A Classical Moment-Based Approach with Bayesian Properties: Econometric Theory and Empirical Evidence from Asset Pricing

Comments welcome

Benjamin Holcblat*

August 15, 2013

Abstract

In this paper, we develop a classical moment-based inference framework with Bayesian properties. We prove that there exists an intensity distribution of the solutions to empirical moment conditions over the parameter space. We approximate it with the empirical saddlepoint (ESP) technique. We call the result the ESP intensity. A higher ESP intensity value indicates a higher estimated probability weight of being a solution to the empirical moment conditions. We propose to use the ESP intensity in the same way as posteriors are used in Bayesian inference to obtain point estimators, confidence regions, and tests. We call this the ESP approach, and explain the rationale behind it. We prove consistency and asymptotic normality of the ESP intensity. The ESP approach provides a unique answer to multiple concerns especially acute in consumption-based asset pricing, such as lack of identification and multiple hypothesis testing on the same data set. It also sheds a new light on consumption-based asset pricing, and, in particular, indicates that consumption-based asset pricing theory is more consistent with data than existing inference approaches suggest.

KEYWORDS: Saddlepoint approximation; Multiple solutions to estimating equations; (Weak) identification; Foundations of statistics; Bayesian inference; Decision theory; Confidence region; Multiple hypothesis testing; Inference of consumption-based asset pricing models.

* Address: BI Norwegian Business School, Department of Financial Economics, 0442 Oslo, Norway. Homepage: http://www.bi.edu/research/employees/?ansattid=A1210114

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, Jay Kadane, Lars-Alexander Kuehn, Vineet Kumar, Chen Li, Artem Neklyudov, Øyvind Norli, Carlos Ramirez, Christoph Rothe, Bryan Routledge, Max Roy, Stefano Sacchetto, Mark Schervish, Teddy Seidenfeld, Steven Shreve, Fallaw Sowell, George Tauchen (discussant) and seminar/conference participants at Carnegie Mellon University, University of Copenhagen (Math), University of Laval, BI Norwegian Business School, University of Stavanger, University of Toulouse (TSE), McGill University, 2012 $(EC)^2$ in Maastricht, Paris December 2012 Finance Meeting, University of Lund, University of Oslo (Math), 2013 Financial Econometrics Conference in Toulouse (TSE) and SMYE 2013 in Aarhus. Any errors are my own. I acknowledge financial support from the William Larimer Mellon fund and the Centre for Asset Pricing Research. This paper is a revised and lightened version of my PhD thesis at Carnegie Mellon University. Parts of it previously circulated under different titles. A first draft dates from August 2009.

"Little experience is sufficient to show that the traditional machinery of statistical processes is wholly unsuited to the needs of practical research. Not only does it take a cannon to shoot a sparrow, but it misses the sparrow! The elaborate mechanism built on the theory of infinitely large samples is not accurate enough for simple laboratory data."

R.A. Fisher (Preface, 1925) cited by P.C.B. Phillips (p.1, 1982)

1. Introduction

The generalized method of moments (GMM), which embeds most of the econometric approaches used, has been shown to perform poorly in several empirical areas. In particular, in its original area of application, 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" (p.210 in Cochrane, 2001). One possible explanation is the inadequacy of consumption-based asset pricing theories. But models are not always rejected (e.g., Vissing-Jorgensen and Attanasio, 2003; Savov, 2011), and simulations point to the insufficiency of the standard classical inference theory for consumption-based asset pricing (e.g., Kocherlakota, 1990a; Hansen, Heaton and Yaron, 1996 and other papers in that issue of JBES).

At least three ways have been historically explored to improve the method of moments generalized by Hansen (1982). The first way is to look for asymptotic refinements. However, asymptotic refinements are about situations where the sample size can be infinitely increased, while practice is necessarily based on bounded sample size. The second way, which was favoured by R.A. Fisher, is to require stronger assumptions, such as normally distributed errors, to derive finite-sample results on which to rely. This second way entails the risk of replacing asymptotic approximations by approximative assumptions, so that it may decrease the gap between asymptotic and practice only to increase the gap between assumptions and the actual structure of data. The third way is to develop inference procedures that rely more on the information contained in the sample at hand and less on asymptotic limits. For example, generalizing Anderson and Rubin (1949), Stock and Wright (2000) derive confidence regions that do not only rely on asymptotic limit of standard statistics but also incorporate information from the global shape of the empirical objective function. In this paper, we go further in this direction. The main *theoretical* contribution of the paper is to define and develop a classical[1] moment-based inference framework that yields point estimators, confidence regions and tests that rely more on the information in the sample at hand and less on asymptotic limits. We call the result the **ESP approach**, as it is based on an empirical saddlepoint (ESP) approximation. This main theoretical contribution yields the main *empirical* contribution of this paper, which is to shed a new light on empirical consumption-based asset pricing through the ESP approach. While the ESP approach explains the difficulties faced by other inference approaches, it suggests that the key equilibrium implication of consumption-based asset pricing theory is more consistent with data than other approaches indicate.

The only difference between the true parameter and other parameter values is that the former one solves the moment conditions. Although analytically unknown, the empirical moment conditions are a finite-sample counterpart of them. Therefore, the idea of the ESP approach is to approximate the distribution of the solutions to the empirical moment conditions thanks to the saddlepoint technique. Different samples imply different empirical moments, and, thus, random solutions to empirical moment conditions. We call the approximation of their distribution the **ESP intensity**. It summarizes in probabilistic terms the approximated uncertainty about the true parameter due to the finiteness of the sample, in the sense that, for an infinite sample, the distribution of the solutions to the empirical moment conditions is the point mass at the true parameter. Thus, we propose to use the ESP intensity in the same way as a posterior is used in Bayesian inference to derive point estimators, confidence regions and tests. We prove that the ESP intensity is consistent and asymptotically normal, which means that it converges to a point mass at the true parameter like a Gaussian distribution with a standard deviation that goes to zero at the rate square root of the sample size. We also show that these results are robust to the presence of multiple solutions to the moment conditions (non-identification), as long as their expected number is finite.

[1] In this paper, the word "classical" is used in opposition to "Bayesian". We characterize as "classical" an approach that does not treat the true parameter as a random variable. The difference and similarity between the theoretical approach here, which is a classical approach, and the Bayesian approach is discussed in section 6. The theoretical approach here is also different from existing classical inference, and the common interpretation of Fisher's fiducial inference (e.g., Seidenfeld, 1992).

This paper relates to several strands of literature. We distinguish five of them.

First, the ESP approach contributes to inference decision theory, which considers inference as a choice of parameter values in the spirit of microeconomic theory under uncertainty. More precisely, an inference decision-theoretic approach is an approach in which an econometrician chooses a utility function (or, equivalently, a loss function)[2] according to an inference purpose, and then makes the inference decision that maximizes the expected utility (or, equivalently, minimizes the expected loss). A decision-theoretic approach does not only provide flexibility through the choice of a utility function, but also provides strong finite-sample foundations. Maximization of expected utility is the *optimal* answer to the inferred uncertainty that comes from estimation, as maximization of expected utility by a consumer is optimal in microeconomic theory. However, a decision-theoretic approach is generally impossible or delicate within existing classical inference theory (e.g., pp.4-5 in Lehmann and Casella, 1983), so that only asymptotic optimality results are typically obtained. In contrast, the ESP approach offers a classical inference framework in which the application of decision theory is straightforward. We prove that for a large class of utility functions the resulting estimators are consistent.

Second, the ESP approach contributes to the saddlepoint literature. The ESP approximation is the empirical counterpart of the saddlepoint approximation. The ESP approach uses the ESP approximation technique in a new way that yields novel theoretical results. The saddlepoint approximation has been developed for a long time to improve on existing inference approaches in econometrics (e.g., Holly and Phillips, 1979; Phillips, 1982) and statistics (e.g., Tingley and Field, 1990; Jensen, 1992; Robinson, Ronchetti and Young, 2003). 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) propose more accurate tests for indirect inference. Sowell (2009) proposes an ESP-based point estimator to automatically correct the higher-order bias of generalized empirical likelihood (GEL) estimators. Aït-Sahalia and Yu (2006) propose a saddlepoint approximation of tran-

[2] For a mathematical reason, we express our decision-theoretic approach in terms of utility functions instead of loss functions because of our emphasis on 0-1 utility functions (see supplemental material). Nevertheless, this utility function should not be confused with the utility function of the representative agent of an asset pricing model. Context indicates which one it is about.

sition density for likelihood-based inference of continuous-time Markov processes. In this paper, we use the ESP approximation to develop an inference framework autonomous from the existing classical approaches. This change of perspective removes several theoretical hurdles to using the saddlepoint approximation for inference. In particular, it removes the dichotomy between the saddlepoint approximation of the distribution of potentially *multiple* solutions to empirical moment conditions and the *uniqueness* of point estimators, which is documented in Skovgaard (1985; 1990), Jensen and Wood (1998), and Almudevar, Field and Robinson (2000). This change of perspective also opens new areas of application to the saddlepoint approximation. For example, it suggests ways to incorporate uncertainty from estimation into the calibration of models. Furthermore, this change of perspective leads to show measure-theoretic, analytical and global asymptotic properties of ESP approximations.

Third, the ESP approach contributes to the identification and weak-identification literatures. Because, unlike the existing saddlepoint literature, it does not build on approaches that rely on identification, there is robustness of the ESP approach to lack of identification. By lack of identification, we designate both situations in which the moment conditions have multiple solutions (non-identification), and situations in which there is not a well-separated extremum of the objective function, although identification holds (weak identification).[3] Lack of identification is a frequent issue in many areas (e.g., Pesaran, 1981; Dominguez and Lobato, 2004; Mavroeidis, 2005) such as consumption-based asset pricing (e.g., Smith, 1999; Stock and Wright, 2000; Neely, Roy and Whiteman, 2001). The weak-identification literature (e.g., Dufour, 1997; Stock and Wright, 2000; Kleibergen, 2005; Guggenberger and Smith, 2005; Otsu, 2006) has developed confidence regions and tests robust to lack of identification for generalized empirical likelihood (GEL). The idea behind them is to deduce probabilistic statements from the asymptotic limit of objective functions instead of from quantities that rely on the asymptotic limit of point estimators. In a similar way,

[3] In this paper we do not follow the asymptotic formalization of weak identification used in the econometric literature (e.g., Stock and Wright, 2000), although the empirical motivation is the same. In this paper, an inference procedure is robust to weak identification if it is robust to the absence of a well-separated extremum. To the author's knowledge, no economic or financial model has been shown to mathematically yield the asymptotic definitions of weak identification used in the econometric literature .

the robustness of the ESP approach to lack of identification derives from the deduction of probabilistic statements from the ESP objective function. However, in contrast to the weak identification literature, the ESP objective function is based on an estimated distribution, the ESP intensity. This difference provides several advantages to the ESP approach, such as much sharper confidence regions and straightforward definition of confidence regions for subvectors of parameters. The ESP approach also offers a complementary approach to the identification literature, which has focused mainly on finding general technical conditions (e.g., Rothenberg, 1971; Komunjer, 2011) such as rank conditions, or model-specific (e.g., Magnac and Thesmar, 2002) conditions to guarantee identification. Despite progresses in the identification literature, identification remains often difficult to prove. Thus, the robustness of the ESP approach to multiple solutions to the moment conditions, when their expected number is finite, can be useful.

Fourth, the ESP approach contributes to the literature about multiple hypothesis testing, which generates two related challenges to existing classical inference theory. The first challenge is the introduction of data-mining biases (e.g., Leamer, 1978, 1983; Lo and MacKinley, 1990). For example, if one tests independently 100 true hypotheses with the exact level of each test equal to $\alpha = 5\%$, one can expect five true hypotheses to be rejected. The literature (e.g., White, 2000; Sullivan, Timmermann and White, 1999, 2001; Romano, Shaikh and Wolf, 2008) has proposed adjustments to control asymptotically the number of simultaneous rejections of true hypotheses. Without requiring specific adjustment, the ESP approach provides an alternative way to control asymptotically the number of simultaneous rejections of true hypotheses. ESP decision-theoretic tests do not lead to any error asymptotically so that the probability of rejection of true hypotheses is asymptotically zero. The second challenge concerns the difficulty of the existing classical hypothesis testing theory, the Neyman-Pearson theory, to deal with tests that are motivated by a first examination of the data set at use. For example, the standard theoretical justification of an asymptotic t-test is that the t-statistic has a probability $1 - \alpha$ (modulo approximation error) to be between the $\alpha/2$ and $1 - \alpha/2$ quantiles of a standard Gaussian distribution under the null hypothesis. However, once computed, the t-statistic is in the region of no-rejection with probability 0 or 1, i.e., it is or it is *not* in the region of no-rejection. Thus, if the result of this first test leads us to compute a second t-test of level α, the corresponding t-statistic has typically a probability different from $1-\alpha$ (modulo approximation error) to be between the $\alpha/2$ and $1-\alpha/2$ quantiles of a standard Gaussian distribution under the null hypothesis. The observation of the value of the first t-statistic has removed a part of the randomness of the second t-statistic. Except in a few cases (e.g., Gouriéroux and Monfort, 1989; Savin, 1984), statistics computed on the same data set are not independent. Thus, the existing classical hypothesis testing theory, the Neyman-Pearson theory, typically requires new hypothesis testing to be carried out on a completely new data set. This is usually impossible in finance and economics, as both fields are essentially non-experimental fields. Developments of the Neyman-Pearson theory have tried to overcome this challenge by assuming that the set of all possible statistics is determined by econometricians with an information set probabilistically independent from the data set. But, such an assumption ignores that the evolution of a field such as consumption-based asset pricing is the result of a hard-to-predict dialogue between theory and empirical studies based on more or less the same data set. In fact, this challenge makes it difficult to justify the use of the Neyman-Pearson theory in economics and finance. The Neyman-Pearson theory relies on probabilistic statements that are valid (modulo approximation error) only before examination of the sample. In contrast, the ESP approach relies on probabilistic statements that are valid (modulo approximation error) before and after examination of a data set so that it is immune to this challenge. The ESP intensity approximates the distribution of the solutions to the empirical moment conditions that one would obtain by drawing an infinite number of samples. To put it differently, the ESP intensity aims at inferring through the observed sample the other solutions that could have been observed.

Fifth, the paper contributes to the empirical consumption-based asset pricing literature. We estimate the relative risk aversion (RRA) of the representative agent using GMM (Hansen, 1982), continuously updated (CU) GMM (Hansen, Heaton and Yaron, 1996), which is an example of generalized empirical likelihood estimators (GEL), CU GMM for lack of identification (Stock and Wright, 2000), and the ESP approach with a 0-1 utility function. 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, such as Barro (2006) or Gabaix (2012). GMM and CU GMM provide almost the same results. They seem to underestimate or overestimate the (non-probabilized) uncertainty about the relative risk aversion of the representative agent. Depending on the sample, they provide relatively tight confidence region, or larger confidence regions which include values inconsistent with standard finance theory implications (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 do not seem informative in practice. The ESP approach explains the difficulties faced by other approaches. The fat and long right tail of the ESP intensity elucidates the large variations and large values of the RRA previously reported in the literature. At the same time, the ESP approach shows that consumption-based asset pricing theory is more consistent with data than other inference approaches suggest. First, in line with implications of finance theory, negative values for the RRA have almost no estimated probability weight. Second, the empirical key moment condition from consumption-based asset pricing theory has an estimated positive probability weight to hold. In addition, ESP point estimates of the relative risk aversion are smaller than the one from the other approaches.

The paper is organized as follow. Section 2 analyzes the problems faced in empirical consumption-based asset pricing, and provides an overview of the ESP approach. Section 3 heuristically explains some of the ideas behind the ESP approximation. Section 4 presents the ESP estimands and estimators, and section 5 the asymptotic behaviour of ESP estimators. Section 6 provides a discussion of the foundation of the ESP framework with respect to existing inference theories. Section 7 introduces a decision-theoretic approach within the ESP framework. Section 8 presents empirical evidence from consumption-based asset pricing. Short proofs and supplementary results are in the Appendix. Supplemental material contains a more complete version of the ESP decision-theoretic approach, and detailed proofs.

2. Motivation and overview

2.1. Analysis of the question

The key equilibrium implication of standard consumption-based asset pricing models is the equality between expected discounted gross return of different assets. More precisely, there is an equilibrium if, at date $t-1$, the expected gross return for date t discounted for risk and time is the same across assets, i.e.,

$$\forall i,j \in [\![1,n]\!], \quad \mathbb{E}_{t-1}\left[M_t(\theta_0)R_t^i\right] = \mathbb{E}_{t-1}\left[M_t(\theta_0)R_t^j\right] \tag{1}$$

where $\mathbb{E}_{t-1}[.]$ denotes the expectation operator conditional on the information available at $t-1$, R_t^j the gross return of asset j between $t-1$ and t, n the number of assets considered and $M_t(\theta_0)$ the stochastic discount factor indexed by the true parameter θ_0. Different consumption-based asset pricing models correspond to different ways of discounting for time and risk through different stochastic discount factors, $M_t(\theta_0)$. Typically, no distributions are assumed except for tractability reasons. Therefore, the standard inference approach in consumption-based asset pricing is GMM (e.g., Jagannathan, Skoulakis and Wang, 2002). Unlike most alternatives,[4] its main assumptions are moment conditions like equations (1).

With the GMM approach (Hansen, 1982), the minimization of a norm of the empirical moment condition first produces a point estimate, i.e., $\hat{\theta}_{gmm}$ minimizes

$$\left\| \frac{1}{T}\sum_{t=1}^{T}\psi(X_t,\theta) \right\| \tag{2}$$

where $\|.\|$ denotes a norm[5] and where $\psi(X_t,\theta) := \left[M_t(\theta)(R_t^1 - R_t^f \quad R_t^2 - R_t^f \cdots \quad R_t^n - R_t^f)'\right] \circledast Y_{t-1}$ with Y_{t-1} an element of the representative agent's information set at date $t-1$, $\circledast$ the Kronecker product and $'$ the transpose symbol. Second, considering that the t-statistic based

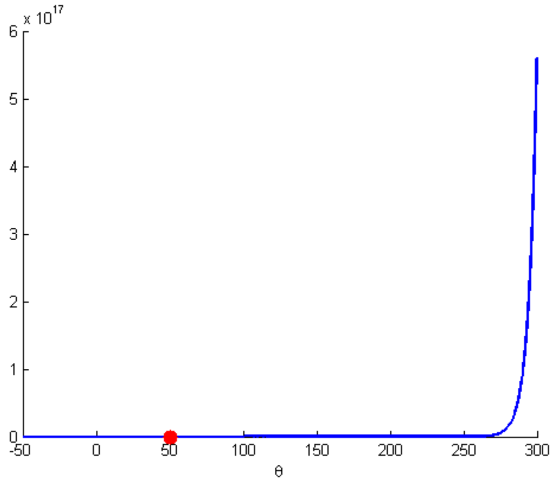
[4] Other moment-based inference approaches, such as the generalized empirical likelihood (GEL) approach (e.g., Newey and Smith, 2004), have been introduced in consumption-based asset pricing. However, without loss of generality, this section 2 focuses on GMM for simplicity. With minor modifications, the analysis applies to these more recent approaches as well.

[5] The norm often depends on data, as in two-step GMM, but this does not affect our analysis.

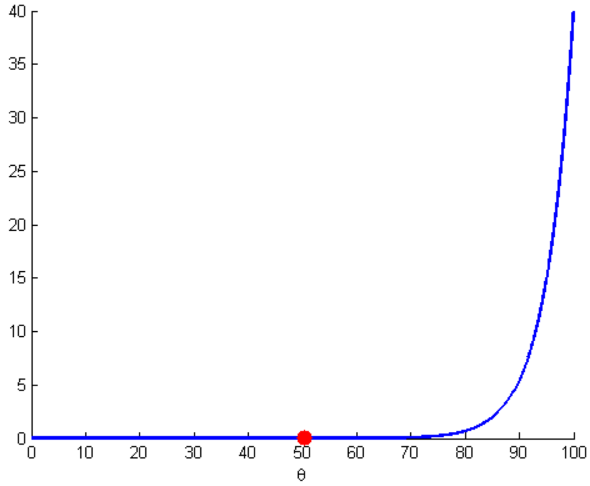
Table 2.1: **GMM inference (1890-2009)**

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

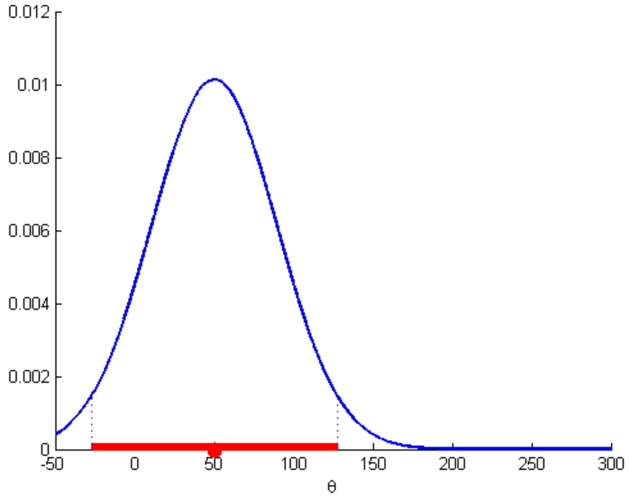
$R_t^m :=$ gross market return, R_t^f :=risk-free asset gross return, $C_t :=$ consumption,
θ :=relative risk aversion,
$\hat{\theta}_{gmm} = 50.3$, $\hat{I}_{.05} = [-26.9, 127.4]$.



(A) GMM objective function and point estimate.



(A zoom) GMM objective function and point estimate.



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

on a kth component $\sqrt{T}\frac{\hat{\theta}_{gmm,k}-\theta_{0,k}}{\hat{\sigma}_{k,k}}$ follows a standard Gaussian distribution, $\mathcal{N}(0,1)$,[6] a confidence region and a set of not-rejected point-hypothesis, $\hat{I}_\alpha = \Big[\hat{\theta}_{gmm,k} - \frac{\hat{\sigma}_{k,k}}{\sqrt{T}}u_{1-\alpha/2},$ $\hat{\theta}_{gmm,k} - \frac{\hat{\sigma}_{k,k}}{\sqrt{T}}u_{\alpha/2}\Big]$ with $u_{\alpha/2}$ the $\alpha/2$ quantile of a $\mathcal{N}(0,1)$, are deduced. Table 2.1(A)(A zoom) shows a GMM objective functions based on a standard moment condition where $M_t(\theta) := \left(\frac{C_t}{C_{t-1}}\right)^{-\theta}$ and $\frac{C_t}{C_{t-1}}$ is the gross growth consumption. The data set and the moment condition are from Julliard and Ghosh (2012). 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; pp.62-64 in Hall, 2005), which generates instable point estimates. However, as shown on Table 2.1(B), standard GMM summarizes inference as if the (non-probabilized) uncertainty[7] 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 Gaussian template used to summarized it. However, different parameter values can have very different theoretical implications. For example, a negative RRA implies a risk-seeking representative agent, while a positive RRA implies a risk-averse representative agent. Progress in consumption-based asset pricing theory will probably exacerbate this problem. Often, the more advanced a model, the larger the space in which the data information is projected and the more convoluted the GMM objective function.

The dichotomy between objective functions without well-separated global minimum and their Gaussian summary is not in contradiction with GMM theory. GMM theory is essentially about the asymptotic limit. GMM theory states that the global minimum of the *asymptotic* objective function corresponds to the true parameter value, but, in a finite sample, it does not even indicate whether the global minimum is the local minimum closest to the true parameter. Similarly, GMM theory states that if we infinitely increased the sample sizes, the t-statistic $\sqrt{T}\frac{\hat{\theta}_{gmm,k}-\theta_{0,k}}{\hat{\sigma}_{k,k}}$ would be distributed according to a $\mathcal{N}(0,1)$, but it does

[6] When the asymptotic distribution of a statistic is chi-square, the reasoning is the same. A chi-square is an inner product of Gaussian distributions.

[7] In this paper, the meaning of uncertainty is the usual one, that is, "quality of being indeterminate as to magnitude or value" (Oxford English Dictionary, 1928). In other words, uncertainty is not necessarily probabilized. In particular, given a sample, in standard classical inference theory, confidence regions summarize an uncertainty without randomness, unlike ESP confidence regions.

not provide information about the finite-sample distribution of the t-statistic. Nevertheless, in practice, any sample size is bounded.[8] Thus, in practice, the weak statistical structure required by standard GMM theory can paradoxically lead to strong statistical restrictions that are justified only asymptotically.[9]

The presence of a large minimal area in the GMM objective functions also suggest an additional and related concern, lack of identification. Lack of identification can occur because the information extracted from the sample through the inference procedure is not sharp enough, or because the model is not identified. Identification means there is only one solution to the moment conditions (1), i.e., the asset pricing equilibrium cannot correspond to multiple RRA values. Typically, such an assumption is unverifiable because the moment conditions are unknown analytically (e.g., section 2.2.3 in Newey and McFadden, 1994). Only with an infinite sample size would the moment conditions be perfectly revealed. However, the Gaussian template from GMM theory is particularly non-robust to non-identification, as the Gaussian distribution is unimodal with exponentially decreasing tails.

2.2. Informal presentation of the ESP approach

The paper aims at addressing the concerns mentioned above. Although there are no ideal finite-sample justifications, asymptotic arguments are not the only way to theoretically compare estimators. From a finite-sample point of view, an ideal point estimator would solve the moment conditions (1), but then no estimation would be needed. However, some inference approaches have higher finite-sample justification than others. For instance, any objective function consisting of the sum of the GMM objective function and a function vanishing asymptotically enjoys the same asymptotic justifications as the GMM objective function. More precisely, estimators induced by the following objective function have the

[8] Although in finance, continuous-time processes are often considered for mathematical tractability, in practice, a sample size is bounded. A computer memory is bounded.

[9] Advocates of Bayesian inference, such as Sims (pp.3-4, 8, 2007a; section III, 2007b), make similar remarks to criticize the classical approach.

same asymptotic properties as GMM estimators

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \psi(X_t, \theta) \right\| + \frac{h(\theta)}{T^k} \tag{3}$$

where $h(.)$ is an arbitrary bounded function and k a large enough constant. But, nobody would accept the objective function (3). For example, $h(.) := c\|. - \overline{\theta}\|$, with c a little larger than the largest number that the computer at use can handle, yields a point estimate close to $\overline{\theta}$ for very different parameter values $\overline{\theta}$, chosen in the parameter space, $\boldsymbol{\Theta}$. The difference between objective function (3) and the GMM objective function (2) is their finite-sample meaning. The GMM point estimate minimizes the norm of the empirical moment conditions, whereas the estimates from objective function (3) does not have a clear finite-sample meaning. More generally, one can use the same device as (3) to create an infinite number of estimates with the same asymptotic properties of the "best" asymptotic estimator available. Therefore, the idea behind the ESP approach is to find an inference approach with a strong finite-sample justification so that it yields inference procedures that rely more on the information contained in the sample at hand and less on asymptotic results. Good asymptotic properties should follow, as an asymptotic performance is the limit of increasing finite-sample performances.

The only difference between θ_0 and other elements of the parameter space is that θ_0 solves the moment conditions (1). The moment conditions (1) are unknown, but the empirical moment condition are their finite-sample counterpart. Different samples imply different empirical moment conditions, and thus different solutions.[10] Thus, we estimate the distribution of the solutions to the empirical moment conditions. The empirical saddlepoint (ESP) technique allows us to approximate this distribution non-parametrically. We call the ESP approximation the **ESP intensity**. Despite its regularity properties, it does not require the introduction of exogenous nuisance parameter, such as a bandwidth parameter, and it does not suffer from the curse of dimensionality usually faced by smooth non-parametric estimators of distributions (Ronchetti and Welsh, 1994). We prove that the ESP intensity converges

[10]To avoid a too cumbersome terminology, we call "empirical moment conditions" both the *ex ante* random empirical moment conditions and the *ex post* realized empirical moment conditions. Context indicates which ones it is about.

Table 2.2: **ESP inference with 0-1 utility function (1890-2009)**

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

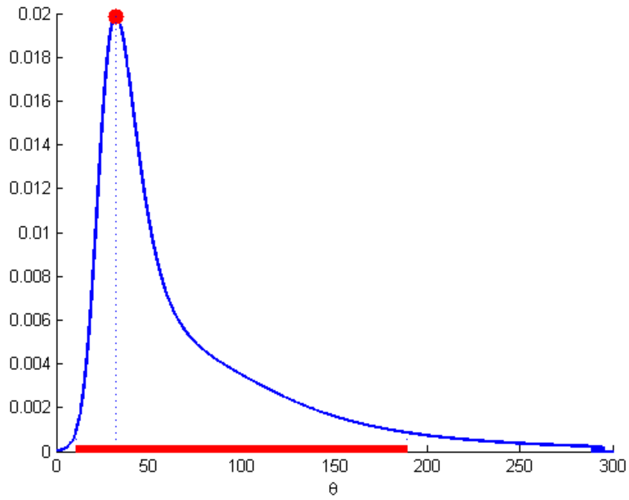
$R_t^m :=$ gross market return, R_t^f :=risk-free asset gross return, $C_t :=$ consumption,

θ :=relative risk aversion,

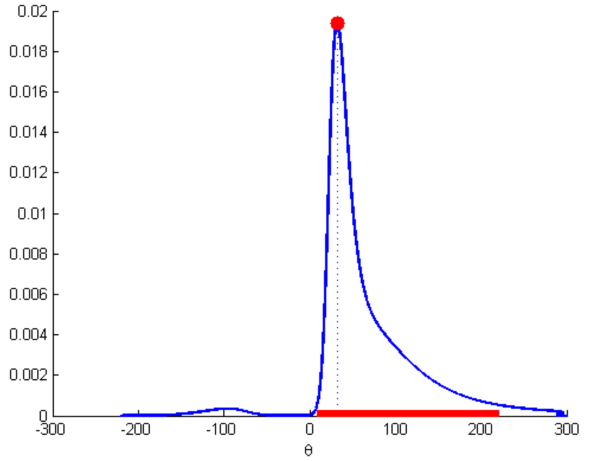
$\hat{\theta}_T^u = 32.21$;

Case with support restricted to $\mathbf{R}_+$: $\hat{I}_{.05} = [10.50, 188.85]$ (stripe on A), ESP support $= [0, 289.0]$

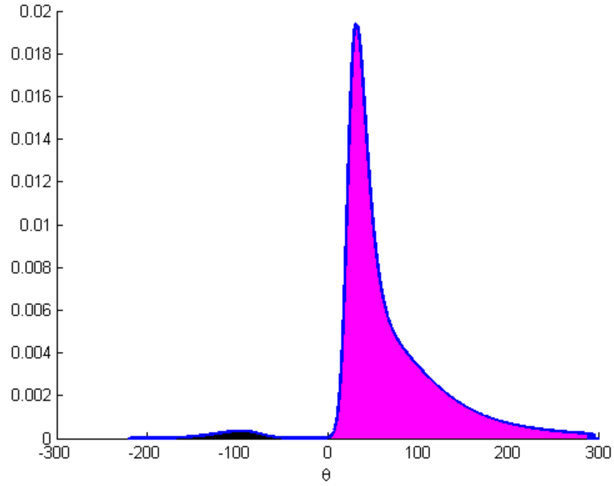
Case without restriction: $\hat{I}_{.05} = [9.0, 220.1]$ (stripe on B), ESP support $= [-218.2, 289.0]$



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



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



(C)Decision-theoretic ESP hypothesis test. $H_0 : \theta_0 > 0$.

to a point mass at the true parameter (or Dirac distribution at the true parameter) like a Gaussian distribution with a standard deviation that goes to zero at the rate square root of the sample size. Thus, the ESP intensity expresses in probabilistic terms the approximated uncertainty about the true parameter due to the finiteness of the sample. Consequently, a decision-theoretic approach is possible. The econometrician can choose a utility function (or, equivalently, a loss function), $u(.,.)$, according to an inference purpose. In practice, the utility function may correspond to the opposite of a financial loss implied by inference imprecision. Thanks to this utility function, we define an **ESP point estimator**, $\hat{\theta}_T^u$, as a maximizer of the ESP expected utility, i.e.,

$$\hat{\theta}_T^u := \arg\max_{\theta_e \in \boldsymbol{\Theta}} \tilde{\mathbb{E}}\left[u(\theta_e, \theta_T^*)\right]$$

where $\tilde{\mathbb{E}}\left[u(\theta_e, \theta_T^*)\right] := \int_{\boldsymbol{\Theta}} u(\theta_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta$ with $\tilde{f}_{\theta_T^*, sp}(.)$ the ESP intensity. By definition, $\hat{\theta}_T^u$ is the *optimal* point estimate with respect to the uncertainty embodied in the ESP intensity. We prove the consistency of $\hat{\theta}_T^u$ for a large class of utility functions. For researchers, a utility function corresponding to absolute preference for finite-sample truth is relevant. In this case, after normalization, utility equals one if θ_e is a solution to the empirical moment conditions and 0 otherwise. The resulting point estimate, which is presented in Table 2.2 for the same data set and moment condition as in Table 2.1, is the mode of the ESP intensity. In other words, it is a parameter value with the highest estimated probability weight of being a solution to the empirical moment conditions. Thus, it is a *maximum-probability* estimate.[11] Such a point estimate aims at taking into account all the samples that could have been observed. In contrast, the GMM point estimate is the realized solution to the empirical moment condition in the comparable just-restricted case (or just-identifying case).[12] Thus, GMM point estimators are backward-looking, while ESP point estimators are not. Because consumption-based asset pricing models are rational expectation models, arguments in the

[11] First, note that it is different from maximum-likelihood estimators (MLE). MLE maximizes the probability weight of the *observed sample*. Loosely speaking, MLE maximizes *plausibility* while maximum ESP aims at maximizing finite-sample *truth*. Second, note also that this is different from the mode of a Bayesian posterior (see section 6).

[12] In the over-restricted case (or over-identified case), GMM is also backward-looking. But, it is not immediately comparable with the ESP approach, because the GMM objective function is not expressed in terms of the dimension of interest, namely parameter values; but in terms of the norm of empirical moment conditions.

spirit of Lucas (1976) suggest that the ESP approach is more appropriate for self-consistency of inference.

We also define confidence regions to assess the stability of ESP point estimators. An **ESP confidence region** of level $1-\alpha$ is a set

$$\tilde{S}_{u,T} := \left\{ \theta_e \in \mathbf{\Theta} : \frac{1}{K_T} \int_{\mathbf{\Theta}} u(\theta_e, \theta) \tilde{f}_{\theta^*_T, sp}(\theta) d\theta \geqslant k_{\alpha,T} \right\}$$

where $k_{\alpha,T}$ is the highest bound satisfying $\int_{\tilde{S}_{u,T}} \frac{1}{K_T} \int_{\mathbf{\Theta}} u(\theta_e, \theta) \tilde{f}_{\theta^*_T, sp}(\theta) d\theta d\theta_e \geqslant 1-\alpha$ and $K_T := \int_{\mathbf{\Theta}^2} u(\theta_e, \theta)\ \tilde{f}_{\theta^*_T, sp}(\theta) d\theta d\theta_e$. We prove that ESP confidence regions converge to their asymptotic counterpart as the sample size increases. Table 2.2 (A) on p.13 shows the ESP confidence region with a 0-1 utility function in the same case as in Table 2.1 on p.9. All the parameter values in the confidence region provide a higher weighted utility for the econometrician than the ones outside. Thus, it captures lack of identification by construction. If the ESP objective function is asymmetric with a fat tail as in Table 2.2, the ESP confidence reflects it. This is not the case with the standard GMM approach because the asymptotic Gaussian distribution is symmetric with exponentially decreasing tails.[13] Thus, standard confidence intervals often underestimate the (non-probabilized) uncertainty about the true parameter. Standard confidence intervals can also simultaneously overestimate the uncertainty in another dimension. They consider the true parameter to be outside the parameter space with a strictly positive probability because the support of a Gaussian distribution is the whole real line. For example, the confidence region of the RRA in Table 2.1 includes negative values, although a negative RRA is often not consistent with standard consumption-based asset pricing theory. ESP confidence regions do not regard values outside the parameter space as possible because the ESP intensity support is included in the parameter space by construction. For example, in Table 2.2(A) we consider the case in which the parameter space is restricted to positive values, while in Table 2.2(B) the parameter space includes the

[13] We write "*standard* GMM approach" because continuously updated GMM confidence regions for lack of identification, S-sets, share similar advantages with ESP confidence regions (Stock and Wright, 2000). However, ESP confidence regions rely even less on asymptotic limit than S-sets. The value of the objective function, from which S-sets are deduced by inversion, is determined by the asymptotic distribution, while it is determined endogenously by the global shape of the ESP intensity in the ESP approach. See section 8 for more comparison.

whole support of the ESP intensity.

In the existing classical inference theory, tests usually correspond to confidence intervals, and thus are subject to the same concerns. Similarly to standard classical inference theory, we can define ESP tests that correspond to confidence regions. However, we also develop ESP decision-theoretic tests that do not correspond to confidence regions. Denote d_H and d_A, respectively, as no-rejection and rejection of a test hypothesis. Given a utility function chosen according to the hypothesis of interest, we define an **ESP decision-theoretic test** as a mapping such that if

$$\tilde{\mathbb{E}}[u(d_H, \theta_T^*)] \geqslant \tilde{\mathbb{E}}[u(d_A, \theta_T^*)] \quad ,$$

then it maps to d_H; and otherwise to d_A. To put it in words, a hypothesis is not rejected if the ESP expected utility provided by retaining the hypothesis is higher than the alternative. ESP decision-theoretic hypothesis testing is more flexible than standard classical testing theory. For instance, testing whether the representative agent is risk averse (i.e., $\theta_0 > 0$) or risk seeking (i.e., $\theta_0 < 0$) is straightforward in the ESP approach. In the case of a 0-1 utility function, we conclude that the representative agent is risk averse because $\int_{-\infty}^{0} \tilde{f}_{\theta_T^*,sp}(\theta)d\theta < \int_{0}^{\infty} \tilde{f}_{\theta_T^*,sp}(\theta)d\theta$ (i.e., black area<magenta area) according to Table 2.2(C). In existing classical inference theory, set-hypothesis tests are usually a challenge (e.g., section 21.D in Gouriéroux and Monfort, 1989). Decision-theoretic ESP tests are also more satisfactory than standard classical tests even from an asymptotic point of view.[14] By construction, a classical tests of level α lead asymptotically to wrongly reject a right hypothesis with probability α. In other words, a perfectly correct consumption-based asset pricing model is asymptotically rejected by a classical test with probability α. This is unsatisfactory because asymptotically the model is perfectly known. Such asymptotic error does not occur with ESP decision-theoretic tests as the ESP intensity converges to a point mass (or Dirac distribution) at the true parameter.[15] In addition, if multiple preference val-

[14]We write "standard classical tests" because there exist examples of classical tests such that the level of the test converges to zero as the sample size increase (see supplemental material).

[15]In the standard classical approach, a typical no-rejection region of a test of level α is $\hat{I}_\alpha = \left[\hat{\theta}_{gmm,k} - \frac{\hat{\sigma}_{k,k}}{\sqrt{T}} u_{1-\alpha/2}, \hat{\theta}_{gmm,k} - \frac{\hat{\sigma}_{k,k}}{\sqrt{T}} u_{\alpha/2}\right]$, and the justification for such a no-rejection

ues of the representative agent yield the same asset pricing equilibrium (non-identification), the standard classical approach is not valid. In contrast, ESP confidence regions and tests are robust to multiple preference values consistent with the moment conditions (1) as long as their expected number is finite. We prove that the ESP intensity converges to a sum of point mass (or Dirac distribution), each centered at a solution to the moment condition. Another more fundamental concern with the existing classical hypothesis testing theory, the Neyman-Pearson theory, is the question of its relevance in economics or finance. For example, if we decide to test whether the representative agent is risk-neutral (i.e., $\theta_0 = 0$) because we observed in Table 2.1(A) that the GMM objective function has large minimal area which includes 0, the standard theory does not allow us to not reject the hypothesis based on a t-test at 5%. Before collection of the data set, the t-statistic had a probability $1-\alpha$ (modulo approximation error) to be contained in the region of no-rejection under the null hypothesis. However, examination of the data set through Table 2.1(A) removed part of its randomness so that its probability of being in the region of no-rejection is not $1-\alpha$ anymore. Thus, most of the existing classical hypothesis testing theory typically requires a new data set probabilistically independent from the previous ones for every new test.[16] Such a requirement cannot be satisfied in economics or finance because they are mainly non-experimental fields. Because of the duality between confidence regions and tests in the Neyman-Pearson theory, existing classical confidence regions face the same challenge. In contrast, the ESP approach is immune to this challenge. While tests à la Neyman-Pearson depend on the *single* observed value of a statistic, ESP tests depends on *all the possible* values of a statistic weighted by their estimated probability weight. ESP tests rely on an approximation of the distribution of the solutions of the empirical moment conditions that one would obtain by drawing an infinite number of samples. An approximation of the distribution of the solutions to the empirical moment conditions, like the ESP intensity, aims at inferring through

region is the following: $\lim_{T\to\infty} \mathbb{P}\left\{\hat{\theta}_{gmm,k} - \frac{\hat{\sigma}_{k,k}}{\sqrt{T}} u_{1-\alpha/2} \leqslant \theta_{0,k} \leqslant \hat{\theta}_{gmm,k} - \frac{\hat{\sigma}_{k,k}}{\sqrt{T}} u_{\alpha/2}\right\} = \lim_{T\to\infty} \mathbb{P}\left\{u_{\alpha/2} \leqslant \sqrt{T}\frac{(\hat{\theta}_{gmm,k} - \theta_{0,k})}{\hat{\sigma}_{k,k}} \leqslant u_{1-\alpha/2}\right\} = 1-\alpha$. In the ESP approach, confidence regions and tests can be disentangled.

[16]Note that this concern does not necessarily invalidate the practice. In fact, the author has a work in progress in which he develops a general ESP-like inference theory that provides the most common hypothesis-testing practices with theoretical justification.

the observed sample the other solutions to the empirical moment conditions that could have been observed.

3. Heuristic derivation of ESP intensity

The purpose of this section is to informally explain some of the basic ideas behind the saddlepoint approximation. ESP intensity is the ESP approximation of the distribution of the solutions to empirical moment conditions. First, we derive heuristically the saddlepoint (SP) intensity under the assumption that the data follow a distribution from a known parametric family. Second, we plug in the empirical distribution and deduce the ESP intensity. For clarity, we consider a one dimensional parameter space (i.e., $m = 1$) in this section.[17]

3.1. The saddlepoint intensity

Denote θ_T^* a solution to the empirical moment conditions, $\frac{1}{T}\sum_{t=1}^{T}\psi(X_t,\theta) = 0$ where $\{X_t\}_{t=1}^{T}$ are univariate i.i.d. data. The probability distribution function (p.d.f.) of data is $f_X(.)$ with θ_0 the true parameter. Denote

$$Z_T := \sqrt{T}(\theta_T^* - \theta_0) \quad .$$

The Edgeworth expansion of the finite-sample distribution of Z_T is

$$f_{Z_T}(z) = \frac{1}{\sigma}\mathfrak{n}\left(\frac{z}{\sigma}\right)\left\{1 + \frac{1}{\sqrt{T}}r_1(z) + \frac{1}{T}r_2(z) + \ldots + \frac{1}{T^{j/2}}r_j(z) + o_p\left(\frac{1}{T^{-j/2}}\right)\right\}$$

where $f_{Z_T}(.)$ denotes the distribution of Z_T, $\mathfrak{n}(.)$ is the standard normal density, $\sigma^2 := \left[\mathbb{E}\frac{\partial\psi(X,\theta_0)}{\partial\theta}\right]^{-1}\mathbb{V}\left[\psi(X,\theta_0)\right]\left[\mathbb{E}\frac{\partial\psi(X,\theta_0)}{\partial\theta}\right]^{-1}$, j is the order of the approximation, $r_1(.)$ is a polynomial without constant term, and $r_j(.)$ are other polynomials. In accordance with the central limit theorem (CLT), the Edgeworth expansion shows that as $T \to \infty$ the distribution of Z_T, $f_{Z_T}(.)$, converges to the Gaussian density $\frac{1}{\sigma}\mathfrak{n}\left(\frac{\cdot}{\sigma}\right)$.

[17] This section is designed to be a self-sufficient introduction to the saddlepoint approximation, which dates back at least to Esscher (1932). More detailed presentations of the saddlepoint approximation include Field and Ronchetti (1990), Jensen (1995), and Goutis and Casella (1999).

The finite-sample quantity of interest is not Z_T, but θ_T^*. By the change of variable $\theta_T^* := T^{-\frac{1}{2}} Z_T + \theta_0$, we obtain the Edgeworth expansion of the distribution of θ_T^*,

$$
\begin{aligned}
f_{\theta_T^*}(\theta) &= \sqrt{T} f_{Z_T}\left(\sqrt{T}(\theta-\theta_0)\right) \\
f_{\theta_T^*}(\theta) &= \frac{\sqrt{T}}{\sigma}\mathfrak{n}\left(\sqrt{T}\frac{\theta-\theta_0}{\sigma}\right)\left\{1+\frac{1}{\sqrt{T}}r_1\left(\sqrt{T}(\theta-\theta_0)\right)+\frac{1}{T}r_2\left(\sqrt{T}(\theta-\theta_0)\right)+\dots\right. \\
&\qquad \left.+\frac{1}{T^{j/2}}r_j\left(\sqrt{T}(\theta-\theta_0)\right)+o_p\left(\frac{1}{T^{j/2}}\right)\right\} \qquad (4)
\end{aligned}
$$

Note that for $\theta = \theta_0$, the first term of the expansion, $\frac{\sqrt{T}}{\sigma}\mathfrak{n}(0)$, provides an accurate approximation of $f_{\theta_T^*}(.)$, because all non-constant monomials equal 0, and even the first polynomial, $r_1(.)$, cancels out. The crux of the SP approximation is to make this be the case for each $\theta \in \mathbf{\Theta}$. For each $\theta \in \mathbf{\Theta}$, $f_{\theta_T^*}(.)$ is recentered at θ_0 in a reversible way, and then only the first term of the expansion is retained. We recenter via a change of measure in the spirit of the Cameron-Martin-Girsanov theorem (e.g., Karatzas and Shreve, 1988, p. 191), termed exponential tilting.[18] In other words, the SP approximation replaces the standard global Gaussian approximation (i.e., CLT) with a continuum of point-wise Gaussian approximations. As a consequence, the error is "squeezed."

The result is the SP intensity

$$
f_{\theta_T^*,sp}(\theta) := \left[\mathbb{E}\mathrm{e}^{\tau(\theta)\psi(X,\theta)}\right]^T \left(\frac{T}{2\pi}\right)^{1/2}\left[\sigma^2(\theta)\right]^{-\frac{1}{2}} \qquad (5)
$$

[18] In finance, the physical distribution is recentered to obtain the risk-adjusted distribution under which there is null expected profit. With the SP approximation, the distribution of data is recentered for each $\theta \in \mathbf{\Theta}$ to better approximate the probability weight of θ satisfying the moment condition. Exponential tilting corresponds to the Radon-Nikodyn derivative $\frac{dP_{\tau(\theta)}}{dP} = \frac{\mathrm{e}^{\tau(\theta)\psi(x,\theta)}}{\mathbb{E}[\mathrm{e}^{\tau(\theta)\psi(X,\theta)}]}$.

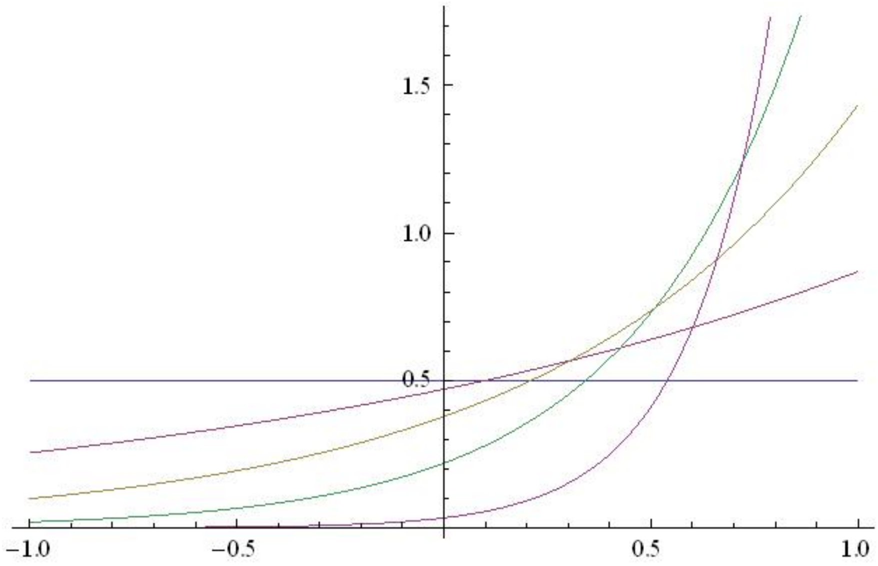


Figure 1: Tilting of $f_X(.) := \mathbf{l}_{[-1,1]}(.)$ for $\frac{1}{T}\sum_{t=1}^{T}\psi(X_t,\theta) := \frac{1}{T}\sum_{t=1}^{T}(X_t-\theta)$ and $T=1$. For θ equals 0, .2, .4, .6, .8 and .95, $\tau(\theta)$, respectively, equals 0, 1.34, 2.4, 5 and 20.

where

$$\sigma^2(\theta) := \left[\int_{\mathbf{R}} \frac{\partial\psi(x,\theta)}{\partial\theta} f_{X,\tau(\theta)}(x)dx\right]^{-1}\left[\int_{\mathbf{R}}\psi(x,\theta)^2 f_{X,\tau(\theta)}(x)dx\right] \times \left[\int_{\mathbf{R}} \frac{\partial\psi(x,\theta)}{\partial\theta} f_{X,\tau(\theta)}(x)dx\right]^{-1} \tag{6}$$

$$f_{X,\tau(\theta)}(x) := \frac{\mathrm{e}^{\tau(\theta)\psi(x,\theta)}}{\mathbb{E}[\mathrm{e}^{\tau(\theta)\psi(X,\theta)}]} f_X(x) \tag{7}$$

$$\tau(\theta) \quad s.t. \quad \int_{\mathbf{R}} \psi(x,\theta)\frac{\mathrm{e}^{\tau\psi(x,\theta)}}{\mathbb{E}[\mathrm{e}^{\tau(\theta)\psi(X,\theta)}]} f_X(x)dx = 0 \quad . \tag{8}$$

The approximation (5) was found by Field (1982), who extended the work of Daniels (1954) for means to Z-estimators (also called M-estimators by an abuse of terminology). The first term of the SP intensity is the exponential tilting term. It comes from recentering. The two other terms correspond to the first term of the Edgeworth expansion (4) for $\theta = \theta_0$. Note that $\mathfrak{n}(0) = \frac{1}{\sqrt{2\pi}}$. The variance $\sigma^2(\theta)$ now depends on θ because it is computed under the new exponentially tilted distribution, $f_{X,\tau(\theta)}(.)$, for each $\theta \in \mathbf{\Theta}$. Equation (7) defines for each $\theta \in \mathbf{\Theta}$ the exponentially tilted distribution under which θ is a solution to the moment condition.The exponentially tilted distribution, $f_{X,\tau(\theta)}(.)$, is indexed by the tilting parameter, $\tau(\theta)$. Equation (8) defines the tilting parameter. It indicates how to tilt the physical p.d.f. $f_X(.)$ to obtain the tilted p.d.f. $f_{X,\tau(\theta)}(.)$. In the case of the estimation of the mean of a uniform distribution over $[-1, 1]$, tilted distributions are displayed on Figure 1 for $T = 1$. The higher is θ, the higher is $\tau(\theta)$, the more tilted is the distribution.

3.2. The ESP intensity

The SP approximation assumes a known parametric family of distribution for data. But, a financial economic model typically does not imply a distribution, except for tractability reasons. The ESP approximation does not need parametric assumptions about the distribution of data.

In the SP intensity (5), substitution of $f_X(.)$ for the empirical distribution yields the following ESP intensity

$$\hat{f}_{\theta_T^*,sp}(\theta) := \exp\left\{T \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)\psi_t(\theta)}\right]\right\}\left(\frac{T}{2\pi}\right)^{1/2}\left[\sigma_T^2(\theta)\right]^{-\frac{1}{2}} \tag{9}$$

where $\psi_t(.) := \psi(X_t, .)$ and

$$\sigma_T^2(\theta) := \left[\sum_{t=1}^{T} \hat{w}_{t,\theta}\frac{\partial \psi_t(\theta)}{\partial\theta}\right]^{-1}\left[\sum_{t=1}^{T} \hat{w}_{t,\theta}\psi_t(\theta)^2\right]\left[\sum_{t=1}^{T} \hat{w}_{t,\theta}\frac{\partial \psi_t(\theta)}{\partial\theta}\right]^{-1},$$

$$\hat{w}_{t,\theta} := \frac{\exp\left[\tau_T(\theta)\psi_t(\theta)\right]}{\frac{1}{T}\sum_{i=1}^{T}\exp\left[\tau_T(\theta)\psi_i(\theta)\right]} \times \frac{1}{T}, \tag{10}$$

$$\tau_T(\theta) \quad s.t. \quad \sum_{t=1}^{T}\psi_t(\theta)\frac{\exp\left[\tau_T(\theta)\psi_t(\theta)\right]}{\frac{1}{T}\sum_{i=1}^{T}\exp\left[\tau_T(\theta)\psi_i(\theta)\right]} \times \frac{1}{T} = 0. \tag{11}$$

The approximation (9) was first studied by Ronchetti and Welsh (1994), who extended the work of Feuerverger (1989) for means to Z-estimators. For a fixed $\theta \in \boldsymbol{\Theta}$, the first term, the exponential tilting term, measures the extent to which the empirical distribution should be tilted so that the finite-sample moment condition (11) is zero. It is the empirical counterpart of $\left[\mathbb{E}\mathrm{e}^{\tau(\theta)\psi(X,\theta)}\right]^T = \exp\left\{T\ln\left[\mathbb{E}\mathrm{e}^{\tau(\theta)\psi(X,\theta)}\right]\right\}$ in (5) . The other terms discount the exponential tilting term according to the level of the variance of the solution to the finite-sample moment condition under the tilted distribution (10).

The SP and ESP approximations have been used to refine existing inference approaches in the same spirit as bootstrap (more precise confidence intervals and bias corrections). In this paper, we use the ESP approximation to develop a novel theoretical framework for inference.

4. The ESP estimand and estimator

This section defines the theoretical framework of the ESP approach.

4.1. The ESP estimand

The ESP estimand is the distribution of the solutions to the empirical moment conditions. We require the following assumptions to define the estimand.

Assumption 1. **(a)** $\{X_t\}_{t=1}^{\infty}$ *is a sequence of random vectors of dimension* p *on the 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* $\mathcal{E} \otimes \mathcal{B}(\mathbf{\Theta})/\mathcal{B}(\mathbf{R}^m)$*-measurable, where* $\mathcal{E} \otimes \mathcal{B}(\mathbf{\Theta})$ *denotes the product* σ*-algebra.* **(d)** *For the sample size at hand* T*, the expectation of the number of solutions to the empirical moment conditions is finite, i.e.,* $\sum_{n=1}^{\infty} n p_{n,T} < \infty$ *where* $p_{n,T}$ *is the probability of having* n *solutions to the empirical moment conditions.*

Assumptions 1(a)(b) are weak and standard. Completeness of the probability space is essential to manipulate negligeable 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 literature. It requires equality between the dimension of the parameter space and number of moment conditions. The reason is simple. In general, if the number of restrictions (moment conditions) is higher than the degrees of freedom (dimension of the parameter space), there is no solution to a system of equations, thus, the probability weight that $\theta \in \mathbf{\Theta}$ is a solution to the empirical moment conditions is zero. Then, an approximation of the finite-sample distribution of the solutions to over-restricting (or over-identifying) empirical moment conditions is generally not useful. The author has a paper in which he shows how one can extend the parameter space to deal with over-restricting moment conditions and perform tests of over-restricting moment conditions. Assumption 1(d), the other departures from the GMM literature, means that the tails of the probability distribution of the number of solutions to the empirical moment conditions are not too thick. It is a mild departure from the GMM literature. Under standard assumptions, Corollary 1 (p. 76) shows the number of solutions to empirical moment conditions to be finite $\mathbb{P}$-a.s. for T big enough. Moreover, 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 empirical moment conditions normalized by the derivative of the latter ones. 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 non-linear moment conditions. Skovgaard (1985; 1990) introduces this notion in the SP literature. However, the existing SP literature has usually attempted to narrow multiplicity to unicity, and thus evacuate point random-field theory at the end. To the author's knowledge, Sowell (2007) is the only paper that considers the ability of the ESP approximation to account for multiple solutions 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 develop an inference framework that allows us to exploit the ability of the ESP approximation to account for multiple solutions to moment conditions.

We specialize the general definition of point random-fields for our purpose.

Definition 4.1 (Point random-field)**.** *Denote $\mathcal{N}_{\boldsymbol{\Theta}}$ the space of finite simple counting measures on $\mathcal{B}(\boldsymbol{\Theta})$, i.e., the space consisting of finite integer-valued measures, N, s.t. (such that) for all $\theta \in \boldsymbol{\Theta}$, $N(\{\theta\}) \in \{0,1\}$. Denote $\mathcal{B}(\mathcal{N}_{\boldsymbol{\Theta}})$ the Borel σ-algebra on $\mathcal{N}_{\boldsymbol{\Theta}}$ generated by the Prohorov metric. A point random-field (or point process) is a measurable mapping from $(\boldsymbol{\Omega}, \mathcal{E}, \mathbb{P})$ to $(\mathcal{N}_{\boldsymbol{\Theta}}, \mathcal{B}(\mathcal{N}_{\boldsymbol{\Theta}}))$.*[19]

In this paper, a point random-field is an application that maps each sample $\{X_t(\omega)\}_{t=1}^{T}$ to the corresponding set of solutions to the empirical moment conditions. More precisely,

[19] 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 finite on bounded sets (e.g., Matthes, Kerstan and Mecke, 1974; Kallenberg, 1975; Daley and Vere-Jones 1988).

for a given sample size T, it maps each realization $\omega \in \mathbf{\Omega}$ to a counting measure, $N_T(\omega, .)$. For all subsets A of $\mathbf{\Theta}$, the counting measure $N_T(\omega, .)$ indicates the number of solutions to the empirical moment conditions contained in A. The following proposition proves that it is actually the case $\mathbb{P}$-a.s. This is the main result of this section 4.1.

Proposition 4.1. *Denote $\#A$ the cardinality of the set A (i.e., the number of elements in A). Under Assumption 1, there exists a point random-field $N_T(.,.)$ such that for all $\omega \in \mathbf{\Omega}$ and $A \in \mathcal{B}(\mathbf{\Theta})$,*

$$N_T(\omega, A) = \#\left\{\theta \in A : \frac{1}{T}\sum_{t=1}^{T} \psi\left(X_t(\omega), \theta\right) = 0\right\} \qquad \mathbb{P}\text{- a.s.}$$

Proof. See Appendix A.1 (p. 69). □

Remark 1. A consequence of Proposition 4.1, 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 empirical moment conditions (see Proposition A.4ii) on p.75). This result generalizes Schmetterer-Jennrich's measurability result (Lemma 2 in Jennrich, 1969) under Assumption 1(d). This generalization should be particularly of interest to the literature about multiple roots of estimating equations (e.g., section 6.4 in Lehmann and Casella, 1983).◇

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 empirical moment conditions corresponds to the intensity measure associated with the point random-field $N_T(.)$. If there can be only one solution to the empirical moment conditions, the intensity measure is the probability distribution of the solution. But in the case of multiple solutions, we should generalize probability measures into intensity measures.

Definition 4.2 (Intensity measure)**.** *Denote $\mathcal{T} := \{\mathcal{T}_n\}_{n\geqslant 1}$ 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 $n \to \infty$.*[20] *The intensity measure of a finite point random field,*

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

N_T, is defined for all $A \in \mathcal{B}(\mathbf{\Theta})$ by

$$\mathbb{F}_T(A) := \lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbb{P}\{N_T(A_{n,i}) = 1\} \quad , \tag{12}$$

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

The idea behind Definition 4.2 is the following. A singleton $\{\theta\}$ can contain at most one solution to the empirical moment conditions, 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 amount of elements, a sequence of increasingly thinner partitions should be introduced to formalize the idea. Definition 4.2 is an adaptation of the general mathematical definition of intensity measures (e.g., Daley and Vere-Jones, 1988) in line with our Definition 4.1 of point random-field.[21]

Lemma A.2 in Appendix A.2 (p. 70) collects results from point random-field theory that ensure the relevance of Definition 4.2. Namely, 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 are shown.

The following proposition clarifies the relation between intensity measures and probability measures. It adapts a result from point random-field theory.

Proposition 4.2. *Under Assumptions 1, for $\theta \in \mathbf{\Theta}$,*

$$\mathbb{F}_T(A_n(\theta)) = \mathbb{P}\left\{N_T(A_n(\theta)) = 1\right\}(1 + \varepsilon_n(\theta)) \qquad \mathbb{F}_T\text{-a.e.}$$

where $\varepsilon_n(\theta) \downarrow 0$ 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. 71). □

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 $\theta_2 \notin A_{n,i}$.

[21] 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., chap 11 in Duffie, 1992).

In accordance with the idea behind Definition 4.2, the intensity measure of a sufficiently small set is approximatively the probability that it contains one solution.

Theorem 1(iii) in Almudevar, Field and Robinson (2000) is a precursor of Proposition 4.2, which is its counterpart in our setup. Almudevar, Field and Robinson (2000) also formalize the point random-field introduced by Skovgaard (p.95, 1985), and thus our section 4.1 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 4.1). Because they construct a point process that discards continuum or accumulation of solutions to estimating equations, their setup do not need to forbid them, 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.

4.2. The ESP estimator

The ESP estimator is the intensity measure induced by the ESP intensity. More precisely, the estimator of the intensity measure of a subset of the parameter space is the integral of the ESP intensity over this subset. In this section, first, we study the properties of ESP intensity given by the approximation (9) (p. 21). We call it the rough ESP intensity. Although the rough ESP intensity seems appropriate in practice, for mathematical reasons we cannot use it directly to develop a theory. Thus, second, we show how we can define the (smooth) ESP intensity by arbitrarily slightly modifying the rough ESP intensity. As in the previous subsection, T remains fixed to the size of the sample at hand.

The use of the approximation (9) (p. 21) to define the rough ESP intensity requires the following assumption.

Assumption 2. *There exists $\varepsilon > 0$ such that 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 $\mathbf{\Theta}$. Assumption 2 allows to apply the implicit function theorem on the boundary of $\mathbf{\Theta}$ when necessary.

Simplification and generalization to m dimensions of the approximation (9) (p. 21) yields the following definition.

Definition 4.3 (Rough ESP intensity)**.** *The rough ESP intensity is*

$$\hat{f}_{\theta_T^*,sp}(\theta) := \exp\left\{T\ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}\left(\frac{T}{2\pi}\right)^{m/2}|\Sigma_T(\theta)|_{det}^{-\frac{1}{2}} \tag{13}$$

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

$$\begin{aligned}
\Sigma_T(\theta) &:= \left[\sum_{t=1}^{T}\hat{w}_{t,\theta}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]^{-1}\left[\sum_{t=1}^{T}\hat{w}_{t,\theta}\psi_t(\theta)\psi_t(\theta)'\right]\left[\sum_{t=1}^{T}\hat{w}_{t,\theta}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}, \\
\hat{w}_{t,\theta} &:= \frac{\exp\left[\tau_T(\theta)'\psi_t(\theta)\right]}{\sum_{i=1}^{T}\exp\left[\tau_T(\theta)'\psi_i(\theta)\right]}, \\
\tau_T(\theta) \quad &\text{s.t.} \quad \sum_{t=1}^{T}\psi_t(\theta)\exp\left[\tau_T(\theta)'\psi_t(\theta)\right] = 0_{m\times 1}\,,
\end{aligned} \tag{14}$$

wherever it exists.

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

Proposition 4.3. *Define the set* $\hat{\mathbf{\Theta}}_T \subset \mathbf{\Theta}$ *where the rough ESP intensity exists*

$$\hat{\mathbf{\Theta}}_T := \left\{\theta\in\mathbf{\Theta} : \exists\tau_T(\theta)\in\mathbf{R}^m \text{ s.t. } \sum_{t=1}^{T}\psi_t(\theta)\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} = 0_{m\times1} \text{ and } |\Sigma_T(\theta)|_{det}\neq 0\right\}.$$

Under Assumptions 1(a)-(c) and 2,

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

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

Proof. See Appendix A.4 (p. 71). □

The continuity of the rough ESP intensity is remarkable for a non-parametric estimate of a distribution obtained without smoothing. Nevertheless, the rough ESP intensity can have two undesirable properties. First, it ignores the information provided by the absence of solution to the tilting equation (14) because the rough ESP intensity 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 (14).

Proposition 4.4. *Denote* $[\![1, T]\!]$ *the integers in* $[1, T]$*. Under Assumptions 1(a)-(c) and 2, for all* $\theta \in \mathbf{\Theta}$*, there exists* $\tau \in \mathbf{R}^m$ *such that* $\sum_{t=1}^{T} \psi_t(\theta) \exp[\tau' \psi_t(\theta)] = 0_{m\times 1}$ *, 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* $t \in [\![1, T]\!]$*, such that* $\sum_{t=1}^{T} \psi_t(\theta) p_t = 0_{m\times 1}$*.*

Proof. See Appendix A.6 (p. 71). □

Proposition 4.4 restates a result used in Theorem 1 in Schennach (2005). It is a direct implication of the duality between the solution to the tilting equation (14) and maximization of entropy under moment conditions (e.g., sec.3(A) in Csiszár, 1975). To the author's knowledge, the duality relation has first been noted in econometrics by Kitamura and Stutzer (1997).

Proposition 4.4 indicates that the particular form of change of measure applied to the empirical distribution through exponential tilting does not really restrict the set of parameter values that admits a solution to the tilting equation (14). In terms of parameter values that have a solution to the tilting equation (14), exponential tilting spans a class of probability measures as rich as the class of probability measures that are equivalent to the empirical probability measure. Put differently, the absence of solution to the tilting equation for a parameter value $\theta \in \mathbf{\Theta}$ means that any reweighting of the data, that does not discard any data point, cannot put the finite-sample moment conditions to zero for this parameter value. Thus, the sample at hand does not provide support for this parameter value being a solution to the empirical moment conditions. Consequently, we set the ESP intensity to zero for parameter values without solution to the tilting equation.[22]

[22] In this way, we are in the spirit of Schennach (2005), in which a function of the parameter, which consists

The second undesirable property of the ESP intensity is that it may not be defined for $\theta \in \mathbf{\Theta}$ such that $|\Sigma_T(\theta)|_{det} = 0$, although there exists $\tau \in \mathbf{R}^m$ such that $\sum_{t=1}^T \psi_t(\theta)\mathrm{e}^{\tau'\psi_t(\theta)} = 0_{m\times 1}$. In this case, according to Proposition 4.4, the ESP intensity would not provide an indication of the probability of being a solution to the empirical moment conditions for a parameter value consistent with the tilted empirical moment conditions. The following assumption rules out such a case.

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

$$\left| \left[\frac{1}{T}\sum_{t=1}^T \frac{\partial \psi_t(\theta)'}{\partial \theta}\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^T \psi_t(\theta)\psi_t(\theta)'\right] \left[\frac{1}{T}\sum_{t=1}^T \frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1} \right|_{det} \neq 0.$$

Assumption 3 means that for a parameter value at which the empirical moment conditions can be set to zero by tilting, the standard variance-covariance estimate at this parameter value must be full-rank. In practice, it is not a stronger assumption than assumptions used in the GMM literature (e.g., Assumption D in Stock and Wright, 2000). Proposition A.1 in Appendix A.5 (p. 71) shows that it is satisfied Lebesgue almost everywhere under reasonable assumptions.

From a mathematical point of view, Assumption 3 is not needed in the sense that no proofs in this paper require it. However, Assumption 3 ensures that the ESP intensity does not ignore relevant parameter values in view of the sample at hand, or, more precisely, that $\hat{\mathbf{\Theta}}_T = \overline{\mathbf{\Theta}}_T$.[23] This combined with Proposition 4.4 allow us to meaningfully set the ESP intensity to zero on the complement of $\hat{\mathbf{\Theta}}_T$. Nevertheless, we also want the ESP intensity to be continuous at the junction between $\hat{\mathbf{\Theta}}_T$ and its complement. This leads to the following

of the product of exponential tilting 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 intensity is an approximation of the distribution of the solutions to the empirical moment conditions, and that we stick to its mathematical definition so that we do not interpret it as a Bayesian posterior. At most we indicate that it can be *interpreted* as an approximation of the distribution of the true parameter conditional on the sample size (see section 6).

[23] $\left\{\theta \in \mathbf{\Theta} : \left| \left[\frac{1}{T}\sum_{t=1}^T \frac{\partial \psi_t(\theta)'}{\partial \theta}\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^T \psi_t(\theta)\psi_t(\theta)'\right] \left[\frac{1}{T}\sum_{t=1}^T \frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1} \right|_{det} \neq 0 \right\} = \{\theta \in \mathbf{\Theta} : \left| \left[\frac{1}{\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \frac{\partial \psi_t(\theta)'}{\partial \theta}\right]^{-1} \left[\frac{1}{\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \psi_t(\theta)\psi_t(\theta)'\right] \left[\frac{1}{\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1} \right|_{det} \neq 0 \text{ with } \tau \in \mathbf{R}^m\}$, as exponential tilting does not alter the initial support of the distribution.

definition.

Definition 4.4 (ESP intensity and intensity measure)**.** *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 4.3, for a small $\eta > 0$,*

i) the ESP intensity (or smooth ESP intensity) is the function $\tilde{f}_{\theta_T^*,sp} : \mathbf{\Theta} \to \mathbf{R}_+$ s.t.

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

where $\bar{f}_T := \sup_{\theta \in \left\{\partial \hat{\mathbf{\Theta}}_T^{-\eta}(\omega)\right\}} \check{f}_{\theta_T^*,sp}(\theta)$ if $\hat{\mathbf{\Theta}}_T^{-\eta}(\omega) \neq \emptyset$, or 0 otherwise;

ii) the ESP intensity measure is the set function $\tilde{\mathbb{F}}_T$ s.t. for all $A \in \mathcal{B}(\mathbf{\Theta})$

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

The idea behind the definition of ESP intensity is the following. In the slightly reduced domain of definition of the rough ESP intensity (i.e., in $\hat{\mathbf{\Theta}}_T^{-\eta}$), the ESP intensity equals the rough ESP intensity. Where the tilting equation (14) does not have a solution (i.e., in $\hat{\mathbf{\Theta}}_T^c$), the ESP intensity equals zero. In between $\hat{\mathbf{\Theta}}_T^c$ and $\hat{\mathbf{\Theta}}_T^{-\eta}$, the values of the ESP intensity are the result of a regularization of the ESP intensity that preserves continuity. We regularize on $\hat{\mathbf{\Theta}}_T \cap \left(\hat{\mathbf{\Theta}}_T^{-\eta}\right)^c$ so that extremal values on the latter set are reached on its common boundary with $\hat{\mathbf{\Theta}}_T^c$ and $\hat{\mathbf{\Theta}}_T^{-\eta}$. Other regularization[24] techniques are possible.

Regularization of the rough ESP intensity is questionable. However, first note that it occurs on an arbitrarily small set. Second, implicit or explicit regularizations are frequent in inference. Even when observations are drawn from an absolutely continuous distribution, the application of the standard maximum likelihood approach implicitly requires a smooth version of a likelihood. For example, different Gaussian densities that are equal Lebesgue

[24] Note that the term "regularization" has a different meaning here from the meaning in ill-posed problems (e.g., Carrasco and Florens, 2000).

almost-everywhere can produce completely different inference results (e.g., section 7.A.2.a. in Gouriéroux and Monfort, 1989). Finally, regularization seems to have a negligeable effect in practice. To the author's knowledge, the need for regularization has never been reported in the SP literature.

The following properties of the ESP intensity follow from Definition 4.4.

Proposition 4.5. *Under Assumptions 1(a)-(c) and 2,*

i) the ESP intensity, $\tilde{f}_{\theta_T^*,sp}(\theta)$, is such that

a) for all $\theta \in \mathbf{\Theta}$, $\omega \mapsto \tilde{f}_{\theta_T^*,sp}(\theta)$ is $\mathcal{E}/\mathcal{B}(\mathbf{R})$-measurable;

b) for all $\omega \in \mathbf{\Omega}$, $\theta \mapsto \tilde{f}_{\theta_T^*,sp}(\theta)$ is a positive continuous function;

ii) the ESP intensity measure is

a) a finite-positive measure on the measurable space $(\mathbf{\Theta}, \mathcal{B}(\mathbf{\Theta}))$ such that

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

Proof. See Appendix A.7 (p.72). □

The value of the ESP intensity indicates the estimated intensity of parameter values. The following proposition clarifies the relation between the estimated intensity of a parameter value $\dot{\theta} \in \mathbf{\Theta}$ and its estimated probability weight of being a solution to the empirical moment conditions.

Proposition 4.6. *Define a point random-field $\tilde{N}(.)$ and a probability measure $\tilde{\mathbb{P}}$ as respective estimates of $N_T(.)$ and $\mathbb{P}$ s.t. they are consistent with $\tilde{\mathbb{F}}_T(.)$, i.e., such that $\tilde{N}(.)$, $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{F}}_T(.)$ satisfy the equation (12) in the Definition 4.2 of intensity measures. Assume that $\dot{\theta} \in \mathbf{\Theta}$ is a Lebesgue point, i.e., there exists $\varepsilon > 0$ such that for all $r > 0$ small enough, $\lambda(\overline{B_r(\dot{\theta})}) > \varepsilon\lambda(\overline{\mathrm{B}_r(\dot{\theta})})$ where $\overline{\mathrm{B}_r(\dot{\theta})}$ denotes the closed ball in $\mathbf{R}^m$ with radius $r > 0$ and center $\dot{\theta}$, $\overline{B_r(\dot{\theta})} := \overline{\mathrm{B}_r(\dot{\theta})} \cap \mathbf{\Theta}$, and where $\lambda(.)$ denotes the Lebesgue measure. Then, under Assumptions 1(a)-(c) and 2,*

$$\tilde{f}_{\theta_T^*,sp}(\dot{\theta}) = \lim_{r \to 0} \frac{\tilde{\mathbb{P}}\left\{\tilde{N}_T(\overline{B_r(\dot{\theta})}) = 1\right\}}{\lambda(\overline{B_r(\dot{\theta})})}$$

Proof. See Appendix A.8 (p. 72). □

Proposition 4.6 means that the estimated intensity of $\theta \in \mathbf{\Theta}$ generally corresponds to the estimated probability weight of θ of being solution to the empirical moment conditions. Under a mild assumption, Lemma A.4 in Appendix A.8 (p.72) ensures the existence of a point random-field $\tilde{N}(.)$ and a probability measure $\tilde{\mathbb{P}}$ such that $\tilde{N}(.)$, $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{F}}_T(.)$ satisfy the equation (12) in the Definition 4.2 of intensity measures. Then, Proposition 4.6 is a direct consequence of the Lebesgue's differentiation theorem and Proposition 4.2. Usually, in applications, all points of the parameter space are Lebesgue points. A point of $\mathbf{\Theta}$ is Lebesgue if the volume of the intersection of $\mathbf{\Theta}$ with a shrinking ball centered at the point does not decrease more quickly than proportionally to the volume of the shrinking ball. Therefore, interior points of the parameter space are necessarily Lebesgue, and points on the boundary are usually Lebesgue because boundaries are typically defined by linear constraints.

5. Asymptotic behavior of the ESP estimator

Whereas in the previous sections T remains fixed to the size of the sample at hand, in this section T goes to infinity. Although the logical value of asymptotic refinements for practice is not necessarily obvious, a first-order asymptotic result can have a logical value for practice. In particular, consistency of an inference procedure ensures that there is learning about the object of study when data correspond to different pieces of information about similar phenomena.

5.1. Consistency and asymptotic normality

In this section, we establish the consistency and asymptotic normality of the ESP intensity measure. By consistency, we mean convergence of the ESP intensity measure to a Dirac at the true parameter. By asymptotic normality, we mean convergence of the standardized ESP intensity measure to a standard normal distribution. The underlying phenomenon behind these results is the one revealed by Laplace's approximation (Laplace, 1774) and revived in inference by Le Cam (1953). For simplicity, our approach relies on basic assumptions, which can be significantly relaxed.

To study the asymptotic behaviour of the estimator, the asymptotic behavior of the estimand should first be fixed. The following assumptions set the asymptotic behavior of the 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* $\mathbb{E}[\psi(X,\theta)] = 0_{m\times 1}$. **(c)** $\mathbb{E}[\sup_{\theta\in\mathbf{\Theta}} \|\psi(X,\theta)\|] < \infty$. **(d)** $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}} \left\|\frac{\partial\psi(X,\theta)}{\partial\theta'}\right\|\right] < \infty$. **(e)** $\left|\mathbb{E}\left[\frac{\partial\psi(X,\theta_0)}{\partial\theta'}\right]\right|_{det} \neq 0$.

Assumptions 4 are basic and standard. Assumption 4(a) ensures the basic requirement for inference, that is, data contains different pieces of information (independence) about the same phenomenon (identically distributed). The conditions "independence and identically distributed" are much stronger than needed, and 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) ensures global identification. It will be relaxed in section 5.2. 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 section set the asymptotic behavior of the estimator. The following assumptions ensure the asymptotic existence of ESP intensity in a set that includes a neighborhood of the true parameter.

Assumption 5. *Define the set*

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

where $\Sigma_\infty(\theta) := \left[\mathbb{E}e^{\tau_\infty(\theta)'\psi(X,\theta)} \frac{\partial\psi(X,\theta)'}{\partial\theta}\right]^{-1} \mathbb{E}\left[e^{\tau_\infty(\theta)'\psi(X,\theta)}\psi(X,\theta)\psi(X,\theta)'\right] \left[\mathbb{E}e^{\tau_\infty(\theta)'\psi(X,\theta)} \frac{\partial\psi(X,\theta)}{\partial\theta'}\right]^{-1}$

(a) *There exists* $\bar{r} > 0$ *such that there exists* $\dot{T} \in \mathbf{N}$*, so that for all* $T \geqslant \dot{T}$, $B_{\bar{r}}(\theta_0) \subset \left\{\hat{\mathbf{\Theta}}_T \cap \hat{\mathbf{\Theta}}_\infty\right\}$. *Define a fixed* $\eta \in]0, \bar{r}[$. **(b)** *For all* $\dot{\theta} \in \hat{\mathbf{\Theta}}_\infty^{-\eta}$*, there exist* $r_1, r_2 > 0$ *such that*

$\mathbb{E}\left[\sup_{(\tau,\theta)\in B_{r_1}(\tau_\infty(\dot\theta))\times B_{r_2}(\dot\theta)}\|\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\|\right]<\infty$.

The set $\hat{\mathbf{\Theta}}_\infty$ corresponds to the parameter values where the limit of rough ESP intensity exists. In particular, the first two conditions ensure that $|\Sigma_\infty(\theta)|_{det}<\infty$ 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 continuity of $\theta\mapsto\tau_\infty(\theta)$ by an implicit function theorem.

The following assumptions ensure the validity of the Laplace's 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 four times continuously differentiable in a neighborhood of θ_0 $\mathbb{P}$-a.s.* **(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)** *For all $k\in[\![1,4]\!]$, there exists $M\geqslant 0$ such that there exist $\dot T\in\mathbf{N}$ and $r>0$, so that for all $T\geqslant\dot T$ and $\theta\in B_r(\theta_0)$ $\left\|D^k\left\{|\Sigma_T(\theta)|_{det}^{-\frac{1}{2}}\right\}\right\|<M$ $\mathbb{P}$-a.s.* **(d)** *For all $k\in[\![1,4]\!]$, there exists $M\geqslant 0$ such that there exist $\dot T\in\mathbf{N}$ and $r>0$, so that for all $T\geqslant\dot T$ and $\theta\in B_r(\theta_0)$, $\left\|D^k\left\{\ln\left[\frac{1}{T}\sum_{i=1}^T\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}\right\|<M$ $\mathbb{P}$-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)(d), adapted from Kass, Tierney and Kadane (1990), essentially ensure the existence and boundedness of the derivatives of ESP intensity terms up to the 4th order 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.

The following assumptions ensure the convergence of 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 $M\geqslant 0$ such that $T\geqslant\dot T$ implies that, for all $\theta\in\hat{\mathbf{\Theta}}_\infty^{-\eta}$, $\mathrm{e}^{-\varepsilon T}\,|\Sigma_T(\theta)|_{det}^{-\frac{1}{2}}\leqslant M$ $\mathbb{P}$-a.s.* **(b)** *For all $\dot\theta\in\hat{\mathbf{\Theta}}_\infty^{-\eta}$, there exist $r_1,r_2>0$ such that* $\mathbb{E}\left[\sup_{(\tau,\theta)\in B_{r_1}(\tau_\infty(\dot\theta))\times B_{r_2}(\dot\theta)}\mathrm{e}^{\tau'\psi(X,\theta)}\right]<$

∞.

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). This is a common type of assumption in entropy-based inference.

Under the assumptions above, we obtain the main result of the paper.

Theorem 5.1 (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., i.e.,*

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

where C_b *denotes the space of continuous bounded functions.*

Proof. See Appendix A.9 (p. 73).□

Theorem 5.1 means that the ESP intensity measure converges to a point mass at the true parameter as the sample size increases. Thus, uncertainty about the solution to the moment conditions vanishes as accumulation of data makes the empirical moment conditions a more precise approximation of the true moment conditions. Theorem 5.1 also means that the estimator, the ESP intensity measure, and the estimand, the intensity measure, converge towards each other as sample size increases. All other consistency results of this paper follow from this theorem.

The counterpart of Theorem 5.1 in Bayesian inference is the consistency of posterior distributions (or Doob's theorem). However, despite the similarities between the two theorems, the theoretical foundations behind them are different, as explained in section 6. Moreover, Theorem 5.1 is stronger than Doob's theorem, in the sense that the ESP intensity integrates to one asymptotically, although it is not divided by its integral. To the author's knowledge, none of the existing applications of the Laplace's approximation to inference (e.g., Chernozhukov and Hong, 2003) share this feature.

A second standard convergence result for Bayesian posterior distributions is asymptotic normality (or Laplace-Bernstein-von Mises' theorem). We also provide its counterpart in our framework.

Theorem 5.2 (Asymptotic Normality). *Let $a, b \in \mathbf{\Theta}$ such that $a \leqslant b$ where "$a \leqslant b$" means that every component of $b - a$ is non-negative. Then, under Assumptions 1(a)-(c),2, 4, 5, 6(a)(c)(d)(e) and 7, as $T \to \infty$,*

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

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* θ_T^* *any solution to the empirical moment conditions, and* $[\Sigma_T(\theta_T^*)]^{\frac{1}{2}}$ *s.t.* $\Sigma_T(\theta_T^*) = \left([\Sigma_T(\theta_T^*)]^{\frac{1}{2}}\right)' [\Sigma_T(\theta_T^*)]^{\frac{1}{2}}$ *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 where* $D(a,b) := \{z : a \leqslant z \leqslant b\}$.

Proof. See Appendix A.9 (p. 73).□

Remark 2. While in "$D_T(a,\theta_T^*,b)$," "θ_T^*" denotes a random variable that maps an $\omega \in \mathbf{\Omega}$ 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 $\omega \in \mathbf{\Omega}$ 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 4.2). We do not introduce two different notations for simplicity. Moreover, the difference between the two meanings disappears when there can be only one solution to the empirical moment conditions.◇

Theorem 5.2 indicates that ESP intensity converges asymptotically to a point mass at the true parameter like a Gaussian distribution with standard deviation that goes to zero at the rate $T^{-\frac{1}{2}}$. Theorem 5.2 is in line with the well-known asymptotic normality of a solution to empirical moment conditions. Theorem 5.2 is close to Theorem 5 in Sowell (2007), although the latter does not provide the asymptotic normality of the ESP intensity. Theorem 5.2 also suggests that confidence regions and tests à la Neyman-Pearson can be derived from the ESP intensity. However, we do not follow this way because of the difficulty in justifying the relevance of the Neyman-Pearson theory to non-experimental fields like economics and finance.

5.2. Robustness to lack of identification

In moment-based inference, identification holds when there is a unique parameter value that solves the moment conditions. However, in practice, moment conditions as a function of the parameter of interest are unknown. Only if we could increase the sample size infinitely would we know them. In addition, robustness to multiple solutions to the moment conditions a fortiori implies robustness to situations where the objective functions consists of several extremal areas, although identification holds (weak-identification). Therefore, robustness to multiple solutions to moment conditions is an important and desirable property.

In finite sample, by construction our inference framework is robust to multiple solutions to moment conditions as long as their expected number is finite. In this section, we show that it is also true asymptotically. More precisely, we establish multi-consistency and multi-asymptotic normality of the ESP intensity measure. By multi-consistency, we mean convergence of the ESP intensity measure to a sum of Dirac distribution each centered at one of the solutions to the moment conditions. By multi-asymptotic normality, we mean that the ESP intensity measure converges to a sum of Dirac like a sum of Gaussian distributions with standard deviation that goes to zero at the rate $T^{-\frac{1}{2}}$.

We adapt assumptions of the previous section to allow for multiple solutions to the moment conditions. Assumptions 4(b) and (e) become the following.

Assumption 8. **(b')** *In the parameter space* $\boldsymbol{\Theta}$*, there exist multiple solutions,* $\left\{\theta_0^{(i)}\right\}_{i=1}^{\overline{n}}$ *with* $\overline{n}$ *the number of solutions,*[25] *to the moment conditions* $\mathbb{E}\left[\psi(X,\theta)\right] = 0_{m\times 1}$ *such that for all*

[25]In accordance with Assumption 1(d), the number of solutions is unbounded but finite $\mathbb{P}$-a.s.

$i \in [\![1, \overline{n}]\!]$, $\theta_0^{(i)} \in \operatorname{int}(\mathbf{\Theta})$. **(e')** *For all* $i \in [\![1, \overline{n}]\!]$, $\left|\mathbb{E}\left[\frac{\partial\psi(X,\theta_0^{(i)})}{\partial\theta'}\right]\right|_{det} \neq 0$.

Assumption 5(a) becomes the following.

Assumption 9. **(a')** *For all* $i \in [\![1, \overline{n}]\!]$, *there exists* $\bar{r}^{(i)} > 0$ *such that there exists* $\dot{T}^{(i)} \in \mathbf{N}$, *so that for all* $T \geqslant \dot{T}^{(i)}$, $B_{\bar{r}^{(i)}}(\theta_0^{(i)}) \subset \left\{\hat{\mathbf{\Theta}}_T \cap \hat{\mathbf{\Theta}}_\infty\right\}$.

Assumption 6 becomes the following.

Assumption 10. *For all* $\theta_0^{(i)}$ *with* $i \in [\![1, \overline{n}]\!]$, *Assumptions 6(a)-(e) are satisfied with* θ_0 *replaced by* $\theta_0^{(i)}$.

Thanks to the above modifications of the assumptions, multi-consistency of the ESP intensity measure follows.

Theorem 5.3 (Multi-consistency)**.** *Under Assumptions 1(a)-(c),2, 4, 5, 6(a)(c)(d)(e) and 7, modified according to Assumptions 8-10, as* $T \to \infty$, *the ESP smooth intensity,* $\tilde{f}_{\theta_T^*,sp}(.)$, *converges in distribution (or narrowly converges) to the sum of Dirac distributions* $\sum_{i=1}^{\overline{n}} \delta_{\theta_0^{(i)}}(.)$ $\mathbb{P}$*-a.s., i.e.,*

$$\forall\varphi \in C_b, \qquad \int_{\mathbf{\Theta}} \varphi(\theta)\tilde{f}_{\theta_T^*,sp}(\theta)d\theta \to \sum_{i=1}^{\overline{n}} \int_{\mathbf{\Theta}} \varphi(\theta)\delta_{\theta_0^{(i)}}(\theta)d\theta \quad \mathbb{P}\text{-a.s.}$$

where C_b *denotes the space of continuous bounded functions.*

Proof. See Appendix A.9 (p. 73).□

Theorem 5.2 becomes the following.

Theorem 5.4 (Multi-asymptotic normality)**.** *Let* $a, b \in \mathbf{\Theta}$ *such that* $a \leqslant b$ *where "*$a \leqslant b$*" means that every component of* $b - a$ *is non-negative. Then, under Assumptions 1(a)-(c),2, 4, 5, 6(a)(c)(d)(e) and 7, modified according to Assumptions 8-10, as* $T \to \infty$

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

where $D_T(a, \theta_T^{*(i)}, b) := \left\{\theta : \theta_T^{*(i)} + T^{-\frac{1}{2}}\left[\Sigma_T(\theta_T^{*(i)})\right]^{\frac{1}{2}} a \leqslant \theta \leqslant \theta_T^{*(i)} + T^{-\frac{1}{2}}\left[\Sigma_T(\theta_T^{*(i)})\right]^{\frac{1}{2}} b\right\}$ *with* $\left\{\theta_T^{*(i)}\right\}_{T\geqslant 1}$ *a sequence of solutions to the empirical moment conditions converging to* $\theta_0^{(i)}$ $\mathbb{P}$*-a.s. and* $n \in [\![1, \overline{n}]\!]$.

Proof. See Appendix A.9 (p. 73).□

Theorems 5.3 and 5.4 have no counterparts in the standard Bayesian framework because the latter deals only with probability measures.

6. Discussion

In this section, we discuss differences and similarities between the ESP framework and existing inference theories.

6.1. Comparison with the Bayesian framework

The main output of both the ESP and Bayesian frameworks is a distribution that summarizes uncertainty about the true parameter. However, the theoretical foundations of ESP and Bayesian frameworks are different. In this section, we explain these differences and their practical implications.

Taken literally, Bayesian theory regards inference as a two-stage game between nature and an econometrician. In the first stage, nature draws the true parameter, θ_0, according to a prior distribution $\pi_{\theta_0}(.)$, and then draws a sample $\{X_t\}_{t=1}^T$ in accordance with a conditional probability distribution (p.d.f.) $l_{X_1,\ldots,X_T|\theta_0}(.|.)$. In the second stage, the econometrican tries to infer the true parameter value θ_0 given the sample at hand. As usual in game theory, the p.d.f. $l_{X_1,\ldots,X_T|\theta_0}(.|.)$ and $\pi_{\theta_0}(.)$ are common knowledge in the game-theoretic sense of the term. Thus, the econometrician updates this prior information, $\pi_{\theta_0}(.)$, thanks to data according to Bayes' formula

$$\pi_{\theta_0|X_1,\ldots,X_T}(\theta|x_1,\ldots,x_T) = \frac{l_{X_1,\ldots,X_T|\theta_0}(x_1,\ldots,x_T|\theta)\pi_{\theta_0}(\theta)}{\int_{\boldsymbol{\Theta}} l_{X_1,\ldots,X_T|\theta_0}(x_1,\ldots,x_T|\theta)\pi_{\theta_0}(\theta)d\theta},$$

to obtain the posterior distribution $\pi_{\theta_0|X_1,\ldots,X_T}(.|.)$. Therefore, as in our framework, Bayesian inference summarizes the uncertainty about the true parameter by means of a distribution.

This similarity between Bayesian and ESP inference should not eclipse their fundamental difference. Bayesian inference produces a distribution that summarizes uncertainty about the true parameter because the true parameter, θ_0, is treated as a random variable. This randomness is necessary to use the Bayes' formula. In other words, Bayesian theory requires an "axiomatic" assimilation of the unknown true parameter, θ_0, to a probabilized uncertainty through a prior $\pi_{\theta_0}(.)$ (p.508 in Robert, 1994). In contrast, in our framework, the randomness that is approximated by ESP intensity comes from data. Different samples imply different empirical moment conditions, and thus different solutions to those empirical moment conditions. The solution to the moment conditions, the true parameter, is not regarded as a random variable. In other words, in our framework, randomness comes from the use of random empirical moment conditions to approximate deterministic moment conditions.

The difference between sources of randomness has practical implications. Typically, the parameters of an economic model of interest are not random. For instance, in consumption-based asset pricing, the RRA of the representative agent is not random. Bayesian inference *transforms* the unknown true parameter into a probabilized uncertainty through two main extra statistical restrictions. First, it needs to specify a prior distribution. An economic model does not imply a specific prior distribution, and the use of non-informative prior distributions is not exempt from criticisms (e.g., section 3.5 in Robert, 1994). Second, it needs to specify the conditional p.d.f. $l_{\underline{X}_T|\theta_0}(.|.)$. Typically an economic model does not imply such family of distributions, except for tractability reasons. From a statistical point of view, these extra statistical restrictions may not matter, and even have been proved useful in many practical situations. But from a structural point of view, they make it difficult to disentangle the part of the inference results due to the empirical relevance of the economic model from the part due to statistical restrictions.[26] Non-parametric Bayesian inference also does not avoid extra statistical assumptions (e.g., Ghosh and Ramamoorthi, 2003; Hjort, Holmes, Müller and Walker, 2010). ESP inference does not require such extra statistical assumptions because source of randomness in the ESP econometric model is the same as in the corresponding economic model, and the distribution of the solutions to the empirical

[26] Assuming a distribution typically 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 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.

moment conditions is estimated non-parametrically.

6.2. ESP and the foundation of probability

In probability,[27] there is a relative consensus about the rules that should be used to compute new probabilities from already defined probabilities. Following Kolmogorov (1933), the rules are those of mathematical measure theory.[28] However, there is no consensus about the way to construct probabilities from a practical situation and interpret them. In this section, we explain why our inference framework is as compatible with the two main conceptions typically advanced to justify existing classical and Bayesian theories as the latter ones. For brevity, we only focus on these two main conceptions of probability, although there exist a lot of other ones (e.g., de Finetti, 1968).

A frequentist conception of probability is typically advanced to justify existing classical theory. It defines a probability as a limit of a frequency. The probability of an event is the limit of the ratio of the number of occurences of the event over the number of experiments (e.g., von Mises, 1928). According to this view, asymptotic classical theory should induce valid probabilistic statements for tests and confidence intervals based on t-statistics because, if we could draw an infinite number of samples, the limit of the proportion of t-statistics in a set would correspond to the standard Gaussian distribution (modulo approximation error).[29] Similarly, if we could draw an infinite number of samples, the limit of the proportion of solutions to the empirical moment conditions in a set would correspond to the ESP intensity (modulo approximation error). Therefore, our inference framework is as compatible with the frequentist conception of probability as the existing asymptotic classical theory.

A subjective conception of probability is typically advanced to justify Bayesian inference (e.g., pp.74-77 in Berger, 1980). It defines a probability as an individual degree of belief in a proposition. It thus abolishes the distinction between unknown and random, and

[27] In this section, by probability we mean probability and its derivative including "intensity."

[28] There are some variants of the Kolmogorov's axiomatic (e.g., finite additivity by de Finetti,1970; infinite probability by Hartigan, 1983).

[29] Note that standard frequentist conceptions of probability does not justify asymptotic theory. Standard frequentist conceptions of probability involve an infinite number of samples, but they do not necessarily involve samples with an infinite number of observations.

it allows us to treat the true parameter as a random variable and then apply Bayes's theorem. However, this conception does not restrict the source of the belief. Therefore, a degree of belief can also stem from an ESP intensity. As a consequence, our inference framework does not contradict the typical Bayesian conception of probability.

6.3. An interpretation of the ESP approach

Mathematically, the ESP intensity is an approximation of the distribution of the solutions to the empirical moment conditions. The ESP intensity summarizes the uncertainty about the true parameter on condition that the empirical moment conditions are proxies for the moment conditions, or more precisely, on condition that the solutions to the empirical moment conditions are proxies of the solution(s) to the moment conditions, the true parameter. From a mathematical point of view, the ESP intensity is *not* an approximation of the distribution of the true parameter.

However, one can *interpret* the ESP intensity divided by its integral over the parameter space as an approximation of the distribution of the true parameter conditional on the sample size, if one considers that (i) the distinction between random and unknown is irrelevant; (ii) given some evidence, probability should express in the language of mathematical measure theory to which extent a proposition is possible with respect to alternative propositions; (iii) the evidence for $\theta \in \mathbf{\Theta}$ being the solution to the moment conditions corresponds to the ESP probability weight that θ is a solution to the empirical moment conditions.[30] From this point of view, the ESP approach offers a way to obtain an output similar to standard Bayesian inference without assuming a prior over the parameter space and a parametric family of distributions for the data. However, this interpretation does not erase fundamental differences between the ESP and Bayesian approach. In particular, this interpretation of the ESP approach assimilates the unknown θ_0 to a probabilized uncertainty through the ESP

[30]This way of interpreting the ESP intensity was inspired by Maher (2010), who criticizes the subjective conception of probability typically advanced to justify Bayesian inference (See section 6.2) . However, similar interpretations have been present in the literature for some time. For example, Gallant and Hong (2007) use the distribution of point estimators to form a distribution of the true parameter that they use similarly to a prior. Moreover, this kind of interpretation might also have been one of the ideas behind fiducial inference (Fisher, 1956).

intensity that is induced by the sample at hand. In contrast, the Bayesian theory assimilates the unknown true parameter, θ_0, to a probabilized uncertainty through an *exogenous* prior $\pi_{\theta_0}(.)$.[31]

Although the ESP approach would remain coherent if the ESP intensity was interpreted as an approximation of the distribution of the true parameter conditional on the sample size, we stick to the mathematical definition of the ESP intensity in this paper for three reasons. First, our point of view in this paper has the advantage to make transparent the underlying mechanism of moment-based inference procedures. Moment-based inference is necessarily based on a finite-sample counterpart of the true parameter that serves as a proxy for the true parameter. Thus, the best we can realistically hope for is an accurate knowledge of this proxy. Second, *disentangling* the true parameter from its proxy removes the typical dichotomy between the nature of the true parameter in the economic or finance model of interest and its nature in Bayesian inference. While the true parameter of an economic or finance model is typically considered constant, Bayesian inference leads to regard the true parameter as a non-degenerate random variable.[32] Third, our point of view does not forbid a user of the ESP approach to *subsequently* and *explicitly* interpret the normalized ESP intensity as an approximation of the distribution of the true parameter conditional on the sample size for the reasons indicated in the previous paragraph.

7. A decision-theoretic approach

In this section, we briefly introduce a decision-theoretic approach within the inference framework of the previous sections. In other words, we regard inference as a choice of parameter values by an econometrician in the spirit of microeconomic theory under (probabilized) uncertainty. The econometrician chooses a utility function (i.e., opposite of a loss

[31]To put it differently, Bayesian inference is a high-level way to reduce and even suppress the gap between econometric asymptotic theory and practice by requiring stronger assumptions, although Fisher rejected it (e.g., p.17 in Fisher, 1956).

[32] This dichotomy is not limited to economics and finance. For example, in order to use Bayesian "tools," physicists typically need to assimilate an "unknown parameter" that "cannot be perceived as resulting from a random experiment" to a "random parameter" with a distribution that "summarizes the available information" (pp.10, 508-510 in Robert, 1994). Moreover, this "dichotomy" looks like schizophrenia from the point of view of the subjective foundations of Bayesian inference theory.

function), $u : (d_e, \theta_T^*) \mapsto u(d_e, \theta)$ where d_e is an inference decision and where $\theta \in \mathbf{\Theta}$ is a potential value of the solution to the empirical moment conditions. The inference decision is typically a parameter value (point estimation) or a subset of the parameter space (hypothesis testing). The utility function indicates the utility provided by decision d_e to the econometrician when a solution to the empirical moment condition is θ. The econometrician makes an inference decision, d_e, that maximizes his ESP expected utility function, $\int_{\mathbf{\Theta}} u(d_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta$. ESP expected utility is a generalization of expected utility defined in microeconomic theory, in the sense that utility functions are integrated w.r.t. an intensity measure that is not necessarily a probability measure.[33] Without loss of utility, the econometrician does not randomize his inference decision (mixed strategy). For the same reason as in Bayesian inference (e.g., Theorem 3.12 on p.147 in Schervish, 1995), randomization cannot improve an optimal non-randomized inference decision (pure strategy). A randomized decision is a weighted average of non-randomized decisions; and the average of elements of a set cannot be bigger than the maximum of the set.

A decision-theoretic approach provides several advantages. First, it provides flexibility through the choice of a utility function. Second, it opens a way to move from statistical statements to economic statements thanks to a utility function that maps inference precision to its economic benefit (e.g., Wald, 1939; McCloskey, 1985). Finally, it provides strong finite-sample foundations. Maximization of expected utility is the *optimal* answer to the estimated uncertainty that comes from inference, in the same way as maximization of expected utility by a consumer is optimal in microeconomic theory. In standard classical inference theory, only some asymptotic optimality is typically obtained.

A decision theoretic approach is generally delicate within the existing classical inference theory. Often it is not possible, as in standard moment-based inference, in which case the objective function is not expressed in terms of the dimension of interest, the parameter values. For example, the objective functions of GMM, empirical likelihood (EL) and exponential tilting (ET), are expressed, respectively, in terms of a norm of the empirical

[33] Note that this extension is mathematically straightforward. Normalizing the ESP intensity to make the ESP intensity a density $\frac{\tilde{f}_{\theta_T^*, sp}(\cdot)}{\int_{\mathbf{\Theta}} \tilde{f}_{\theta_T^*, sp}(\theta) d\theta}$ does not affect the definitions below. However, the set of decision-theoretic axioms used should be modified. This is left for future research.

moment conditions, the probability weight of the observed sample, and the informational content (defined as entropy) of the sample. When a decision-theoretic approach is possible, it typically does not produce a complete ranking of inference decisions. Given two decision rules $d_{e_1}(.)$ and $d_{e_2}(.)$, the risk functions $\theta \mapsto \mathbb{E}_\theta\left[u(d_{e_1}(X), \theta)\right]$ and $\theta \mapsto \mathbb{E}_\theta\left[u(d_{e_2}(X), \theta)\right]$ typically cross each other (e.g., pp.194-195 in Wasserman, 2003). In Bayesian theory, integration of the classical risk functions w.r.t. the prior makes a decision-theoretic approach possible. In the ESP approach, integration of the utility function w.r.t. the ESP intensity makes a decision-theoretic approach possible.

The ESP decision-theoretic approach presented in this paper is formally close to the one for Bayesian inference. However, there are fundamental differences. From a Bayesian perspective, an econometrician faces a *known* probabilized uncertainty *of* the true parameter, while from an ESP perspective an econometrician faces an *estimated* probabilized uncertainty *about* the true parameter due to the finiteness of the sample. In other words, the output of the ESP decision-theoretic approach is not claimed to be the optimal answer to the true uncertainty of the true parameter, but only to be the optimal answer to the estimated uncertainty of a finite-sample proxy of the true parameter (see section 6). As a particular consequence, the main decision-theoretic justification of Bayesian procedures is irrelevant to the ESP approach. The Wald's theorem (1950), according to which for a non-Bayesian point estimator there exists a Bayesian point estimator that provides higher expected utility, requires among other conditions the p.d.f. of data conditional on the true parameter to be known (e.g., chap.8 in Robert, 1994). The ESP approach does not require such knowledge.

In this section, for brevity, we only present the elements of the decision-theoretic approach that are used in the empirical section, and we omit the proofs. A more complete version with proofs is in the supplemental material.

7.1. Point estimator

Similarly to Bayesian decision-theory, a maximizer of an ESP expected 0-1 utility function is the mode of the ESP intensity (see supplemental material). For brevity, we use this characterization as a definition in the main body of this paper.

Definition 7.1 (Maximum-ESP point estimator)**.** *A maximum-ESP point estimator, $\hat{\theta}_T$, is a $\mathcal{E}/\mathcal{B}(\mathbf{\Theta})$-measurable maximizer of the ESP intensity ,i.e.,*

$$\hat{\theta}_T := \arg\max_{\theta_e \in \mathbf{\Theta}} \tilde{f}_{\theta_T^*,sp}(\theta_e).$$

Definition 7.1 provides an alternative interpretation of maximizers of an ESP expected 0-1 utility function. A maximum-ESP point estimator is the parameter value with the highest estimated probability weight of being a solution to the empirical moment conditions. In this sense, it is a maximum-probability estimator. A maximum-probability estimator is different from a maximum-likelihood estimator (see footnote 11 on p. 14).

Definition 7.1 also shows that our maximum-ESP point estimator corresponds to the point estimator introduced in Sowell (2009) to correct the higher-order bias of exponential tilting estimators (ET). Sowell (2009) shows that the logarithm of ESP intensity divided by the sample size corresponds to the exponential tilting objective function plus two terms that vanish asymptotically. He deduces that maximum-ESP estimators share the same first-order asymptotic properties as ET estimators, but are higher-order bias corrected thanks to the extra two terms of the objective function.

In accordance with Sowell (2009), the following proposition states that maximum-ESP estimators exist, and are consistent.

Proposition 7.1. *Under Assumptions 1(a)-(c) and 2,*

i) there exists a maximum-ESP point estimator $\hat{\theta}_T$;

ii) under the additional Assumptions 4, 5, 6(a)(c)(d)(e) and 7, a maximum-ESP point estimator converges $\mathbb{P}$-a.s. to the true parameter, i.e.,

$$\lim_{T\to\infty} \hat{\theta}_T = \theta_0 \quad \mathbb{P}\text{-a.s.}$$

The first part of the proposition follows from Schmetterer-Jennrich's measurability result and the continuity of the ESP intensity. We deduce the second part of the proposition from the consistency of the ESP intensity, unlike Sowell (2009) who deduces it from the consistency of ET estimators.

7.2. Confidence regions

As for continuous utility functions, we define confidence regions.

Definition 7.2 (Maximum-ESP confidence region)**.** *A maximum-ESP confidence region of level $1-\alpha$ with $\alpha \in [0,1]$ is a $\mathcal{B}(\mathbf{\Theta})$-measurable set*

$$\tilde{S}_T := \left\{ \theta_e \in \mathbf{\Theta} : \frac{1}{K_T} \tilde{f}_{\theta_T^*,sp}(\theta_e) \geqslant k_{\alpha,T} \right\}$$

where $k_{\alpha,T}$ is the highest bound satisfying $\frac{1}{K_T} \int_{\tilde{S}_T} \tilde{f}_{\theta_T^*,sp}(\theta) d\theta \geqslant 1-\alpha$ and $K_T := \int_{\mathbf{\Theta}} \tilde{f}_{\theta_T^*,sp}(\theta) d\theta$.

A maximum-ESP confidence region is also by construction an indicator of the reliability of a point estimate. Nevertheless, it also has an interpretation on its own. By Proposition 4.6, all elements in the maximum-ESP confidence region have a higher probability weight of being a solution to the empirical moment conditions than the ones outside it. In this sense, it is a maximum-probability based confidence region. We do not require the ESP confidence region level to be exactly equal to $1-\alpha$ in order to ensure existence. As for continuous utility functions, marginal, conditional, two-sided symmetric confidence regions and two-sided equal-tailed ESP confidence regions can also be defined.

To the author's knowledge, Sowell (2007) is the only one to use this type of confidence region in the saddlepoint literature. The main differences between the confidence regions in Sowell (2007) and the ones in the definition just above are the following. Sowell (2007) uses the ESP technique to approximate the distribution of the local minima of the second step GMM objective function, while we use it to approximate the distribution of the solutions to the empirical moment conditions. He proposes to use the obtained confidence region for the GMM estimate, while we use it for the maximum-ESP estimate.

The Bayesian counterpart of maximum-ESP confidence regions are typically called "highest posterior density" (HPD) regions (e.g., p.327 in Schervish, 1995). However, they are different. From a Bayesian perspective, an HPD region of level α is the smallest set that contains the true parameter with probability at least $1-\alpha$. In the most comparable case in which there can be only one solution to the empirical moment conditions, a maximum-ESP confidence region is the smallest closed set that contains the finite-sample proxy of the true parameter (the solution to the empirical moment condition) with estimated probability at least $1-\alpha$.

The following proposition ensures existence and consistency of maximum-ESP confidence regions.

Proposition 7.2. *For all $\alpha \in]0,1[$, if $\hat{\boldsymbol{\Theta}}_T^{-\eta} \neq \emptyset$, under Assumptions 1(a)-(c) and 2,*

i) there exists a maximum-ESP confidence region, $\tilde{S}_T$, which is a non-empty compact subset of $\mathbf{R}^m$;

ii) under the additional Assumptions 4, 5, 6(a)(c)(d)(e) and 7, for all $\theta_e \in \boldsymbol{\Theta} \setminus \{\theta_0\}$, for T big enough, $\theta_e \notin \tilde{S}_T$.

The above proposition does not correspond to the Neyman-Pearson criterium of validity of confidence regions, (asymptotic) coverage. Nevertheless, Theorem 5.2 suggests that asymptotic coverage might also be proved. For example, adaptation of the arguments from Chernozhukov and Hong (2003) provides asymptotic coverage for equal-tailed maximum-ESP confidence region.

However, we do not follow this direction as coverage has a limited interest in practice and theory. First, the frequentist conception of probability advanced by the advocates of the Neyman-Pearson theory does not particularly justify the property of asymptotic coverage (see section 6.2). Second, the property of coverage, which defines a confidence region à la Neyman-Pearson, is often misleading. This definition hides the fact that classical moment-based inference necessarily relies on the use of a finite-sample counterpart of the true parameter as a proxy for the latter one. The construction of a confidence interval à la Neyman-Pearson itself typically presupposes that the finite-sample counterpart of a true parameter is the best guess for the true parameter. For example, a standard GMM confidence interval is centered at the GMM point estimate. Third, confidence intervals that satisfies the Neyman-Pearson criterium are typically not used because of this property. While coverage is not about estimates, confidence intervals are typically used in econometrics as an indicator of the *confidence* we can have in an estimate.[34] Fourth, if coverage was the reason why practitioners use confidence intervals, it would be of limited interest. A confidence region with asymptotic coverage $1 - \alpha$ is a random set that has probability $1 - \alpha$ (modulo approximation error) to contain the true parameter before examination of the dataset. But, once a confidence region à la Neyman-Pearson is computed, it contains the true parameter with probability 0 or 1, i.e., it contains or does not contain the true parameter. Thus, coverage is rather a requirement about the expected performance of the procedure used to compute confidence intervals than a requirement about the actual confidence interval from which practical conclusions are drawn.[35]

A fifth and more fundamental concern, which also applies to tests à la Neyman-Pearson, is the general theoretical invalidation of reported coverage of confidence regions (modulo approximation error) by prior knowledge about the data set at use. When prior knowledge is not probabilistically independent from the data set in use, it removes parts of its randomness so that the probability of a usual confidence region to contain the true parameter is not $1-\alpha$. Thus, the computation of every new confidence region typically requires a completely new data set. This is a serious challenge to the use of the Neyman-Pearson theory in finance and economics, as both fields are essentially non-experimental fields.

Although Bayesian inference should, in theory, be immune to this last concern through the incorporation of any prior information into the prior distribution, in fact, the situation is as acute as for classical inference à la Neyman-Pearson. Bayesian inference theory does not consider that an econometrician assumes a prior distribution, but that the econometrician knows it *exactly*. Bayesian theory regards inference as a game between nature and an econometrician, in which the prior distribution is *common knowledge* (see section 6.1). However, econometricians are not born with a general Bayesian model of the world that they reduce through Bayesian updating along their life, and elicitation of a prior can at most result in

[34]For example, the wikipedia page about “confidence interval”says in its first sentence that it “is used to indicate the reliability of an estimate” (wikipedia website, February 6th 2013). This fact has also been acknowledged by theoreticians (e.g., p.136 in Hansen, 2013; Gleser and Hwang, 1987), although it is not in line with the existing classical definition.

[35]This point about confidence intervals à la Neyman-Pearson has long been made by advocates of Bayesian inference (e.g., sec. 2 chap. 17 in Savage, 1954; pp.1-2 in Sims, 2007a).

an approximation. Nobody can elicit the *exact* probability of an event among the *infinite* possibilities offered by the all the real numbers between 0 and 1, especially when an *infinite* number of events should be considered. Thus, the assumption of exact econometrician's knowledge about the prior distribution, which incorporates all prior information, is a challenge to the theoretical justification of Bayesian inference.[36] Savage, one of the founders of modern Bayesian theory, acknowledged it when he developed his "personalistic" point of view on Bayesian inference (pp.59-60 in Savage,1954).

In contrast, the ESP approach is by construction immune to theoretical invalidations induced by prior knowledge. Unlike Bayesian inference, the ESP approach does not require a prior distribution that corresponds to prior information. The ESP intensity arises endogenously from the data through the ESP technique. Morever, the ESP approach relies on probabilistic statements that are valid *before* and *after* examination of a data set. Probabilistic statements about solutions to empirical moment conditions are valid before and after examination of a data set because they are about all the solutions to the empirical moment conditions that could have been observed and their probability weights of occurring.

Despite all these differences between classical confidence regions à la Neyman-Pearson, Bayesian HPD region and maximum-ESP confidence regions, under standard assumptions, CLT, Laplace-Bernstein-von Mises' theorem and Theorem 5.2 indicate that they behave similarly asymptotically.

8. Empirical evidence from asset pricing

8.1. Setup

In this section, we present empirical evidence from consumption-based asset pricing using the ESP approach and the main existing moment-based methods. This section complements section 2. For simplicity and clarity, we only estimate the relative risk aversion of the representative agent. We rely on the following key moment condition of consumption-

[36]The impossibility for a Bayesian to regard his model as an approximation, or more precisely his impossibility to look at his model from "outside," also generates practical challenges such as the difficulty to develop Bayesian tests of goodness-of-fit.

based asset pricing theory

$$\mathbb{E}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_t^m - R_t^f)\right] = 0 \tag{15}$$

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 (15) 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). It corresponds to standard US data at yearly frequency. The first sample, which is from Shiller's website, spans from 1890 to 2009. The second sample spans from 1930 to 2009. See Julliard and Ghosh (2012) for a more detailed data description.

We estimate the relative risk aversion using GMM (Hansen, 1982), continuously updated (CU) GMM (Hansen, Heaton and Yaron, 1996), which is an example of generalized empirical likelihood estimators (GEL), CU GMM for lack of identification (Stock and Wright, 2000), and the ESP approach with a 0-1 utility function. 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 (15) corresponds to a martingale difference) and empirically (e.g., pp. 86-87 in Hall, 2005).

8.2. Empirical evidence

Inference results are similar for both samples. Tables 8.1 and 8.3 show the results from CU GMM. The point estimate and the standard confidence region based are almost identical to the GMM ones.[37] However, on Table 8.1(C) and 8.3(C), the confidence region for weak identification, the S-sets, are quite different. By definition, in our case, an S-set is

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

[37] 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, while the ESP point estimate is generically different even if a 0-1 utility function is chosen. The ESP intensity incorporates additional information through the estimated variance of the solution to the tilted empirical moment conditions.

Table 8.1: **Continuously updated (CU) GMM inference (1890-2009)**

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

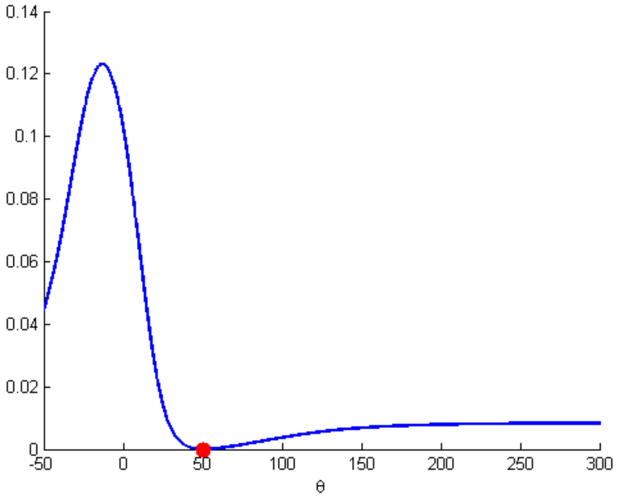
$R_t^m :=$ gross market return, R_t^f :=risk-free asset gross return, $C_t :=$ consumption,

θ :=relative risk aversion,

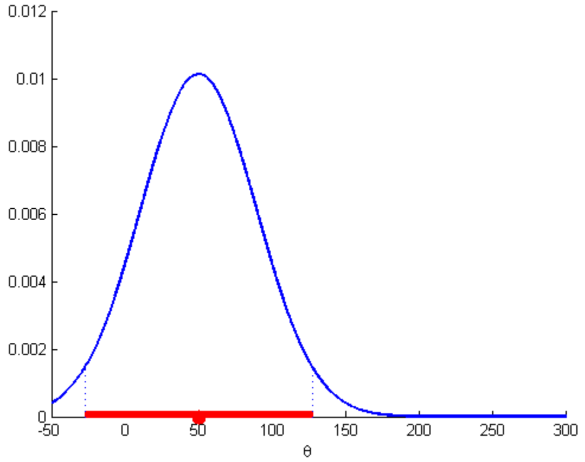
$\hat{\theta}_{cu} = 50.3$ (bullet), $\hat{I}^{cu}_{.05} = [-26.9, 127.4]$ (stripe in B),

$\hat{I}^{S}_{.05} = [18.2, \mathit{3890}]$ (lower bound in italic; stripe in C)

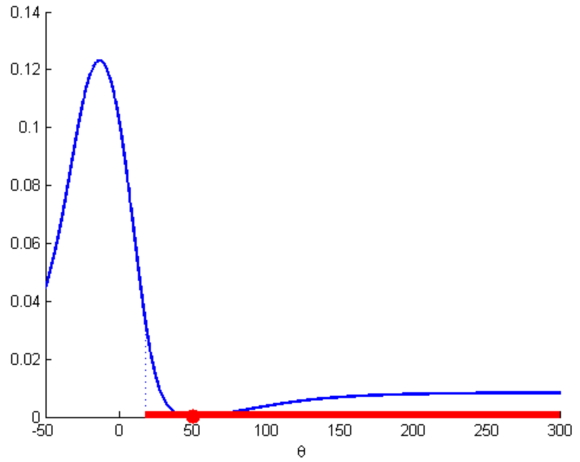
Rk: We constrain the numerical search for point estimate to discard large values of θ.



(A) Objective function and point estimate.



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

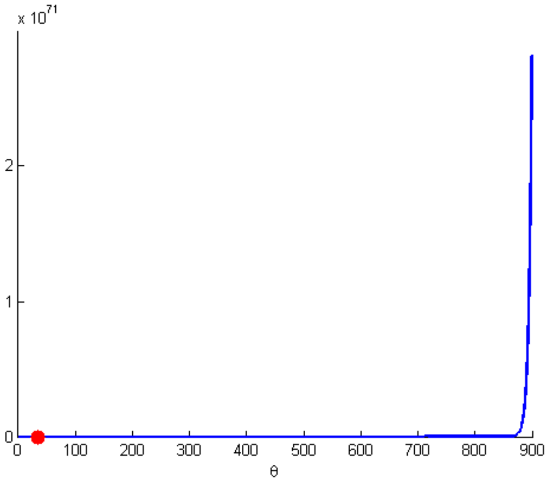


(C) Right-side truncated S-set.

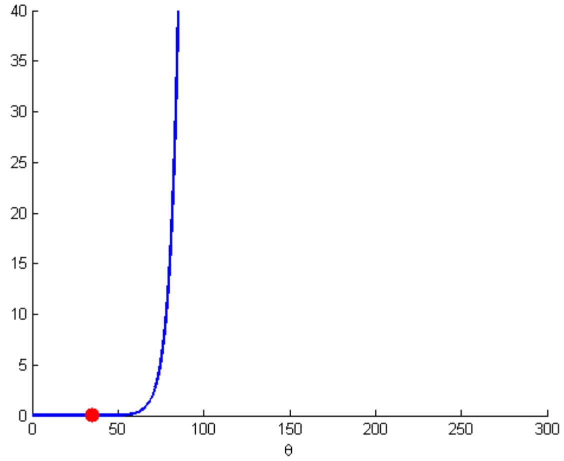
Table 8.2: **GMM inference (1930-2009)**

$$\frac{1}{2009-1930}\sum_{t=1930}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_t^m - R_t^f)\right] = 0,$$

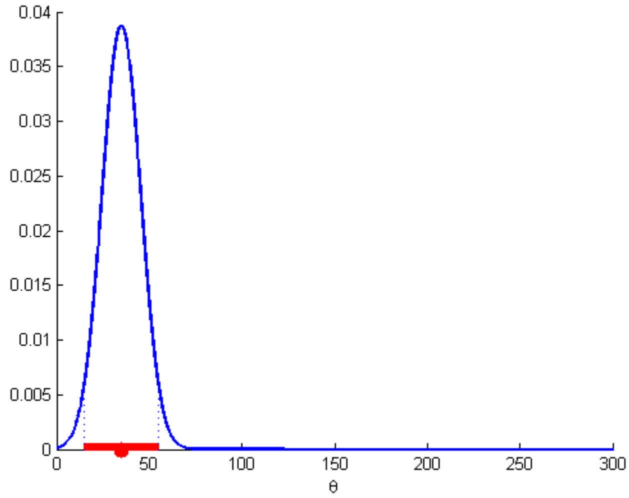
$R_t^m :=$ gross market return, R_t^f :=risk-free asset gross return, $C_t :=$ consumption,
θ :=relative risk aversion,
$\hat{\theta}_{gmm} = 35.0$ (bullet), $\hat{I}_{.05} = [14.8, 55.1]$ (stripe)



(A) Objective function and point estimate.



(A zoom) Objective function and point estimate.



(C) Gaussian distribution, point estimate and 5% confidence interval.

Table 8.3: **Continuously updated (CU) GMM inference (1930-2009)**

$\frac{1}{2009-1930}\sum_{t=1930}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_t^m - R_t^f)\right] = 0,$

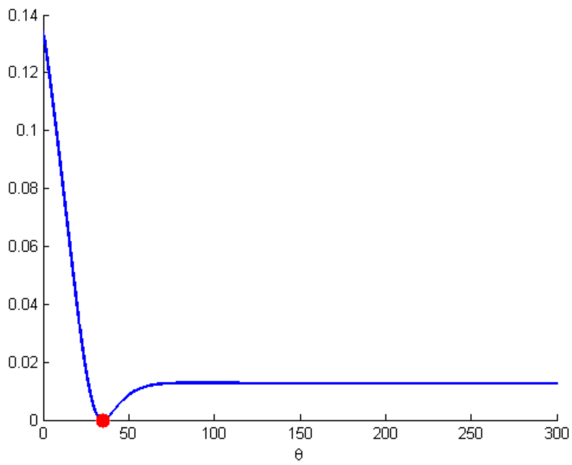
$R_t^m :=$ gross market return, R_t^f :=risk-free asset gross return, $C_t :=$ consumption,

θ :=relative risk aversion,

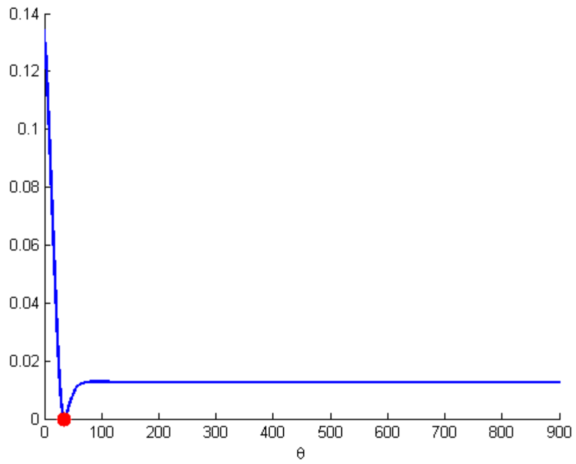
$\hat{\theta}_{cu} = 35.0$ (bullet), $\hat{I}_{.05}^{cu} = [14.8, 55.1]$ (stripe in B),

$\hat{I}_{.05}^{S} = [17.9, \mathit{3616}]$ (lower bound in italic; stripe in C),

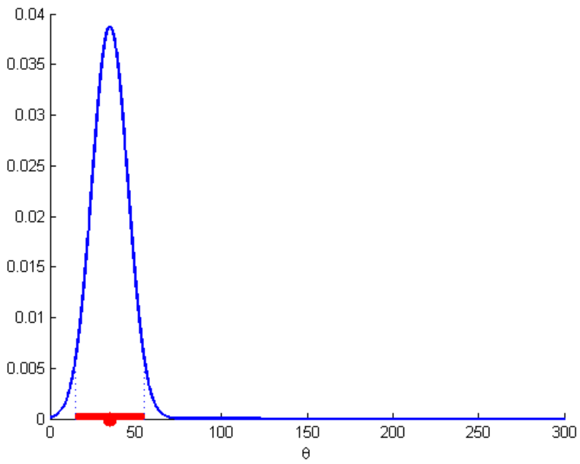
Rk: We constrain the numerical search for point estimate to discard large values of θ.



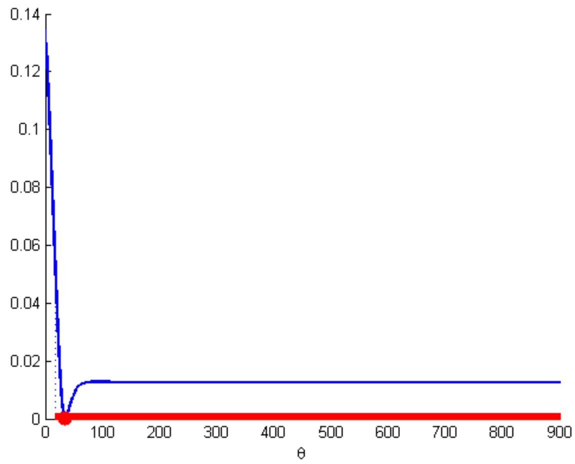
(A zoom) Objective function and point estimate.



(A) Objective function and point estimate.



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



(C) Right-side truncated S-set.

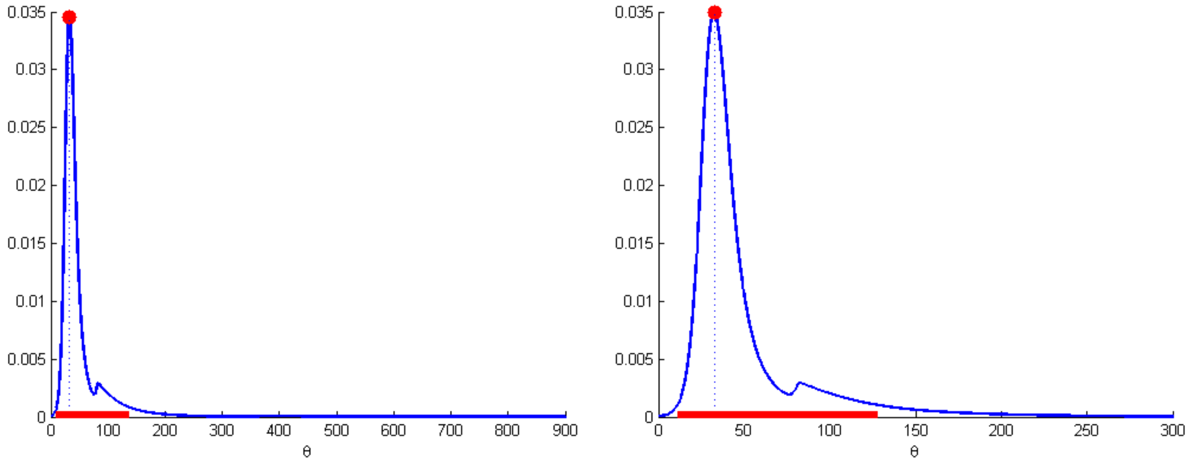
Table 8.4: **ESP inference with 0-1 utility (1930-2009)**

$\frac{1}{2009-1930}\sum_{t=1930}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_t^m - R_t^f)\right] = 0,$

$R_t^m :=$ gross market return, R_t^f :=risk-free asset gross return, $C_t :=$ consumption,

θ :=relative risk aversion,

$\hat{\theta}_T = 32.5,\ \hat{I}_{.05} = [11.1, 127.8]$, ESP intensity support= $[0, 813.3]$.



(A)(A zoom) ESP intensity, point estimate and confidence interval.

where c_α is the α quantile of a chi-square of degree one and $Q_{T,CU}(.)$ is the CU GMM objective function, i.e., $Q_{T,CU}(\theta) := \left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right]'\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right]$. 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 on 8.1(C) and 9.2(C) so that it is not very informative. In the less favourable case with over-restricting moment conditions, S-sets are generally empty (e.g., Stock and Wright, 2000) so that they are not very informative either.

Table 2.2 on p.13 and Table 8.4 display the results for the ESP approach with a 0-1 utility function. In both cases, the ESP intensity 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 as the ESP intensity is not flat around its mode. Besides, both samples yield almost the same point estimate. At the same time, ESP results indicate that consumption-based asset pricing theory is more consistent with data than other inference approaches suggest. First, in line with finance theory, negative values for the RRA have almost no estimated probability weight, while confidence interval from other approaches often include negative values (e.g., Table 2.1 on p.9; 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 4.4 on p.28 indicates if the moment condition was inconsistent with data, the ESP intensity would be zero everywhere. These findings are all the more encouraging for consumption-based asset pricing theory because the moment condition (15) do not resort to Epstein-Zin-Weil preferences (Epstein and Zin, 1989) or other advanced preferences, which yield more flexible stochastic discount factors.

9. Conclusion

Several areas such as empirical consumption-based asset pricing have been a challenge to existing inference approaches. This paper proposes the ESP approach to tackle this challenge.

The starting point of the ESP framework is the acknowledgement that inference practice relies on samples with a bounded number of observations. More precisely, the starting point is the acknowledgement that moment-based inference is based on the use of a finite-sample counterpart of the true parameter as a proxy for the latter one. Then, the idea of the ESP approach is to approximate the distribution of the finite-sample counterpart of the true parameter thanks to the saddlepoint technique. The result of this approximation, the ESP intensity, summarizes in probabilistic terms the estimated uncertainty about the true parameter due to the finiteness of the sample. Thus, an econometrician can choose a utility function (or, equivalently, a loss function) according to the inference purpose, and make inference decisions that maximize the ESP expected utility.

The ESP approach combines strengths of the Bayesian and standard classical approaches. The ESP framework is the result of a search for stronger finite-sample foundations for inference. The largely exogenous Gaussian template, which is typically used to summarize information from data, is replaced by an ESP intensity, which relies more on the information from data and less on asymptotic structure. Nevertheless, we prove that the ESP framework enjoys good asymptotic properties. In addition, we show the inherent robustness of the

ESP approach to lack of identification. We also explain why the ESP approach provides a unique answer to multiple concerns that are faced by existing classical and Bayesian inference, such as multiple hypothesis testing. Empirical evidence from consumption-based asset pricing confirm the practical relevance of the theoretical strength of the ESP approach. At the same time, the ESP approach sheds some light in consumption-based asset pricing. While the ESP approach explains the difficulties faced by other inference approaches, it suggests that the key equilibrium implication of consumption-based asset pricing theory is more consistent with data than other approaches indicate.

All this contributes to the literature in several directions. Overall, the ESP approach appears to be a more accurate cannon with stronger foundations, to continue Fisher's metaphor.

References

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

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

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, *Quaterly Journal of Economics* **121**(3), 823–866.

Berger, J. O.: (1980) 2006, *Statistical decision theory and Bayesian analysis*, Statistics, second edn, Springer.

Bhattacharya, R. N. and Ramaswamy, R. R.: (1976) 1986, *Normal Approximation and Asymptotic Expansions*, second edn, Robert E. Krieger Publishing Company, Inc.

Carrasco, M. and Florens, J.-P.: 2000, Generalization of GMM to a continuum of moment conditions, *Econometric Theory* **16**(6), 797–834.

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.

Chernoff, H.: 1952, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *The Annals of Mathematical Statistics* **23**(4), 493–507.

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.

de Finetti, B.: 1937, La prévision: ses lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincaré* **7**(1), 1–68.

Translated in English in "Studies in Subjective Probability" by Henry E. Kyrburgh and Howard E. Smokler, 1964.

de Finetti, B.: 1968, *International Encyclopedia of the Social Sciences*, New York Macmillan, chapter "Probability: Interpretations", pp. 496–505.

de Finetti, B.: (1970) 1990, *Theory of Probability. A critical introductory treatment*, Wiley Classics, John Wiley and Sons.

Translated by Antonio Machi and Adrian Smith.

DeGroot, M. H.: 2004 (1970), *Optimal Statistical Decisions*, Wiley classics library, John Wiley & Sons.

Domínguez, M. A. and Lobato, I. N.: 2004, Consistent estimation of models defined by conditional moment restrictions, *Econometrica* **72**(5), 1601–1615.

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

Dufour, J.-M.: 1997, Some impossibility theorems in econometrics with applications to structural and dynamic models, *Econometrica* **65**(6), 1365–1387.

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

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

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.

Fisher, R. A.: 1925, *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd.

Fisher, R. A.: 1973 (1956), *Statistical Methods and Scientific Inference*, third edn, Hafner Press.

Reprinted in "Statistical Methods Experimental Design and Scientific Inference" by John Henry Bennett.

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

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

Gagliardini, P., Gouriéroux, C. and Renault, E.: 2011, Efficient derivative pricing by the extended method of moments, *Econometrica* **79**(4), 11811232.

Gallant, R. A. and Hong, H.: 2007, A statistical inquiry into the plausibility of recursive utility, *Journal of Financial Econometrics* **5**, 523–559.

Ghosh, A., Julliard, C. and Taylor, A. P.: 2010, What is the consumption-CAPM missing? An information-theoretic framework for the analysis of asset pricing models, *working paper* .

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

Gleser, L. J. and Hwang, J. T.: 1987, The nonexistence of 100(1-α)% confidence sets of finite expected diameter in errors-in-variables and related models, *The Annals of Statistics* **15**(4), 1351–1362.

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

Goutis, C. and Casella, G.: 1999, Explaining the saddlepoint approximation, *The American Statistician* **53**(3), 216–214.

Guggenberger, P. and Smith, R. J.: 2005, Generalized empirical likelihood estimators and tests under partial, weak, and strong identification, *Econometric Theory* **21**, 667–709.

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

Hansen, B. E.: January 18th 2013, Econometrics.
Available on the author’s webpage.

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

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.

Hartigan, J. A.: 1983, *Bayes Theory*, Series in Statistics, Springer.

Hazewinkel, M. (ed.): 2002, *The Online Encyclopaedia of Mathematics*, Springer.
available at http://eom.springer.de.

Hiriart-Urruty, J.-B. and Lemaréchal, C.: (1993) 1996, *Convex Analysis and Minimization Algorithms 1. Fundamentals*, number 305 in *Comprehensive Studies in Mathematics*, Springer.

Hirirart-Urruty, J.-B. and Lemaréchal, C.: (1993) 2010, *Convex Analysis and Minimization Algorithms 2. Advanced Theory and Bundle Methods*, number 306 in *Comprehensive Studies in Mathematics*, Springer.

Hjort, N. L., Holmes, C., Mller, P. and Walker, S. G. (eds): 2010, *Bayesian Nonparametrics*, Cambridge University Press.

Hoeffding, W.: 1965, Asymptotically optimal tests for multinomial distributions, *The Annals of Mathematical Statistics* **36**(2), 369–401.

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.

Itô, K.: 1970, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3, University of California Press, chapter "Poisson Point Processes Attached to Markov Processes", pp. 225–239.

Jagannathan, R., Skoulakis, G. and Wang, Z.: 2002, Generalized method of moments: Applications in finance, *Journal of Business and Economics Statistics* **20**(4), 470–481.

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

Jensen, J. L.: 1992, The modified signed likelihood statistic and saddlepoint approximations, *Biometrika* **79**(4), 693–703.

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.: (2007) 2012, Can rare events explain the equity premium puzzle?, *Forthcoming in Review of Financial Studies* .

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

Karatzas, I. and Shreve, S. E.: (1988) 2005, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, second edn, Springer.

Corrected eighth printing.

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.

Kitamura, Y., Santos, A. and Shaikh, A. M.: 2012, On the asymptotic optimality of empirical likelihood for testing moment restrictions, *Econometrica* **80**(1), 413–423.

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

Kleibergen, F.: 2005, Testing parameters in GMM without assuming that they are identified, *Econometrica* **73**(4), 1103–1123.

Kocherlakota, N. R.: 1990a, On tests of representative consumer asset pricing models, *Journal of Monetary Economics* **26**(2), 286–304.

Kocherlakota, N. R.: 1990b, On the discount factor in growth economies, *Journal of Monetary Economics* **25**(1), 43–47.

Kolmogorov, A. N.: (1933) 1936, *Osnovnie poniatya teorii veroyatnostei*, ONTI NKTP SSSR.

Translated in English by Nathan Morrison under the title “Foundations of probability”.

Komunjer, I.: 2011, Global identification in nonlinear models with moment restrictions, *Econometric Theory (forthcoming)* .

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

Laplace, P.-S.: (1774) 1878, *Œuvres complètes de Laplace*, Vol. 8, Gauthier-Villars, chapter "Memoire sur la probabilité des causes par les événements", pp. 31–71.

Leamer, E. E.: 1978, *Specification searches. Ad Hoc Inference with Nonexperimental Data*, John Wiley & Sons.

Available on the homepage of the author.

Leamer, E. E.: 1983, Let's take the con out of econometrics, *The American Economic Review* **73**(1), 31–43.

LeCam, L.: 1953, On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates, *University of California Publications in Statistics* **1**(11), 277–330.

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

Lo, A. and MacKinley, C.: 1990, Data snooping biases in tests of financial asset pricing, *Review of Financial Studies* **3**(3), 431–468.

Lucas, R. E.: 1976, Econometric policy evaluation: A critique, *Carnegie Rochester Conference Series on Public Policy* **1**, 19–46.

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

Ma, Y. and Ronchetti, E.: 2011, Saddlepoint test in measurement error models, *Journal of the American Statistical Association* **106**(493), 147–156.

Magnac, T. and Thesmar, D.: 2002, Identifying dynamic discrete decision processes, *Econometrica* **70**(2), 801–816.

Maher, P.: 2010, Bayesian probability, *Synthese* **172**(1), 119–127.

Mas-Colell, A., Whinston, M. D. and Green, J. R.: 1995, *Microeconomic Theory*, Oxford University Press.

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

Mavroeidis, S.: 2005, Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve, *Journal of Money, Credit and Banking* **37**(3), 421–448.

McCloskey, D. N.: 1985, The loss function has been mislaid: The rhetoric of significance tests, *The American Economic Review, Papers and Proceedings* **75**(2), 201–205.

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

Neely, C. J., Roy, A. and Whiteman, C.: 2001, Risk aversion versus intertemporal substitution: A case study of identification failure in the intertemporal consumption CAPM, *Journal of Business and Economic Statistics* **19**(4), 395–403.

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.

Newey, W. K. and Smith, R. J.: 2004, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica* **72**, 219–255.

Neyman, J.: 1977, Frequentist probability and frequentist statistics, *Synthese* **36**(1), 97–131.

Otsu, T.: 2006, Generalized empirical likelihood inference for nonlinear and time series models under weak identification, *Econometric Theory,* **22**, 513–527.

Oxford English Dictionary: (1928) 2009, CD-ROM (v.4.0), second edn, Oxford University Press.

Parker, J. A. and Julliard, C.: 2005, Consumption risk and the cross section of expected returns, *Journal of Political Economy* **113**(1), 185–222.

Pesaran, M. H.: 1981, Identification of rational expectations models, *Journal of Econometrics* **16**(3), 375–398.

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* .

Robert, C. P.: (1994) 2007, *The Bayesian Choice, From Decision-Theoretic Foundations to Computational Implementation*, Texts in statistics, second edn, Springer.

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.

Romano, J. P., Shaikh, A. M. and Wolf, M.: 2008, Formalized data snooping based on generalized error rates, *Econometric Theory* **24**, 404–447.

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.

Rothenberg, T. J.: 1971, Identification in parametric models, *Econometrica* **39**(3), 577–591.

Savage, L. J.: (1954) 1972, *The Foundations of Statistics*, Dover.

Savin, N. E.: 1980, The Bonferroni and the Scheffé multiple comparison procedures, *Review of Economic Studies* **47**(1), 255–273.

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.

Schwartz, L.: (1950-1951) 2010, *Théorie des distributions*, third edn, Hermann.

Schwartz, L.: (1955) 1998, *Méthodes mathématiques pour les sciences physiques*, Hermann.

Ecrit avec le concours de Denise Huet. Published in English under the title "Mathematics For the Physical Sciences".

Seidenfeld, T.: 1992, R.A. Fisher's fiducial argument and Bayes' theorem, *Statistical Science* **7**(3), 358–368.

Sims, C. A.: 2007a, Bayesian methods in applied econometrics, or, why econometrics should always and everywhere be Bayesian, *Discussion paper* .

Sims, C. A.: 2007b, Thinking about instrumental variables, *Discussion paper* .

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.

Sowell, F.: 1996, Optimal tests of parameter variation in the generalized method of moments framework, *Econometrica* **64**(5), 1085–1108.

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.

Sullivan, R., Timmermann, A. and White, H.: 1999, Data-snooping, technical trading rule performance, and the bootstrap, *The Journal of Finance* **54**(5), 1647–1691.

Sullivan, R., Timmermann, A. and White, H.: 2001, Dangers of data mining: The case of calendar effects in stock returns, *Journal of Econometrics* **105**, 249–286.

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.: 1998, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Villanacci, A., Carosi, L., Benevieri, P. and Battinelli, A.: 2002, *Differential Topology and General Equilibrium with Complete and Incomplete Markets*, Kluwer Academic Publishers.

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

von Mises, R.: (1928) 1957, *Probability, Statistics and Truth*, Dover.

Wald, A.: 1939, Contributions to the theory of statistical estimation and testing hypotheses, *Annals of Mathematical Statistics* **10**(4), 299–326.

Wald, A.: 1949, Note on the consistency of the maximum likelihood estimate, *Annals of Mathematical Statistics* **20**(4), 595–601.

Wald, A.: 1950, *Statistical Decision Functions*, John Wiley & Sons.

Wasserman, L.: 2004, *All of Statistics: A Concise Course in Statistical Inference*, Texts in statistics, Springer.

White, H.: 2000, A reality check for data snooping, *Econometrica* **68**(5), 1097–1126.

A. Supplementary results and proofs

For brevity, this appendix does not contain all the details of the proofs. However, when possible, we have tried to identify and give references in which similar methods of proof are explained in more length. Moreover, detailed proofs are available in the supplemental material.

A.1. Proof of Proposition 4.1

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. Proposition 9.1.VIII in Daley and Vere-Jones, 1988) it is sufficient to prove that there exists a function $\omega \mapsto N_T(\omega, .)$ such that 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 the Lemma A.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, putting $P := \Psi_T^{-1}(\{0\})$ 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 a $\mathcal{E}/\mathcal{B}(\mathbf{N})$-measurable function $\omega \mapsto N_T(\omega, A)$ such that

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

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

Lemma A.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

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

Apply the monotone class Theorem 3.1 in Rogers and Williams (1979) 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 $\gamma_1 \mapsto \int_A g(\gamma_1,\gamma_2)\nu(d\gamma_2)$ is $\mathcal{B}\left(\mathbf{\Gamma}_1\right)/\mathcal{B}(\overline{\mathbf{N}})$-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 $\mathcal{B}\left(\mathbf{\Gamma}_1\right)/\mathcal{B}(\overline{\mathbf{N}})$-measurable as $\nu(P_{\gamma_1} \cap A) = \int_A \mathbf{1}_P(\gamma_1,\gamma_2)\nu(d\gamma_2)$ because $\forall P \in \mathcal{B}(\mathbf{\Gamma}_1) \otimes \mathcal{B}(\mathbf{\Gamma}_2)$, $\mathbf{1}_P(.)$ is $\mathcal{B}(\mathbf{\Gamma}_1) \otimes \mathcal{B}(\mathbf{\Gamma}_2)/\mathcal{B}(\mathbf{R})$-measurable and bounded. □

A.2. Lemma A.2

Lemma A.2. *Under Assumptions 1,*

i) there exists a dissecting systems of $(\mathbf{\Theta}, \mathcal{B}(\mathbf{\Theta}))$ *;*

ii) if $\mathcal{T} := \{\mathcal{T}_n\}_{n\geqslant 1}$ *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 is definition-chasing. *iii)* It is a consequence of Assumption 1(d) and Khinchin's existence Theorem (e.g. Proposition 9.3.IX in Daley and Vere-Jones, 1988). □

A.3. Proof of Proposition 4.2

It is a consequence of equation (9.3.24) on p.48 in Daley and Vere-Jones (1988). □

A.4. Proof of Proposition 4.3

i) Apply implicit function theorem to $\frac{\partial\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\mathrm{e}^{\tau'\psi_t(\theta)}\right]}{\partial\tau} = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)'\mathrm{e}^{\tau'\psi_t(\theta)}$.

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

A.5. Proposition A.1

Proposition A.1. *Under Assumptions 1 and 2, if for all* $\dot{\theta} \in \overline{\mathbf{\Theta}}_T$, *with* $\dot{\tau}$ *s.t.* $\sum_{t=1}^{T}\psi_t(\dot{\theta})\mathrm{e}^{\dot{\tau}\psi_t(\dot{\theta})} = 0_{m\times 1}$, *the rank of the* $m \times 2m$ *matrix*

$$\left.\frac{\partial\left[\sum_{t=1}^{T}\psi_t(\theta)\mathrm{e}^{\tau\psi_t(\theta)}\right]}{\partial(\theta,\tau)'}\right|_{\substack{\tau=\dot{\tau}\\ \theta=\dot{\theta}}} \tag{16}$$

equals m *, then* $\lambda(\overline{\mathbf{\Theta}}_T \setminus \hat{\mathbf{\Theta}}_T) = 0$, *where* $\lambda(.)$ *denotes the Lebesgue measure.*

Proof. Apply transversality theorem (e.g. Theorem 26 p.151 in Villanacci, Carosi, Benevieri and Battinelli, 2002). □

A square matrix is generically non-singular. Here the additional m columns, makes the singularity of the matrix (16) even more difficult.

A.6. Proof of Proposition 4.4

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., Proposition XII.2.4.1(iii) in Hiriart-Urruty and Lemaréchal, 1993).

A.7. Proof of Proposition 4.5

A.7.1 Proof of Proposition 4.5i)a)

The proof relies on multiple applications of Lemma A.1 on p.69, Lemma 2 in Jennrich (1969) and the idea of the proof of the later. It is skipped because of its length. However, as for the other proofs, a detailed version is available in the supplemental material.

A.7.2 Proof of Proposition 4.5i)b)

Continuity and positivity are obtained by construction. □

A.7.3 Proof of Proposition 4.5ii)a)

Continuity and positivity of $\tilde{f}_{\theta^*_T,sp}(.)$ over the compact space $\mathbf{\Theta}$ implies finiteness and positivity of the set function $\tilde{\mathbb{F}}_T(.) := \int_. \tilde{f}_{\theta^*_T,sp}(\theta)d\theta$. Fubini-Tonelli theorem implies σ-additivity of $\tilde{\mathbb{F}}_T(.)$..

A.7.4 Proof of Proposition 4.5ii)b)

For all $\omega \in \mathbf{\Omega}$, $\theta \mapsto \tilde{f}_{\theta^*_T,sp}(\theta)$ is $\mathcal{B}(\mathbf{\Theta})/\mathcal{B}(\mathbf{R})$-measurable and bounded on $\mathbf{\Theta}$ by Proposition 4.5i)a) and b), respectively. Thus, by Lemma A.3 below, $\forall A \in \mathcal{B}(\mathbf{\Theta})$, $\omega \mapsto \int_A \tilde{f}_{\theta^*_T,sp}(\theta)d\theta$ is $\mathcal{E}/\mathcal{B}(\mathbf{R})$-measurable. □

Lemma A.3. *If, for all $\omega \in \mathbf{\Omega}$, a function $g : \mathbf{\Omega} \times \mathbf{\Theta} \to \mathbf{R}$ is $\mathcal{B}(\mathbf{\Theta})/\mathcal{B}(\mathbf{R})$-measurable and bounded on $\mathbf{\Theta}$, then, for all $A \in \mathcal{B}(\mathbf{\Theta})$, $\omega \mapsto \int_A g(\omega,\theta)d\theta$ is $\mathcal{E}/\mathcal{B}(\mathbf{R})$-measurable.*

Proof. Follow the idea of the proof of Lemma A.1 on p.69. □

A.8. Proof of Proposition 4.6 and Lemma A.4

Lemma A.4. *Denote $\mathcal{T} := \{\mathcal{T}_n\}_{n\geqslant 1}$ a dissecting system of $\mathbf{\Theta}$. Let $\tilde{\mathbb{F}}_T(.)$ be a finite positive measure. Under Assumptions 1-2, if there exists a random variable, Y, from the probability space $(\mathbf{\Omega},\mathcal{E},\mathbb{P})$ to $(\mathbf{N},\mathcal{B}(\mathbf{N}))$ with expectation $\mathbb{E}(Y) = \tilde{\mathbb{F}}_T(\mathbf{\Theta})$, then there exists a point* *random-field, $\tilde{N}_T(.)$, and a probability measure, $\tilde{\mathbb{P}}$, such that for all $A \in \mathcal{B}(\boldsymbol{\Theta})$*

$$\tilde{\mathbb{F}}_T(A) := \lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbb{P}\{\tilde{N}_T(A_{n,i}) = 1\} \quad , \tag{17}$$

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

Proof. Follow the idea of the proof of Theorem 4.2 in Itô (1970) by defining the point random-field $\tilde{N}_T(.) := Y\tilde{\mathbb{F}}_T(.)/\tilde{\mathbb{F}}_T(\boldsymbol{\Theta})$.□

The existence of a random variable, Y on the probability space $(\boldsymbol{\Omega}, \mathcal{E}, \mathbb{P})$ with expectation $\mathbb{E}(Y) = \tilde{\mathbb{F}}_T(\boldsymbol{\Theta})$ is a reasonable assumption. For example, the existence of a random variable distributed according to a uniform distribution on the unit interval $[0, 1]$ is a sufficient condition for this assumption (e.g., Lemma 3.22 p.56 in Kallenberg, 1997).

Proof of Proposition 4.6. Apply Lebesgue's differentiation theorem (e.g., Theorem 3.21 in Folland, 1984), and Proposition 4.2 (p.25).□

A.9. Proof of Theorems

A.9.1 Preliminary results

This subsection contains some results needed for Theorems 5.1 and 5.2. Most of them are variants of results already known, but not necessarily easy to find in the literature.

Measurability and convergence results

Lemma A.5. *Let $\{A_T\}_{T\geqslant 1}$ a sequence of square matrices converging to A as $T \to \infty$. Then*

i) if A is an invertible matrix, then there exists $\dot{T} \in \mathbf{N}$ such that $T \geqslant \dot{T}$ 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 (n-d.m), then there exists $\dot{T} \in \mathbf{N}$ such that $T \geqslant \dot{T}$ implies A_T is n-d.m.

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

ii) Note $\max \text{sp} A_T = \max_{z:\|z\|=1} z' A_T z$ where $\text{sp} A_T$ 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 introduce a set of assumptions and new notations to derive generic results which are used several times.

Assumption 11. (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 the measurable space $(\boldsymbol{\Gamma}, \mathcal{B}(\boldsymbol{\Gamma}))$ such that $\boldsymbol{\Gamma} \subset \mathbf{R}^m$ is compact and $\mathcal{B}(\boldsymbol{\Gamma})$ is the Borel σ-algebra.* **; (c)** *Let $h : \mathbf{R}^p \times \boldsymbol{\Gamma} \mapsto \mathbf{R}^q$ with $q \in \mathbf{N}$ be a function such that $\forall x \in \mathbf{R}^p$, $\gamma \mapsto h(x, \gamma)$ is continuous, and $\forall \gamma \in \boldsymbol{\Gamma}$, $x \mapsto h(x, \gamma)$ is $\mathcal{B}(\mathbf{R}^p)/\mathcal{B}(\mathbf{R}^q)$-measurable.*; **(d)** $\mathbb{E}\left[\sup_{\gamma \in \boldsymbol{\Gamma}} \|h(X, \gamma)\|\right] < \infty$. **; (e)** *In the parameter space $\boldsymbol{\Gamma}$, there exists a unique $\gamma_0 \in \text{int}(\boldsymbol{\Gamma})$ such that $\mathbb{E}\left[h(X, \gamma_0)\right] = 0_{m \times 1}$*; **(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 A.2 (Uniform-strong LLN). *Under Assumptions 11(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 $T \to \infty$, i.e., there exists $E \in \mathcal{E}$ such that $\mathbb{P}\left\{E\right\} = 0$ and*

$$\forall \omega \in \boldsymbol{\Omega} \setminus E, \quad \sup_{\gamma \in \boldsymbol{\Gamma}} \left\| \frac{1}{T} \sum_{t=1}^T h(X_t, \gamma) - \mathbb{E}\left[h(X, \gamma)\right] \right\| \to 0 \quad \text{as } T \to \infty. \tag{18}$$

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

Hereafter, we do not mention negligeable sets associated with properties that holds a.s., because they result from the application of a countable number of properties that hold a.s. In particular, measurability of suprema of functions is obtained from .

Proposition A.3 (Existence of solutions to empirical moment conditions). *Under the Assumptions 11(a)-(c)(e)-(h), if*

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

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

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

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 FOC $\left[\frac{1}{T}\sum_{t=1}^{T} \frac{\partial h(X_t,\gamma)'}{\partial \gamma}\right] \left[\frac{1}{T}\sum_{t=1}^{T} h(X_t, \gamma)\right] = 0_{m\times 1}$ with $\left[\frac{1}{T}\sum_{t=1}^{T} \frac{\partial h(X_t,\gamma)'}{\partial \gamma}\right]$ invertible.

ii) It 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 (Lemma 2 in Jennrich, 1969) ensures the measurability of only one of them.

Proposition A.4 (Measurability of solutions to empirical moment conditions)**.** *Under the assumptions of Proposition A.3 and Assumption 11(f), there exists* $\dot{T} \in \mathbf{N}$ *so that* $T \geqslant \dot{T}$ *implies*

i) the number of solutions to the empirical moment conditions is finite;

ii) each of the solutions to the empirical moment conditions are $\mathcal{E}/\mathcal{B}(\mathbf{\Gamma})$*-measurable* $\mathbb{P}$*-a.s., i.e., if* γ_T^* *is such that* $\frac{1}{T}\sum_{t=1}^{T} h(X_t, \gamma_T^*) = 0_{m\times 1}$*, then* $\gamma_T^*(.)$ *is* $\mathcal{E}/\mathcal{B}(\mathbf{\Gamma})$*-measurable* $\mathbb{P}$*-a.s.*

Proof . *i)* It follows from the Proposition A.3ii), and the definition of compactness.

ii) Apply Lemma 9.1.XIII on p.16 in Daley and Vere-Jones (2008). □

The following proposition is a standard result.

Proposition A.5 (Consistency of solutions to empirical moment conditions)**.** *Under the assumptions of Propositions A.4, 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, i.e.,*

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

Proof . It follows from Proposition A.3 and A.4. □

Corollary 1. *Under the Assumptions 1(a)-(c), 2 and 4, Propositions A.3, A.4 and A.5 apply to solutions to the empirical moment conditions*

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

Proof . Check that the assumptions of Proposition A.4 are satisfied. □

Lemma A.6. *Under Assumptions 1(a)-(c), 2,5(a)(b),*

i) for all $\theta \in \hat{\mathbf{\Theta}}_\infty^{-\eta}$, there exists a unique $\tau_\infty(\theta)$ such that $\mathbb{E}\left[\psi(X,\theta)\mathrm{e}^{\tau_\infty(\theta)'\psi(X,\theta)}\right] = 0$

ii) $\tau_\infty : \hat{\mathbf{\Theta}}_\infty^{-\eta} \to \mathbf{R}^m$ is continuous.

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

Laplace's approximation

Laplace's approximation is a well-known method originally presented by Laplace (Laplace, 1774). Here, we adapt the version presented in Chen (1985) and Kass, Tierney and Kadane (1990) for our purpose.[38]

Assumption 12 (Laplace's regularity)**. (a)** *Let $\left\{\dot{\theta}_T\right\}_{T=1}^{\infty}$ with $\dot{\theta}_T \in \mathbf{\Theta}\ \forall T \geqslant 1$ a sequence converging in the interior of $\mathbf{\Theta}$.* **(b)** *Let $\{h_T(.)\}_{T\geqslant 1}$ a sequence of real-valued functions. There exists $r_h > 0$ and $T_h \in \mathbf{N}$ such that*

[38] Kass, Tierney and Kadane (1990) explicit the Laplace's approximation used in Chen (1985). The differences between Kass, Tierney and Kadane's theorem and our propostion 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.

i) $\forall T \geqslant T_h$, $h_T(.) \in C^4\left(B_{r_h}(\dot{\theta}_T)\right)$*;*

ii) there exists $M_h \geqslant 0$ *so that* $\forall T \geqslant T_h$, $\forall k \in [\![1,4]\!]$, $\forall \theta \in B_{r_h}(\dot{\theta}_T)$, $\left\|D^k h_T(\theta)\right\| < M_h$, *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* $\{b_T(.)\}_{T\geqslant 1}$ *a sequence of real-valued functions such that there exists* $r_b > 0$, $M_b \geqslant 0$ *and* $T_b \in \mathbf{N}$ *so that*

i) $\forall T \geqslant T_b$, $b_T(.) \in C^3\left(B_{r_b}(\dot{\theta}_T)\right)$*;*

ii) $\forall T \geqslant T_b$, $\forall k \in [\![1,3]\!]$, $\forall \theta \in B_{r_b}(\dot{\theta}_T)$, $\left\|D^k b_T(\theta)\right\| < M_b$.

Proposition A.6. *Under Assumptions 1(b) and 12, 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*

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

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

Lemma A.7. *Under Assumptions 1(b) and 12,*

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

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

Proposition A.7. *Under Assumptions 1(b) and 12, there exists* T_1 *and* $r > 0$ *such that for all* $T \geqslant T_1$

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

and the RHS and the LHS are well-defined.

***Proof*.** Combine Proposition A.6 and Lemma A.7. □

A.9.2 Proof of Theorems 5.1 and 5.2

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 2 on p.36. The context indicates which object it is about.

Around θ_T^*: application of Laplace's approximation

Proposition A.8. *Under Assumptions 1(a)-(c),2, 4, 5(a), 6(a)(c)-(e), Laplace's approximations corresponding to Propositions A.6 and A.7 can be applied $\mathbb{P}$-a.s. to $\int_{B_r(\theta_T^*)} \tilde{f}_{\theta_T^*,sp}(\theta)d\theta$ with small enough $r > 0$ by putting*

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

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 1. 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 12. Corollary 1 and Lemmas below ensure that it is the case. □

Lemma A.8. *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 $\tau_T(.)$.□

Lemma A.9. *Under the assumptions of Lemma A.8, 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. □

Lemma A.10. *Under the assumptions of Lemma A.8 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.,

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

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

Proof . Differentiate and apply Lemma A.8. □

Outside a neighborhood of θ_T^*

Proposition A.9. *Under the assumptions of Lemma A.6 and A.11 and Assumptions 1(a)-(c),2, 4(a),7(b) for all small enough $r>0$ there exists $\varepsilon>0$ and $\dot{T}\in\mathbf{N}$ such that*

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

Proof. Check assumptions of Proposition A.2 for application to $\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_\infty(\theta)'\psi_t(\theta)}$ in $B_{r_{\dot\theta}}(\dot\theta)$ with $r_{\dot\theta} > 0$ and $\dot\theta \in \hat{\mathbf{\Theta}}_\infty^{-\eta}$. By Proposition A.2, for all $\dot\theta \in \hat{\mathbf{\Theta}}_\infty^{-\eta}$, there exists $r_{\dot\theta} > 0$ such that as $T \to \infty$,

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

Now, because $\hat{\mathbf{\Theta}}_\infty^{-\eta}$ is compact, there exists $\left\{\dot\theta_k\right\}_{k=1}^{K}$ such that $\hat{\mathbf{\Theta}}_\infty^{-\eta} = \bigcup_{k=1}^{K} B_{\dot r_k}(\dot\theta_k)$. Thus, as $T \to \infty$,

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

Thus, for small enough $\varepsilon > 0$, there exists $\dot T$ such that for all $T \geqslant \dot T$

$$\sup_{\theta \in \hat{\mathbf{\Theta}}_\infty^{-\eta}} \left\| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_\infty(\theta)'\psi_t(\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau_\infty(\theta)'\psi(X,\theta)}\right] \right\| < \varepsilon. \tag{19}$$

Moreover, by Lemma A.11, for small enough $\varepsilon > 0$, there exists, $r_3 > 0$ such that $\forall \theta \in \hat{\mathbf{\Theta}}_\infty^{-\eta} \setminus B_{r_3}(\theta_0)$,

$$\mathbb{E}\left[\mathrm{e}^{\tau_\infty(\theta)'\psi(X,\theta)}\right] < 1 - 2\varepsilon \quad , \tag{20}$$

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

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

*(a)*By triangle inequality, $\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_\infty(\theta)'\psi_t(\theta)} \leqslant \left\| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_\infty(\theta)'\psi_t(\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau_\infty(\theta)'\psi(X,\theta)}\right] \right\| +$

$\left\|\mathbb{E}\left[\mathrm{e}^{\tau_\infty(\theta)'\psi(X,\theta)}\right]\right\| \leqslant 1-2\varepsilon+\varepsilon = 1-\varepsilon$ because of inequalities (20) and (19); *(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 $\frac{\partial^2\left\{\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\right\}}{\partial\tau\partial\tau'} = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)'\mathrm{e}^{\tau'\psi_t(\theta)}$. □

Lemma A.11. *Under Assumptions 1(a)-(c), 4(b) and 5, for all $\theta \in \hat{\boldsymbol{\Theta}}_\infty^{-\eta}\setminus\{\theta_0\}$*

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

Proof. By definition of $\hat{\boldsymbol{\Theta}}_\infty^{-\eta}$ and a standard result on Laplace's transform (e.g. Theorem 3 pp.182-183 in Monfort, 1996) $\tau_\infty(\theta)$ is the unique minimum of a strictly convex function. Therefore, the result follows.□

Conclusion of the proofs

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

i) as $T\to\infty$, $\int_{B_r(\theta_T^*)}\tilde{f}_{\theta_T^*,sp}(\theta)d\theta \to 1$ $\mathbb{P}$-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 $\theta\in\boldsymbol{\Theta}\setminus B_r(\theta_T^*)$,

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

Proof. *i)* By Proposition A.8, apply Proposition A.7, combined with Lemma A.10.

ii) By Proposition A.9 and Assumptions 7(a), for all small enough $r>0$ the result follows. □

Conclusion of the proof of Theorem 5.1 .

For all $\varphi \in C_b$ and for all $r > 0$,

$$\begin{aligned}
&\left| \int_{\boldsymbol{\Theta}} \varphi(\theta) \tilde{f}_{\theta_T^*,sp}(\theta) d\theta - \varphi(\theta_0) \right| \\
\leqslant & \left| \int_{B_r(\theta_T^*)} \varphi(\theta_0) \tilde{f}_{\theta_T^*,sp}(\theta) d\theta - \varphi(\theta_0) \right| + \left| \int_{B_r(\theta_T^*)} \left[\varphi(\theta) - \varphi(\theta_0) \right] \tilde{f}_{\theta_T^*,sp}(\theta) d\theta \right| \\
& + \left| \int_{\boldsymbol{\Theta} \setminus B_r(\theta_T^*)} \varphi(\theta) \tilde{f}_{\theta_T^*,sp}(\theta) d\theta \right|
\end{aligned}$$

Therefore, by Corollary 2, for small enough $r > 0$, for all $\varepsilon > 0$, for T big enough,

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

which is the result needed. □

Conclusion of the proof of Theorem 2. Use Proposition A.6 and apply the change of variable $z := \sqrt{T} \Sigma_T(\theta_T^*)^{-\frac{1}{2}}(\theta - \theta_T^*)$.□

A.9.3 Proof of Theorems 5.3 and 5.4

It follows immediately from the proof of Theorem 5.1 and 5.2. Choose a partition of the parameter space such that each element of the partition contains only one solution to the moment conditions. Then, apply Theorem 5.1 and 5.2 to each element of the partition.