1999, Finite sample properties of tests of the Epstein-Zin asset pricing model, *Journal of Econometrics* **93**(1), 113–148.

Sowell, F.: 2007, The empirical saddlepoint approximation for GMM estimators, *working paper, Tepper School of Business, Carnegie Mellon University* .

Sowell, F.: 2009, The empirical saddlepoint likelihood estimator applied to two-step GMM, *working paper, Tepper School of Business, Carnegie Mellon University* .

Stock, J. H. and Wright, J. H.: 2000, GMM with weak identification, *Econometrica* **68**(5), 1055–1096.

Tingley, M. and Field, C. A.: 1990, Small-sample confidence intervals, *Journal of the American Statistical Association* **85**, 427–434.

van der Vaart, A. W. and Wellner, J. A.: 1996, *Weak Convergence and Empirical Processes*, Springer.

Vissing-Jørgensen, A. and Attanasio, O. P.: 2003, Stock-market participation, intertemporal substitution, *American Economic Review, AEA Papers and Proceedings* **93**(2).

Notes and acknowledgements

Helpful comments were provided by Anthony L. Almudevar, Petyo Bonev (discussant), Paul-Antoine Chevalier, Jérôme Detemple (discussant), Paul Ehling, Dennis Epple, Anisha Ghosh, Gautam Iyer, Rick Green, Steffen Grønneberg, Lars-Alexander Kuehn, Vineet Kumar, Chen Li, Jesper Lund (discussant), Artem Neklyudov, Carlos Ramirez, Bryan Routledge, Max Roy, Stefano Sacchetto, Teddy Seidenfeld, Steven Shreve, Fallaw Sowell, George Tauchen (discussant) and seminar/conference participants where this paper and previous versions of it have been presented. Any errors are my own. I acknowledge financial support from the Centre for Asset Pricing Research.

Appendix A. Proofs

For brevity, this appendix contains only a condensed version of the proofs. Detailed proofs are available in the supplemental material.

A.1. **Proof of Proposition 1.** Denote $\nu(.)$ the counting measure, $\underline{X}_T := (X_t)_{t=1}^T$ and $\Psi_T(\underline{X}_T(\omega), \theta) := \frac{1}{T}\sum_{t=1}^T \psi(X_t(\omega), \theta)$. By a standard result about random measures (e.g. Daley and Vere-Jones, 2008, prop. 9.1.VIII) it is sufficient to prove that there exists a function $\omega \mapsto N_T(\omega, .)$ s.t. for any given $A \in \mathcal{B}(\mathbf{\Theta})$, $\omega \mapsto N_T(\omega, A)$ is $\mathcal{E}/\mathcal{B}(\mathbf{N})$-measurable and $N_T(\omega, A) = \nu\{\theta \in A: \ \Psi_T(\underline{X}_T(\omega), \theta) = 0\}$ $\mathbb{P}$-a.s. Fix $A \in \mathcal{B}(\mathbf{\Theta})$.

By Lemma 1 below with $\mathbf{\Gamma}_1 := \mathbf{R}^{pT}$ and $\mathbf{\Gamma}_2 := \mathbf{\Theta}$, if a set $P \in \mathcal{B}(\mathbf{R}^{pT}) \otimes \mathcal{B}(\mathbf{\Theta})$, then $\underline{x}_T \mapsto \nu(P_{\underline{x}_T} \cap A)$ is $\mathcal{B}\left((\mathbf{R}^p)^T\right)/\mathcal{B}(\overline{\mathbf{N}})$-measurable, where $P_{\underline{x}_T} := \{\theta \in \mathbf{\Theta} : (\underline{x}_T, \theta) \in P\}$ and $\overline{\mathbf{N}} := \mathbf{N} \cup \{\infty\}$. Then, setting $P := \Psi_T^{-1}(\{0\})$, we have $\omega \mapsto \nu\left(\{\theta \in A : \Psi_T(\underline{X}_T(\omega), \theta) = 0\}\right)$ $\mathcal{E}/\mathcal{B}(\overline{\mathbf{N}})$-measurable because the composition of measurable functions is a measurable function. Now Assumption 1(d) implies that the number of solutions to the empirical moment conditions is finite $\mathbb{P}$-a.s. and Assumption 1(a) states that $(\mathbf{\Omega}, \mathcal{E}, \mathbb{P})$ is complete. Thus, there exists an $\mathcal{E}/\mathcal{B}(\mathbf{N})$-measurable function $\omega \mapsto N_T(\omega, A)$ s.t.

$$N_T(\omega, A) := \begin{cases} \nu\left(\{\theta \in A : \Psi_T(\underline{X}_T(\omega), \theta) = 0\}\right) & \text{if } \omega \in \mathbf{\Omega}\backslash F \\ 0 & \text{if } \omega \in F, \end{cases}$$

where $F := \{\omega \in \mathbf{\Omega} : \nu\left(\{\theta \in A : \Psi_T(\underline{X}_T(\omega), \theta) = 0\}\right) = \infty\}$ and $\mathbb{P}\{F\} = 0$ (e.g. Kallenberg, 1997/2002, Lemma 1.25). □

Lemma 1. *Let $\mathbf{\Gamma}_1 \subset \mathbf{R}^n$ and $\mathbf{\Gamma}_2 \subset \mathbf{R}^q$ with $(n, q) \in \mathbf{N}^2$. For all $A \in \mathcal{B}(\mathbf{\Gamma}_2)$, $\forall P \in \mathcal{B}(\mathbf{\Gamma}_1) \otimes \mathcal{B}(\mathbf{\Gamma}_2)$, $\gamma_1 \mapsto \nu(P_{\gamma_1} \cap A)$, where $P_{\gamma_1} := \{\gamma_2 \in \mathbf{\Gamma}_2 : (\gamma_1, \gamma_2) \in P\}$, is $\mathcal{B}(\mathbf{\Gamma}_1)/\mathcal{B}(\overline{\mathbf{N}})$-measurable.*

Proof. Let $A \in \mathcal{B}(\mathbf{\Gamma}_2)$. Define for this proof

$$\mathcal{H}_A := \left\{ h(.) : \begin{array}{c} h(.) \text{ is bounded} \\ h(.) \text{ is } \mathcal{B}(\mathbf{\Gamma}_1) \otimes \mathcal{B}(\mathbf{\Gamma}_2)/\mathcal{B}(\mathbf{R})\text{-measurable} \\ \gamma_1 \mapsto \int_A h(\gamma_1, \gamma_2)\nu(\mathrm{d}\gamma_2) \text{ is } \mathcal{B}(\mathbf{\Gamma}_1)/\mathcal{B}(\overline{\mathbf{R}})\text{-measurable} \end{array} \right\}$$

Apply a functional form of Sierspinki monotone class theorem (e.g., Theorem 3.1 in Rogers and Williams,1979/2008) with the set of measurable rectangles, $\mathcal{I} := \{R = R_{\mathbf{\Gamma}_1} \times R_{\mathbf{\Gamma}_2} \text{ s.t. } R_{\mathbf{\Gamma}_1} \in \mathcal{B}(\mathbf{\Gamma}_1) \wedge R_{\mathbf{\Gamma}_2} \in \mathcal{B}(\mathbf{\Gamma}_2)\}$, as the π-system to show that if a function $g(.)$ is $\mathcal{B}(\mathbf{\Gamma}_1) \otimes \mathcal{B}(\mathbf{\Gamma}_2)/\mathcal{B}(\mathbf{R})$-measurable and bounded, $g \in \mathcal{H}_A$, and thus $\gamma_1 \mapsto \int_A g(\gamma_1, \gamma_2)\nu(\mathrm{d}\gamma_2)$ is $\mathcal{B}(\mathbf{\Gamma}_1)/\mathcal{B}(\overline{\mathbf{N}})$-measurable. Deduce that $\forall P \in \mathcal{B}(\mathbf{\Gamma}_1) \otimes \mathcal{B}(\mathbf{\Gamma}_2)$, $\gamma_1 \mapsto \nu(P_{\gamma_1} \cap A)$ is $\mathcal{B}(\mathbf{\Gamma}_1)/\mathcal{B}(\overline{\mathbf{N}})$-measurable as $\nu(P_{\gamma_1} \cap A) = \int_A \mathbf{l}_P(\gamma_1, \gamma_2)\nu(\mathrm{d}\gamma_2)$ because $\forall P \in \mathcal{B}(\mathbf{\Gamma}_1) \otimes \mathcal{B}(\mathbf{\Gamma}_2)$, $\mathbf{l}_P(.)$ is $\mathcal{B}(\mathbf{\Gamma}_1) \otimes \mathcal{B}(\mathbf{\Gamma}_2)/\mathcal{B}(\mathbf{R})$-measurable and bounded. □

A.2. **Lemma 2.**

Lemma 2. *Under Assumptions 1,*

i) *there exists a dissecting systems of* $(\mathbf{\Theta}, \mathcal{B}(\mathbf{\Theta}))$ *;*
ii) *if* $\mathcal{T} := (\mathcal{T}_n)_{n\geqslant 1}$ *is a dissecting system of* $\mathbf{\Theta}$*, then, for any bounded Borel sets* A*,* $\mathcal{T}(A) := (\mathcal{T}_n(A))_{n\geqslant 1}$ *with* $\mathcal{T}_n(A) := \{A_{n,i} \cap A : i = 1, \ldots, k_n$ *and* $A_{n,i} \in \mathcal{T}_n\}$ *is a dissecting system;*
iii) $\mathbb{F}_T(.)$ *is* $\mathbb{P}$*-a.s. a finite measure on* $(\mathbf{\Theta}, \mathcal{B}(\mathbf{\Theta}))$ *that does not depend on the dissecting system.*

Proof. *i)* Take partitions consisting of hypercubes whose corners or faces have been removed when necessary to make intersections empty. *ii)* It follows from the definition. *iii)* This is a consequence of Assumption 1(d) and Khinchin's existence theorem (e.g. Daley and Vere-Jones, 1988/2008, prop. 9.3.IX). □

A.3. **Proof of Proposition 2.** This is a consequence of equation (9.3.24) in Daley and Vere-Jones (1988/2008, p. 48). □

A.4. **Proof of Proposition 3.** *i)* Apply the implicit function theorem to $\frac{1}{T}\sum_{t=1}^{T} \psi_t(\theta)\mathrm{e}^{\tau'\psi_t(\theta)} = 0_{m\times 1}$.

ii) Continuity follows from the implicit function theorem. A proof by contradiction implies uniqueness, as a convex function cannot have two distinct strict local minima (e.g., Roberts and Varberg, 1973, theo. A, p. 123). □

A.5. **Proof of Proposition 4.** The "if" part is straightforward. The "only if" part is an implication of duality theory, and the duality with the maximization of entropy under moment conditions (e.g., Hiriart-Urruty and Lemaréchal, 1993/1996, prop. XII.2.4.1(iii)). □

A.6. **Proof of Proposition 5i).**

Proposition 6. *Under Assumptions 1(a)–(c) and 2, there exist*

i) *an ESP approximation,* $\tilde{f}_{\theta_T^*,sp}(.)$*;*
ii) *an ESP intensity s.t.*
   a) *a finite-positive measure on the measurable space* $(\mathbf{\Theta}, \mathcal{B}(\mathbf{\Theta}))$ *s.t.*

b) *for all* $A \in \mathcal{B}(\boldsymbol{\Theta})$, $\omega \mapsto \tilde{\mathbb{F}}_T(A)$ *is* $\mathcal{E}/\mathcal{B}(\mathbf{R})$*-measurable.*

Proof. Set

$$\tilde{f}_{\theta_T^*,sp}(\theta) := \begin{cases} \hat{f}_{\theta_T^*,sp}(\theta) & \text{if } \theta \in \hat{\boldsymbol{\Theta}}_T^{-\eta} \\ \min\left[\bar{f}_T, \hat{f}_{\theta_T^*,sp}(\theta)\right] \frac{1}{\eta}\rho(\theta, \hat{\boldsymbol{\Theta}}_T^c) & \text{if } \theta \in \hat{\boldsymbol{\Theta}}_T \cap \left(\hat{\boldsymbol{\Theta}}_T^{-\eta}\right)^c \\ 0 & \text{if } \theta \in \hat{\boldsymbol{\Theta}}_T^c, \end{cases}$$

where $\bar{f}_T := \sup_{\theta \in \left\{\partial \hat{\boldsymbol{\Theta}}_T^{-\eta}(\omega)\right\}} \hat{f}_{\theta_T^*,sp}(\theta)$ if $\hat{\boldsymbol{\Theta}}_T^{-\eta}(\omega) \neq \emptyset$, or 0 otherwise. By construction $\tilde{f}_{\theta_{T,sp}^*}(.)$ is positive and continuous. Thus, the rest of the proof, which is tedious and long, essentially consists in checking measurability: See supplemental material. □

A.7. **Proof of Proposition 5ii).** On can build a point random field with an intensity that corresponds to any given finite measure. (e.g., Fristedt and Gray, 1997, p. 587, Lemma 9).

A.8. **Proof of Theorems 1 and 2.**

A.8.1. *Preliminary results.* This subsection contains some results needed for Theorems 1 and 2. Most of them are variants of results already known, but not necessarily easy to find in the literature.

Measurability and convergence results.

Lemma 3. *Let* $(A_T)_{T \geqslant 1}$ *be a sequence of square matrices converging to a square matrix* A *as* $T \to \infty$*:* $\lim_{T\to\infty} \|A_T - A\| = 0$*. Then*

i) *if* A *is an invertible matrix, then there exists* $\dot{T} \in \mathbf{N}$ *s.t.* $T \geqslant \dot{T}$ *implies* A_T *is invertible;*

ii) *if* $(A_T)_{T\geqslant 1}$ *is a sequence of symmetric matrices and* A *is a negative-definite matrix, then there exists* $\dot{T} \in \mathbf{N}$ *s.t.* $T \geqslant \dot{T}$ *implies* A_T *is a negative-definite matrix.*

Proof. *i)* The determinant function $|.|_{\det}$ is a continuous function.

ii) Note $\max \operatorname{sp} A_T = \max_{z:\|z\|=1} z' A_T z$ where $\operatorname{sp} A_T$ denotes the set of eigenvalues of A; and prove $\sup_{z:\|z\|=1} |z' A_T z - z' A z| \to 0$, as $T \to \infty$. □

We now introduce a set of assumptions and new notations to derive generic results that are used several times.

Assumption 8. **(a)** $\underline{X}_\infty := (X_t)_{t=1}^\infty$ *is a sequence of i.i.d. random vectors of dimension* p *on the complete probability sample space* $(\boldsymbol{\Omega}, \mathcal{E}, \mathbb{P})$*.* **(b)** *Let* $(\boldsymbol{\Gamma}, \mathcal{B}(\boldsymbol{\Gamma}))$ *be the measurable space* *s.t.* $\mathbf{\Gamma} \subset \mathbf{R}^m$ *is compact and* $\mathcal{B}(\mathbf{\Gamma})$ *is the Borel* σ*-algebra.* **(c)** *Let* $h : \mathbf{R}^p \times \mathbf{\Gamma} \mapsto \mathbf{R}^q$ *with* $q \in \mathbf{N}$ *be a function s.t.* $\forall x \in \mathbf{R}^p$, $\gamma \mapsto h(x, \gamma)$ *is continuous, and* $\forall \gamma \in \mathbf{\Gamma}$, $x \mapsto h(x, \gamma)$ *is* $\mathcal{B}(\mathbf{R}^p)/\mathcal{B}(\mathbf{R}^q)$*-measurable.* **(d)** $\mathbb{E}\left[\sup_{\gamma \in \mathbf{\Gamma}} \|h(X, \gamma)\|\right] < \infty$. **(e)** *In the parameter space* $\mathbf{\Gamma}$*, there exists a unique* $\gamma_0 \in \mathrm{int}(\mathbf{\Gamma})$ *s.t.* $\mathbb{E}\left[h(X, \gamma_0)\right] = 0_{m \times 1}$. **(f)** *For all* $x \in \mathbf{R}^p$, $\gamma \mapsto h(x, \gamma)$ *is continuously differentiable.* **(g)** $\left|\mathbb{E}\left[\frac{\partial h(X, \gamma_0)}{\partial \gamma'}\right]\right|_{\det} \neq 0$. **(h)** $q = m$.

Proposition 7 (Uniform-strong LLN)**.** *Under Assumptions 8(a)–(d) ,* $\frac{1}{T}\sum_{t=1}^{T} h(X_t, \gamma)$ *converges* $\mathbb{P}$*-a.s. to* $\mathbb{E}\left[h(X, \gamma)\right]$ *uniformly w.r.t.* γ *as* $T \to \infty$*: There exists* $E \in \mathcal{E}$ *s.t.* $\mathbb{P}\{E\} = 0$ *and*

$$\forall \omega \in \mathbf{\Omega} \setminus E, \quad \sup_{\gamma \in \mathbf{\Gamma}} \left\| \frac{1}{T} \sum_{t=1}^{T} h(X_t, \gamma) - \mathbb{E}\left[h(X, \gamma)\right] \right\| \to 0 \text{ as } T \to \infty.$$

Proof. This is a standard result (e.g., Ghosh and Ramamoorthi, 2003, theo. 1.3.3 pp. 24-25). □

Hereafter, we do not mention negligible sets associated with properties that holds a.s., because they result from the application of a countable number of properties that hold a.s.

Proposition 8 (Existence of solutions to empirical moment conditions)**.** *Under the Assumptions 8(a)–(c)(e)–(h), if*

(a) *as* $T \to \infty$, $\sup_{\gamma \in \mathbf{\Gamma}} \left\| \frac{1}{T} \sum_{t=1}^{T} h(X_t, \gamma) - \mathbb{E}\left[h(X, \gamma)\right] \right\| \to 0 \quad \mathbb{P}$*-a.s.*

(b) *as* $T \to \infty$, $\sup_{\gamma \in \mathbf{\Gamma}} \left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\partial h(X_t, \gamma)}{\partial \gamma'} - \mathbb{E}\left[\frac{\partial h(X, \gamma)}{\partial \gamma'}\right] \right\| \to 0 \quad \mathbb{P}$*-a.s.,*

then, for all $r > 0$*, there exists* $\dot{T} \in \mathbf{N}$*, so that* $T \geqslant \dot{T}$ *implies*

i) *there exists* $\mathbb{P}$*-a.s. a solution to the empirical moment conditions: There exists* γ_T^* *s.t.*

$$\frac{1}{T} \sum_{t=1}^{T} h(X_t, \gamma_T^*) = 0_{m \times 1};$$

ii) *all solutions to the empirical moment conditions are in* $B_r(\gamma_0)$*.*

Proof. *i)* For T big enough, a solution to the empirical moment conditions solves the following first-order condition $\left[\frac{1}{T}\sum_{t=1}^{T} \frac{\partial h(X_t, \gamma)'}{\partial \gamma}\right] \left[\frac{1}{T}\sum_{t=1}^{T} h(X_t, \gamma)\right] = 0_{m \times 1}$ with $\left[\frac{1}{T}\sum_{t=1}^{T} \frac{\partial h(X_t, \gamma)'}{\partial \gamma}\right]$ invertible.

ii) This follows from assumption (a). □

The next proposition ensures $\mathbb{P}$-a.s. the measurability of all the solutions to the empirical moment conditions. By regarding solutions to the empirical moment conditions as minima of $\gamma \mapsto \|\frac{1}{T}\sum_{t=1}^{T} h(X_t,\gamma)\|$, the Schmetterer-Jennrich's measurability result (Jennrich, 1969, Lemma 2) ensures the measurability of only one of them.

Proposition 9 (Measurability of solutions to empirical moment conditions)**.** *Define the event* $\dot{F}^c := \left\{\omega \in \mathbf{\Omega} : 1 \leqslant \#\{\gamma \in \mathbf{\Gamma} : \frac{1}{T}\sum_{t=1}^{T} h\left(X_t(\omega),\gamma\right) = 0\} < \infty\right\}$*. Under Assumptions 8(a)-(c) and (i),* $\mathbb{P}$*-a.s., each of the solutions to the empirical moment conditions is* $\mathcal{E}/\mathcal{B}(\mathbf{\Gamma})$*-measurable, i.e., for all* $\omega \in \dot{F}^c$*, if* $\dot{\gamma} \in \mathbf{\Gamma}$ *is such that* $\frac{1}{T}\sum_{t=1}^{T} h(X_t(\omega),\dot{\gamma}) = 0_{m\times 1}$ *, then there exits* γ_T^* $\mathcal{E}/\mathcal{B}(\mathbf{\Gamma})$*-measurable s.t.* $\gamma_T^*(\omega) = \dot{\gamma}$ *and, for all* $\tilde{\omega} \in \dot{F}^c$*,* $\frac{1}{T}\sum_{t=1}^{T} h(X_t(\tilde{\omega}),\gamma_T^*(\tilde{\omega})) = 0_{m\times 1}$.

Proof. Check the assumptions of Proposition 1. Then, under Assumption 8(b), by isomorphism (e.g., Kallenberg, 1997/2002, Theorem A.1.2), w.l.o.g., we can assume that $\mathbf{\Gamma}$ is a Borel subset of $[0,1]$, so that we have a simple point process except on the $\mathbb{P}$-null set F. Then, one can extract the jump points (i.e., solutions to the estimating equations), which correspond to stopping times, and thus are measurable. □

The following proposition is a standard result.

Proposition 10 (Consistency of solutions to empirical moment conditions)**.** *Under the assumptions of Propositions 8 and 9, every sequence of solutions to the empirical moment conditions,* $\{\gamma_T^*\}_{T\geqslant 1}$*, converges* $\mathbb{P}$*-a.s. to the population parameter,* γ_0*:*

$$\lim_{T\to\infty} \gamma_T^* = \gamma_0 \quad \mathbb{P}\text{-}a.s.$$

Proof. This follows from Propositions 8 and 9. □

Corollary 3. *Under the Assumptions 1(a)–(c), 2 and 4, Propositions 8, 9 and 10 apply to solutions to the empirical moment conditions:*

$$\frac{1}{T}\sum_{t=1}^{T} \psi(X_t,\theta) = 0_{m\times 1}.$$

Proof. Confirm that the assumptions of Propositions 8, 9 and 10 are satisfied. □

Lemma 4. *Under Assumptions 1(a)–(c), 2,5(a)(b),*

i) *for all* $\theta \in \hat{\mathbf{\Theta}}^{-\eta}$*, there exists a unique* $\tau(\theta)$ *s.t.* $\mathbb{E}\left[\psi(X,\theta)\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] = 0$

ii) $\tau : \hat{\mathbf{\Theta}}^{-\eta} \to \mathbf{R}^m$ *is continuous.*

Proof. Prove both results at once by application of the sufficiency part of Kumagai's (1980) implicit function theorem. □

Laplace's approximation. Laplace's approximation is a well-known method originally presented by Laplace (1774/1878). Here, we adapt the version presented in Chen (1985) and Kass, Tierney and Kadane (1990) for our purposes.[7]

Assumption 9 (Laplace's regularity)**.** **(a)** *Let* $\left(\dot{\theta}_T\right)_{T=1}^{\infty}$ *with* $\dot{\theta}_T \in \mathbf{\Theta}$ $\forall T \geqslant 1$ *be a sequence converging in the interior of* $\mathbf{\Theta}$*.* **(b)** *Let* $(h_T(.))_{T\geqslant 1}$ *be a sequence of real-valued functions. There exists* $r_h > 0$ *and* $T_h \in \mathbf{N}$ *s.t.*

i) $\forall T \geqslant T_h$, $h_T(.) \in C^3\left(B_{r_h}(\dot{\theta}_T)\right)$;

ii) *there exists* $M_h \geqslant 0$ *so that* $\forall T \geqslant T_h$, $\forall \theta \in B_{r_h}(\dot{\theta}_T)$, $\left\|D^3 h_T(\theta)\right\| < M_h$, *where* D^k *denotes the differential operator of order* k*;*

iii) $\forall T \geqslant T_h$, $h_T(\dot{\theta}_T) = 0$ *and* $\frac{\partial h_T(\dot{\theta}_T)}{\partial \theta'} = 0_{1\times m}$*;*

(c) *The sequence of symmetric matrices* $\left(\frac{\partial^2 h_T(\dot{\theta}_T)}{\partial\theta\partial\theta'}\right)_{T\geqslant T_h}$ *converges to a negative-definite matrix.*
(d) *Let* $(b_T(.))_{T\geqslant 1}$ *be a sequence of real-valued functions s.t. there exists* $r_b > 0$, $M_b \geqslant 0$ *and* $T_b \in \mathbf{N}$ *so that*

i) $b_T(.) \in C^1\left(B_{r_b}(\dot{\theta}_T)\right)$;

ii) $\forall T \geqslant T_b$, $\forall \theta \in B_{r_b}(\dot{\theta}_T)$, $\|Db_T(\theta)\| < M_b$.

Proposition 11. *Under Assumptions 1(b) and 9, there exists* $r > 0$ *so that, for any neighborhood of* $\dot{\theta}_T$, $V_r(\dot{\theta}_T)$, *included in* $B_r(\dot{\theta}_T)$, *we have*

$$\int_{V_r(\dot{\theta}_T)} b_T(\theta)\mathrm{e}^{[Th_T(\theta)]}\mathrm{d}\theta = \int_{V_r(\dot{\theta}_T)} \exp\left\{\frac{T}{2}(\theta-\dot{\theta}_T)'\frac{\partial^2 h_T(\dot{\theta}_T)}{\partial\theta\partial\theta'}(\theta-\dot{\theta}_T)\right\}\mathrm{d}\theta\left[b_T(\dot{\theta}_T) + O\left(\frac{1}{T}\right)\right]$$

Proof. Adapt the proof from Kass, Tierney and Kadane (1990). □

Lemma 5. *Under Assumptions 1(b) and 9,*

$$\int_{V_r(\dot{\theta}_T)} \exp\left\{\frac{T}{2}(\theta-\dot{\theta}_T)'\frac{\partial^2 h_T(\dot{\theta}_T)}{\partial\theta\partial\theta'}(\theta-\dot{\theta}_T)\right\}\mathrm{d}\theta \underset{\infty}{\sim} \left(\frac{2\pi}{T}\right)^{m/2}\left|\left(-\frac{\partial^2 h_T(\dot{\theta}_T)}{\partial\theta\partial\theta'}\right)\right|_{\det}^{-1/2}$$

[7] Kass, Tierney and Kadane (1990) make explicit the Laplace's approximation used in Chen (1985). The differences between Kass, Tierney and Kadane's theorem and our proposition are the following. In our case, $b_T(.)$ depends on T. Their assumptions do not seem to ensure the convergence of the Hessian $\frac{\partial^2 h_T(\dot{\theta}_T)}{\partial\theta\partial\theta'}$. Their assumptions are stronger, because they provide a higher-order expansion.

where given two functions $f(.)$ and $g(.)$ with domain D and $a \in D$, $f(z) \underset{a}{\sim} g(z)$ means that there exists a function $\varphi(.)$ defined on D s.t. $f(.) = g(.)\varphi(.)$ and $\lim_{z \to a} \varphi(z) = 1$.

Proof. Apply the Lebesgue-dominated convergence theorem and use the definition of multivariate Gaussian densities. □

Proposition 12. *Under Assumptions 1(b) and 9, there exists T_1 and $r > 0$ s.t. for all $T \geqslant T_1$*

$$\int_{V_r(\dot{\theta}_T)} b_T(\theta) \exp\left[T h_T(\theta)\right] \mathrm{d}\theta = \left(\frac{2\pi}{T}\right)^{m/2} \left| \left(-\frac{\partial^2 h_T(\dot{\theta}_T)}{\partial\theta\partial\theta'} \right) \right|_{\det}^{-1/2} \left[b_T(\dot{\theta}_T) + O\left(\frac{1}{T}\right) \right]$$

and the RHS and the LHS are well defined.

Proof. Combine Proposition 11 and Lemma 5. □

A.8.2. *Proof of Theorems 1 and 2.* Note that "θ_T^*" can denote a random variable that maps an $\omega \in \mathbf{\Omega}$ to *one* of the potentially multiple solutions to the empirical moment conditions, or the random correspondence that maps an $\omega \in \mathbf{\Omega}$ to the *set* of solutions to the empirical moment conditions. See Remark 1 on p. 16. The context indicates the meaning.

Around θ_T^*: application of Laplace's approximation.

Proposition 13. *Under Assumptions 1(a)–(c),2, 4, 5(a), 6(a)(c)–(e), Laplace's approximations corresponding to Propositions 11 and 12 can be applied $\mathbb{P}$-a.s. to $\int_{B_r(\theta_T^*)} \tilde{f}_{\theta_T^*,sp}(\theta)\mathrm{d}\theta$ with small enough $r > 0$ by setting*

$$\begin{aligned}
\dot{\theta}_T &:= \theta_T^* \\
h_T(\theta) &:= \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] \\
b_T(\theta) &:= \left|\Sigma_T(\theta)\right|_{\det}^{-\frac{1}{2}},
\end{aligned}$$

where the RHS are well-defined for T big enough.

Proof. First, note $\mathbb{P}$-a.s. for T big enough the RHS exist in $B_r(\theta_T^*)$ by Assumption 5(a) and Corollary 3. Second, check the assumptions of Laplace's approximation. Lemma 3 in Jennrich (1969) ensures that the Taylor expansions with a mean-value form of the remainder used to prove Laplace's approximation preserve measurability. Thus, it is now sufficient to show that the above quantities satisfy Assumption 9. Corollary 3 and lemmas below ensure that this is the case. □

Lemma 6. *Under Assumptions 1(a)–(c),2, 4, and 5(a), for T big enough $\mathbb{P}$-a.s.,*

$$\left.\frac{\partial \tau_T(\theta)}{\partial \theta'}\right|_{\theta=\theta_T^*} = -\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta_T^*)}{\partial \theta'}\right],$$

where the LHS and RHS are well defined.

Proof. Check the assumptions of the implicit-function theorem to apply it to the tilting equation defining $\tau_T(.)$. Then note that $\tau_T(\theta_T^*) = 0_{m\times 1}$. □

Lemma 7. *Under the assumptions of Lemma 6, for T big enough $\mathbb{P}$-a.s.*

$$\left.\frac{\partial \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]}{\partial \theta}\right|_{\theta=\theta_T^*} = 0_{m\times 1},$$

where the LHS is well defined.

Proof. Differentiate and note that $\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*) = 0_{m\times 1}$. □

Lemma 8. *Under the assumptions of Lemma 6 and Assumption 6(a),*

i) *for T big enough, $\mathbb{P}$-a.s.,*

$$\left.\frac{\partial^2 \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]}{\partial \theta \partial \theta'}\right|_{\theta=\theta_T^*} = -\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta_T^*)'}{\partial \theta}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta_T^*)}{\partial \theta'}\right],$$

where the RHS and the LHS are well defined;

ii) *under the additional Assumption 6(e), as $T\to\infty$, $\mathbb{P}$-a.s.,*

$$\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta_T^*)'}{\partial \theta}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta_T^*)}{\partial \theta'}\right]^{-1} \to \Sigma(\theta_0),$$

where $\Sigma(\theta_0) := \left[\mathbb{E}\frac{\partial \psi(X,\theta_0)'}{\partial \theta}\right]^{-1}\mathbb{E}\left[\psi(X,\theta_0)\psi(X,\theta_0)'\right]\left[\mathbb{E}\frac{\partial \psi(X,\theta_0)}{\partial \theta'}\right]^{-1}$ is a positive-definite matrix.

Proof. *i)* Differentiate, and note that $\tau_T(\theta_T^*) = 0_{m\times 1}$, and apply Lemma 6. *ii)* Apply uniform LLN. □

Outside a neighborhood of θ_T^*.

Proposition 14. *Under the assumptions of Lemma 4 and 9 and Assumptions 1(a)–(c),2, 4(a),7(b), for all small enough $r > 0$, there exists $\varepsilon > 0$ and $\dot{T} \in \mathbf{N}$ s.t.*

$$\forall T \geqslant \dot{T},\, \forall \theta \in \hat{\boldsymbol{\Theta}}_T^{-\eta} \setminus B_r(\theta_0), \quad \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)\psi_t(\theta)} < 1 - \varepsilon \quad \mathbb{P}\text{-}a.s.$$

Proof. Check the assumptions of Proposition 7 for application to $\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)}$ in $B_{r_{\dot{\theta}}}(\dot{\theta})$ with $r_{\dot{\theta}} > 0$ and $\dot{\theta} \in \hat{\boldsymbol{\Theta}}^{-\eta}$. By Proposition 7, for all $\dot{\theta} \in \hat{\boldsymbol{\Theta}}^{-\eta}$, there exists $r_{\dot{\theta}} > 0$ s.t. as $T \to \infty$,

$$\sup_{\theta \in \overline{B_{r_{\dot{\theta}}}(\dot{\theta})}} \left\| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] \right\| \to 0.$$

Now, because $\hat{\boldsymbol{\Theta}}^{-\eta}$ is compact, there exists $\left\{\dot{\theta}_k\right\}_{k=1}^{K}$ s.t. $\hat{\boldsymbol{\Theta}}^{-\eta} = \bigcup_{k=1}^{K} B_{\dot{r}_k}(\dot{\theta}_k)$. Thus, as $T \to \infty$,

$$\sup_{\theta \in \hat{\boldsymbol{\Theta}}^{-\eta}} \left\| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] \right\| \to 0.$$

Thus, for small enough $\varepsilon > 0$, there exists $\dot{T}$ s.t. for all $T \geqslant \dot{T}$,

$$\sup_{\theta \in \hat{\boldsymbol{\Theta}}^{-\eta}} \left\| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] \right\| < \varepsilon. \tag{5}$$

Moreover, by Lemma 9, for small enough $\varepsilon > 0$, there exists, $r_3 > 0$ s.t. $\forall \theta \in \hat{\boldsymbol{\Theta}}^{-\eta} \setminus B_{r_3}(\theta_0)$,

$$\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] < 1 - 2\varepsilon, \tag{6}$$

because $\theta \mapsto \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right]$ is continuous as a uniform limit of continuous functions $\theta \mapsto \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)}$. Consequently, for all $T \geqslant \dot{T}$, $\forall \theta \in \hat{\boldsymbol{\Theta}}^{-\eta} \setminus B_{r_3}(\theta_0)$,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] + \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] \\
&\overset{(a)}{\Rightarrow} \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)} \leqslant 1 - \varepsilon \\
&\overset{(b)}{\Rightarrow} \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} \leqslant 1 - \varepsilon
\end{aligned}$$

(a) By the triangle inequality, $\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)} \leqslant \left\| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau(\theta)'\psi_t(\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] \right\| + \left\| \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] \right\| \leqslant 1 - 2\varepsilon + \varepsilon = 1 - \varepsilon$ because of inequalities (6) and (5); *(b)* $\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} = \min_{\tau \in \mathbf{R}^m} \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}$ because $\frac{\partial^2 \left\{ \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)} \right\}}{\partial \tau \partial \tau'} = \frac{1}{T}\sum_{t=1}^{T} \psi_t(\theta)\psi_t(\theta)'\mathrm{e}^{\tau'\psi_t(\theta)}$. $\square$

Lemma 9. *Under Assumptions 1(a)–(c), 4(b) and 5, for all* $\theta \in \hat{\mathbf{\Theta}}_{\infty}^{-\eta} \setminus \{\theta_0\}$

$$\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X,\theta)}\right] < 1.$$

Proof. By the definition of $\hat{\mathbf{\Theta}}^{-\eta}$ and a standard result on Laplace's transform (e.g. Monfort, 1980/1996, theo. 3 pp. 182–183), $\tau(\theta)$ is the unique minimum of a strictly convex function. Therefore, the result follows. □

Conclusion of the proofs.

Corollary 4. *Under Assumptions 1(a)–(c), 2, 4, 5, 6(a)(c)(d)(e) and 7, for all small enough* $r > 0$*,*

i) *as* $T \to \infty$, $\int_{B_r(\theta_T^*)} \tilde{f}_{\theta_T^*,sp}(\theta)\mathrm{d}\theta \to 1$ $\mathbb{P}$*-a.s.;*

ii) *there exist* $\dot{T} \in \mathbf{N}$, $M \geqslant 0$ *and* $\varepsilon > 0$ *s.t. for all* $T > \dot{T}$ *and for all* $\theta \in \mathbf{\Theta} \setminus B_r(\theta_T^*)$*,*

$$\tilde{f}_{\theta_T^*,sp}(\theta) < \exp\{-T\varepsilon\} M \quad \mathbb{P}\text{-}a.s.$$

Proof. *i)* By Proposition 13, apply Proposition 12, combined with Lemma 8.

ii) By Proposition 14 and Assumptions 7(a), for all $r > 0$ small enough, the result follows.

□

Conclusion of the proof of Theorem 1 . For all $\varphi \in C_b$ and for all $r > 0$, $\left|\int_{\mathbf{\Theta}} \varphi(\theta)\tilde{f}_{\theta_T^*,sp}(\theta)\mathrm{d}\theta - \varphi(\theta_0)\right|$ $\leqslant \left|\int_{B_r(\theta_T^*)} \varphi(\theta_0)\tilde{f}_{\theta_T^*,sp}(\theta)\mathrm{d}\theta - \varphi(\theta_0)\right| + \left|\int_{B_r(\theta_T^*)} [\varphi(\theta) - \varphi(\theta_0)]\, \tilde{f}_{\theta_T^*,sp}(\theta)\mathrm{d}\theta\right| + \left|\int_{\mathbf{\Theta}\setminus B_r(\theta_T^*)} \varphi(\theta)\tilde{f}_{\theta_T^*,sp}(\theta)\mathrm{d}\theta\right|$. Therefore, by Corollary 4, for all $\varepsilon > 0$, for $r > 0$ small enough, for T big enough,

$$\left|\int_{\mathbf{\Theta}} \varphi(\theta)\tilde{f}_{\theta_T^*,sp}(\theta)\mathrm{d}\theta - \varphi(\theta_0)\right| \leqslant \varepsilon \quad \mathbb{P}\text{-a.s.},$$

which is the result needed. □

Conclusion of the proof of Theorem 2. Use Proposition 11 and apply the change of variable $s := \sqrt{T}\Sigma_T(\theta_T^*)^{-\frac{1}{2}}(\theta - \theta_T^*)$. □

BI Norwegian Business School, Oslo, Norway 0484

E-mail address: Benjamin.Holcblat@bi.no