A Classical Moment-Based Inference Framework with Bayesian Properties: Econometric Theory and Simulation Evidence from Asset Pricing

Benjamin Holcblat

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at
Carnegie Mellon University
David A. Tepper School of Business
Pittsburgh, Pennsylvania

Dissertation Committee:
Fallaw Sowell (Chair)
Anisha Ghosh
Gautam Iyer
Lars-Alexander Kuehn

Spring 2012

KEYWORDS: Empirical saddlepoint; Point random-field; Decision theory; Weak identification; Identification; Exponential tilting; Foundations of statistics; Asymptotic theory; Bayesian inference; Multiple hypothesis testing; Over-restricting moment conditions; Consumption-based asset pricing.

Abstract

Dynamic, empirical, consumption-based asset pricing and other areas have been a challenge to existing inference theories. The main contribution of this dissertation is to define and develop an inference approach to tackle this challenge. We call it the ESP approach, as it is based on the empirical saddlepoint (ESP) technique. The idea is to provide a moment-based framework that generates point estimates, confidence regions and tests that rely more on the information in the sample at hand and less on asymptotic limits. The result is an inference approach that provides a unique answer to multiple theoretical and practical concerns faced by existing inference approaches.

Three main steps have been undertaken to reach this objective.

A first step is to put the ESP technique into a general mathematical framework autonomous from standard classical inference. We prove that there exists an intensity distribution of solutions to the empirical moments over the parameter space. Then, we use the ESP technique to approximate this intensity distribution. We call the result the ESP intensity. We prove that it is consistent and asymptotically normal, that is to say that it converges to a point mass at the population parameter like a Gaussian distribution with a standard deviation that goes to zero at rate square root of the sample size. These results are robust to the presence of multiple solutions to the moment conditions (non-identification), as long as their number is finite.

A second step is the development of a decision-theoretic approach within the ESP inference framework. In other words, we propose to choose a loss function according to an inference purpose, and then make the inference decision that minimizes the expected loss. Minimization of expected loss is the optimal answer to the estimated uncertainty that comes from inference, 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., p.4-5 in Lehmann and Casella, 1998), so that only asymptotic optimality results are typically obtained. For a large class of loss functions, we provide ESP point estimate, confidence region, and prove that they are consistent. Simulations of a consumption-based asset pricing model suggest that ESP point estimates and confidence regions perform similarly to, or clearly outperform, the best existing moment-based inference approaches.

A third step develops tests within the ESP framework. In standard classical inference theory, tests usually correspond to confidence regions. Similar tests can be defined in the ESP framework. As an example, we provide a test of over-restricting moment conditions in this spirit. However, we also propose straightforward decision-theoretic point-hypothesis and set-hypothesis tests, which does not correspond to confidence regions. Set-hypothesis tests are typically non-trivial in classical inference theory. Unlike standard classical tests, we prove that ESP decision-theoretic tests do not lead to any asymptotic error. We study their robustness to the presence of multiple solutions to the moment conditions. Unlike standard classical tests, multiple hypothesis testing on the same data set does not undermine the theoretical validity of confidence-region based and decision-theoretic ESP tests. Simulations explore the performance of ESP tests in the context of consumption-based asset pricing.

Acknowledgements

....

Contents

1. Introduction

One of the fundamental problems in classical[1] inference is the gap between practice which is necessarily based on bounded sample size, and theory whose main results are about situations where the sample size can be infinitely increased. This gap has both a qualitative and a quantitative dimension: it is, respectively, a logical and an accuracy problem. While the qualitative dimension of the problem is most of the time acute, the quantitative dimension of the problem is more or less acute depending on the situation. The moment based estimation setting generalized by Hansen (1982) embeds most of the econometric approaches used; and applications and simulations have revealed that it can often lead situations where the quantitative gap between standard asymptotic theory and practice in samples of usual sizes cannot be ignored.

For example, in the original area of application of GMM, consumption-based asset pricing (Hansen and Singleton, 1982), the literature has found little common ground about the values of the representative agent's relative risk aversion (RRA) and elasticity of intertemporal substitution (EIS). Point estimates from economically similar moment conditions are generally outside of each other's confidence intervals. 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, 2010), 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).

There are at least two main ways to try to reduce the gap between econometric asymptotic theory and practice, which relies on bounded samples. One way, which is often put forward, is to look for better asymptotic properties with the hope that they will induce a better finite-sample behaviour (e.g., Newey and Smith, 2004). Another way is to develop inference procedures that rely more on the information contained in the sample at hand and less on asymptotic results. For example, generalizing Anderson and Rubin (1949), Stock and Wright (2000) derive confidence regions that incorporate information from the global shape of the empirical objective function instead of relying on the asymptotic limit of standard statistics. In this dissertation, we go further in this direction. The main contribution

[1] In this dissertation, the word "classical" is used in opposition to "Bayesian". We characterize as "classical" an approach that does not treat the population 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 the common interpretation of fiducial statistics (e.g., Seidenfeld, 1992).

of the paper is to define and develop a moment-based inference framework that yields point estimates, 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.

The only difference between the population parameter and other parameter values is that the former one solves the moment conditions. Although analytically unknown, the empirical moment conditions are their finite-sample counterpart. 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 uncertainty about the population parameter due to the finiteness of the sample. Thus, we propose to use it in the same way as a posterior is used in Bayesian inference to derive estimate and confidence region. We show that the ESP approach satisfies criteria similar to the ones advanced to justify existing classical inference theories. In particular, we prove that the ESP intensity is consistent and asymptotic normal. We also show that these results are robust to the presence of multiple solutions to the moment conditions (non-identification), as long as their number is finite.

The ESP approach leads to contributions in several strands of literature. We distinguish three 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 provides 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 estimated 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., p.4-5 in Lehmann and Casella,

[2] 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 section 7). To avoid any confusion in this dissertation between the utility function of a representative agent and the one chosen by an econometrician, we reserve the term preferences for the former and and utility for the latter.

1998), 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 estimates 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. Following the statistical literature (e.g., Tingley and Field, 1990; Jensen, 1992; Robinson, Ronchetti and Young, 2003), the saddlepoint approximation has been used to improve on existing inference approaches in econometrics. 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 estimator to automatically correct the higher-order bias of generalized empirical likelihood (GEL) estimators. Aït-Sahalia and Yu (2006) propose a saddlepoint approximation of transition density for likelihood-based inference of continuous-time Markov processes. In this dissertation, 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 the use of 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 estimates, which is documented in Skovgaard (1985; 1990), Jensen and Wood (1998), and Almudevar, Field and Robinson (2000). Our 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, the ESP approach is robust 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 the objective function behaves as if the moment conditions had multiple solutions, although they have only one (weak identification). Lack of identification is a frequent issue in many areas (e.g., Pesaran, 1981; Rust, 1994; 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 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 estimates. In a similar way, 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 shorter confidence regions and straightforward construction 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. Nonetheless, identification remains often difficult to prove. Thus, the robustness of the ESP approach to multiple solutions to the moment conditions can be useful.

The dissertation is organized as follow. Section 2 analyzes the problem faced by empirical consumption-based asset pricing, and provides an overview of the ESP approach. Section 3 presents heuristically the idea behind the ESP approximation. Section 4 presents the ESP estimands and estimators, 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 presents a decision-theoretic approach within the ESP framework. Section 8 derives ESP tests. Section 9 extends the ESP framework to deal with over-restricting moment conditions. Section 10 presents simulation results from a consumption-based asset pricing model. Proofs and supplementary results are in the Appendix.

2. Motivation and overview

2.1. Analysis of the question

The key equilibrium implication of standard consumption-based asset pricing models is zero expected discounted profit for the representative agent. More precisely, there is an equilibrium if \$1 invested at date t in any asset j minus its expected gross return for date $t+1$ discounted for risk and time equals 0 i.e.

$$\forall j \in [\![1, n]\!], \quad 1 - \mathbb{E}_t \left[M_{t+1}(\theta_0) R_{j,t+1} \right] = 0 \tag{1}$$

where $\mathbb{E}_t[.]$ denotes the expectation operator conditional on the information available at t, $R_{j,t+1}$ the gross return of asset j between t and $t+1$, n the number of assets considered and $M_{t+1}(\theta_0)$ the stochastic discount factor indexed by the population parameter θ_0. Different consumption-based asset pricing models correspond to different ways of discounting for time and risk through different stochastic discount factors, $M_{t+1}(\theta_0)$. Typically, no distributions are assumed except for tractability reasons. Therefore, the standard inference approach in consumption-based asset pricing is the generalized method of moments (GMM). Unlike most alternatives,[3] 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[4] and where $\psi(X_t, \theta) := [(1 \quad 1 \cdots 1)' - M_{t+1}(\theta)(R_{1,t+1} \quad R_{2,t+1} \cdots R_{n,t+1})'] \circledast Y_t$ with Y_t an element of the representative agent's information set at date t, $\circledast$ the Kronecker product and $'$ the transpose symbol. Second, considering that the t-statistic based on a kth component $\sqrt{T}\frac{\hat{\theta}_{gmm,k} - \theta_{0,k}}{\hat{\sigma}_{k,k}}$ follows a standard normal distribution, $\mathcal{N}(0,1)$,[5] a confidence set estimate and the set of point-hypothesis not rejected, $\hat{I}_\alpha = \left[\hat{\theta}_{gmm,k} - \frac{\hat{\sigma}_{k,k}}{\sqrt{T}} u_{1-\alpha/2}, \hat{\theta}_{gmm,k} - \frac{\hat{\sigma}_{k,k}}{\sqrt{T}} u_{\alpha/2} \right]$, are deduced. $u_{\alpha/2}$ denotes the $\alpha/2$ quantile of a $\mathcal{N}(0,1)$. Figures 1 and 2 show GMM objective functions multiplied by $T^{3/4}$ of consumption-based asset pricing models with constant relative risk aversion (CRRA) and Epstein-Zin-Weil (EZW) preferences, respectively. In Figure 1, although multiplying the objective function by $T^{3/4}$ magnifies level differences,[6] there is a clear canyon in the risk aversion dimension, γ,

[3] Other moment-based inference approaches, such as the generalized empirical likelihood (GEL) approach, have been introduced in consumption-based asset pricing. However, without loss of generality, the introduction focuses on GMM for simplicity. With minor modifications, the analysis applies to these more recent approaches as well.

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

[5] 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.

[6] Multiplying the GMM objective function by $T^{1-\varepsilon}$ is a technique often used in practice to find the global minimum (e.g., Hall, 2005). Asymptotically, for $0 < \varepsilon < 1$ with W_T a positive-definite

around a time discount factor, β, slightly below 1. For the more general Epstein-Zin-Weil preferences, there are even more canyons, as shown in Figure 2. In both cases, there is no clear global minimum. Small variations of the objective function caused by slight data modifications often yield very different point estimates. Reported standard deviations often do not account for such variations. Standard GMM theory summarizes inference as if the uncertainty about the population parameter corresponded to a relatively peaked Gaussian density centered at the global minimum.[7] Paradoxically, the weak statistical structure required by standard GMM theory often leads in empirical consumption-based asset pricing to strong statistical restrictions and thus to overestimation of the inference precision.[8] Nonetheless, different parameter values can have very different theoretical implications. For example, whereas Guvenen (2009) argues for an EIS smaller than one, Bansal and Yaron (2004) require an EIS higher than one so that the intertemporal substitution effect dominates the wealth effect. Progress in consumption-based asset pricing theory will probably exacerbate this problem. As illustrated by Figures 1 and 2, 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.

Careful examination of GMM theory explains this unaccounted instability of GMM estimates. Although any sample size is bounded,[11] the theoretical justifications put forward for GMM estimates and tests assume infinitely increasing sample sizes. This is only the global minimum of the *asymptotic* objective function that corresponds to the population parameter value. In a finite sample, the global minimum is not necessarily even the local minimum closest to the population parameter. Similarly, only the *asymptotic limit* of $\sqrt{T}\frac{\hat{\theta}_{gmm,k}-\theta_{0,k}}{\hat{\sigma}_{k,k}}$ is distributed $\mathcal{N}(0,1)$. Moreover, there is no means to compute a bound for the error implied by the use of these asymptotic results.

The GMM objective functions also suggest an additional problem, lack of identification. Multiplicity of local minima is often the symptom of identification failure. Identification means there is only one solution to the moment conditions (1), i.e., the asset pricing equilibrium cannot correspond to multiple

---

matrix (p-d.m.) whose limit is also a p-d.m. $\lim_{T\to\infty} T^{1-\varepsilon}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right]' W_T \left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right] = \begin{cases}\lim_{T\to\infty} T^{-\varepsilon}\sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right] W_T\sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right] = 0 & \text{if } \theta=\theta_0\\ \infty & \text{if } \theta\neq\theta_0\end{cases}$ where $\psi_t(\theta) := \psi(X_t,\theta)$.

[7] In this dissertation, we maintain the distinction between "uncertain" and "random." In particular, given a sample, in standard classical inference theory, confidence regions summarize an uncertainty without randomness, unlike ESP confidence regions. See the second paragraph of section 9.2.2 for more explanations.

[8] Sims (p. 3-4, 8, 2007a; section III, 2007b) makes a similar remark to justify the Bayesian approach w.r.t. the classical approach.

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

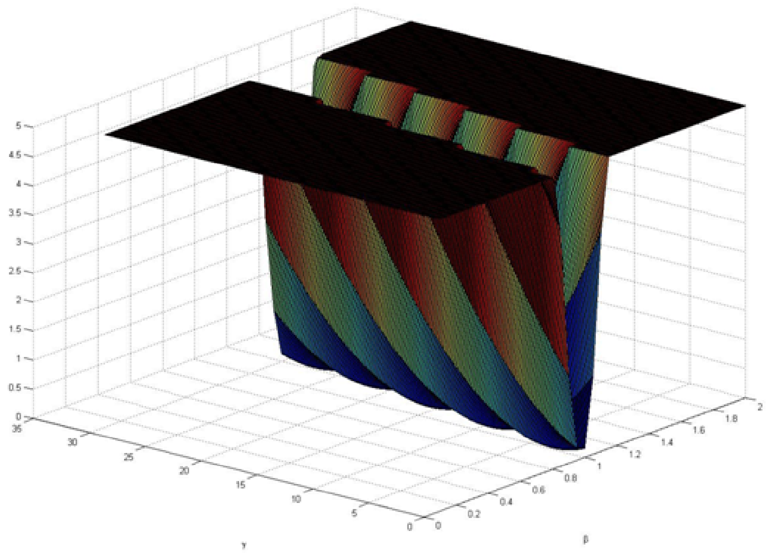

Figure 1: GMM objective function multiplied by $T^{3/4}$ of the consumption-based asset pricing model with constant relative risk aversion (CRRA) preferences with fixed weighting matrix. Instruments are: $\frac{C_t}{C_{t-1}}$ and cay_t. Values of the objective function superior to 5 are set to 5.

EIS and 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.

2.2. Informal presentation of the ESP approach

The dissertation 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 estimates. From a finite-sample point of view, an ideal estimate 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, estimates induced by the following objective function have the same asymptotic properties as GMM estimates :

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

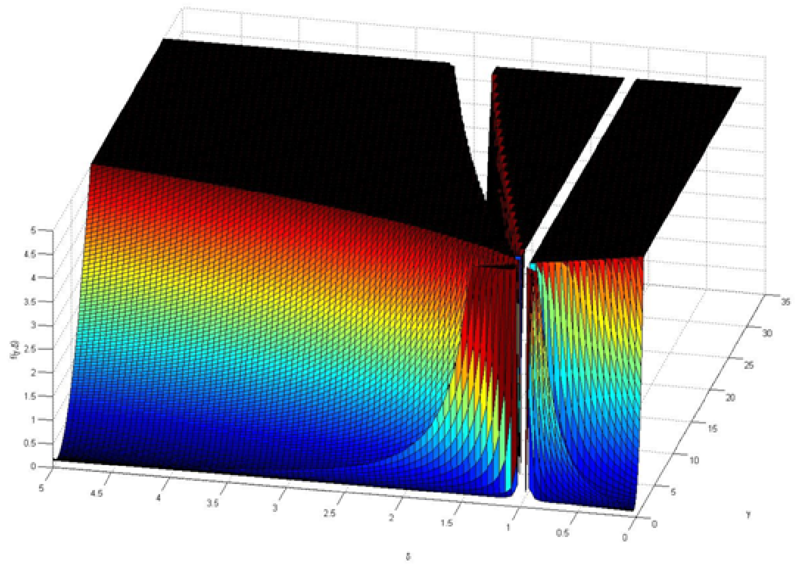

Figure 2: GMM objective function multiplied by $T^{3/4}$ of the consumption-based asset pricing model with Epstein-Zin-Weil preferences for $\beta = 0.95$, and with fixed weighting matrix.[10] γ and $1/\delta$ are respectively known as the relative risk aversion (RRA) and as the elasticity of intertemporal substitution (EIS). Instruments are: $\frac{C_t}{C_{t-1}}$ and cay_{t-1}. Values of the objective function superior to 5 are set to 5.

where $h(.)$ is an arbitrary bounded function and k a large enough constant. However, nobody would accept the objective function (3). $h(.) := c\|. - \overline{\theta}\|$ 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 lot of 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 the strongest possible 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). But the moment conditions (1) are unknown. Nevertheless, because the empirical moment condition approximates the moment conditions (1), 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. Different samples imply different empirical moment conditions, and thus different solutions.[12] We call the ESP approximation the **ESP intensity**. Despite its regularity properties, it does not suffer from the curse of dimensionality usually faced by smooth non-parametric estimates of distributions (Ronchetti and Welsh, 1994). We prove that as the sample size increases it converges to a point mass at the population parameter (or Dirac distribution at the population parameter). The ESP intensity summarizes in probabilistic terms the uncertainty coming from the imperfect knowledge of the moment condition. 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 an inference imprecision. Thanks to this utility function, we define an **ESP point estimate**, θ_T^u, as a maximizer of the ESP expected utility i.e.

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

where $\tilde{\mathbb{E}} \left[u(\theta_e, \theta_T^*) \right] := \int_{\mathbf{\Theta}} u(\theta_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta$ with $\tilde{f}_{\theta_T^*, sp}(.)$ the ESP intensity. By definition, $\hat{\theta}_T^u$ is an *optimal* point estimate for 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, an econometrician utility function corresponding to absolute preference for finite-sample "truth" is pertinent. In this case, after normalization, utility equals 1 if θ_e is a solution to the empirical moment conditions and 0 otherwise. The resulting point estimate 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.[13] Such an estimate aims at taking into account all the possible samples. In contrast, the GMM point estimate is the solution to the empirical moment condition in the comparable just-restricted case (or just-identifying case).[14] Thus, GMM point estimates are backward-looking, while ESP point estimates are not. Since consumption-based asset pricing models are rational expectation models, the ESP approach is more appropriate for self-consistency of inference.

---

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

[13] 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 p.34).

[14]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, parameter values; but in terms of the norm of empirical moment conditions.

We also define confidence regions to assess the stability of ESP point estimates. 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. All the parameter values in the confidence region provide a higher weighted utility for the econometrician than the ones outside. Thus, a small variation of the ESP intensity caused by a slight data modification leads to a new point estimate belonging to the original ESP confidence . If the ESP intensity is multimodal, the ESP confidence can consist of a union of disjoint sets. This is not the case with the standard GMM approach because the asymptotic Gaussian distribution is unimodal.[15] Standard confidence intervals often underestimate the uncertainty about the population parameter. Standard confidence intervals also overestimate the uncertainty in another dimension. They consider the population 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, in consumption-based asset pricing, a time discount factor potentially higher than one is implied, although a consumption-based asset pricing model is not necessarily defined for such values.[16] 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.

In standard classical inference theory, tests usually correspond to confidence intervals, and thus are subject to the same concerns. Similarly to standard classical inference theory, we define ESP tests based on confidence regions. However, we also develop ESP decision-theoretic that are not based on confidence regions. Denote d_H and d_A, respectively, as acceptance and rejection of a test hypothesis. We define an **ESP test** as a mapping such that if

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

[15] We write "*standard* GMM approach" because continuously updated GMM confidence regions for lack of identification (LCU) share similar advantages with ESP confidence regions (Stock and Wright, 2000). However, LCU confidence regions do not allow us to handle point estimate instability in terms of decision-making. Moreover, they tend to be huge w.r.t. ESP confidence regions. See the simulations in section 10.3 p. 69.

[16] When the time discount factor is greater than one, the value function of an infinitely-lived representative agent may explode to infinity. However, Kocherlakota (1990b) provides examples of economy in which $\beta > 1$ is reasonnable.

then it maps to d_H; and otherwise to d_A. In other words, a hypothesis is accepted if the ESP expected utility provided by the acceptance hypothesis is higher than the one of the alternative. ESP hypothesis testing is more flexible than classical testing theory. For instance, testing whether the EIS of the representative agent is greater than one is straightforward in the ESP approach. In classical inference theory, set-hypothesis tests are usually a challenge (e.g., section 21.D in Gouriéroux and Monfort, 1989). Point-hypothesis ESP tests are also more satisfactory than classical tests even from an asymptotic point of view. By construction, classical tests of level α lead asymptotically to wrongly rejecting a right hypothesis with probability α. In other words, a perfectly correct consumption-based asset pricing theory 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 the ESP approach as the ESP intensity converges to a point mass (or Dirac distribution) at the population parameter.[17] In addition, if multiple preference values 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). 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.

3. Heuristic derivation of ESP intensity

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] In the standard classical approach, a typical acceptance 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 an acceptance 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 are disentangled.

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 $f_{X,\theta_0}(.)$ with θ_0 the population parameter. Denote

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

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 $\frac{1}{\sigma}\mathfrak{n}\left(\frac{\cdot}{\sigma}\right)$.

The 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^*,

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

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

[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

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."

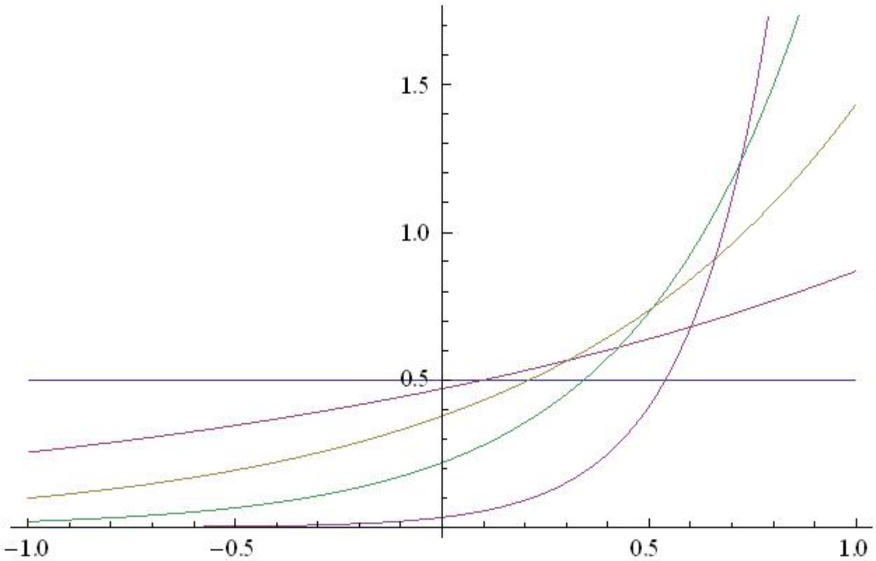

Figure 3: Tilting of $f_X(.) := \mathbf{1}_{[-1,1]}(.)$ for $\frac{1}{T}\sum_{t=1}^{T}\psi(X_t,\theta) := \frac{1}{T}\sum_{t=1}^{T}(X_t-\theta)$ and $T=1$. For θ equals 0, .2, .4, .6, .8 and .95, $\tau(\theta)$, respectively, equals 0, 1.34, 2.4, 5 and 20.

The result is the SP intensity

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

where

$$\begin{aligned}
\sigma^2(\theta) \quad &:= \quad \left[\int_{\mathbf{R}}\frac{\partial\psi(x,\theta)}{\partial\theta}f_{X,\tau(\theta)}(x)dx\right]^{-1}\left[\int_{\mathbf{R}}\psi(x,\theta)^2 f_{X,\tau(\theta)}(x)dx\right] && (6)\\
&\qquad\times\left[\int_{\mathbf{R}}\frac{\partial\psi(x,\theta)}{\partial\theta}f_{X,\tau(\theta)}(x)dx\right]^{-1} &&\\
f_{X,\tau(\theta)}(x) \quad &:= \quad \frac{f_{X,\theta}(x)}{\exp\left\{T\ln\left[\mathbb{E}\mathrm{e}^{\tau(\theta)\psi(X,\theta)}\right]\right\}} && (7)\\
\tau(\theta) \quad &s.t. \quad \int_{\mathbf{R}}\psi(x,\theta)\frac{f_{X,\theta}(x)}{\exp\left\{T\ln\left[\mathbb{E}\mathrm{e}^{\tau(\theta)\psi(X,\theta)}\right]\right\}}dx = 0 && (8)
\end{aligned}$$

The approximation (5) was found by Field (1982), who extended the work of Daniels (1954) for means to Z-estimators (or 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

probability weight of θ satisfying the moment condition. Exponential tilting corresponds to the Radon-Nikodyn derivative $\frac{dP_{\tau(\theta)}}{dP} = \mathrm{e}^{\tau(\theta)X}$.

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, $\tau(\theta)$. The line (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 3 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 the 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 $f_{X,\theta}(.)$ 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

$$\begin{aligned}
\sigma_T^2(\theta) \quad &:= \quad \left[\sum_{t=1}^{T} \hat{w}_{t,\theta}\frac{\partial\psi_t(\theta)}{\partial\theta}\right]^{-1}\left[\sum_{t=1}^{T} \hat{w}_{t,\theta}\psi_t(\theta)^2\right]\left[\sum_{t=1}^{T} \hat{w}_{t,\theta}\frac{\partial\psi_t(\theta)}{\partial\theta}\right]^{-1}, \\
\hat{w}_{t,\theta} \quad &:= \quad \frac{\exp\left[\tau_T(\theta)\psi_t(\theta)\right]}{\frac{1}{T}\sum_{i=1}^{T}\exp\left[\tau_T(\theta)\psi_i(\theta)\right]} \times \frac{1}{T}, \\
\tau_T(\theta) \quad &s.t. \quad \sum_{t=1}^{T}\psi_t(\theta)\frac{\exp\left[\tau_T(\theta)\psi_t(\theta)\right]}{\frac{1}{T}\sum_{i=1}^{T}\exp\left[\tau_T(\theta)\psi_i(\theta)\right]} \times \frac{1}{T} = 0.
\end{aligned}$$

The approximation (9) was first studied by Ronchetti and Welsh (1994), who extended the work of Feuerverger (1989) for means to Z-estimators.

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 dissertation, 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* $(\mathbf{\Theta}, \mathcal{B}(\mathbf{\Theta}))$ *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. Section 9 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. 91) 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 narrow multiplicity to unicity, and thus evacuate point random-field theory at the end. To the knowledge of the author, Sowell (2007) is the only dissertation 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 a such theoretical development. In this dissertation, we take advantage of point random-field theory to develop an inference framework that allows 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 $\mathcal{N}_{\mathbf{\Theta}}$ the space of finite simple counting measures on $\mathcal{B}(\mathbf{\Theta})$, i.e., the space consisting of integer-valued measures, N, such that for all $\theta \in \mathbf{\Theta}$, $N(\{\theta\}) \in \{0,1\}$. Denote $\mathcal{B}(\mathcal{N}_{\mathbf{\Theta}})$ the Borel σ-algebra on $\mathcal{N}_{\mathbf{\Theta}}$ generated by the Prohorov metric. A point random-field (or point process) is a measurable mapping from $(\mathbf{\Omega}, \mathcal{E}, \mathbb{P})$ to $(\mathcal{N}_{\mathbf{\Theta}}, \mathcal{B}(\mathcal{N}_{\mathbf{\Theta}}))$.*[19]

In this dissertation, 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, 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 number of elements in the set A (or cardinal). Under Assumption 1, there exists $\mathbb{P}$- a.s. a point random-field $N_T(.,.)$ such that for all $\omega \in \mathbf{\Omega}$ and $A \in \mathcal{B}(\mathbf{\Theta})$,*

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

Proof. See Appendix A.1 (p. 81). □

Hereafter, for simplicity, we drop the dependence of the point random-field on $\omega \in \mathbf{\Omega}$.

[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).

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

$$\mathbb{F}_T(A) := \lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbb{P}\{N_T(A_{n,i}) = 1\} \quad , \tag{10}$$

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 only contain a unique solution to the empirical moment conditions. 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 of intensity measure 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.

Lemma A.2 in Appendix A.2 (p. 82) 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.

---

[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

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

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

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

where $\varepsilon_n \downarrow 0$ 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. 83). □

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

Proposition 4.2 can be regarded as the counterpart of Theorem 1 (iii) from Almudevar, Field and Robinson (2000) 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 grant 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 discreet, their setup does not hold in contrast to ours.

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, we first study the properties of ESP intensity given by the approximation (9) (p. 20). 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 (or 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. 20) 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 $\{\theta \in \mathbf{R}^m : \|\theta - \mathbf{\Theta}\| < \varepsilon\}$ where $\|\theta - \mathbf{\Theta}\| = \inf_{\dot{\theta} \in \mathbf{\Theta}} \|\theta - \mathbf{\Theta}\|$.*

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. 20) yields the following definition.

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

$$\hat{f}_{\theta_T^*, sp}(\theta) := \exp\left\{T \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}\left(\frac{T}{2\pi}\right)^{m/2} |\Sigma_T(\theta)|_{det}^{-\frac{1}{2}} \tag{11}$$

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

$$\begin{aligned}
\Sigma_T(\theta) &:= \left[\sum_{t=1}^{T} \hat{w}_{t,\theta} \frac{\partial \psi_t(\theta)'}{\partial \theta}\right]^{-1} \left[\sum_{t=1}^{T} \hat{w}_{t,\theta} \psi_t(\theta)\psi_t(\theta)'\right] \left[\sum_{t=1}^{T} \hat{w}_{t,\theta} \frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1}, \\
\hat{w}_{t,\theta} &:= \frac{\exp\left[\tau_T(\theta)'\psi_t(\theta)\right]}{\sum_{i=1}^{T} \exp\left[\tau_T(\theta)'\psi_i(\theta)\right]}, \\
\tau_T(\theta) \quad &\text{s.t.} \quad \sum_{t=1}^{T} \psi_t(\theta) \exp\left[\tau_T(\theta)'\psi_t(\theta)\right] = 0_{m\times 1}, \qquad (12)
\end{aligned}$$

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*

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

Under Assumptions 1 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.

Proof. See Appendix A.4 (p. 85). $\square$

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 is not defined for θ such that $|\Sigma_T(\theta)|_{det} = 0$. Moreover, in the neighborhood of such points $|\Sigma_T(\theta)|_{det}^{-\frac{1}{2}}$ goes to ∞, and thus swamps any information contained in the other term of the ESP intensity. The following assumption rules out such a possibility.

Assumption 3. *For any set $A \subset \mathbf{\Theta}$, denote $A^{-\eta} := \{a \in A : \|a - \partial A \cap (\partial \mathbf{\Theta})^c\| \geqslant \eta\}$ with $\eta > 0$ where A^c and ∂A, respectively, denote the complement of A in $\mathbf{\Theta}$ and its boundary. Define the set*

$$\check{\mathbf{\Theta}}_T \quad := \quad \left\{ \theta \in \mathbf{\Theta} : \left| \left[\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \psi_t(\theta)'}{\partial \theta} \right]^{-1} \left[\frac{1}{T} \sum_{t=1}^{T} \psi_t(\theta) \psi_t(\theta)' \right] \left[\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta'} \right]^{-1} \right|_{det} = 0 \right\}$$

For the sample at hand, for all $\eta > 0$ small enough, the sets $\check{\mathbf{\Theta}}_T$ and $\hat{\mathbf{\Theta}}_T^{-\eta}$ do not have any common elements, i.e., $\check{\mathbf{\Theta}}_T \cap \hat{\mathbf{\Theta}}_T^{-\eta} = \emptyset$ where $\emptyset$ denotes the empty set.

Assumption 3 means that $|\Sigma_T(\theta)|_{det}^{-\frac{1}{2}}$ cannot go to ∞ in an arbitrarily slightly reduced domain of the rough ESP intensity $\hat{\mathbf{\Theta}}_T^{-\eta}$. Note that $\check{\mathbf{\Theta}}_T \cap \{\theta \in \mathbf{\Theta} : \text{tilting equation (12) has a solution}\} = \{\theta \in \mathbf{\Theta} : |\Sigma_T(\theta)|_{det} = 0\}$, as exponential tilting does not alter the support of the initial distribution. Proposition A.1 in Appendix A.5 (p. 86) shows the assumption to be satisfied Lebesgue almost everywhere under reasonnable assumptions. Assumption 3 is not, in practice, stronger than assumptions used in the GMM literature (e.g., Assumption D in Stock and Wright, 2000).

The ignorance of the information provided by the absence of a solution to the tilting equation (12) is a second undesirable property of the rough ESP intensity, which does not exist when the tilting equation (12) does not have a solution. Nonetheless, the absence of a solution for a parameter $\theta \in \Theta$ means that even by reweighting the data, the empirical moment conditions cannot be set to zero for this parameter value. To put it differently, the sample at hand does not provide support for this parameter value being a solution to the empirical moment conditions. Thus, we set the ESP intensity to zero for these parameter values. Nevertheless, we also want the ESP intensity to be continuous. This leads to the following definition.

Definition 4.4 (ESP intensity and intensity measure)**.** *Under Assumptions 1-3, for $\eta > 0$ small enough,* *i) the ESP intensity (or smooth ESP intensity) is the function* $\tilde{f}_{\theta_T^*,sp} : \mathbf{\Theta} \to \mathbf{R}_+$ *s.t.*

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

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 (12) 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 an extension by continuity of the ESP intensity. We extend by continuity to $\hat{\mathbf{\Theta}}_T \cap \left(\hat{\mathbf{\Theta}}_T^{-\eta}\right)^c$ so that extremal values on the latter set are reached on the common boundary of $\hat{\mathbf{\Theta}}_T \cap \left(\hat{\mathbf{\Theta}}_T^{-\eta}\right)^c$ with $\hat{\mathbf{\Theta}}_T^c$ and $\hat{\mathbf{\Theta}}_T^{-\eta}$. Other regularization[21] 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 members of an equivalent class of Gaussian densities produces completely different inference results (e.g., section 7.A.2.a. in Gouriéroux and Monfort, 1989). Finally, regularization has not been necessary in practice. To the knowledge of the author, the need for regularization has never been reported in the SP literature. Nor our simulations, in section 10, do regularizations appear necessary.

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

Proposition 4.4. *Under Assumptions 1-3,*

i) the ESP intensity, $\tilde{f}_{\theta_T^*,sp}(.)$*, is a* $\mathcal{E}\otimes\mathcal{B}(\mathbf{\Theta})/\mathcal{B}(\mathbf{R})$*-measurable positive function continuous with*

[21] Note that the term regularization has a different meaning here from the meaning in ill-posed problems.

respect to (w.r.t.) θ;

ii) the ESP intensity measure is a finite positive measure on the measurable space $(\boldsymbol{\Theta}, \mathcal{B}(\boldsymbol{\Theta}))$.

Proof. See Appendix A.6 (p.86). $\square$

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

Proposition 4.5. *Define a point random-field $\tilde{N}(.)$ and a probability measure $\tilde{\mathbb{P}}$ as respective estimates of $N_T(.)$ and $\mathbb{P}$ consistent with $\tilde{\mathbb{F}}_T(.)$ s.t. $\tilde{N}(.)$, $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{F}}_T(.)$ satisfy the equation (10) in the Definition 4.2 of intensity measures. Assume that $\dot{\theta} \in \boldsymbol{\Theta}$ 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 $\dot{\theta}$, $\overline{B_r(\dot{\theta})} := \overline{\mathrm{B}_r(\dot{\theta})} \cap \boldsymbol{\Theta}$, and where $\lambda(.)$ denotes the Lebesgue measure. Then, under Assumptions 1-3, for all $\dot{\theta} \in \boldsymbol{\Theta}$ Lebesgue*

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

Proof. See Appendix A.7 (p. 86). $\square$

Proposition 4.5 means that the estimated intensity of $\theta \in \boldsymbol{\Theta}$ generally corresponds to the estimated probability weight of θ of being solution to the empirical moment conditions. Under a mild assumption, Lemma A.6 in Appendix A.7 (p. 86) 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 (10) in the Definition 4.2 of intensity measures. Then, Proposition 4.5 is an immediate 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 $\boldsymbol{\Theta}$ is Lebesgue if the volume of the intersection of $\boldsymbol{\Theta}$ 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

5.1. Consistency and asymptotic normality

Whereas in the previous sections T remains fixed to the size of the sample at hand, in this subsection T goes to infinity. In this section, we study the asymptotic behaviour of the ESP intensity measure. More precisely, 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 population 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.

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* $\boldsymbol{\Theta}$*, there exists a unique solution* $\theta_0 \in \mathrm{int}\,(\boldsymbol{\Theta})$ *to the moment conditions* $\mathbb{E}\left[\psi(X,\theta)\right] = 0_{m\times 1}$. **; (c)** $\mathbb{E}\left[\sup_{\theta\in\boldsymbol{\Theta}} \|\psi(X,\theta)\|\right] < \infty$ **; (d)** $\mathbb{E}\left[\sup_{\theta\in\boldsymbol{\Theta}} \left\|\frac{\partial\psi(X,\theta)}{\partial\theta'}\right\|\right] < \infty$ **; (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, accumulation of 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 population 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 population parameter.

Assumption 5. *Define the set*

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

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

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 ESP intensity is asymptotically well-defined in a fixed neighborhood of the population 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 population 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* $\theta \in B_r(\theta_0)$ $\left\| D^k \left\{ |\Sigma_T(\theta)|_{det}^{-\frac{1}{2}} \right\} \right\| < M$**;** **(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* $\theta \in B_r(\theta_0)$*,* $\left\| D^k \left\{ \ln\left[\frac{1}{T}\sum_{i=1}^T \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] \right\} \right\| < M$ **; (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)-(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 population parameter. 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 population 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 for all $\theta \in \hat{\mathbf{\Theta}}_{\infty}^{-\eta}$, $\mathrm{e}^{-\varepsilon T} \left|\Sigma_T(\theta)\right|_{det}^{-\frac{1}{2}} \leqslant M$;* **(b)** *For all $\dot{\theta} \in \hat{\mathbf{\Theta}}_{\infty}^{-\eta}$, there exist $r_1, r_2 > 0$ such that $\mathbb{E}\left[\sup_{(\tau,\theta)\in B_{r_1}(\tau_\infty(\dot{\theta}))\times B_{r_2}(\dot{\theta})} \mathrm{e}^{\tau' \psi(X,\theta)}\right] < \infty$.*

Assumptions 7 correspond 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 dissertation.

Theorem 5.1 (Consistency)**.** *Under Assumptions 1-7, 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. 108).□

Theorem 5.1 means the ESP intensity measure converges to a point mass at the population 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 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. This theorem also shows that asymptotically the ESP integrates to one, although ESP intensity does not typically integrate to one for any T. All other consistency results of this dissertation 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 their similarities, their theoretical foundations are different, as explained in section 6. A second standard convergence result for posterior distributions is asymptotic normality (or 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 1-7, as $T \to \infty$*

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

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. 108).□

Theorem 5.2 indicates that ESP intensity converges asymptotically to a point mass at the population 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 can be derived along the line of Chernozhukov and Hong (2003).[22] However, we do not follow this way because we want to preserve robustness to lack of identification.

5.2. Robustness to lack of identification

In moment-based inference, identification characterizes a situation in which if we knew the moment conditions as a function of the parameter of interest, we could deduce the population parameter value. 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 behaves as if the moment conditions had multiple solutions, although they have only one (weak-identification). Therefore, robustness to multiple solutions to moment conditions is an important and desirable property.

In finite sample, our inference framework is robust to multiple solutions to moment conditions by construction. In this section, we show that it is also true asymptotically. More precisely, we establish

[22] The author has a work in progress in which he introduces intensity and other ideas presented in this dissertation to the generalized Bayesian framework provided in Chernozhukov and Hong (2003).

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. *Denote $[\![1, n]\!]$ the integers in $[1, n]$.* **(b')** *In the parameter space $\mathbf{\Theta}$, there exist multiple solutions, $\left\{\theta_0^{(i)}\right\}_{i=1}^{\overline{n}}$ with $\overline{n}$ the number of solutions,[23] to the moment conditions $\mathbb{E}\left[\psi(X,\theta)\right] = 0_{m\times 1}$ such that for all $i \in [\![1, \overline{n}]\!]$, $\theta_0^{(i)} \in \operatorname{int}(\mathbf{\Theta})$* **; (e')** *For all $i \in [\![1, \overline{n}]\!]$, $\left|\mathbb{E}\left[\frac{\partial\psi(X,\theta_0^{(i)})}{\partial\theta'}\right]\right|_{det} \neq 0$.*

Asumption 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 $T \geqslant \dot{T}^{(i)}$, $B_{\bar{r}^{(i)}}(\theta_0^{(i)}) \subset \hat{\mathbf{\Theta}}_T$.*

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 1-7 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. 108).□

Theorem 5.2 becomes the following.

[23] In accordance with Assumption 1(d), the number of solutions is unbounded but finite.

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 1-7 modified according to Assumptions 8-10, as* $T \to \infty$

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

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. 108).□

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 population 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 population 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.) $l_{X_1,\dots,X_T|\theta_0}(.|.)$. In the second stage, the econometrican tries to infer the population parameter value θ_0 given the sample at hand. As usual in game theory, the p.d.f. $l_{X_1,\dots,X_T|\theta_0}(.|.)$ and $\pi_{\theta_0}(.)$ are common knowledge. 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 population parameter by means of a distribution.

This similarity between Bayesian and ESP inferences should not eclipse their fundamental difference. Bayesian inference produces a distribution that summarizes uncertainty about the population parameter because the population 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" transformation of the unknown, θ_0, into a probabilizable uncertain 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. However, the solution to the moment conditions, the population 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 and EIS of the representative agent are not random. Bayesian inference transforms the unknown population parameter into a probabilizable uncertain 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 a 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.[24] Non-parametric Bayesian analysis also does not avoid extra

[24]Assuming a distributions corresponds to imposing an infinite number of extra moment restrictions. A characteristic function uniquely determines probability distribution; and if the characteristic function of a random variable X is analytic in the neighborhood of zero, then it can expanded at zero into an infinite Taylor series $\mathbb{E}\left(e^{iuX}\right) = \sum_{j=0}^{\infty} \frac{(iu)^j}{j!}\mathbb{E}\left(X^j\right)$ where i denotes here the imaginary unit.

statistical assumptions (e.g., Ghosh and Ramamoorthi, 2003). 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 moment conditions is estimated non-parametrically.

6.2. ESP and the foundation of probability

In probability,[25] there is 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.[26] 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 theory as the latter ones. For brevity, we only focus on these two main conceptions of probability, although there exist 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 because, if the Gaussian approximation is accurate and if we could draw an infinite number of samples, the limit of the proportion of t-statistics in a set would approximatively correspond to the standard Gaussian distribution.[27] Similarly, if the ESP approximation is accurate and 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 approximatively correspond to the ESP intensity. Therefore, our inference framework appears as compatible with the frequentist conception of probability as the existing asymptotic classical theory. Note, however, that this type of compatibility is achieved through different ways. In asymptotic classical theory, the same Gaussian approximation yields probabilistic statements for all sample sizes. In the ESP framework, an approximation of a finite-sample distribution, which by

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

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

[27] 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.

construction is different for each sample size, yields probabilistic statements.

A subjective conception of probability is typically advanced to justify Bayesian inference (e.g., p. 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 it allows us to treat the population 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 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 population 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 for the solution to the moment conditions, the population parameter. From a mathematical point of view, the ESP intensity is not an approximation of the distribution of the population parameter.

However, one can *interpret* the ESP intensity divided by its integral over the parameter space as the distribution of the population parameter, if he considers that (i) the distinction between random and unknown is irrelevant; (ii) probability should express in the language of mathematical measure theory to which extent a proposition is possible with respect to alternative propositions given the evidence available; (iii) the evidence available for $\theta \in \mathbf{\Theta}$ being the solution to the moment conditions corresponds to the ESP estimated probability weight that θ is a solution to the empirical moment conditions.[28] From this point 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 distribution 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 transforms the unknown θ_0 into a probabilizable uncertain through the ESP intensity that is induced by the sample at hand, while the Bayesian theory transforms the unknown, θ_0, into a probabilizable uncertain through an *exogenous* prior $\pi_{\theta_0}(.)$.

Although the ESP approach would remain coherent if the ESP intensity was regarded as a dis-

[28]This way of interpreting the ESP intensity was inspired by Maher (2010), who criticises the subjective conception of probability typically advanced to justify Bayesian inference (See section 6.2).

tribution of the population parameter, we do not follow this interpretation in this dissertation for two reasons. First, our point of view in this dissertation has the advantage to keep transparent the underlying mechanism of moment-based inference procedures. Moment-based inference is necessarily based on a finite-sample counterparts of the population parameter that serves as proxies for the population parameter. Thus, the best we can realistically hope for is an accurate knowledge of these proxies. Second, our point of view does not forbid a user of the ESP approach to *subsequently* and *explicitly* interpret the normalized ESP intensity as a distribution of the population parameter for the reasons indicated in the previous paragraph.

7. A decision-theoretic approach

In this section 7, we present 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 uncertainty. The econometrician chooses a utility function (i.e., opposite of a loss 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.[29] 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 eco-

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

nomic statement 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, 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 standard 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 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 $\theta \mapsto \mathbb{E}_\theta\left[u(d_e(X), \theta)\right]$ and $\theta \mapsto \mathbb{E}_\theta\left[u(d_e(X), \theta)\right]$ typically cross each others (e.g., section 2.D in Gouriéroux and Monfort, 1989). In Bayesian theory, integration of the classical risk functions w.r.t. the posterior 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 dissertation is close to the one for Bayesian inference. However, the fundamental differences analyzed in section 6 remain.

In the next section, an inference decision by the econometrician is a subset of the parameter space (hypothesis-testing). In this section, we focus on the case in which an inference decision by the econometrician is an element of the parameter space (point estimation). In the remaining of the present section, we treat separately continuous utility functions and 0-1 utility functions for clarity. However, combination of the two are possible and relevant as shown in the case of over-restricting moment conditions in section 9.2.2.

7.1. Continuous utility functions

In this section, we consider the case in which the utility function chosen by the econometrician is continuous. In this case, we require the following assumptions.

Assumption 11. **(a)** *$u(.,.)$ is continuous* **; (b)** *For all $\dot{\theta} \in \mathbf{\Theta}$ and $\theta \in \mathbf{\Theta} \setminus \{\dot{\theta}\}$, $u(\theta, \dot{\theta}) < u(\dot{\theta}, \dot{\theta})$* **; (c)** *For all $\theta, \dot{\theta} \in \mathbf{\Theta}^2$, $u(\theta, \theta) = u(\dot{\theta}, \dot{\theta})$.* **; (d)** *For all $\theta_e, \theta \in \mathbf{\Theta}$, $0 \leqslant u(\theta, \theta)$.*

Assumption 11(a) is standard in decision theory (e.g., Definition 3.C.1 and Proposition 3.C.1 on pp.46-47 in Mas-Collel, Whinston and Green, 1995). See section 7.2 below for a relevant case where the utility function is not continuous. Since the parameter space $\mathbf{\Theta}$ is compact, continuity implies boundedness, and thus rules out Saint-Petersburg type paradoxes (e.g., p.185 in Mas-Collel, Whinston and Green, 1995). Boundedness also ensures that the ESP estimated expected utility is always well-defined, i.e., $\int_{\mathbf{\Theta}} u(\theta_e, \theta) f_{\theta_T^*, sp}(\theta) d\theta < \infty$ for all $\theta_e \in \mathbf{\Theta}$. Assumption 11(b) formalizes the econometrician preference for accuracy. This means the econometrician is strictly better off when his point estimate equals a solution to the empirical moment conditions than otherwise. Assumption 11(c) means the econometrician's preference for accuracy is independent of the actual values of the solutions to the empirical moment conditions. Assumption 11(d) is the opposite of the standard convention in decision theory for inference. Usually decision-theory for inference is expressed in terms of loss functions (i.e., opposite of utility functions) instead of utility functions (e.g., p.52,60 in Robert, 1994). In this dissertation, we use the latter ones because of our emphasis on 0-1 utility functions. 0-1 utility functions do not have a formal counterpart in terms of loss functions when integrated with respect to continuous distributions (e.g., p.166 in Robert, 1994) because 0-1 utility functions are not mathematical functions in this case (see section 7.2). To avoid any confusion in this dissertation between the utility function of a representative agent and the one chosen by the econometrician, we reserve the term preferences for the first one and and utility for the second one.

Point estimate

Once we have characterized the utility function, the definition of corresponding point estimates follows.

Definition 7.1 (ESP point estimate)**.** *Given a utility function $u(.,.)$, an ESP point estimate, $\hat{\theta}_T^u$, is a $\mathcal{E}/\mathcal{B}(\mathbf{\Theta})$-measurable maximizer of the ESP expected utility, i.e.,*

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

where $\tilde{\mathbb{E}}[u(\theta_e, \theta_T^*)] := \int_{\mathbf{\Theta}} u(\theta_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta$.[30]

The following proposition presents finite-sample properties of maximization of ESP expected util-

[30]Note that this notation corresponds the usual notation only if there can be only one solution to the empirical moment conditions.

ity.

Proposition 7.1. *Under Assumptions 1-3;11 ,*

i) $\theta_e \mapsto \int_{\mathbf{\Theta}} u(\theta_e, \theta)\tilde{f}_{\theta^*_T,sp}(\theta)d\theta$ *is continuous over* $\mathbf{\Theta}$*;*

ii) there exists an ESP point estimate $\hat{\theta}^u_T$*.*

Proof. See Appendix A.9 (p.108). □

Proposition 7.1i) means the preference relation generated by maximization of the ESP expected utility is continuous[31], i.e., if for two converging sequences of parameter values, $\left\{\theta_n^{(1)}\right\}_{n\geqslant 1}$ $\left\{\theta_n^{(2)}\right\}_{n\geqslant 1}$, $\theta_n^{(1)}$ is always preferred to $\theta_n^{(2)}$, then preference cannot be reversed at the limit. (e.g., p.46 in Mas-Collel, Whinston and Green, 1995). Proposition 7.1ii) is a consequence of Proposition 7.1i) and of the compactness of the parameter space.

The following proposition presents asymptotic properties of maximization of ESP expected utility.

Proposition 7.2. *Under Assumptions 1-7;11, as* $T \to \infty$*,*

i) $\sup_{\theta_e \in \mathbf{\Theta}} \left\| \int_{\mathbf{\Theta}} u(\theta_e, \theta)\tilde{f}_{\theta^*_T,sp}(\theta)d\theta - u(\theta_e, \theta_0) \right\| \to 0$ $\mathbb{P}$*-a.s.;*

ii) an ESP point estimate converges $\mathbb{P}$*-a.s. to the population parameter, i.e.,*

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

Proof. See Appendix A.10 (p.109). □

Proposition 7.2i) means the preference relation corresponding to the ESP expected utility is consistent, i.e., the preference relation corresponding to the ESP expected utility converges to the preference relation corresponding to the utility function with knowledge of the population parameter. Proposition 7.2ii) is an immediate consequence of Proposition 7.2i).

Confidence regions

Point estimates are not necessarily stable. The typical symptom of instability is the absence of a unique well separated maximum of the objective function. Confidence regions provide an indication of the stability of point estimates.

We define ESP confidence region as follows.

[31] Continuity of the utility function is different from continuity of the expected utility.

Definition 7.2 (ESP confidence region)**.** *Given a utility function $u(.,.)$, an ESP confidence region of level $1-\alpha$ with $\alpha \in [0,1]$ is a $\mathcal{B}(\boldsymbol{\Theta})$-measurable set*

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

where $k_{\alpha,T}$ is the highest bound satisfying $\int_{\tilde{S}_T^u} \frac{1}{K_T^u} \int_{\boldsymbol{\Theta}} u(\theta_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta d\theta_e \geqslant 1-\alpha$ and $K_T^u := \int_{\boldsymbol{\Theta}^2} u(\theta_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta d\theta_e$.

By definition, all the elements of the parameter space contained in $\tilde{S}_T^u$ provide a higher ESP expected utility than any elements of the parameter space outside $\tilde{S}_T^u$. In other words, ESP confidence regions correspond to the parameter values which are the closest to maximize the ESP expected utility. Thus it is the smallest set satisfying a constraint of ESP expected utility level $1-\alpha$. A small connected ESP confidence region indicates well-separated maximizer of the ESP expected utility, and thus a reliable ESP point estimate. For the opposite reason, a large ESP confidence region or a ESP confidence region that consists of the union of disjoint sets indicates an unreliable point estimate. Although Definition 7.2 correspond to joint confidence regions, marginal and conditional ESP confidence regions can also be defined. Note also that two-sided symmetric confidence regions and two-sided equal-tailed ESP confidence regions can also be defined. For brevity, we focus only on short ESP confidence region.

In Definition 7.2 of ESP confidence region, we do not require the ESP expected utility level to be exactly equal $1-\alpha$ in order to ensure their existence. If the ESP intensity is locally perfectly flat, the ESP expected utility level over the ESP confidence cannot equal $1-\alpha$.

To the knowledge of the author, such confidence regions have not been studied in Bayesian theory except in the case of 0-1 utility function. Usual Bayesian and classical confidence regions consider parameter values only from a probabilistic point of view. ESP confidence regions takes into account an additional dimension through the utility function.

Since the integral $\int_{\boldsymbol{\Theta}^2} u(\theta_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta d\theta_e$ can take an arbitraty positive value, we normalize the ESP expected utility to define ESP confidence regions. The following assumption ensures the possibility to normalize, i.e., $K_T^u \neq 0$.

Assumption 12. *The domain of definition of the rough ESP intensity is not empty, i.e., $\hat{\boldsymbol{\Theta}}_T \neq \emptyset$.*

Assumption 12 is mild. If $\hat{\boldsymbol{\Theta}}_T$ is empty either there is not support for the model of interest or the sample size is too small. For T big enough, $\hat{\boldsymbol{\Theta}}_T$ is not empty. By Corollary 1 on p.91 in appendix, for T big enough there exists a consistent solution to the empirical moment conditions. Thus, for T big enough, $\hat{\mathbf{\Theta}}_T$ contains a neighborhood of a solution to the empirical moment conditions by Assumption 6(e).

To study the consistency of ESP confidence regions, we introduce a notion of convergence.

Definition 7.3 (Convergence of sets)**.** *Let* $\operatorname{int}(A)$ *denote the interior of a set* A*. A sequence of sets* $\{A_T\}_{T\geqslant 1}$ *converges to a set* A *if and only if for all* $a_1 \in \operatorname{int}(A)$ *and* $a_2 \in \operatorname{int}(A^c)$ *there exists* $\dot{T} \in \mathbf{N}$ *s.t.* $T \geqslant \dot{T}$ *implies* $a_1 \in \operatorname{int}(A_T)$ *and* $a_2 \in \operatorname{int}(A_T^c)$*. It is denoted* $A_T \rightsquigarrow A$*.*

Definition 7.3 means that a sequence of sets converges to a limiting sets if the interior of the sets matches asymptotically. Using this definition, we can prove that ESP confidence regions converge to their asymptotic counterpart. The following proposition ensures existence and consistency of ESP confidence regions.

Proposition 7.3. *Define an asymptotic ESP confidence region of level* $1-\alpha$ *as a measurable set*

$$\tilde{S}^u_\infty := \left\{\theta_e \in \mathbf{\Theta} : \frac{1}{K_\infty} u(\theta_e, \theta_0) \geqslant k_{\alpha,\infty}\right\}$$

where $k_{\alpha,\infty}$ *is the highest bound satisfying* $\frac{1}{K_\infty}\int_{\tilde{S}^u_T} u(\theta_e,\theta_0)d\theta_e \geqslant 1-\alpha$ *and* $K_\infty := \int_{\mathbf{\Theta}} u(\theta_e,\theta_0)d\theta_e$*. For all* $\alpha \in [0,1]$*,*

i) under Assumptions 1-3;11-12 there exist an ESP confidence region, $\tilde{S}^u_T$*, and an asymptotic ESP confidence region of level* $1-\alpha$*;*

ii) under Assumptions 1-7;11-12, as $T \to \infty$*,*

$$\tilde{S}^u_T \rightsquigarrow \tilde{S}^u_\infty \quad \mathbb{P}\text{-a.s.}$$

Proof. See Appendix A.11 (p.110). □

Asymptotic ESP confidence regions correspond to the parameter values that provide the most weighted utility. Asymptotic ESP confidence does not only include the population parameter with continuous utility functions. Parameter values different from the population parameter also provide utility to the econometrician. Proposition 7.3 is a consequence of Proposition 7.1iii)

7.2. 0-1 utility functions

A 0-1 utility function equals one by normalization when the inference decision is right (*i.e.*, when θ_e is a solution to the empirical moment conditions) and zero otherwise. In other words, the use of a 0-1 utility function yields maximization of the expected finite-sample "truth." By finite-sample "truth" we mean solution to the empirical moment conditions, while in the existing literature the word "truth" is reserved for solution to the moment condition (i.e., population parameter).

Point estimate

In point inference, the 0-1 utility function is not a usual function. On the one hand, Assumption 1(d) rules out situations with a continuum of solutions to the empirical moment conditions. On the other hand, by construction, the ESP intensity measure is absolutely continuous w.r.t. the Lebesgue measure, which ignores points. Therefore, in point inference, the 0-1 utility function is a "generalized" function which corresponds to a family of Dirac distributions indexed by θ_e, $\{\delta_{\theta_e}(.)\}_{\theta_e \in \mathbf{\Theta}}$.[32]

Definition 7.4 (Maximum ESP point estimate)**.** *A maximum ESP point estimate, $\hat{\theta}_T$, is an ESP estimate that maximizes the ESP expected 0-1 utility function, i.e.,*

$$\hat{\theta}_T := \arg\max_{\theta_e \in \mathbf{\Theta}} \tilde{\mathbb{E}}[\delta_{\theta_e}(\theta_T^*)]$$

where $\delta_{\theta_e}(.)$ is the Dirac distribution at θ_e.

The following immediate proposition clarifies the meaning of Definition 7.4.

Proposition 7.4. *Under Assumptions 1-3, the Definition 7.4 is equivalent to each of the following properties*

i) if there exists a unique $\hat{\theta}_T \in \mathrm{int}(\mathbf{\Theta})$, for small enough $r > 0$, $\hat{\theta}_T = \arg\max_{\theta_e \in \mathbf{\Theta}} \int_{\mathbf{\Theta}} \mathbf{1}_{B_r(\theta_e)}(\theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta$;

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

---

[32] See section 1 on pp.13-20 in chap.1 in Schwartz (1950-1951) for a discussion about the differences between a (Schwartz) distributions and functions. We still use the word function for 0-1 utility function to avoid a too cumbersome terminology. This misuse of language is already well-spread in the Bayesian decision theory. Note also that the meaning of Dirac distributions in this section is different from the one in Theorems 5.1 and 5.3. Here Dirac distributions formalize absolute preference of the econometrician for finite-sample "truth", while in Theorem 5.1 and 5.2 they formalize probabilistic distributions.

Proof. See Appendix A.12 (p.116). □

Proposition 7.4i) provides an alternative, but equivalent formalization of absolute preference for finite-sample "truth." This formalization is adapted from Robert (p.166, 1994). Proposition 7.4ii) provides an alternative interpretation of maximum ESP point estimates. A maximum ESP point estimate 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 estimate. A maximum-probability estimate is different from a maximum-likelihood estimator. See footnote 14 on p. 15. Proposition 7.4ii) also shows that our maximum ESP point estimate corresponds to the point estimate introduced in Sowell (2009) to correct the higher-order bias of exponential tilting estimates (ET). Sowell (2009) shows the logarithm of ESP intensity divided by the sample size to correspond to the exponential tilting objective function plus two terms that vanish asymptotically. He deduces that maximum ESP estimates share the same first-order asymptotic properties as ET estimates, 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 estimates are consistent.

Proposition 7.5. *Under Assumptions 1-3,*

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

ii) under additional Assumptions 4-7 and 12, a maximum ESP point estimates converges $\mathbb{P}$-a.s. to the population parameter, i.e.,

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

Proof. See Appendix A.13 (p.116) □

Proposition 7.5 i) follows from Lemma 2 in Jennrich (1969) and the continuity of the ESP intensity. We deduce Proposition 7.5 ii) from the consistency of the ESP intensity, unlike Sowell (2009) who deduces it from the consistency of ET estimates.

Confidence regions

As for continuous utility functions, we define confidence regions.

Definition 7.5 (Maximum ESP confidence region)**.** *A maximum ESP confidence region of level* $1-\alpha$ *with* $\alpha \in [0,1]$ *is a* $\mathcal{B}(\boldsymbol{\Theta})$*-measurable set*

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

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* $K_T := \int_{\boldsymbol{\Theta}} \tilde{f}_{\theta_T^*,sp}(\theta)d\theta$.

Since $\int_{\boldsymbol{\Theta}} \delta_{\theta_e}(\theta)\tilde{f}_{\theta_T^*,sp}(\theta)d\theta = \tilde{f}_{\theta_T^*,sp}(\theta_e)$, Definition 7.5 is in line with Definition 7.2; and thus the same interpretation still holds. By Proposition 4.5, 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. In this sense, they are maximum-probability based. As for continuous utility functions, 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. In particular, Definition 9.3 on p.65 provides an example of marginal ESP confidence region in the case of over-restricting moment conditions.

To the knowledge of the author, 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 Definition 7.5 are the following. He uses the ESP technique to approximate the distribution of the local minima of the second step GMM objective function, while we use it 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, HPD regions involve the population parameter unlike maximum ESP confidence region. Because in the ESP approach randomness comes from data and not from treating the population parameter as a random variable (see section 6), ESP confidence regions do not *formally* involves the population parameter, θ_0.

The formal absence of the population parameter in Definition 7.5 is also one of the differences w.r.t. standard classical confidence regions. By definition, a usual classical confidence region of level $1-\alpha$ should contain the population parameter with probability $1-\alpha$ before observation of the sample. But, given a sample at hand, a usual classical confidence region does not provide a probabilistic statement about the population parameter. Given a sample at hand, a classical confidence region is fixed, and thus it has probability one or zero to contain the population parameter. In contrast, ESP confidence regions *formally* involve only the finite sample counterpart of population parameter, as the ESP approach acknowledges that practice relies on finite samples. Moreover, the induced probabilistic statements about the final counterpart of the population parameter does not disappear once the ESP confidence region is computed. Another difference between maximum ESP confidence regions and usual classical confidence regions is that the latter ones only requires to report standard errors. In standard classical inference, whatever is the sample size, the same Gaussian approximation is used, whereas in the ESP approach an approximation of a finite-sample distribution, which by construction is different for each sample size, yields probabilistic statements.

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

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

Proposition 7.6. *For all* $\alpha \in]0,1[$,

i) under Assumptions 1-3,12, there exists a maximum ESP confidence region, $\tilde{S}_T$*;*

ii) under Assumptions 1-7,12, as $T \to \infty$*,*

$$\tilde{S}_T \rightsquigarrow \{\theta_0\} \quad \mathbb{P}\text{-}a.s.$$

Proof. See Appendix A.14 (p.116). $\square$

Proposition 7.6i) follows from the same arguments as Proposition 7.3 i).

7.3. Robustness to lack of identification

In this section, we present how the multiplicity of solutions to the moment conditions affects the decision-theoretic approach of sections 7.1 and 8. For clarity, the structure of the section is similar to sections 7.1 and 8. For brevity, we try to only indicate the necessary changes w.r.t sections 7.1 and 8. Proofs are adaptation of the proofs in the case of identification.

By definition, point estimation is not relevant in the case of multiple solutions to the moment conditions. Therefore, like existing point estimates, in this case, ESP point estimates are only locally consistent i.e. they are consistent when the parameter space is restricted to a subset containing a unique solution. However, ESP confidence sets reflect the lack of reliability of ESP point estimates. We show that ESP confidence sets are globally consistent in the presence of multiple solutions to the moment conditions. In contrast, standard classical confidence sets are not consistent. By construction, they consider the uncertainty about the population parameter corresponds to a Gaussian density centered at the point estimate. Thus, standard point estimates contaminate standard confidence sets.

7.3.1 Continuous utility functions

Point estimate

Definition 7.1 becomes.

Definition 7.6. *Denote $\mathcal{P}(\mathbf{\Theta}) := \{\mathbf{\Theta}_i\}_{i=1}^{\overline{n}}$ a partition of $\mathbf{\Theta}$ such that for all $i \in [\![1, \overline{n}]\!]$, $\mathbf{\Theta}_i$ contains a unique solution to the moment conditions $\theta_0^{(i)} \in \mathrm{int}(\mathbf{\Theta}_i)$. Given a utility function $u(.,.)$ and a subset $\mathbf{\Theta}_i$, a local ESP point estimate, $\hat{\theta}_T^{u,(i)}$, is a $\mathcal{E}/\mathcal{B}(\mathbf{\Theta})$-measurable maximizer of the ESP expected utility over $\mathbf{\Theta}_i$ i.e*

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

where $\tilde{\mathbb{E}}[u(\theta_e, \theta_T^*)] := \int_{\mathbf{\Theta}} u(\theta_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta$.

Note we still integrate the utility function over the whole parameter space. Proposition 7.1 remains valid for local ESP point estimates after obvious change. Proposition 7.2 becomes the following.

Proposition 7.7. *Under Assumptions 1-3,8-10,11, for all $i \in [\![1, \overline{n}]\!]$, as $T \to \infty$,*

i) $\sup_{\theta_e \in \mathbf{\Theta}} \left\| \int_{\mathbf{\Theta}} u(\theta_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta - \sum_{i=1}^{\overline{n}} u(\theta_e, \theta_0^{(i)}) \right\| \to 0 \quad \mathbb{P}$*-a.s.;*

ii) a local ESP point estimate converges locally $\mathbb{P}$-a.s. to its corresponding solution to the moment conditions i.e.

$$\lim_{T \to \infty} \hat{\theta}_T^{u,(i)} = \theta_0^{(i)} \quad \mathbb{P}\text{-a.s.}.$$

Proof. Adapt proof of Proposition 7.2. □

Proposition 7.7 shows that multiplicity of solutions to moment conditions implies multimodal ESP intensity. Similarly, multiplicity of local minima in a GMM objective function may be a symptom of multiple solution to the moment conditions. Theorem 4.1.2 in Amemiya (1985) has a spirit similar to Proposition 7.7 .

Confidence region

Definitions 7.2, 7.3 and Proposition 7.3i) remain valid. Proposition 7.3ii) becomes the following.

Proposition 7.8. *Define an asymptotic ESP confidence set of level $1-\alpha$ as a measurable set*

$$\tilde{S}^u_\infty := \left\{ \theta_e \in \mathbf{\Theta} : \frac{1}{K_\infty} \sum_{i=1}^{\overline{n}} u(\theta_e, \theta_0^{(i)}) \geqslant k_{\alpha,\infty} \right\}$$

where $k_{\alpha,\infty}$ is the highest bound satisfying $\frac{1}{K_\infty} \sum_{i=1}^{\overline{n}} \int_{\tilde{S}^u_T} u(\theta_e, \theta_0^{(i)}) d\theta_e \geqslant 1-\alpha$ and $K_\infty := \sum_{i=1}^{\overline{n}} \int_{\mathbf{\Theta}} u(\theta_e, \theta_0^{(i)}) d\theta_e$. For all $\alpha \in [0,1]$, under Assumptions 1-3,8-10 ,11-12, as $T \to \infty$,

$$\tilde{S}^u_T \rightsquigarrow \tilde{S}^u_\infty \quad \mathbb{P}\text{-a.s.}$$

Proof. Adapt proof of Proposition 7.3. □

The definition of an asymptotic ESP confidence set is in line with the definition in Proposition 7.3 by Proposition 7.7i).

7.3.2 0-1 utility functions

Point estimate

Definition 7.4 becomes the following.

Definition 7.7. *A local maximum ESP point estimate , $\hat{\theta}_T$, is an ESP estimate that maximizes the ESP expected 0-1 utility function over $\mathbf{\Theta}_i$ i.e.*

$$\hat{\theta}_T := \arg\max_{\theta_e \in \mathbf{\Theta}_i} \tilde{\mathbb{E}}[\delta_{\theta_e}(\theta^*_T)]$$

where $\delta_{\theta_e}(.)$ is the Dirac distribution at θ_e.

After obvious modifications, Propositions 7.4 and 7.5i) remain valid for local maximum ESP point estimate. Proposition 7.5ii) becomes the following.

Proposition 7.9. *Under Assumptions 1-3,8-10, a local maximum ESP point estimates converges locally $\mathbb{P}$-a.s. to its corresponding solution to the moment conditions i.e. for all $i \in [\![1, \overline{n}]\!]$*

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

Proof. Adapt proof of Proposition 7.5. □

Confidence region

Definition 7.2 and Proposition 7.6i) remain valid for the same reason, but Proposition 7.6ii) becomes the following.

Proposition 7.10. *Under Assumptions 1-3,8-10 ,11-12, for all $\alpha \in]0, 1[$, as $T \to \infty$,*

$$\tilde{S}_T \rightsquigarrow \bigsqcup_{i=1}^{\overline{n}} \left\{ \theta_0^{(i)} \right\} \quad \mathbb{P}\text{-}a.s.$$

where $\bigsqcup$ denotes a union of disjoint sets.

Proof. Adapt proof of Proposition 7.5. □

8. ESP hypothesis testing

In classical inference theory, there is usually a duality between tests and confidence regions, in the sense that, the set of point-hypothesis that would not be rejected corresponds to a confidence region. We can also define ESP tests that are based on this duality. For brevity, we do not present them formally in this section. We only present two examples of them, later, in the section 9.2.2, in the case of over-restricting moment conditions.

In this section, we present formally a decision-theoretic approach to derive ESP tests, which are not based on confidence regions. A such approach is not possible in standard classical inference frameworks.

8.1. Notations and definitions

The following definition sets the notations for tests.

Definition 8.1 (Test). *Define the subsets $\mathbf{\Theta}_H \subset \mathbf{\Theta}$ and $\mathbf{\Theta}_A \subset \mathbf{\Theta}$ such that $\mathbf{\Theta}_H \cap \mathbf{\Theta}_A = \varnothing$. Define the measurable decision space $(\mathbf{D}, \mathcal{D})$ where $\mathbf{D} := \{d_H, d_A\}$, with d_H and d_A and where $\mathcal{D}$ is the power set of $\mathbf{D}$. The decisions d_H and d_A respectively correspond to acceptance of $\mathbf{\Theta}_H$ and rejection of $\mathbf{\Theta}_H$. Given a sample size T, a test is a $\mathcal{E}/\mathcal{D}$-measurable function $d_T(.)$.*

As in point estimation, the decision which maximizes the ESP expected utility is retained. Thus, we define an ESP test as follows.

Definition 8.2 (ESP decision-theoretic test). *Given a utility function $u(.,.)$, an ESP hypothesis test is a $\mathcal{E}/\mathcal{D}$-measurable function, d_T, such that for all $\omega \in \mathbf{\Omega}$ if*

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

then $d_T(\omega) = d_H$; and otherwise $d_T(\omega) = d_A$.

ESP tests solve two problems faced by classical tests. First, standard classical tests often imply that the population parameter can be outside the parameter space with a strictly positive probability, although the economic model is typically not defined for these parameter values. For example, in consumption-based asset pricing, standard confidence intervals and tests for the time discount factor consider it can take values higher than one. The support of the Gaussian distribution is the whole real line. However, for values higher than one, the consumption-based asset pricing is typically not defined. The value function of a dynastic representative agent explodes to infinity for time discount factor higher than one. In contrast, like ESP confidence regions, ESP tests do not regard values outside the parameter space as possible because the ESP intensity is not defined outside the parameter space by construction.

The second problem is about the asymptotic properties of standard classical tests. In the standard classical theory, a test is consistent if the probability of rejection of the alternative goes to one when $\theta_0 \in \mathbf{\Theta}_A$ as the sample size increases to infinity (e.g., p.553 in Gouriéroux and Monfort, 1989). However, a consistent test typically leads to asymptotically reject the hypothesis of the test when $\theta_0 \in \mathbf{\Theta}_H$ with a probability equal to the level of the test, although asymptotically a model is perfectly known. As shown below, such asymptotic mistake does not occur with decision-theoretic ESP tests. We introduce the notion of double-consistency to characterize this property.

Definition 8.3 (Double consistency). *A test $d_T(.)$ is doubly-consistent $\mathbb{P}$-a.s. if and only if*

$$\lim_{T\to\infty} d_T = \begin{cases} d_H & \text{if } \theta_0 \in \mathbf{\Theta}_H \\ d_A & \text{if } \theta_0 \in \mathbf{\Theta}_A \end{cases} \quad \mathbb{P}\text{-a.s.}$$

In a test, there are two possible inference decisions (acceptance and rejection of the test hypothesis) and two possible right propositions (the hypothesis is correct and the alternative is correct). Therefore, in hypothesis testing, a utility function takes at most four different values. Consequently, instead of distinguishing between continuous and 0-1 utility functions as for point estimation, we distinguish between set and point hypothesis for purpose of clarity. However, combinations of the two are possible as shown, later, in section 9.2.1 in the case of the test of over-restricting conditions.

8.2. Set hypothesis

The following assumption sets notations for the utility function.

Assumption 13. *For all $(d, \theta) \in \mathbf{D} \times \mathbf{\Theta}$, $u : (d, \theta) \mapsto c_d \mathbf{l}_{\mathbf{\Theta}_H}(\theta) + b_d \mathbf{l}_{\mathbf{\Theta}_A}(\theta)$ with $c_{d_H} > c_{d_A}$ and $b_{d_A} > b_{d_H}$.*

The strict inequality conditions on the values of the utility function ensure that a right inference decision provide a strictly higher expected utility to the econometrician strictly than the wrong ones. The maximum ESP (or maximum expected "truth") approach corresponds to $c_{d_H} = b_{d_A} = 1$ and $c_{d_A} = b_{d_H} = 0$.

The following proposition reformulates conveniently ESP tests in the case of set hypothesis.

Proposition 8.1. *Under Assumptions 1-3 and 13, the ESP hypothesis test in Definition 8.2 is equivalent to the test d_T, such that for all $\omega \in \mathbf{\Omega}$ if*

$$c\tilde{\mathbb{F}}_T(\mathbf{\Theta}_H) \geqslant \tilde{\mathbb{F}}_T(\mathbf{\Theta}_A) \tag{13}$$

with $c := \frac{c_{d_H} - c_{d_A}}{b_{d_A} - b_{d_H}}$, then $d_T(\omega) = d_H$; and otherwise $d_T(\omega) = d_A$.

Proof. See Appendix A.15 (p.117). □

Proposition 8.1 is the immediate counterpart of standard result in Bayesian inference (e.g., p.218 in Schervish, 1995). In the case of a maximum ESP approach (i.e., $c = 1$), the meaning is clear. We accept the hypothesis if the estimated intensity measure that solutions to the empirical moment conditions are in $\mathbf{\Theta}_H$ is higher than it is in $\mathbf{\Theta}_A$. If there can only be one solution to the empirical moment conditions, we accept the most probable hypothesis. Despite this appealing meaning, Proposition 8.1 also shows that the hypothesis with the biggest volume is favoured.

The following proposition ensures the existence and the double-consistency of an ESP set-hypothesis test.

Proposition 8.2. *Given a utility function $u(.,.)$,*

i) under Assumptions 1-3 and 13, there exists an ESP set-hypothesis test;

ii) under Assumptions 1-7 and 13, an ESP set-hypothesis is doubly consistent $\mathbb{P}$-a.s.

Proof. See Appendix A.16 (p.117). □

Proposition 8.2i) is immediate. Proposition 8.2ii) is a consequence of the convergence of the ESP intensity measure to a Dirac distribution centered at the population parameter. Unlike in the standard classical approach, there is no uncertainty asymptotically; and thus no mistake occurs.

8.3. Point hypothesis

In the case of point-hypothesis (i.e., $\mathbf{\Theta}_H := \{\theta_H\}$), we derive results similar to set-hypothesis tests. The counterpart of Assumption 13 is the following assumption.

Assumption 14. *For all $(d, \theta) \in \mathbf{D} \times \mathbf{\Theta}$, $u : (d, \theta) \mapsto c_d \delta_{\theta_H}(\theta) + b_d \mathbf{l}_{\mathbf{\Theta}_A}(\theta)$ with $c_{d_H} > c_{d_A}$ and $b_{d_A} > b_{d_H}$.*

Since $\mathbf{\Theta}_H$ is a parameter value, the utility function is expressed in terms of Dirac distribution for same reason as in section 7.2. A maximum ESP approach also corresponds to $c_{d_H} = b_{d_A} = 1$ and $c_{d_A} = b_{d_H} = 0$.

The counterpart of Proposition 8.1 is the following proposition.

Proposition 8.3. *Under Assumptions 1-3 and 14, the ESP hypothesis test in Definition 8.2 is equivalent to the test d_T, such that for all $\omega \in \mathbf{\Omega}$ if*

$$c\tilde{f}_{\theta_T^*, sp}(\theta_H) \geqslant \tilde{\mathbb{F}}_T(\mathbf{\Theta}_A) \tag{14}$$

with $c := \frac{c_{d_H} - c_{d_A}}{b_{d_A} - b_{d_H}}$, then $d_T(\omega) = d_H$; and otherwise $d_T(\omega) = d_A$.

Proof. See Appendix A.17 (p.117). □

In the case of a maximum ESP approach (i.e., $c = 1$), an ESP test does not have the same straightforward meaning as in Proposition 8.1 . The LHS of equation (14) is in terms of intensity weight, while the RHS is in terms of intensity measure. The test hypothesis is accepted when the estimated intensity (or probability by Proposition 4.5) weight of θ_H being a solution to the empirical moment conditions is higher than the intensity measure of $\mathbf{\Theta}_A$. In a maximum ESP approach, Proposition 8.3 also shows some similarity with Jeffreys' Bayes factors (e.g., section 4.2.2 in Schervish, 1995). However, Jeffrey's approach requires to choose a prior over $\mathbf{\Theta}_A$ given that $\theta_0 \neq \theta_H$.

The following proposition is the point-hypothesis counterpart of Proposition 8.2.

Proposition 8.4. *Given a utility function $u(.,.)$,*

i) under Assumptions 1-3 and 14, there exists an ESP point-hypothesis test;

ii) under Assumptions 1-7 and 14, an ESP point-hypothesis is doubly-consistent $\mathbb{P}$-a.s.

Proof. See Appendix A.18 (p.117). □

8.4. Robustness to lack of identification

Standard classical tests correspond to standard classical confidence intervals. Thus they are not robust to the presence of multiple solutions the moment conditions. ESP tests presents some robustness to this situation. They take into account the uncertainty due to the multiplicity of solutions.

Definitions 8.1 and 8.2 remain relevant unlike Definition 8.3.

8.4.1 Set hypothesis

Proposition 8.1 and 8.2i) remain valid for the same reasons, but Proposition 8.2ii) becomes the following.

Proposition 8.5. *Given a utility function $u(.,.)$, under Assumptions 1-3,8-10 and 13, as $T \to \infty$, $\mathbb{P}$-a.s.*

$$\lim_{T\to\infty} d_T = \begin{cases} d_H & \textit{if } c\#\left\{\theta_0^{(i)} : i \in [\![1,\overline{n}]\!] \textit{ and } \theta_0^{(i)} \in \mathbf{\Theta}_H\right\} \geqslant \#\left\{\theta_0^{(i)} : i \in [\![1,\overline{n}]\!] \textit{ and } \theta_0^{(i)} \in \mathbf{\Theta}_A\right\} \\ d_A & \textit{otherwise} \end{cases}$$

Proof. Adapt proof of Proposition 8.2ii) . □

In other words, if the number of solutions to the moment conditions in $\mathbf{\Theta}_H$ weighted by c is higher that the one in $\mathbf{\Theta}_A$, the hypothesis is accepted. In the case of a maximum ESP approach, $c = 1$.

8.4.2 Point hypothesis

Propositions 8.3 and 8.4i) remain valid for the same reasons, but Proposition 8.4ii) becomes

Proposition 8.6. *Given a utility function $u(.,.)$, under Assumptions 1-3,8-10 and 13, as $T \to \infty$, $\mathbb{P}$-a.s.*

$$\lim_{T\to\infty} d_T = \begin{cases} d_H & \textit{if } \theta_H \in \left\{\theta_0^{(i)}\right\}_{i=1}^{\overline{n}} \\ d_A & \textit{otherwise} \end{cases}$$

Proof. Adapt proof of Proposition 8.4ii) . □

According to Proposition 8.6, if the hypothesis corresponds to a solution to the moment conditions, the hypothesis is accepted. Thus, the solution to the moment conditions which corresponds to the point-hypothesis is favoured.

9. Over-restricting moment conditions

In previous sections, we assume that the number of moment conditions equals the dimension of the parameter space (just-restricting moment conditions). See Assumption 1(c) on p.21. In this section, we consider the case in which the number of moment conditions is strictly greater than the dimension of the parameter space (over-restricting moment conditions). Then, thanks to the additional moment conditions, we propose tests of goodness-of-fit or, more precisely, tests of over-restricting moment conditions.

For at least three reasons, the case with over-restricting moment conditions is an important case to deal with. First, an economic model often implies a number of moment conditions greater than the dimension of the parameter space (e.g., Carrasco and Florens, 2000). In rational expectation models, one can derive as many moment conditions as he wants from the orthogonality between the error of prediction concerning the next period and the present information. For example, in the general asset pricing framework of section 2.1, under the assumption that $\mathbb{E}\left\{[1 - M_{t+1}(\theta_0)R_{j,t+1}]^2\right\} < \infty$, if we denote Y_t an element of the information set at date t, the asset pricing equation (1) on p.11 implies

$\mathbb{E}\left\{[1 - M_{t+1}(\theta_0)R_{j,t+1}]h(Y_t)\right\} = 0$ for all functions $h(.)$ measurable such that $\mathbb{E}\left[h(Y_t)^2\right] < \infty$. Second, the more moment conditions, the more information from economic theory is incorporated into inference so that its results are typically sharper. In inference, information comes either from the structure imposed by the econometric model or from data. Third, over-restricting moment conditions allow the derivation of tests of goodness-of-fit, which aim at assessing the agreement between an whole econometric model and data. Tests previously presented in section 8 aim at assessing the agreement between a restriction on the parameter space and data.

In standard classical theory, tests of goodness-of-fit, such as the so-called test of over-identifying restrictions (Hansen, 1982), are widely developed and used. In contrast, testing goodness-of-fit is theoretically delicate in Bayesian inference (e.g., Robert, 1994, p.374), although successful practice of Bayesian inference often resorts to indirect measures of goodness-of-fit (e.g., Rubin, 1984; Gelman and Shalizi, 2011). As explained in section 6.1, taken *literally*, Bayesian theory regards inference as a game between nature and an econometrician, in which the prior over the parameter space, $\pi_{\theta_0}(.)$, and the probability distribution of data conditional on population parameter, $l_{X_1,\ldots,X_T|\theta_0}(.|.)$, are common knowledge. In other words, Bayesian theory considers that the econometrician knows the true prior and the true conditional p.d.f. according to which data are generated. Therefore, from a literal Bayesian perspective, the econometric model fits the data by construction so that an econometrician can only learn about parameter values. This is one of the consequences of the fundamental differences between Bayesian and ESP inference.

9.1. From over-restricting to just-restricting moment conditions

On the one hand, identification requires that the number of moment conditions is greater than, or equal to, the dimension of the parameter space. On the other hand, as explained in the justification of Assumption 1(c) on p.21, if the number of moment conditions is greater than the dimension of the parameter space, there is generically no solution to empirical moment conditions, so that the intensity distribution is zero. Therefore, the idea of this section is to reduce the over-restricted case to the just-restricted case of the previous sections by increasing the dimension of the parameter space. We simultaneously undertake two tasks to implement this idea. We adapt the setup of sections 4.1, 4.2 and 5.1 to the case with over-restricting moment conditions, and we show that this modified setup satisfies the assumptions of sections 4.1, 4.2 and 5.1 so that their results still hold. For clarity, the structure of this section is similar to the one of sections 4.1, 4.2 and 5.1.

9.1.1 ESP estimand

In the setup with over-restricting moment conditions, Assumption 1(a) remains unchanged, while Assumptions 1(b)(c) and (d) become Assumption 15(b*)(c*) and (d*).

Assumption 15. **(b*)** *Let the measurable space* $(\mathbf{\Phi}, \mathcal{B}(\mathbf{\Phi}))$ *such that* $\mathbf{\Phi} \subset \mathbf{R}^q$ *is compact and* $\mathcal{B}(\mathbf{\Phi})$ *denotes the Borel* σ*-algebra on* $\mathbf{\Phi}$ **(c*)** *The moment function* $g : \mathbf{R}^p \times \mathbf{\Phi} \to \mathbf{R}^m$ *is* $\mathcal{E} \otimes \mathcal{B}(\mathbf{\Phi})/\mathcal{B}(\mathbf{R}^m)$*-measurable, where* $\mathcal{E} \otimes \mathcal{B}(\mathbf{\Theta})$ *denotes the product* σ*-algebra and* $q < m$ **; (d*)** *For the sample size at hand* T*, there exists* $g^{(1)} : \mathbf{R}^p \times \mathbf{\Phi} \to \mathbf{R}^q$ *and* $g^{(2)} : \mathbf{R}^p \times \mathbf{\Phi} \to \mathbf{R}^{m-q}$ *such that* $g(.) =: (g^{(1)}(.) \quad g^{(2)}(.))'$ *and the expectation of the number of solutions to the empirical moment conditions based on* $g^{(1)}(.)$*, 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* **; (e*)** $\sup_{(x,\phi) \in \aleph \times \mathbf{\Phi}} \|g^{(2)}(x, \phi)\| < \infty$ *where* $\aleph$ *denotes the support of the distribution of the data,* X*.*

Assumption 15(e*) is the only completely new assumption. It guarantees the parameter space to remain compact, while we expand it to obtain just-restricting moment conditions. From a mathematical point of view, Assumption 15(e*) is strong, but it is innocuous in practice because a computer can only handle bounded quantities.

Thanks to the notations introduced in Assumption 15, we transform the over-restricting moment function, $g : \mathbf{R}^p \times \mathbf{\Phi} \to \mathbf{R}^m$, into the just-restricting moment function, $\psi : \mathbf{R}^p \times \mathbf{\Theta} \to \mathbf{R}^m$ with $\mathbf{\Theta} := \mathbf{\Phi} \times \mathbf{\Xi}$, such that

$$\psi(X, \theta) := \begin{pmatrix} g^{(1)}(X, \phi) \\ g^{(2)}(X, \phi) - \xi \end{pmatrix} \tag{15}$$

where $\theta := (\phi' \quad \xi')'$. Our transformation is the same as the one used by Newey and McFadden (p.2232, 1994) for a different purpose. Other transformations of over-restricting moment conditions into just-restricting systems have been introduced to the GMM and saddlepoint literature. Newey and McFadden (1994), Imbens (1997), and Ronchetti and Trojani (2003) use a transformation based on an extended FOC of the GMM objective function. Following Sowell (1996), Sowell (2007, 2009) proposes a transformation based on an orthonormalized extended FOC of the second-step GMM objective function. Czellar and Ronchetti (2010) and Holcblat (2009) use Sowell's transformation respectively to apply saddlepoint approximation to indirect inference tests and to apply bootstrap to GMM. The three main differences between these transformations and ours are the following. First, our transformation requires the econometrician to split the moment conditions into two groups, while the other transformations automatically select q dimensions of the empirical moment conditions through the derivatives of the latter ones. Second, the asymptotic variance of the square root of the sample size multiplied by the difference between a solution to the empirical moment conditions and the population parameter will be greater with our transformation than with the alternatives. This asymptotic advantage of the alternative transformation has been shown to be often a disadvantage in finite sample. One of the main conclusions of the July 1996 issue of the Journal of Business and Economic Statistics is that the identity weighting matrix generally outperforms the optimal weighting matrix for GMM in finite samples. Third, our transformation captures the solutions to empirical moment conditions that correspond to $g^{(1)}(.,.)$, while the existing transformations capture all the local extrema of the GMM objective function so that, in the presence of non-linear moment conditions, indicator functions should be added to discard local maxima (e.g., Sowell, 2007). However, the inclusion of indicator functions breaks the regularity properties of the moment function that we require to define the ESP approximation.[33] Moreover, it does not also discard local minima that do not converge to a global minimum so that conditions for identification are difficult to characterize.

Our transformation (15) combined with Assumption 15 yield Assumption 1 to hold.

Proposition 9.1. *Under Assumption 1 modified according to Assumption 15, Assumption 1 is satisfied for the just-restricting moment function defined in equation (15). In particular, there exists a convex compact set, $\mathbf{\Xi}$, that includes $\left\{g^{(2)}(x,\phi) : (x,\phi) \in \aleph \times \mathbf{\Phi}\right\}$ so that for all $t \in \mathbf{N}$, $\phi \in \mathbf{\Phi}$, and $\omega \in \mathbf{\Omega}$, we can find $\dot{\xi} \in \operatorname{int}(\mathbf{\Xi})$ such that $\frac{1}{T}\sum_{t=1}^{T} g^{(2)}(X_t(\omega),\phi) - \dot{\xi} = 0_{(m-q)\times 1}$.*

Proof. See Appendix A.19 (p. 117). □

Proposition 9.1 mainly means that all the results of section 4.1 hold for over-restricting moment conditions after transformation (15). In particular, the ESP estimand is well-defined. Without loss of generality, we require a large enough compact set $\mathbf{\Xi}$ for mathematical convenience.

Hereafter, for simplicity,we drop reference to transformation (15). The context indicates whether we refer to an original just-restricting moment function $\psi(.,.)$, or to a just-restricting moment function $\psi(.,.)$ built from an over-restricting moment function, $g(.,.)$, according to transformation (15).

[33] A solution to maintain regularity properties would be to introduce mollifiers, but it would create theoretical and practical complications.

9.1.2 ESP estimator

To have the results of section 4.2 to hold, we modify its assumption as follow. Assumption 2 becomes

Assumption 16. *There exists $\varepsilon > 0$ such that for all $x \in \aleph$, $\theta \mapsto g(x,\phi)$ is continuously differentiable in $\{\phi \in \mathbf{R}^q : \|\phi - \mathbf{\Phi}\| < \varepsilon\}$.*

Assumption 3 becomes the following.

Assumption 17. *Define the sets*

$$\begin{aligned}
\hat{\mathbf{\Phi}}_T &:= \left\{ \phi \in \mathbf{\Phi} : \exists \tau_T(\phi) \in \mathbf{R}^q \text{ s.t. } \begin{array}{r} \sum_{t=1}^T g_t^{(1)}(\phi) \mathrm{e}^{\tau_T(\phi)' g_t^{(1)}(\phi)} = 0_{q\times 1} \\ \left|\left[\frac{1}{T}\sum_{t=1}^T \frac{\partial g_t^{(1)}(\phi)'}{\partial \phi}\right]\right|_{det} \neq 0 \\ \left|\left[\frac{1}{T}\sum_{t=1}^T g_t(\phi) g_t(\phi)'\right]\right|_{det} \neq 0 \end{array} \right\} \\
\check{\mathbf{\Phi}}_T &:= \left\{ \phi \in \mathbf{\Phi} : \left|\left[\frac{1}{T}\sum_{t=1}^T \frac{\partial g_t^{(1)}(\phi)'}{\partial \phi}\right]\right|_{det} = 0 \text{ and } \left|\left[\frac{1}{T}\sum_{t=1}^T g_t(\phi) g_t(\phi)'\right]\right|_{det} = 0 \right\}
\end{aligned}$$

For the sample at hand, for all $\eta > 0$ small enough, the sets $\check{\mathbf{\Phi}}_{g,T}$ and $\hat{\mathbf{\Phi}}_{g,T}^{-\eta}$ do not have any common elements, i.e., $\check{\mathbf{\Phi}}_T \cap \hat{\mathbf{\Phi}}_T^{-\eta} = \emptyset$.

The second and third restriction on $\phi \in \mathbf{\Phi}$ in the definition of $\hat{\mathbf{\Phi}}_T$ corresponds to the restriction $|\Sigma_T(\theta)|_{det} \neq 0$ in the definition of $\hat{\mathbf{\Theta}}_T$ in Proposition 4.3 on p.25. Similarly, the two restrictions on $\phi \in \mathbf{\Phi}$ in the definition of $\check{\mathbf{\Phi}}_T$ correspond to the restriction $|\Sigma_T(\theta)|_{det} = 0$ in the definition of $\check{\mathbf{\Theta}}_T$ in Assumption 3 on p.26.

Under the above adapted assumptions we obtain the following.

Proposition 9.2. *Let $\mathbf{\Xi}$ be the compact set introduced in Proposition 9.1. Define the set $\hat{\mathbf{\Xi}}_T \subset \mathbf{R}^{m-q}$ such that*

$$\hat{\mathbf{\Xi}}_T := \left\{ \frac{\sum_{t=1}^T g_t^{(2)}(\phi) \mathrm{e}^{\tau' g_t(\phi)}}{\sum_{t=1}^T \mathrm{e}^{\tau' g_t(\phi)}} : (\phi, \tau) \in \hat{\mathbf{\Phi}}_T \times \mathbf{R}^m \text{ s.t. } \sum_{t=1}^T g_t^{(1)}(\phi) \mathrm{e}^{\tau' g_t(\phi)} = 0_{q\times 1} \right\}.$$

Under Assumptions 1-3 modified according to Assumptions 15-17,

i) there exists $\varepsilon > 0$ such that for all $x \in \mathbf{R}^p$, $\theta \mapsto \psi(x,\theta)$ is continuously differentiable in $\{\theta \in \mathbf{R}^m : \inf(\theta, \mathbf{\Theta}) < \varepsilon\}$;

i) $\hat{\boldsymbol{\Xi}}_T \subset \boldsymbol{\Xi}$ *and* $\hat{\boldsymbol{\Phi}}_T \times \hat{\boldsymbol{\Xi}}_T = \hat{\boldsymbol{\Theta}}_T$*;*

ii) $\check{\boldsymbol{\Phi}}_T \times \boldsymbol{\Xi} = \check{\boldsymbol{\Theta}}_T$.

Then, under the latter assumptions, Assumptions 2-3 hold.

Proof. See Appendix A.20 (p.118). □

Proposition 9.2 means that all the results of section 4.2 hold for over-restricting moment conditions after transformation (15). In particular, the ESP estimator is well-defined. Now, we turn to its asymptotic behaviour.

9.1.3 Asymptotic behavior

Assumption 4(b)-(e) becomes the following.

Assumption 18. **(b*)** *In the parameter space* $\boldsymbol{\Phi}$*, there exists a unique solution* $\phi_0 \in \operatorname{int}(\boldsymbol{\Phi})$ *to the moment conditions* $\mathbb{E}\left[g^{(1)}(X,\phi)\right] = 0_{q\times 1}$ **; (c*)** $\mathbb{E}\left[\sup_{\phi\in\boldsymbol{\Phi}} \|g(X,\phi)\|\right] < \infty$ **; (d*)** $\mathbb{E}\left[\sup_{\phi\in\boldsymbol{\Phi}} \left\|\frac{\partial g(X,\phi)}{\partial\phi'}\right\|\right] < \infty$ **; (e*)** $\left|\mathbb{E}\left[\frac{\partial g^{(1)}(X,\phi)}{\partial\phi'}\right]\right|_{det} \neq 0$.

Assumption 18 means that in terms of identification the assumption is the same as the one to apply the ESP approach to the first q moment conditions. In terms of boundedness, the assumption is the same as the one for a just-restricting moment function.

Assumption 5 becomes the following.

Assumption 19. *Define the set*

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

$$\hat{\boldsymbol{\Xi}}_\infty := \left\{ \frac{\mathbb{E}\left[g^{(2)}(X,\phi) e^{\tau' g(X,\phi)}\right]}{\mathbb{E}\left[e^{\tau' g(X,\phi)}\right]} : (\phi,\tau) \in \hat{\boldsymbol{\Phi}}_\infty \times \mathbf{R}^m \text{ s.t.} \mathbb{E}\left[g^{(1)}(X,\phi) e^{\tau' g(X,\phi)}\right] = 0_{q\times 1} \right\}.$$

(a*) *There exists* $\bar{r} > 0$ *such that there exists* $\dot{T} \in \mathbf{N}$*, so that for all* $T \geqslant \dot{T}$*,* $B_{\bar{r}}(\phi_0) \subset \hat{\boldsymbol{\Phi}}_T$*, and* $B_{\bar{r}}(\xi_0) \subset \hat{\boldsymbol{\Xi}}_T$ *where* $\xi_0 := \mathbb{E}\left[g^{(2)}(X,\phi_0)\right]$*. Define a fixed* $\eta \in]0,\bar{r}[$ **; (b*)** *For all* $\dot{\phi} \in \hat{\boldsymbol{\Phi}}_\infty^{-\eta}$*, for*

$\dot{\tau} \in \mathbf{R}^m$ *and* $\dot{\xi} \in \mathbf{R}^{m-q}$ *s.t.* $\dot{\xi} = \mathbb{E}\left[g^{(2)}(X,\dot{\phi})e^{\dot{\tau}' g(X,\dot{\phi})}\right] / \mathbb{E}\left[e^{\dot{\tau}' g(X,\dot{\phi})}\right]$ *and* $\mathbb{E}\left[g^{(1)}(X,\dot{\phi})e^{\dot{\tau}' g(X,\dot{\phi})}\right] = 0_{q\times 1}$*, there exists* $r_1, r_2 > 0$ *so that for all* $\tau \in B_{r_1}(\dot{\tau})$*,* $\mathbb{E}\left[\sup_{\phi \in B_{r_2}(\dot{\phi})} \|g(X,\phi)e^{\tau' g(X,\phi)}\|\right] < \infty$.

Assumption 6(a)(b) becomes the following.

Assumption 20. **(a*)** *For all* $x \in \mathbf{R}^p$*, the function* $\phi \mapsto g(X,\phi)$ *is four times continuously differentiable in a neighborhood of* ϕ_0 $\mathbb{P}$*-a.s.* **; (b*)** *For all* $k \in [\![1,2]\!]$*, there exists* $r > 0$*, such that* $\mathbb{E}\left[\sup_{\phi \in B_r(\phi_0)} \|D^k g(X,\phi)\|\right] < \infty$ *where* D^k *denotes the differential operator w.r.t.* ϕ *of order* k **;** **(e*)** *There exists* $r > 0$ *such that* $\left\|\mathbb{E}\left[\sup_{\phi \in B_r(\phi_0)} g(X,\phi)g(X,\phi)'\right]\right\| < \infty$.

Assumption 7 becomes the following.

Assumption 21. *Let* $\eta > 0$ *be defined as in Assumption 19(a*).* **(b*)** *For all* $\dot{\phi} \in \hat{\mathbf{\Phi}}_{\infty}^{-\eta}$*, for* $\dot{\tau} \in \mathbf{R}^m$ *and* $\dot{\xi} \in \mathbf{R}^{m-q}$ *s.t.* $\dot{\xi} = \mathbb{E}\left[g^{(2)}(X,\dot{\phi})e^{\dot{\tau}' g(X,\dot{\phi})}\right] / \mathbb{E}\left[e^{\dot{\tau}' g(X,\dot{\phi})}\right]$ *and* $\mathbb{E}\left[g^{(1)}(X,\dot{\phi})e^{\dot{\tau}' g(X,\dot{\phi})}\right] = 0_{q\times 1}$*, there exist* $r_1, r_2 > 0$ *so that* $\mathbb{E}\left[\sup_{(\tau,\phi) \in B_{r_1}(\dot{\tau}) \times B_{r_2}(\dot{\phi})} e^{\tau' g(X,\phi)}\right] < \infty$.

Under the modified assumptions, the original assumptions still hold.

Proposition 9.3. *Under Assumptions 1-7 modified according to Assumptions 15-21, Assumptions 5-7 hold.*

Proof. See Appendix A.21 (p. 119). □

Proposition 9.3 means that consistency and asymptotic normality of the ESP intensity (i.e., Theorems 5.1 and 5.2) hold for over-restricting moment conditions after extension of the parameter space. Robustness to lack of identification can also be derived along the lines of section 5.2.

9.2. Tests of over-restricting moment conditions

In this section, we presents test of over-restricting moment conditions. The idea is to reduce tests of over-restricting moment conditions to tests of a restriction on the parameter space, which are introduced in section 8. More precisely, the idea is to define a test of over-restricting moment conditions as a test of the hypothesis $\xi_0 = 0_{(m-q)\times 1}$, where $\xi_0 := \mathbb{E}\left[g^{(2)}(X,\phi_0)\right]$. First, we implement the idea in the case of decision-theoretic tests . Second, we implement the idea in the case of confidence-region based tests. In this section, we follow the notations and definitions of section 8.1.

We want to define tests of over-restricting moment conditions as tests of the hypothesis $\xi_0 = 0_{(m-q)\times 1}$. But, if the moment conditions are not consistent with data, the auxiliary parameter space, Ξ, which is defined in Proposition 9.1, may not contain $0_{(m-q)\times 1}$ so that the hypothesis is not well-defined. Thus the following assumption is needed.

Assumption 22. *There exists* $r>0$ *such that* $\mathrm{B}_r(0_{(m-q)\times 1})\subset\boldsymbol{\Xi}$.

Assumption 22 does not entail any loss of generality because we can always expand $\boldsymbol{\Xi}$ so that the assumption is satisfied.

9.2.1 Decision-theoretic tests

In this section, we derive tests of over-restricting moment conditions following the ESP testing decision-theoretic framework of section 8. The test of over-restricting moment conditions corresponds to $\boldsymbol{\Theta}_H=\boldsymbol{\Phi}\times\left\{0_{(m-q)\times 1}\right\}$ and $\boldsymbol{\Theta}_A=\boldsymbol{\Phi}\times\left\{\boldsymbol{\Xi}\setminus\left\{0_{(m-q)\times 1}\right\}\right\}$. Thus, a utility functions for a test of over-restricting moment conditions has the following form.

Assumption 23. *For all* $(d,\xi)\in\mathbf{D}\times\boldsymbol{\Xi}$, $u:(d,\theta)\mapsto\left[c_d\delta_{\left\{0_{(m-q)\times 1}\right\}}(\xi)+b_d\mathbf{l}_{\boldsymbol{\Xi}_A}(\xi)\right]\mathbf{l}_{\boldsymbol{\Phi}}(\phi)$ *with* $c_{d_H}>c_{d_A}$, $b_{d_A}>b_{d_H}$, *and* $\boldsymbol{\Xi}_A:=\left\{\boldsymbol{\Xi}\setminus\left\{0_{(m-q)\times 1}\right\}\right\}$.

Assumption 23 is the adaptation of Assumption 14 to tests of over-restricting moment conditions. In Assumption 23, a utility function does not depend on ϕ because the test hypothesis put a constraint only on ξ.

Under Assumption 23, the following proposition reformulates conveniently decision-theoretic ESP tests of over-restricting moment conditions.

Proposition 9.4 (Decision-theoretic ESP test of over-restricting moment conditions)**.** *Denote* $\tilde{f}_{\xi_T^*,sp}(\xi):=\int_{\boldsymbol{\Phi}}\tilde{f}_{\theta_T^*,sp}(\theta)d\phi$. *Under Assumptions 1-3 modified according to Assumptions 15-17 and Assumptions 22-23, the ESP hypothesis test in Definition 8.2 is equivalent to the test*

$$c\tilde{f}_{\xi_T^*,sp}(0_{1\times(m-q)})\geqslant\int_{\boldsymbol{\Phi}\times\boldsymbol{\Xi}_A}\tilde{f}_{\theta_T^*,sp}(\theta)d\theta \tag{16}$$

with $c:=\frac{c_{d_H}-c_{d_A}}{b_{d_A}-b_{d_H}}$, *then* $d_T(\omega)=d_H$; *and otherwise* $d_T(\omega)=d_A$.

Proof. See Appendix A.22 (p.122).□

The meaning of Proposition 9.4 is similar to the one of Propositions 8.3 in section 8. The maximum-ESP approach corresponds to $c=1$.

The following proposition ensures the existence and the double-consistency of decision-theoretic ESP test of over-restricting moment conditions.

Proposition 9.5. *Given a utility function $u(.,.)$,*

i) under Assumptions 1-3 modified according to Assumptions 15-17 and Assumptions 22-23, there exists a decision-theoretic ESP test of over-restricting moment conditions;

ii) under Assumptions 1-7 modified according to Assumptions 15-21 and Assumptions 22-23, a decision-theoretic ESP test of over-restricting moment conditions is doubly-consistent $\mathbb{P}$-a.s.

Proof. See Appendix A.23 (p.122). □

Proposition 9.5 is the counterpart of Proposition 8.4 for the marginal ESP intensity of ξ_T^*, $\tilde{f}_{\xi_T^*,sp}(.)$. Following the arguments of Proposition 8.6, robustness of decision-theoretic ESP tests of over-restricting moment conditions to lack of identification can be established.

9.2.2 Confidence-region based tests

For ESP decision-theoretic tests, each hypothesis requires a specific utility function, which is different from the one used for point estimation. In contrast, ESP confidence-region based tests relies on a unique utility function that correspond to the one used for point estimation. In a ESP confidence-region based test, a point-hypothesis is rejected, if it is far from being the econometrician optimal point estimate. More precisely, in a ESP confidence-region based tests, a point-hypothesis is rejected, if it is outside the corresponding confidence region. In the case of tests of over-restricting moment conditions, which is presented in this section, the hypothesis $\xi_0 = 0_{(m-q)\times 1}$ is rejected, if the marginal ESP confidence region of ξ_T^* does not contain $0_{(m-1)\times 1}$.

In basic standard classical inference, the point-hypotheses outside a confidence region are also the ones rejected. However, standard classical tests are fundamentally different from ESP confidence-region based tests. First, unlike with ESP, the relation between tests and confidence regions is essentially a mathematical coincidence in standard classical inference. The rationale behind standard classical confidence region and tests are different. An ideal standard classical confidence region is a random set that has approximatively a probability $1-\alpha$ to contain the population parameter before collection of the sample, while the hypothesis of a standard classical test is accepted if its fixed region of acceptance contains the corresponding random statistic. Second, in standard classical inference, tests rely on probabilistic statements that hold only before collection of the sample, while ESP tests rely on probabilistic statements that hold before and after collection of a sample. In particular, unlike in standard classical inference, multiple hypothesis testing on the same date set does not undermine the theoretical validity of ESP tests.

For clarity, the structure of this section is similar to the one of section 7, which considers separately continuous utility functions and 0-1 utility function.

Continuous utility functions

By definition confidence-region based tests of over-restricting moment conditions rely on confidence regions, which in turn rely the utility function used used for point estimation. If the utility function is continuous with respect to both the parameter of interest , ϕ, and the auxiliary parameter, ξ, then the ESP confidence region for test of over-restricting moment conditions is the marginal of the confidence region in Definition 7.2 on p.42. However, the choice of a continuous utility function for the inference precision of ξ can be delicate. ξ measures the goodness-of-fit of the model, but it does not come from the economic model under study. Thus, we focus on a utility function that is continuous w.r.t the parameter of interest, ϕ, and 0-1 w.r.t. the auxiliary parameter ξ. In other words, we consider in this section the composite utility function $u_c\left(\theta_e, \theta\right) := u(\phi_e, \phi)\delta_{\xi_e}(\xi)$ for all $(\theta_e, \theta) \in \mathbf{\Theta}^2$. Based on this utility function, we define the following marginal confidence region.

Definition 9.1 (ESP confidence-region for test of over-restricting moment conditions)**.** *Given a utility function $u(.,.)$, define an ESP confidence region of level $1-\alpha$ for over-restricting moment conditions with $\alpha \in [0,1]$ as a $\mathcal{B}(\mathbf{\Theta})$-measurable set*

$$\tilde{S}^{u_c}_{T,\xi} := \left\{ \xi_e \in \mathbf{\Xi} : \frac{1}{K^{u_c}_{T,\xi}} f^{u_c}_{\xi^*_T, sp}(\xi_e) \geqslant k_{\alpha,T,\xi} \right\}$$

where $k_{\alpha,T,\xi}$ is the highest bound satisfying $\frac{1}{K^{u_c}_T}\int_{\tilde{S}^{u_c}_{T,\xi}} f^{u_c}_{\xi^*_T,sp}(\xi_e)d\xi_e \geqslant 1-\alpha$, $K^{u_c}_{T,\xi} := \int_{\mathbf{\Xi}} f^{u_c}_{\xi^*_T,sp}(\xi_e)d\xi_e$, and $f^{u_c}_{\xi^*_T,sp}(.) := \int_{\mathbf{\Phi}^2} u(\phi_e,\phi)\tilde{f}_{\theta^*_T,sp}(\phi,.)d\phi d\phi_e$.

$f^{u_c}_{\xi^*_T,sp}(.)$ corresponds to the the utility-weighted marginal ESP intensity of ξ^*_T. Definition of ESP confidence regions for over-restricting moment conditions is in line with Definition 7.2 and Definition 7.5 of ESP confidence regions for continuous and 0-1 utility functions because $\int_{\mathbf{\Theta}} u(\phi_e,\phi)\delta_{\xi_e}(\xi)\tilde{f}_{\theta^*_T,sp}(\theta)d\theta = \int_{\mathbf{\Phi}} u(\phi_e,\phi)\tilde{f}_{\theta^*_T,sp}(\phi,\xi_e)d\phi$.

Proposition 9.6. *For all* $\alpha \in]0,1[$,

i) under Assumptions 1-3 modified according to Assumptions 15-17, and Assumptions 11- 12,22, there exists an ESP confidence region for over-restricting moment conditions, $\tilde{S}^{u_c}_{T,\xi}$*;*

ii) under Assumptions 1-7 modified according to Assumptions 15-21, and Assumptions 11- 12,22, as $T \to \infty$*,*

$$\tilde{S}^{u_c}_{T,\xi} \rightsquigarrow \{\xi_0\} \quad \mathbb{P}\text{-a.s.}$$

Proof. See Appendix A.24 (p.122). □

Proposition 9.6 is the counterpart of Proposition 7.6.

Definition 9.2 (Confidence-region based test of over-restricting moment conditions)**.** *Under Assumptions 1-3 modified according to Assumptions 15-17 and Assumptions 11- 12, a confidence-region based ESP test of over-restricting moment conditions is the test* $d_T(.)$ *such that for all* $\omega \in \mathbf{\Omega}$ *if*

$$0_{(m-q)\times 1} \in \tilde{S}^{u_c}_{T,\xi}$$

then $d_T(\omega) = d_H$*, and otherwise* $d_T(\omega) = d_A$*.*

Definition means that, given a utility function, that a model is rejected at level α, if the marginal ESP confidence region of ξ of level α does not contain $0_{(m-1)\times 1}$. In other words, there is rejection, if considering that the over-restricting moment conditions are satisfied is far from being an optimal answer to the uncertainty summarized by the ESP intensity.

0-1 utility functions

Confidence-based ESP tests of over-restricting moment conditions are simpler when the utility function for the parameter of interest is a 0-1 utility function. The counterpart of Definition 9.1 is the following.

Definition 9.3 (Maximum ESP confidence region for test of over-restricting moment conditions)**.** *A maximum ESP confidence region of level* $1-\alpha$ *with* $\alpha \in [0,1]$ *is a* $\mathcal{B}(\mathbf{\Theta})$*-measurable set*

$$\tilde{S}_{T,\xi} := \left\{ \xi_e \in \mathbf{\Xi} : \frac{1}{K_{T,\xi}} \tilde{f}_{\xi^*_T,sp}(\xi_e) \geqslant k_{\alpha,T} \right\}$$

where $k_{\alpha,T}$ *is the highest bound satisfying* $\frac{1}{K_{T,\xi}}\int_{\tilde{S}_{T,\xi}}\tilde{f}_{\xi_T^*,sp}(\xi)d\xi \geqslant 1-\alpha$*,* $K_{T,\xi} := \int_{\mathbf{\Xi}}\tilde{f}_{\xi_T^*,sp}(\xi)d\xi$*, and* $\tilde{f}_{\xi_T^*,sp}(\xi_e) := \int_{\mathbf{\Phi}}\tilde{f}_{\theta_T^*,sp}(\phi,\xi_e)d\phi$*.*

The counterpart of Proposition 9.6 is the following.

Proposition 9.7. *For all* $\alpha \in]0,1[$*,*

i) under Assumptions 1-3 modified according to Assumptions 15-17, and Assumptions 11- 12,22, there exists an ESP confidence region for over-restricting moment conditions, $\tilde{S}_{T,\xi}$*;*

ii) under Assumptions 1-7 modified according to Assumptions 15-21, and Assumptions 11- 12,22,, as $T \to \infty$*,*

$$\tilde{S}_{T,\xi} \rightsquigarrow \{\xi_0\} \quad \mathbb{P}\text{-a.s.}$$

Proof. See Appendix A.25 (p.123). □

The counterpart of Definition 9.2 for 0-1 utility function is the following.

Definition 9.4 (Maximum ESP confidence-region based test of over-restricting moment conditions)**.** *Under Assumptions 1-3 modified according to Assumptions 15-17 and Assumptions 11- 12, a confidence-region based ESP test of over-restricting moment conditions is the test* $d_T(.)$ *such that for all* $\omega \in \mathbf{\Omega}$ *if*

$$0_{(m-q)\times 1} \in \tilde{S}_{T,\xi}$$

then $d_T(\omega) = d_H$*, and otherwise* $d_T(\omega) = d_A$*.*

Definition 9.4 means that there is rejection if the ESP probability weight that over-restricting moment conditions are satisfied is low.

10. Simulation and inference of a consumption-based asset pricing model

In this section, we illustrate the benefits provided by the ESP approach with respect to existing inference approaches in the case of a stylized consumption-based asset pricing model.

As explained in section 2, the classical inference theory used in consumption-based asset pricing is *logically* irrelevant in practice. Asymptotic classical inference theory is about situations in which the sample size can be infinitely increased, although practice relies on bounded samples.[34] The ESP approach is also *logically* irrelevant in practice. An approximation of the distribution of the solutions to the empirical moment conditions is *logically* irrelevant for an actual solution to the moment conditions, the population parameter.

However, asymptotic classical inference theory has been proved helpful in many practical situations. The previous sections suggest that we can expect the same from the ESP approach. Consequently, the appropriate question is to determine whether the ESP approach is *practically* more relevant than the classical inference theory. The simulations in this section suggest that it is the case, although the comparison is based on the criteria of standard classical inference theory.

10.1. Model

We consider a standard consumption-based asset pricing model. An infinitely-lived agent represents the economy. He is endowed with CRRA preferences. Thus, he maximizes

$$\max_{\{C_t\}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\theta}}{1-\theta}\right] \tag{17}$$

where $\beta \in]0,1[$ denotes in this section his time discount factor, θ is his RRA and C_t is his consumption in period t. At each period, the representative agent can consume and invest in n different assets according to his wealth. More precisely, he faces the following budget constraint

$$C_t + \sum_{j=1}^{n} P_{j,t} Q_{j,t} \leqslant \sum_{j=1}^{n} (P_{j,t} + D_{t,j}) Q_{j,t-1} \tag{18}$$

where $Q_{j,t}$ is the quantity of asset j held at the end of date t, and $P_{j,t}$ and $D_{j,t}$ are, respectively, the price and dividend of asset j at date t. Under the usual technical conditions (e.g., Radner, 1972; Lucas, 1978), the solution to the maximization program of the representative agent (17)-(18) satisfies the following Euler equations

$$\forall j \in [\![1,n]\!], \qquad \mathbb{E}_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\theta} R_{j,t+1} - 1\right] = 0 \tag{19}$$

[34] van der Vaart (p. 3, 1998) makes a similar remark. He writes “In fact, strictly speaking, most asymptotic results that are currently available are logically useless. This is because most asymptotic results are limit results, rather than approximations consisting of an approximating formula plus an accurate error bound.”

where $R_{j,t+1} := \frac{P_{j,t+1}}{P_{j,t}}$ is the gross return of asset j between date t and $t+1$. Euler equations (19) mean that the expected gross return of every asset discounted for time and risk equals \$1. This relation corresponds to the moment conditions (1) (p. 11) with $M_{t+1}(\theta) = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\theta}$. This model has been used extensively in the empirical consumption-based asset pricing literature since the seminal paper by Hansen and Singleton (1982).

10.2. Specifications

We want to simulate simple specifications of the model (17)-(18) so that the moment conditions (19) can be solved in closed-form. This is necessary in order to know whether the model is identified. We also want our moment conditions to correspond to moment conditions already used in the literature to assess inference approaches. The resulting specifications, albeit not necessarily realistic, allow a transparent illustration of the abilities of the ESP approach.

We assume the logarithm of gross growth consumption to be independent and identically distributed (i.i.d.) according to a centered Gaussian distribution with variance σ^2, i.e., $\ln\left(\frac{C_{t+1}}{C_t}\right) \hookrightarrow \mathcal{N}(0, \sigma^2)$. We assume the gross return of only one asset (asset 1) to be observed. We assume the logarithm of the gross return of the asset to be i.i.d. according to a centered Gaussian distribution with variance $9\sigma^2$, i.e., $\ln\left(R_{1,t+1}\right) \hookrightarrow \mathcal{N}(0, 9\sigma^2)$. Gross growth consumption and the gross return of the asset are independent. We set $\sigma^2 = .2$. We also set the RRA and time discount factor of the representative agent, respectively, to 3 and $\mathrm{e}^{-\frac{\theta_0^2\sigma^2}{2}}$, i.e., $\theta_0 = 3$ and $\beta = \mathrm{e}^{-\frac{(3\times.2)^2}{2}} \simeq .83$.

We assume the econometrician knows the value of the time discount factor. The econometrician wants to estimate the RRA of the representative agent. We distinguish two cases, a case with lack of identification, and a well-identified case.

Well-identified case

We assume that the econometrician only observes the logarithm of gross growth consumption with a noise equal to the logarithm of the gross return of the asset divided by 3. We also assume that the econometrician uses, as an instrument, a variable equal to the noise. In other words, although the econometrician thinks his moment condition is (19) with $j = 1$, his actual moment condition is

$$\mathbb{E}\left\{\left[\beta_0\left(\widehat{\frac{C_{t+1}}{C_t}}\right)^{-\theta} R_{t+1} - 1\right] Y_t\right\} = 0$$

where $\log\widehat{\left(\frac{C_{t+1}}{C_t}\right)} := \log\left(\frac{C_{t+1}}{C_t}\right) + \frac{1}{3}\log\left(R_{t+1}\right)$ and $Y_t := \left[\frac{1}{3}\log\left(R_{t+1}\right)\right]$. Despite the noise, the population parameter (i.e., the RRA of the representative agent) is the only solution to the moment condition.[35] This moment condition has also been used extensively in the econometric literature (e.g., Hall and Horowitz, 1996; Gregory, Lamarche and Smith, 2002).

Case with lack of identification

We keep the same specification as in the case with identification, except there is no instrument. In other words, although the econometrician thinks his moment condition is (19) with $j = 1$, his actual moment condition is

$$\mathbb{E}\left[\beta_0\left(\widehat{\frac{C_{t+1}}{C_t}}\right)^{-\theta} R_{t+1} - 1\right] = 0 \tag{20}$$

with $\log\widehat{\left(\frac{C_{t+1}}{C_t}\right)} := \log\left(\frac{C_{t+1}}{C_t}\right) + \frac{1}{3}\log\left(R_{1,t+1}\right)$. The population parameter is a solution to the moment condition (20); however, there is another solution $\theta = 0$.[36] The existence of two solutions to the moment conditions is necessary to produce a GMM objective function with multiple local minima in a simple model. The moment condition (20) has been used extensively in the econometric literature (e.g., Hall and Horowitz, 1996; Gregory, Lamarche and Smith, 2002).

10.3. Simulations results

In each case, we draw $1,000$ samples. For each sample, we apply standard GMM (see Hansen, 1982), standard continuously updated GMM (see Hansen, Heaton and Yaron, 1996), which we denote CU, continuously updated GMM for lack of identification (see Stock and Wright, 2000), which we denote LCU (low continuously updated), and ESP. The regularization of the ESP intensity for $\theta \in$

[35] This can be shown thanks to Laplace transforms of Gaussian distributions.

[36] See footnote 36 on p. 69.

$\hat{\mathbf{\Theta}}_T \cap \left(\hat{\mathbf{\Theta}}_T^{-\eta}\right)^c$ does not appear necessary (see Definition 4.4 on p. 26). We use the same fixed grid for each approach with increment .01. The bounds of the grid are determined such that ESP intensity is negligeable beyond.

Well-identified case

The grid goes from -15 to 20. We report the average of the bias $\hat{\theta} - \theta_0$ and the average of the square error $(\hat{\theta} - \theta_0)^2$. We also consider confidence intervals of level 10% and 5%. We report their average length and the proportion of those that do not contain the population parameter. These are standard criteria of assessment of classical inference approaches (e.g., Hall and Horowitz, 1996; Gregory, Lamarche and Smith, 2002). For LCU, we can only report a lower bound of the average length because, by construction, the maximum length is the distance between the two bounds of the grid.

The results are reported in Table 1. The different approaches perform similarly, except for LCU. The average length of LCU is significantly larger than that of any other approaches for a comparable empirical level. As documented in the literature (e.g., Hansen, Heaton and Yaron, 1996), CU objective functions tend to be flat and low in the tail. The LCU confidence intervals are

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

where c_α is the α quantile of a chi-square of degree 1 and $Q_{T,CU}(.)$ is the continuously updated GMM objective function, i.e., $Q_{T,CU}(\theta) := \left[\frac{1}{T}\sum_{t=1}^T \psi_t(\theta)\right]' \left[\frac{1}{T}\sum_{t=1}^T \psi_t(\theta)\psi_t(\theta)'\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^T \psi_t(\theta)\right]$. Therefore, the LCU confidence intervals are usually huge.

				α=10		α=5	
T	Method	Bias	MSE	Level	Length	Level	Length
50	GMM	0.089	0.303	12.2	1.618	7.5	1.928‡
	CU	0.111	0.620	12.2	1.711	7.5	2.039
	LCU	-	-	9.2‡	*16.562*	4.2‡	*22.123*
	ESP	-0.060‡	0.189‡	11.7	1.603‡	6.2	2.008
100	GMM	0.038‡	0.127	10.4	1.095‡	6.2	1.305‡
	CU	0.046	0.701	10.5	1.212	6.3	1.444
	LCU	-	-	9.3	*11.928*	3.4	*17.064*
	ESP	-0.042	0.097‡	10.1‡	1.100	5.2‡	1.348

Table 1: Monte Carlo evaluation in the well-identified case. The symbol ‡ highlights the best performing method. Lengths in italics are lower bounds. Levels are in percentage.

Case with lack of identification

T	Method	α=10 Level	Length	α=5 Level	Length
50	GMM	51.1	141.309	48.1	168.380
	CU	52.7	29.391	49.7	35.022
	LCU	11.5	*20.026*	5.9	*27.183*
	ESP	9.2‡	4.932‡	5.2‡	5.993‡
100	GMM	53.25	17.686	51.25	21.074
	CU	52.95	9.453	51.15	11.263
	LCU	9.25‡	*11.993*	5.65	*17.009*
	ESP	8.4	3.567‡	4.85‡	4.292‡

Table 2: Monte Carlo evaluation in the case with lack of identification. The symbol ‡ highlights the best performing method. Lengths in italics are lower bounds. Levels are in percentage.

The grid goes from -17 to 22. Given the presence of two solutions to the moment conditions, we do not report performance criteria for point estimates. We report confidence intervals of level 5% and 10%.

The large average lengths reported for GMM confidence intervals are due to infrequent draws with large estimated variance. The estimated variance is large when the objective function is locally almost flat. This phenomenon also occurs with CU confidence intervals to a lesser extent. Even if we ignore these problems, standard GMM and CU confidence intervals perform poorly. They do not contain the population RRA with a much higher probability than the nominal level. In contrast, the empirical levels of LCU and ESP confidence intervals remain close to the nominal levels. However, as in the well-identified case, the ESP confidence intervals clearly outperform the LCU confidence intervals in terms of length.

11. Conclusion

Several areas such as empirical consumption-based asset pricing have been a challenge to standard moment-based inference approaches. This dissertation 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 bounded size. More precisely, the starting point is the acknowledgement that moment-based inference is based on the use of a finite-sample counterpart of the population 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 population parameter thanks to the saddlepoint technique. The result of this approximation, the ESP intensity, summarizes in probabilistic terms the estimated uncertainty about the population 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. Nevertheless, we prove that the ESP framework enjoys good asymptotic properties. In addition, we prove the inherent robustness of the ESP approach to lack of identification. We also explain why the ESP approach provides a unique answer to multiple theoretical concerns, such as asymptotic testing error, and practical concerns, such as confidence region outside the domain of definition of a model, that are faced by standard classical inference. Simulations confirm the practical relevance of the theoretical properties of the ESP approach. All this contributes to the literature in several directions.

Nevertheless, more can be achieved. From an empirical point of view, it would be interesting to apply the ESP approach to real data in consumption-based asset pricing and other areas. There is a work in progress in which consumption-based asset pricing models are estimated. From a theoretical point of view, the flexibility of the ESP framework opens several possibilities. In particular, the author has a work in progress in which he shows how to introduce exogenous information and robustness considerations into inference with a view to portfolio choice. The ESP framework and the ideas behind it seem to be promising avenues for further research.

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, *Annals of Mathematical Statistics* **20**(1), 46–63.

Backus, D. K., Routledge, B. R. and Zin, S. E.: September 2008 (first draft June 2007), Who holds risky assets?, *working paper* .

Backus, D. K. and Zin, S. E.: 1994, Reverse engineering the yield curve, *NBER working paper No. 4676* .

Bansal, R., Kiku, D. and Yaron, A.: 2007, Risks for the long run: Estimation and inference, *working paper* .

Bansal, R. and Yaron, A.: 2004, Risks for the long run: a potential resolution of asset pricing puzzles, *Journal of Finance* **59**(4), 1481–1509.

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.

Berk, R. H.: 1972, Consistency and asymptotic normality of MLE's for exponential models, *The Annals of Mathematical Statistics* **43**(1), 193–204.

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

Boguth, O. and Kuehn, L.-A.: 2011, Consumption volatility risk, *working paper* .

Brown, B. W. and Newey, W. K.: 2002, Generalized method of moments, efficient bootstrapping, and improved inference, *Journal of Business and Economic Statistics* **20**(4), 507–517.

Calin, O. L., Chen, Y., Cosimano, T. F. and Himonas, A. A.: 2005, Solving asset pricing models when the price-dividend function is analytic, *Econometrica* **73**(3), 961–982.

Campbell, J. Y. and Cochrane, J. H.: 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *The Journal of Political Economy* **107**(2), 205–251.

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.

Chen, X., Favilukis, J. and Ludvigson, S. C.: 2008, An estimation of economic models with recursive preferences, *working paper* .

Chen, X. and Ludvigson, S. C.: 2009, Land of addicts? An empirical investigation of habit-based asset pricing models, *Journal of Applied Econometrics* **24**, 1057–1093.

Chen, Y., Cosimano, T. F. and Himonas, A. A.: 2008a, Analytic solving of asset pricing models: The by force of habit case, *Journal of Economic Dynamics and Control* **32**, 3631–3660.

Chen, Y., Cosimano, T. F. and Himonas, A. A.: 2008b, Continuous time one-dimensional asset pricing models with analytic price-dividend functions, *working paper (forthcoming in Economic Theory)* .

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

Constantinides, G. M. and Ghosh, A.: 2008, Asset pricing tests with long run risks in consumption growth, *working paper* .

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.

D'Haultfoeuille, X.: 2009, *Essai sur quelques problèmes d'identification en économie*, PhD thesis, Université Paris 1 Panthéon-Sorbonne and Paris School of Economics.

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.

Epstein, L. G. and Zin, S. E.: 1991, Risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis, *The Journal of Political Economy* **99**(2), 263–286.

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

Flodén, M.: 2008, A note on the accuracy of Markov-chain approximations to highly persistent ar(1) processes, *Economics Letters* **99**, 516–520.

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

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.

Gelman, A. and Shalizi, C. R.: 2011, Philosophy and the practice of Bayesian statistics, *Forthcoming in British Journal of Mathematical and Statistical Psychology* .

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.

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.

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

Grossman, S. J. and Shiller, R. J.: 1981, The determinants of the variability of stock market prices, *American Economic Review* **71**(2), 222–227.

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

Guvenen, F.: 2009, A parsimonous macroeconomic model for asset pricing, *Econometrica* **77**(6), 1711–1750.

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

Hall, P. and Horowitz, J. L.: 1996, Bootstrap critical values for tests based on generalized-method-of-moments estimators, *Econometrica* **64**(4), 891–916.

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

Hansen, L. P. and Heaton, J. C.: 2008, Consumption strikes back? measuring long-run risk, *Journal of Political Economy* **116**(21), 260–302.

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.

Hansen, P. L., Heaton, J. C., Lee, J. and Roussanov, N.: 2007, *Handbook of Econometrics*, Vol. 6A, North-Holland, chapter 61, pp. 3967–4056.

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.

Holcblat, B.: 2009, Bootstrap for linear GMM without recentering, *Working paper* .

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.

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

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

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

Judd, K. L.: 1998, *Numerical Methods in Economics*, MIT Press.

Julliard, C. and Ghosh, A.: (2007) 2010, Can rare events explain the equity premium puzzle?, *working paper* .

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. 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.

Kydland, F. E. and Prescott, E. C.: 1982, Time to build and aggregate fluctuations, *Econometrica* **50**(6), 1345–1370.

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.

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.

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.

Mehra, R. and Prescott, E. C.: 1985, The equity premium a puzzle, *Journal of Monetary Economics* **14**, 145–161.

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

Nakamura, E., Steinsson, J., Barro, R. and Ursúa, J.: 2010, Crises and recoveries in an empirical model of consumption disasters, *working paper* .

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.

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.

Radner, R.: 1972, Existence of equilibrium of plans, prices, and price expectations in a sequence of markets, *Econometrica* **40**(2), 289–303.

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

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.

Rubin, D. B.: 1984, Bayesianly justifiable and relevant frequency calculations for the applied statistician, *Annals of Statistics* **12**(4), 1151–1172.

Rust, J.: 1994, *Handbook of econometrics*, Vol. 4, Elsevier Science, chapter 51 "Structural Estimation of Markov Decision Processes", pp. 3082–3143.

Saporta, G.: (1990) 2006, *Probabilités, Analyse des données et Statistique*, second edn, Technip.

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

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* .

Singleton, K. J.: 2006, *Empirical Dynamic Asset Pricing, Model Specification and Econometric Assessment*, Princeton University Press.

Skiadas, C.: 2009, *Asset Pricing Theory*, Princeton Series in Finance, Princeton University Press.

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.

Tauchen, G.: 1986, Statistical properties of generalized method-of-moments estimators of structural parameters, *Journal of Business and Economic Statistics* **4**(4), 397–416.

van Binsbergen, J. H., Fernández-Villaverde, J., Koijen, R. S. and Rubio-Ramírez, J. F.: 2010, The term structure of interest rates in a DSGE model with recursive preferences, *working paper* .

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.

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

A. Supplementary results and proofs

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, 2008) it is sufficient to prove that there exists a function $\omega \mapsto N_T(\omega, .)$ such that for any given $A \in \mathcal{B}(\boldsymbol{\Theta})$, $\omega \mapsto N_T(\omega, A)$ 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 $A \in \mathcal{B}(\boldsymbol{\Theta})$.

By the Lemma A.1 below, if a set $P \in \mathcal{B}(\mathbf{R}^{pT}) \otimes \mathcal{B}(\boldsymbol{\Theta})$, 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, $P_{\underline{x}_T} := \{\theta \in \boldsymbol{\Theta} : (\underline{x}_T, \theta) \in P\}$ 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, since 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.e. and Assumption 1(a) presents that $(\boldsymbol{\Omega}, \mathcal{E}, \mathbb{P})$ is complete. Thus, there exists a $\mathcal{E}/\mathcal{B}(\mathbf{N})$-measurable function $\omega \mapsto N_T(\omega, A)$ such that

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

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

Lemma A.1. *For all $A \in \mathcal{B}(\boldsymbol{\Theta})$, $\forall P \in \mathcal{B}(\mathbf{R}^{pT}) \otimes \mathcal{B}(\boldsymbol{\Theta})$, $\underline{x}_T \mapsto \nu(P_{\underline{x}_T} \cap A)$ is $\mathcal{B}\left((\mathbf{R}^p)^T\right)/\mathcal{B}(\overline{\mathbf{N}})$-measurable.*

Proof . Let $A \in \mathcal{B}(\boldsymbol{\Theta})$. Define

$$\mathcal{H}_A := \left\{ h(.) : \begin{array}{c} h(.) \text{ is bounded} \\ h(.) \text{ is } \mathcal{B}(\mathbf{R}^{pT}) \otimes \mathcal{B}(\boldsymbol{\Theta})/\mathcal{B}(\mathbf{R})\text{-measurable} \\ \underline{x}_T \mapsto \int_A h(\underline{x}_T, \theta)\nu(d\theta) \text{ is } \mathcal{B}(\mathbf{R}^{pT})/\mathcal{B}(\overline{\mathbf{R}})\text{-measurable} \end{array} \right\}$$

Obviously, $\mathcal{H}_A$ is a $\mathbf{R}$-vector space, and it contains the constant function 1. Moreover, $\{h_n(.)\}_{n \geqslant 1}$ is a sequence of non-negative functions in $\mathcal{H}_A$ such that $h_n(.) \uparrow h(.)$ where $h(.)$ is a bounded function on $\mathbf{R}^{pT} \times \boldsymbol{\Theta}$, then $h(.) \in \mathcal{H}_A$ by the preservation of measurability under limit and the Lebesgue monotone convergence Theorem. In addition, $\mathcal{H}_A$ contains the indicator function of every set in the π-system consisting of measurable rectangles, $\mathcal{I} := \left\{R : R := R_{\mathbf{R}^{pT}} \times R_{\mathbf{\Theta}} \text{ with } R_{\mathbf{R}^{pT}} \in \mathcal{B}(\mathbf{R}^{pT}) \wedge R_{\mathbf{\Theta}} \in \mathcal{B}(\mathbf{\Theta})\right\}$ (an intersection of two measurable rectangles is a measurable rectangle) ; because

$$
\begin{aligned}
\mathbf{l}_R(\underline{x}_T, \theta) &= \mathbf{l}_{R_{\mathbf{R}^{pT}}}(\underline{x}_T)\mathbf{l}_{R_{\mathbf{\Theta}}}(\theta) \\
\int_A \mathbf{l}_R(\underline{x}_T, \theta)\nu(d\theta) &= \mathbf{l}_{R_{\mathbf{R}^{pT}}}(\underline{x}_T) \int_A \mathbf{l}_{R_{\mathbf{\Theta}}}(\theta)\nu(d\theta) \\
\nu(R_{\underline{x}_T}) &= \mathbf{l}_{R_{\mathbf{R}^{pT}}}(\underline{x}_T)\nu(R_{\mathbf{\Theta}} \cap A)
\end{aligned}
$$

and because $\sigma(\mathcal{I}) = \mathcal{B}(\mathbf{R}^{pT}) \otimes \mathcal{B}(\mathbf{\Theta})$ (e.g., paragraph II.11 in Rogers and Williams, 2000) . Consequently, by the monotone class Theorem 3.1 from Rogers and Williams (2000), $\forall P \in \mathcal{B}(\mathbf{R}^{pT}) \otimes \mathcal{B}(\mathbf{\Theta})$, $\mathbf{l}_P(.) \in \mathcal{H}_A$, which in turn implies that $\underline{x}_T \mapsto \nu(P_{\underline{x}_T} \cap A)$ is $\mathcal{B}\left((\mathbf{R}^p)^T\right) / \mathcal{B}(\overline{\mathbf{N}})$-measurable. $\square$

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* $\mathcal{T}_n(A) := \{A_{n,i} \cap A : i = 1, \dots, k_n \text{ and } A_{n,i} \in \mathcal{T}_n\}$ *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, 2008). For completeness, we provide a proof adapted to our framework.

By Definition 8.3 of an intensity measure,

$$
\begin{aligned}
\mathbb{F}_T(A) &= \lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbb{P}\{N_T(A_{n,i}) = 1\} = \lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbb{E}\left[\mathbf{l}_{\{N_T(A_{n,i})=1\}}\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbf{l}_{\{N_T(A_{n,i})=1\}}\right].
\end{aligned}
$$

where the last equality comes from Fubini-Tonelli theorem. Now, for all A and B in $\mathcal{B}(\mathbf{\Theta})$,

$$\mathbf{1}_{\{N_T(A\cup B)=1\}} \leqslant \mathbf{1}_{\{N_T(A)=1\}} + \mathbf{1}_{\{N_T(B)=1\}}. \tag{21}$$

Thus, apply Lebesgue monotone convergence theorem to deduce

$$\mathbb{F}_T(A) \;=\; \mathbb{E}\left[\lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbf{1}_{\{N_T(A_{n,i})=1\}}\right] = \mathbb{E}\left[N_T(A)\right], \tag{22}$$

where the last equality comes from Lemma A.3 below. Obviously, $\mathbb{F}_T(.)$ does not depend on the dissecting system chosen. Since $N_T(.)$ is a measure, $\mathbb{F}_T(\emptyset) = \mathbb{E}0 = 0$. Moreover, by the σ-additivity of $N_T(.)$, and Fubini-Tonelli theorem, $\mathbb{F}_T(.)$ is σ-additive. Thus, $\mathbb{F}_T(.)$ is a measure. By Assumption 1(d), $\mathbb{F}_T(.)$ is also finite $\mathbb{P}$a.s.$\square$

Lemma A.3. *Under Assumptions 1,* $\forall A \in \mathcal{B}(\mathbf{\Theta})$, $\lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbf{1}_{\{N_T(A_{n,i})=1\}} = N_T(A)$ $\mathbb{P}$*-a.s.*

***Proof* .** W.l.o.g., for all $n \geqslant 1$, change the numbering of $A_{n,i}$ w.r.t. i so that

$$\begin{cases} \sum_{i:An,i\in\mathcal{T}_n(A)} \mathbf{1}_{\{N_T(A_{n,i})=1\}} = \sum_{i=1}^{N_T(A)} \mathbf{1}_{\{N_T(A_{n,i})=1\}}\,, \quad \forall n \geqslant 1; \\ A_{n+1,i} \subset A_{n,i}\,, \quad \forall i \in [\![1, N_T(A)]\!],\, \forall n \geqslant 1. \end{cases}$$

Thus, $\forall i \in [\![1, N_T(A)]\!]$, $A_{n,i} \downarrow \{\theta_i\}$ where $\theta_i \in A$ and $\frac{1}{T}\sum_{t=1}^{T} \psi_t(\theta_i) = 0$. Then, by the separating property of dissecting systems, $\forall i \in [\![1, N_T(A)]\!]$, $\exists n_i \in \mathbf{N}$ such that $\forall n \geqslant n_i$, $\mathbf{1}_{\{N_T(A_{n,i})=1\}} = 1 = N_T(\theta_i)$. It means that $\forall i \in [\![1, N_T(A)]\!]$, $\mathbf{1}_{\{N_T(A_{n,i})=1\}} \to \mathbf{1}_{\{N_T(\theta_i)=1\}} = N_T(\theta_i)$ as $n \to \infty$. Now, by Assumption 1(d), $N_T(A)$ is finite $\mathbb{P}$-a.s. The result follows immediately. $\square$

A.3. Proof of Proposition 4.2

It is a consequence of equation (9.3.24) p.48 in Daley and Vere-Jones (2008). For completeness, a similar proof is given. If $\mathbb{F}_T(\mathbf{\Theta}) = 0$, the result is immediate by Lemma A.4 below. If $\mathbb{F}_T(\mathbf{\Theta}) > 0$, the idea is to prove the existence of the limit of $\mathbb{P}\left\{N_T(A_n(\theta)) = 1\right\}/\mathbb{F}_T(A_n(\theta))$ as $n \to \infty$ by Doob's martingale convergence theorem, and then to show the limit can only equal 1.

Define for this proof the probability space $(\mathbf{\Theta}, \mathcal{B}(\mathbf{\Theta}), \mathrm{I\!P})$ with $\mathrm{I\!P} := \mathbb{F}_T(.)/\mathbb{F}_T(\mathbf{\Theta})$ and the following sequence of random variables on it

$$Y_n(\theta) := \sum_{i\in I_n} \mathbf{1}_{A_{n,i}}(\theta)\frac{\mathbb{P}\left\{N(A_{n,i})=1\right\}}{\mathbb{F}_T(A_{ni})}$$

where $I_n := \{j \in [\![1, k_n]\!] : \mathbb{F}_T(A_{n,i}) > 0\}$. Denote $\{\mathcal{F}_n\}_{n\geqslant 1}$ the filtration $\mathcal{F}_n := \sigma\{Y_k : k \in [\![1, n]\!]\}$.

Next show that $(\{Y_n\}_{n\geqslant 1}, \{\mathcal{F}_n\}_{n\geqslant 1}, \mathbb{P})$ is an L^1-bounded submartingale. By construction $\{Y_n\}$ is $\{\mathcal{F}_n\}$-adapted. It remains to show that $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] \geqslant Y_n$ $\mathbb{P}$-a.s.. For all $i \in I_n$, denote $y_{n,i} := \mathbb{P}\{N(A_{n,i}) = 1\}/\mathbb{F}_T(A_{n,i})$. For all $\theta \in \mathbf{\Theta}$

$$\begin{aligned}
\mathbb{E}[Y_{n+1}|Y_n](\theta) &= \sum_{i\in I_n}\sum_{j\in I_{n+1}:A_{n+1,j}\subset A_{n,i}} y_{n+1,j}\mathbb{P}\{Y_{n+1,j}=y_{n+1,j}|Y_n=y_{n,j}\}\mathbf{1}_{A_{n,i}}(\theta)\\
&= \sum_{i\in I_n}\sum_{j\in I_{n+1}:A_{n+1,j}\subset A_{n,i}} y_{n+1,j}\frac{\mathbb{F}_T(A_{n+1,j})}{\mathbb{F}_T(A_{n,i})}\mathbf{1}_{A_{n,i}}(\theta)
\end{aligned}$$

because for $j \in I_{n+1}$ s.t. $A_{n+1,j} \subset A_{n,i}$, $\mathbb{P}\{Y_{n+1,j} = y_{n+1,j}|Y_n = y_{n,j}\} = \frac{\mathbb{P}\{Y_{n+1,j}=y_{n+1,j}\cap Y_n=y_{n,j}\}}{\mathbb{P}\{Y_n=y_{n,j}\}} = \frac{\mathbb{P}\{Y_{n+1,j}=y_{n+1,j}\}}{\mathbb{P}\{Y_n=y_{n,j}\}} = \frac{\mathbb{F}_T(A_{n+1,j})}{\mathbb{F}_T(A_{n,i})}$. Now, for all $i \in I_n$,

$$\begin{aligned}
\sum_{j\in I_{n+1}:A_{n+1,j}\subset A_{n,i}} y_{n+1,j}\frac{\mathbb{F}_T(A_{n+1,j})}{\mathbb{F}_T(A_{n,i})} &= \sum_{j\in I_{n+1}:A_{n+1,j}\subset A_{n,i}} \frac{\mathbb{P}\{N(A_{n+1,j})=1\}}{\mathbb{F}_T(A_{n+1,j})}\frac{\mathbb{F}_T(A_{n+1,j})}{\mathbb{F}_T(A_{n,i})}\\
&\geqslant \frac{\mathbb{P}\{N(A_{n,i})=1\}}{\mathbb{F}_T(A_{n,i})} \text{ by Lemma A.4 below.}
\end{aligned}$$

Sum over $i \in I_n$ to deduce $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] \geqslant Y_n$. Now, by Lemma A.4 below, for all $n \geqslant 1$, $|Y_n| \leqslant 1$. Therefore, $\{\mathrm{Y}_n\}$ is an L^1-bounded submartingale.

Apply Doob's martingale convergence theorem (e.g. Theorem 7.18 in Kallenberg, 2001) to deduce $\lim_{n\to\infty} Y_n$ exists $\mathbb{P}$-a.s., and thus $\mathbb{F}_T$-a.e. Since for all $n \geqslant 1$, $|Y_n| \leqslant 1$, by Lebesgue dominated convergence theorem

$$\mathbb{F}_T(\mathbf{\Theta}) = \lim_{n\to\infty}\int_{\mathbf{\Theta}} Y_n\mathbb{F}_T(d\theta) = \int_{\mathbf{\Theta}}\lim_{n\to\infty} Y_n\mathbb{F}_T(d\theta) \leqslant \int_{\mathbf{\Theta}} 1\mathbb{F}_T(d\theta) = \mathbb{F}_T(\mathbf{\Theta})$$

where the first equality comes from the Definition 8.3 of the intensity measure. It follows that $Y_n \to 1$ $\mathbb{F}_T$ a.e. as $n \to \infty$, which in turn implies the result by Lemma A.4 below. □

Lemma A.4. *Let* $\mathbb{F}(.)$ *an intensity measure associated with the point random-field* $N(.)$ *over* $\mathbf{\Theta}$*. Then* $\forall A \in \mathcal{B}(\mathbf{\Theta})$, $\mathbb{F}_T(A) \geqslant \mathbb{P}\left\{N(A) = 1\right\}$.

Proof . Use equation (21) and Definition 8.3 of an intensity measure. □

A.4. Proof of Proposition 4.3

i) Note $\frac{\partial\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\mathrm{e}^{\tau'\psi_t(\theta)}\right]}{\partial\tau} = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)'\mathrm{e}^{\tau'\psi_t(\theta)}$. Thus by Assumption 2, the implicit function theorem is valid in $\hat{\mathbf{\Theta}}_T$. Let $\dot{\theta} \in \hat{\mathbf{\Theta}}_T$. By the implicit function theorem, there exists $r_1 > 0$ so that $\tau_T : B_{r_1}(\dot{\theta}) \mapsto \mathbf{R}^m$ is a C^1 mapping s.t. $\sum_{i=1}^{T}\psi_i(\theta)\mathrm{e}^{\tau_T(\theta)\psi_i(\theta)} = 0_{m\times 1}$, where $B_{r_1}(\dot{\theta})$ denotes an open ball of radius r_1 centered at $\dot{\theta}$ in $\mathbf{\Theta}$. Now, by the stability of continuity under composition, there also exists $B_{r_2}(\dot{\theta})$ with $r_2 > 0$ s.t. $\forall\theta \in B_{r_2}(\dot{\theta})$ $\left|\sum_{t=1}^{T}\psi_t(\theta)\psi_i(\theta)'\mathrm{e}^{\tau_T(\theta)\psi_t(\theta)}\right|_{det} \neq 0$. Thus $B_{\min(r_1,r_2)}(\dot{\theta}) \subset \hat{\mathbf{\Theta}}_T$, which in turn implies that that $\hat{\mathbf{\Theta}}_T$ is open in $\mathbf{\Theta}$.

ii) First, any $\tau_T(.)$ is a C^1 function, by the implicit function theorem and Assumption 2. Thus, $\hat{f}_{\theta_T^*,sp}(.)$ is continuous.

Second, prove uniqueness by contradiction. Assume there exists $\theta \in \hat{\mathbf{\Theta}}_T$ with $\tau^{(1)}, \tau^{(2)} \in \mathbf{R}^m$ such that $\tau^{(1)} \neq \tau^{(2)}$ and for $j = 1, 2$

$$\begin{cases} \sum_{t=1}^{T}\psi_t(\theta)\mathrm{e}^{\tau^{(j)\prime}\psi_t(\theta)} = 0_{m\times 1} \\ \left|\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)'\mathrm{e}^{\tau^{(j)\prime}\psi_t(\theta)}\right|_{det} \neq 0. \end{cases}$$

Then, by Lemma A.5 below, $\tau^{(1)}$ and $\tau^{(2)}$ are two strict local minima of the convex function of $\tau \mapsto \sum_{t=1}^{T}\exp\left[\tau'\psi_t(\theta)\right]$. Now, there is contradiction since a convex function cannot have two distinct strict local minima (e.g. Theorem A p.123 in Roberts and Varberg, 1973) . □

Lemma A.5. *Under Assumptions 1-2, $\tau_T(\theta)$ is a minimum of the convex function $\tau \mapsto \sum_{t=1}^{T}\exp\left[\tau'\psi_t(\theta)\right]$.*

Proof. By Assumption 2, it is sufficient to show that $\frac{\partial\left\{\sum_{t=1}^{T}\exp[\tau'\psi_t(\theta)]\right\}}{\partial\tau\partial\tau'} = \sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)'\mathrm{e}^{\tau'\psi_t(\theta)}$ is a positive semi-definite symmetric matrix (p.s.m).

For all $t \in [\![1, T]\!]$, for all $y \in \mathbf{R}^m$, $y'\psi_t(\theta)\psi_t(\theta)'y$ is a scalar squared. Thus, $\psi_t(\theta)\psi_t(\theta)'$ is a p.s.m. Now a weighted sum of p.s.m with positive weight is a p.s.m. Thus, $\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)'\mathrm{e}^{\tau'\psi_t(\theta)}$ are also p.s.m. □

A.5. Proposition A.1

Proposition A.1. *Define respectively the set where the rough ESP intensity, $f_{\theta_T^*,sp}(.)$, exists, and the set where the empirical moment conditions can be recentered*

$$\overline{\mathbf{\Theta}}_T \quad := \quad \left\{\theta \in \mathbf{\Theta} : \exists \tau_{T,\theta} \in \mathbf{R}^m \text{ s.t. } \sum_{t=1}^{T} \psi_t(\theta) \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} = 0\right\}.$$

Under Assumptions 1 and 2, if for all $\dot{\theta} \in \overline{\mathbf{\Theta}}_T$ the rank of the $m \times 2m$ matrix

$$\left.\frac{\partial\left[\sum_{t=1}^{T} \psi_t(\theta) \mathrm{e}^{\tau\psi_t(\theta)}\right]}{\partial(\theta,\tau)'}\right|_{\substack{\tau=\tau_T(\dot{\theta})\\ \theta=\dot{\theta}}} \tag{23}$$

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 (23) even more difficult.

A.6. Proof of Proposition 4.4

i) To prove measurability write $\tilde{f}_{\theta_T^*,sp}(.)$ with the help of indicator functions. Continuity is obtained by construction.

ii) Continuity of $\tilde{f}_{\theta_T^*,sp}(.)$ over the compact space $\mathbf{\Theta}$ implies finiteness 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. Proof of Proposition 4.5 and Lemma A.6

Lemma A.6. *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 1-3, 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 $A \in \mathcal{B}(\mathbf{\Theta})$*

$$\tilde{\mathbb{F}}_T(A) := \lim_{n\to\infty} \sum_{i:A_{n,i}\in\mathcal{T}_n(A)} \mathbb{P}\{\tilde{N}_T(A_{n,i}) = 1\} \quad , \tag{24}$$

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). For all $\omega \in \mathbf{\Omega}$, for all $A \in \mathcal{B}(\mathbf{\Omega})$,

$$\tilde{N}_T(A) := \frac{\tilde{\mathbb{F}}_T(A)}{\tilde{\mathbb{F}}_T(\mathbf{\Theta})} Y$$

is a point random-field that satisfies equation (24) by equality (22) p.83.□

The existence of a random variable, Y on the probability space $(\mathbf{\Omega}, \mathcal{E}, \mathbb{P})$ with expectation $\mathbb{E}(Y) = \tilde{\mathbb{F}}_T(\Theta)$ 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 4.5. By Lebesgue's differentiation theorem (e.g., Theorem 3.21 in Folland, 1984),

$$\begin{aligned}
\tilde{f}_{\theta_T^*, sp}(\theta) &= \lim_{r \to 0} \frac{\tilde{\mathbb{F}}_T(\overline{B_r(\theta)})}{\lambda(\overline{B_r(\theta)})} \\
&= \lim_{r \to 0} \frac{\tilde{\mathbb{P}}\left\{\tilde{N}_T(\overline{B_r(\theta)}) = 1\right\}[1 + \varepsilon(r)]}{\lambda(\overline{B_r(\theta)})} \\
&= \lim_{r \to 0} \frac{\tilde{\mathbb{P}}\left\{\tilde{N}_T(\overline{B_r(\theta)}) = 1\right\}}{\lambda(\overline{B_r(\theta)})}
\end{aligned}$$

where $\varepsilon(.)$ is a positive function such that $\lim_{r \to 0} \varepsilon(r) = 0$ by Proposition 4.2 p.23. □

A.8. Proof of Theorems

A.8.1 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.7. *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)* By Assumption $|A|_{det} > 0$. The determinant function $|.|_{det}$ is a multilinear function, and thus a continuous function. Thus, $\lim_{T\to\infty} A_T = A$ implies the result.

ii) On the one hand, A_T is a n-d.m. if and only if all its eigenvalues are strictly negative. On the other hand, $\max \mathrm{sp} A_T = \max_{z:\|z\|=1} z' A_T z$ where $\mathrm{sp} A_T$ denotes the set of eigenvalues of A. Thus, it is sufficient to prove that $\lim_{T\to\infty} \max_{z:\|z\|=1} z' A_T z = \max_{z:\|z\|=1} z' A z$, which in turn implies that it is sufficient to prove that $\sup_{z:\|z\|=1} |z' A_T z - z' A z| \to 0$, as $T \to \infty$. We prove this last result by contradiction.

Assume that $\sup_{z:\|z\|=1} |z' A_T z - z' A z|$ does not converge to 0 as $T \to \infty$. Then, there exists $\varepsilon > 0$ and an increasing function $\alpha_1 : \mathbf{N} \mapsto \mathbf{N}$ defining a subsequence of vectors of norm 1, $\left\{z_{\alpha_1(T)}\right\}_{T\geqslant 1}$, and a subsequence of matrices, $\left\{A_{\alpha_1(T)}\right\}_{T\geqslant 1}$, such that

$$\begin{aligned}
\varepsilon &< \left| z'_{\alpha_1(T)} A_{\alpha_1(T)} z_{\alpha_1(T)} - z'_{\alpha_1(T)} A z_{\alpha_1(T)} \right| \\
&= \left| z'_{\alpha_1(T)} \left(A_{\alpha_1(T)} - A \right) z_{\alpha_1(T)} \right| \leqslant \sum_{(k,l)\in[\![1,m]\!]^2} \left| \left[a^{(k,l)}_{\alpha_1(T)} - a^{(k,l)} \right] z^{(k)}_{\alpha_1(T)} z^{(l)}_{\alpha_1(T)} \right| \\
&\leqslant m^2 \times \max_{(k,l)\in[\![1,m]\!]^2} \left| a^{(k,l)}_{\alpha_1(T)} - a^{(k,l)} \right|
\end{aligned}$$

where m is the size of the matrix A and $a^{(k,l)}$ denotes the component of the matrix A in the kth row and lth column. Now, by Assumption, $\max_{(k,l)\in[\![1,m]\!]^2} \left| a^{(k,l)}_{\alpha_1(T)} - a^{(k,l)} \right| \to 0$ as $T \to \infty$. Thus, there is a contradiction.$\square$

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

Assumption 24. **(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* $(\mathbf{\Omega}, \mathcal{E}, \mathbb{P})$ **; (b)** *Let the measurable space* $(\mathbf{B}, \mathcal{B}(\mathbf{B}))$ *such that* $\mathbf{B} \subset \mathbf{R}^m$ *is compact and* $\mathcal{B}(\mathbf{B})$ *is the Borel σ-algebra.* **; (c)** *Let* $h : \mathbf{R}^p \times \mathbf{B} \mapsto \mathbf{R}^q$ *with* $q \in \mathbf{N}$ *be a function such that* $\forall x \in \mathbf{R}^p$, $\beta \mapsto h(x,\beta)$ *is continuous, and* $\forall \beta \in \mathbf{B}$, $x \mapsto h(x,\beta)$ *is* $\mathcal{B}(\mathbf{R}^p)/\mathcal{B}(\mathbf{R}^q)$*-measurable.*; **(d)** $\mathbb{E}\left[\sup_{\beta\in\mathbf{B}} \|h(X,\beta)\|\right] < \infty$. **; (e)** *In the parameter space* $\mathbf{B}$*, there exists a unique* $\beta_0 \in \mathrm{int}(\mathbf{B})$ *such that* $\mathbb{E}[h(X,\beta_0)] = 0_{m\times 1}$**; (f)** *For all* $x \in \mathbf{R}^p$, $\beta \mapsto h(x,\beta)$ *is continuously differentiable***; (g)** $\left|\mathbb{E}\left[\frac{\partial h(X,\beta_0)}{\partial \beta'}\right]\right|_{det} \neq 0$.

Proposition A.2 (Uniform-strong LLN). *Under Assumptions 24(a)-(d) ,* $\frac{1}{T}\sum_{t=1}^{T} h(X_t,\beta)$ *converges*

$\mathbb{P}$*-a.s. to* $\mathbb{E}[h(X,\beta)]$ *uniformly w.r.t .*β *as* $T \to \infty$ *i.e. there exists* $E \in \mathcal{E}$ *such that* $\mathbb{P}\{E\} = 0$ *and*

$$\forall \omega \in \mathbf{\Omega} \setminus E, \quad \sup_{\beta \in \mathbf{B}} \left\| \frac{1}{T} \sum_{t=1}^{T} h(X_t, \beta) - \mathbb{E}[h(X,\beta)] \right\| \to 0 \quad \textit{as } T \to \infty. \tag{25}$$

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.

Proposition A.3 (Existence of solutions to empirical moment conditions)**.** *Under the Assumptions 24(a)-(c)(e)-(g), if*

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

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

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

i) *there exists* $\mathbb{P}$*-a.s. a solution to the empirical moment conditions i.e. there exists* β_T^* *such that*

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

ii) *all solutions to the empirical moment conditions are in* $B_r(\beta_0)$*.*

Proof . *i)* By Assumption 24(e) and the continuity of uniform limits of continuous functions for $\varepsilon > 0$ small enough there exists $r > 0$ such that $\forall \beta \in \mathbf{B} \setminus B_r(\beta_0)$,

$$\begin{aligned}
&\Leftrightarrow \quad 2\varepsilon < \|\mathbb{E}[h(X,\beta)]\| \\
&\Leftrightarrow \quad 2\varepsilon < \left\| \mathbb{E}[h(X,\beta)] - \frac{1}{T}\sum_{t=1}^{T} h(X_t,\beta) \right\| + \left\| \frac{1}{T}\sum_{t=1}^{T} h(X_t,\beta) \right\| \\
&\Leftrightarrow \quad 2\varepsilon - \left\| \mathbb{E}[h(X,\beta)] - \frac{1}{T}\sum_{t=1}^{T} h(X_t,\beta) \right\| < \left\| \frac{1}{T}\sum_{t=1}^{T} h(X_t,\beta) \right\|
\end{aligned}$$

where the second inequality comes from the triangle inequality. Now, by Assumption (a), there exists $\dot{T} \in \mathbf{N}$ such that $T \geqslant \dot{T}$ implies that $\forall \beta \in \mathbf{B}$ $\|\mathbb{E}[h(X,\beta)] - \frac{1}{T}\sum_{t=1}^{T} h(X_t,\beta)\| < \varepsilon$ $\mathbb{P}$-a.s. which implies $-\varepsilon < -\|\mathbb{E}[h(X,\beta)] - \frac{1}{T}\sum_{t=1}^{T} h(X_t,\beta)\|$. Therefore, on the one hand, $\forall \beta \in \mathbf{B} \setminus B_r(\beta_0)$, $\varepsilon < \|\frac{1}{T}\sum_{t=1}^{T} h(X_t,\beta)]\|$ $\mathbb{P}$-a.s.; and on the other hand, by Assumption (a) for T_1

$\forall\beta \in B_r(\beta_0)$, $\|\frac{1}{T}\sum_{t=1}^T h(X_t,\beta)\| < \varepsilon$. Now, the function $\beta \mapsto \|\frac{1}{T}\sum_{t=1}^T h(X_t,\beta)\|$ has a minimum, since it is continuous over a compact set $\mathbf{B}$ by Assumption 24(b)(c). Thus, for T big enough, $\dot{\beta}_T := \arg\min_{\beta\in\mathbf{B}} \|\frac{1}{T}\sum_{t=1}^T h(X_t,\beta)\| \in B_r(\beta_0)$ $\mathbb{P}$-a.s. Now, the smaller is ε, the smaller is r, since by the continuity of uniform limits of continuous functions, Assumption (b) and Lemma A.7 $\frac{1}{T}\sum_{t=1}^T \frac{\partial h(X_t,\beta)'}{\partial\beta}$ is invertible in a neighborhood of β_0 $\mathbb{P}$-a.s. Thus for ε small enough, $\mathbb{P}$-a.s. that for $T \geqslant \dot{T}$, $2\left[\frac{1}{T}\sum_{t=1}^T \frac{\partial h(X_t,\dot{\beta}_T)'}{\partial\beta}\right]\left[\frac{1}{T}\sum_{t=1}^T h(X_t,\dot{\beta}_T)\right] = 0_{m\times 1}$ with $\left[\frac{1}{T}\sum_{t=1}^T \frac{\partial h(X_t,\dot{\beta}_T)'}{\partial\beta}\right]$ invertible. Thus $\frac{1}{T}\sum_{t=1}^T h(X_t,\dot{\beta}_T) = 0_{m\times 1}$ $\mathbb{P}$-a.s. which implies the result by putting $\beta_T^* := \dot{\beta}_T$.

ii) Immediate from the proof of *i)*. □

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 $\beta \mapsto \|\frac{1}{T}\sum_{t=1}^T h(X_t,\beta)\|$, the 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, 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) all solutions to the empirical moment conditions are measurable $\mathbb{P}$-a.s. i.e. if β_T^* is such that $\frac{1}{T}\sum_{t=1}^T h(X_t,\beta_T^*) = 0$, then $\beta_T^*(.)$ is $\mathcal{E}/\mathcal{B}(\mathbf{B})$-measurable $\mathbb{P}$-a.s.

Proof . *i)* According to the proof of Proposition A.3, for $r > 0$ small enough there exists $\dot{T} \in \mathbf{N}$ so that for all $T \geqslant \dot{T}$ all solutions to the empirical moment conditions, $\left\{\dot{\beta}_T^{(v)}\right\}_{v\in V}$, lies in $\overline{B_r(\beta_0)}$ with $\frac{1}{T}\sum_{t=1}^T \frac{\partial h(X,\beta)'}{\partial\beta}$ invertible $\forall\beta \in B_r(\beta_0)$ $\mathbb{P}$-a.s. Thus by the inverse function theorem applied for all $v \in V$ to $\frac{1}{T}\sum_{t=1}^T h(X_t,\dot{\beta}_T^{(v)}) = 0_{m\times 1}$ by Assumption 24c), there exists $\varepsilon_v > 0$ such that $\dot{\beta}_T^{(v)}$ is the unique solution to the empirical moment conditions in $B_{\varepsilon_v}(\dot{\beta}_T^{(v)})$. Thus, $\bigcup_{v\in V} B_{\varepsilon_v}(\dot{\beta}_T^{(v)})$ is an open covering of the compact set $\overline{B_r(\beta_0)}$. Now, any open cover of a compact set contains a finite open cover. Thus, there exists $\left\{\dot{\beta}_T^{(k)}\right\}_{k=1}^K \in B_r(\beta_0)^K$ and $\{\varepsilon_k\}_{k=1}^K$ positive such that $\overline{B_r(\beta_0)} \subset \bigcup_{k=1}^K B_{\varepsilon_k}(\dot{\beta}_T^{(k)})$, where $\dot{\beta}_T^{(k)}$ is the unique solution to the empirical moment conditions in $B_{\varepsilon_k}(\dot{\beta}_T^{(k)})$. Therefore, the number of solutions to the empirical moment conditions is finite.

ii) Assumptions 24(a)-(c),(h) correspond to Assumption 1. Thus by Proposition 4.1, there exists a finite (simple) point random-field $N_T(.,.)$ such that $\forall\omega \in \mathbf{\Omega} \setminus E$, $N_T(\omega,.) = \#\{\beta \in . : \frac{1}{T}\sum_{t=1}^T h(X_t,\beta) = 0\}$ where $\mathbb{P}\{E\} = 0$. Now, for any point random-field $N(.,.)$ on a complete separable metric space $\mathcal{Y}$, there exists a sequence of measurable random elements, $\{Y_k\}_{k=1}^{\infty}$,[37] such that

$$\forall B \in \mathcal{B}(\mathcal{Y}), \quad N(\omega, B) = \sum_k \delta_{Y_k(\omega)}(B)$$

(e.g. Lemma 9.1.XIII p.16 in Daley and Vere-Jones, 2008). The result follows. $\square$

The following proposition is a standard result.

Proposition A.5 (Consistency of solutions to empirical moment conditions)**.** *Under the assumptions of Propositions A.3, every sequence of solutions to the empirical moment conditions, $\{\beta_T^*\}_{T\geqslant 1}$, converges $\mathbb{P}$-a.s. to the population parameter, β_0, i.e.*

$$\lim_{T\to\infty} \beta_T^* = \beta_0 \quad \mathbb{P}\text{-}a.s.$$

Proof . Proposition A.4 ensures the $\mathcal{E}/\mathcal{B}(\mathbf{B})$-measurability of β_T^* for T big enough. For all $r > 0$, we can choose ε small enough in the proof of Proposition A.3*i*) to make solutions to empirical moment conditions in $B_r(\beta_0)$ $\mathbb{P}$-a.s. for T big enough. $\square$

Corollary 1. *Under the Assumptions 1(a)-(c) and 5(a)-(b),(e), 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 the assumptions of Proposition A.3 are satisfied. Assumptions 1(a) and 4(a) provide Assumption 24(a). Assumption 1(b)(c),4(b), 2 and 4(e) respectively provide Assumptions 24(b)(c)(e)(f)(g).

Application of Proposition A.2 to $\frac{1}{T}\sum_{t=1}^{T} \psi(X_t, \theta)$ and $\frac{1}{T}\sum_{t=1}^{T} \frac{\partial\psi(X_t,\theta)}{\partial\theta'}$ provides respectively assumptions (a) and (b) of Proposition A.3. In both cases, Assumptions 1(a) and 4(a) provide Assumption 24(a);and Assumptions 1(b)(c),4(c) respectively provide Assumptions 24(b)(c)(d). In the case of $\frac{1}{T}\sum_{t=1}^{T} \psi(X_t, \theta)$, Assumption 4(c) provides Assumption 24(d). In the case of $\frac{1}{T}\sum_{t=1}^{T} \frac{\partial\psi(X_t,\theta)}{\partial\theta'}$, Assumption 4(d) provides Assumption 24(d). $\square$

[37] Since the number of solutions to the empirical moment conditions is finite $\mathbb{P}$-a.s. (Assumption 1(d)), there exists k_0 such that $k \geqslant k_0$ $\forall \omega \in \mathbf{\Omega} \setminus E$, $Y_k(\omega) = \emptyset$ with $\mathbb{P}\{E\} = 0$.

Lemma A.8. *Under Assumptions 1(a)-(c), 2,5(a)(b),*

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

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

Proof . *i)* By definition of $\hat{\mathbf{\Theta}}_{\infty}^{-\eta}$ and a standard result on Laplace's transform (e.g. Theorem 3 p183 in Monfort, 1996) $\mathbb{E}\left[\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\right] = 0$ is the FOC of the convex function $\tau \mapsto \mathbb{E}\left[\mathrm{e}^{\tau\psi(X,\theta)}\right]$. In addition, for all $\theta \in \hat{\mathbf{\Theta}}_{\infty}^{-\eta}$, $\mathbb{E}\left[\psi(X,\theta)\psi(X,\theta)'\mathrm{e}^{\tau'\psi(X,\theta)}\right]$ is a symmetric p-d.m by the definition of $\hat{\mathbf{\Theta}}_{\infty}^{-\eta}$. Thus, $\tilde{\tau}_{\infty}(\theta)$ is unique as the solution to the FOC of a strictly convex function.

ii) By the Assumptions 2,5(b) and the Lebesgue dominated convergence theorem $\theta \mapsto \mathbb{E}\left[\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\right]$ is continuous. As a convex function $\tau \mapsto \mathbb{E}\left[\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\right]$ is continuous (e.g., Theorem 3.1.2 p.174 in Hiriart-Urruty and Lemaréchal, 1996). Thus, by *i)* and a version of the implicit function theorem (Kumagai, 1980) ,$\tau_{\infty}(.)$ is continuous.□

Lemma A.9. *Under the assumptions of Lemma A.8, define the compact set*

$$\hat{C}_{\infty}^{-\eta} := \tau_{\infty}\left(\hat{\mathbf{\Theta}}_{\infty}^{-\eta}\right).$$

Under the assumptions of Lemma A.8 and Assumption 4(a), for all $\theta \in \hat{\mathbf{\Theta}}_{\infty}^{-\eta}$*, as* $T \to \infty$

$$\sup_{\tau \in \hat{C}_{\infty}^{-\eta}} \left\| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi(X_t,\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau'\psi(X,\theta)}\right] \right\| \to 0 \quad \mathbb{P}\text{-}a.s.$$

$$\sup_{\tau \in \hat{C}_{\infty}^{-\eta}} \left\| \frac{1}{T}\sum_{t=1}^{T} \psi(X_t,\theta)\mathrm{e}^{\tau'\psi(X_t,\theta)} - \mathbb{E}\left[\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\right] \right\| \to 0 \quad \mathbb{P}\text{-}a.s.$$

$$\sup_{\tau \in \hat{C}_{\infty}^{-\eta}} \left\| \frac{1}{T}\sum_{t=1}^{T} \psi(X_t,\theta)\psi(X_t,\theta)'\mathrm{e}^{\tau'\psi(X_t,\theta)} - \mathbb{E}\left[\psi(X,\theta)\psi(X,\theta)'\mathrm{e}^{\tau'\psi(X,\theta)}\right] \right\| \to 0 \quad \mathbb{P}\text{-}a.s.$$

Proof . First, note $\hat{C}_{\infty}^{-\eta}$ is well-defined. $\tau_{\infty}(.)$ can be extended by continuity to $\hat{\mathbf{\Theta}}_{\infty}^{-\eta}$; and the image of a compact by a continuous function is compact. Second, prove the claims. By definition of $\hat{\mathbf{\Theta}}_{\infty}^{-\eta}$ and a standard result about Laplace transforms (e.g. Theorem 3 p183 in Monfort, 1996), the expectations there $\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X,\theta)}\right]$, $\mathbb{E}\left[\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\right]$ and $\mathbb{E}\left[\psi(X,\theta)\psi(X,\theta)'\mathrm{e}^{\tau'\psi(X,\theta)}\right]$ are finite. Thus, by Assumption 4(a) and the LLN, there is point-wise convergence. Since $\tau \mapsto \frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau\psi(X_t,\theta)}$, $\tau \mapsto \frac{1}{T}\sum_{t=1}^{T}\psi(X_t,\theta)\mathrm{e}^{\tau'\psi(X_t,\theta)}$ and $\tau \mapsto \frac{1}{T}\sum_{t=1}^{T}\psi(X_t,\theta)\psi(X_t,\theta)'\mathrm{e}^{\tau'\psi(X_t,\theta)}$ are convex, the result follows (e.g., Theorem 3.1.5. p.177 in Hiriart-Urruty and Lemaréchal, 1996). □

Lemma A.10. *Under the assumptions of Lemma A.9 , for all $\theta \in \operatorname{int}\left(\hat{\mathbf{\Theta}}_{\infty}^{-\eta}\right)$, as $T \to \infty$, $\tau_T(\theta) \to \tau_\infty(\theta)$ $\mathbb{P}$-a.s.*

Proof . Check the assumptions of Proposition A.3 are satisfied. Assumptions 1(a) and 4(a) provide to Assumption 24(a). Assumption 1(b)(c) respectively provide Assumption 24(b)(c). Lemma A.8i), whose assumptions are the same as Lemma A.9, provides Assumption 24(e). Assumption 24(f) is immediate. Definition of $\hat{\mathbf{\Theta}}_{\infty}^{-\eta}$ ensures Assumption 24(g). The second and third results of Lemma A.9 corresponds to assumptions (a)(b) of Proposition A.3. Then, apply Proposition A.3 to the empirical moment conditions $\sum_{t=1}^{T} \psi(X_t, \theta) \mathrm{e}^{\tau' \psi(X_t,\theta)} = 0_{m\times 1}$ to ensure the existence of $\tau_T(\theta)$ for T big enough for all $\theta \in \hat{\mathbf{\Theta}}_{\infty}^{-\eta}$. Proposition A.4, whose assumptions are the same as Proposition A.3 , ensures the $\mathcal{E}/\mathcal{B}(\mathbf{R}^m)$-measurability of $\tau_T(\theta)$.[38] Then, the result follows from Proposition A.5, whose assumptions are the same as Proposition A.3.$\square$

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.[39]

Assumption 25 (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*

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)**

[38] Here, we can also use Lemma 2 from Jennrich (1969).

[39] 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.

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 25, 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 . In the proof, we implicitly always assume that $T \in \mathbf{N}$ is big enough and $\varepsilon > 0$ small enough so that all written quantities are well-defined by the Assumptions. For clarity, all Taylor expansions are written for $m = 1$.

By Assumption 25(b)i), using obvious notations, a Taylor with a mean-value form of the remainder of $h_T(.)$ at $\dot{\theta}_T$ yields

$$\begin{aligned} & h_T(\theta) \\ = \; & h_T(\dot{\theta}_T) + h_T^{(1)}(\dot{\theta}_T)(\theta-\dot{\theta}_T) + \frac{1}{2}h_T^{(2)}(\dot{\theta}_T)(\theta-\dot{\theta}_T)^2 + \frac{1}{6}h_T^{(3)}(\dot{\theta}_T)(\theta-\dot{\theta}_T)^3 + \overline{h}_{1,T}(\theta)(\theta-\dot{\theta}_T)^4 \\ = \; & \frac{1}{2}h_T^{(2)}(\dot{\theta}_T)(\theta-\dot{\theta}_T)^2 + \frac{1}{6}h_T^{(3)}(\dot{\theta}_T)(\theta-\dot{\theta}_T)^3 + \overline{h}_{1,T}(\theta)(\theta-\dot{\theta}_T)^4. \end{aligned}$$

where $h_T^{(k)}(.)$ denotes the kth derivative, $\overline{h}_{1,T}(.)$ is a continuous function and where Assumption 25(b)iii) is used. Plugging the expansion into the exponential function, we obtain

$$\begin{aligned} & \exp\left[Th_T(\theta)\right] \\ = \; & \exp\left\{\frac{T}{2}h_T^{(2)}(\dot{\theta}_T)(\theta-\dot{\theta}_T)^2\right\}\exp\left\{\frac{T}{6}h_T^{(3)}(\dot{\theta}_T)(\theta-\dot{\theta}_T)^3 + T\overline{h}_{1,T}(\theta)(\theta-\dot{\theta}_T)^4\right\}. \end{aligned} \tag{26}$$

Now, we compute in details $\exp\left\{\frac{T}{6}h_T^{(3)}(\dot{\theta}_T)(\theta-\dot{\theta}_T)^3 + T\overline{h}_{1,T}(\theta)(\theta-\dot{\theta}_T)^4\right\}b_T(\theta)$. A Taylor with a mean-value form of the remainder expansion of $x \mapsto \mathrm{e}^x$ yields

$$\begin{aligned}
& \exp\left\{\frac{T}{6}h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3 + T\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4\right\} \\
= \; & 1+\frac{T}{6}h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3 + T\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4 + \overline{h}_{2,T}(\theta)\left[\frac{T}{6}h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3 + T\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4\right]^2 \\
= \; & 1+\frac{T}{6}h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3 + T\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4 + \overline{h}_{2,T}(\theta)\frac{T^2}{36}h_T^{(3)}(\dot\theta_T)^2(\theta-\dot\theta_T)^6 \\
& +\overline{h}_{2,T}(\theta)2\frac{T}{6}h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4 + \overline{h}_{2,T}(\theta)T^2\overline{h}_{1,T}(\theta)^2(\theta-\dot\theta_T)^8.
\end{aligned}$$

Moreover, by Assumption 25(c)i), a Taylor with a mean-value form of the remainder of $b_T(.)$ yields

$$b_T(\theta) = b_T(\dot\theta_T) + b_T^{(1)}(\dot\theta_T)(\theta-\dot\theta) + \overline{b}_T(\theta)(\theta-\dot\theta_T)^2.$$

Therefore, multiplying the two expansions we get

$$\begin{aligned}
& \exp\left\{\frac{T}{6}h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3 + T\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4\right\}b_T(\theta) \\
= \; & b_T(\dot\theta_T) + b_T^{(1)}(\dot\theta_T)(\theta-\dot\theta_T) + \overline{b_T}(\theta)(\theta-\dot\theta_T)^2 \\
& +\frac{T}{6}h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3 b_T(\dot\theta_T) + \frac{T}{6}h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3 b_T^{(1)}(\dot\theta_T)(\theta-\dot\theta_T) \\
& \quad +\frac{T}{6}h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3\overline{b}_T(\theta)(\theta-\dot\theta_T)^2 \\
& +T\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4 b_T(\dot\theta_T) + T\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4 b^{(1)}(\dot\theta_T)(\theta-\dot\theta_T) \\
& \quad +T\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4\overline{b}_T(\theta)(\theta-\dot\theta_T)^2 \\
& +\frac{T^2}{36}\overline{h}_{2,T}(\theta)h_T^{(3)}(\dot\theta_T)^2(\theta-\dot\theta_T)^6 b_T(\dot\theta_T) + \frac{T^2}{36}\overline{h}_{2,T}(\theta)h_T^{(3)}(\dot\theta_T)^2(\theta-\dot\theta_T)^6 b_T^{(1)}(\dot\theta_T)(\theta-\dot\theta_T) \\
& \quad +\frac{T^2}{36}\overline{h}_{2,T}(\theta)h_T^{(3)}(\dot\theta_T)^2(\theta-\dot\theta_T)^6\overline{b}_T(\theta)(\theta-\dot\theta_T)^2 \\
& +2\frac{T}{6}\overline{h}_{2,T}(\theta)h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4 b_T(\dot\theta_T) \\
& \quad +2\frac{T}{6}\overline{h}_{2,T}(\theta)h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4 b^{(1)}(\dot\theta_T)(\theta-\dot\theta_T) \\
& \quad +2\frac{T}{6}\overline{h}_{2,T}(\theta)h_T^{(3)}(\dot\theta_T)(\theta-\dot\theta_T)^3\overline{h}_{1,T}(\theta)(\theta-\dot\theta_T)^4\overline{b}_T(\theta)(\theta-\dot\theta_T)^2 \\
& +T^2\overline{h}_{2,T}(\theta)\overline{h}_{1,T}(\theta)^2(\theta-\dot\theta_T)^8 b_T(\dot\theta_T) + T^2\overline{h}_{2,T}(\theta)\overline{h}_{1,T}(\theta)^2(\theta-\dot\theta_T)^8 b_T^{(1)}(\dot\theta_T)(\theta-\dot\theta_T) \\
& \quad +\overline{h}_{2,T}(\theta)T^2\overline{h}_{1,T}(\theta)^2(\theta-\dot\theta_T)^8\overline{b}_T(\theta)(\theta-\dot\theta_T)^2 \\
:= \; & b_T(\dot\theta_T) + I_T(\theta,\dot\theta_T) + R_T(\theta,\dot\theta_T)
\end{aligned}$$

where $I_T(\theta,\dot\theta_T) := b_T^{(1)}(\dot\theta_T)(\theta-\dot\theta_T) + \frac{T}{6}h_T^{(3)}(\dot\theta_T)b_T(\dot\theta_T)(\theta-\dot\theta_T)^3 + \frac{T}{6}h_T^{(3)}(\dot\theta_T)b_T^{(1)}(\dot\theta_T)(\theta-\dot\theta_T)^4$.

Now, by equation (26),

$$\int_{V_r(\dot\theta_T)} b_T(\theta) \mathrm{e}^{[T h_T(\theta)]} d\theta = \int_{V_r(\dot\theta_T)} \exp\left\{\frac{T}{2} h_T^{(2)}(\dot\theta_T)(\theta-\dot\theta_T)^2\right\}\left\{b_T(\dot\theta_T)+I_T(\theta,\dot\theta_T)+R_T(\theta,\dot\theta_T)\right\} d\theta$$

where the integral is well-defined as an integral of continuous functions (Assumption 25(a)i) and (b)i)) over a compact set. In addition, by Lemma A.11 below, we have

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

where given two functions $f(.)$ and $g(.)$ whose domain is D, $f(z) \underset{a}{\sim} g(z)$ with $a \in D$ means that there exists a function $\varphi(.)$ defined on D such that $f(.) = g(.)\varphi(.)$ and $\lim_{z\to a}\varphi(z) = 1$. Thus, by Assumption 25(c), it is sufficient to show that

$$\int_{V_r(\dot\theta_T)} \exp\left\{\frac{T}{2} h_T^{(2)}(\dot\theta_T)(\theta-\dot\theta_T)^2\right\}\left\{I_T(\theta,\dot\theta_T)+R_T(\theta,\dot\theta_T)\right\} d\theta \underset{\infty}{\sim} \frac{O(T^{-1})}{T^{\frac{m}{2}}}.$$

Deal with the integrals of $I_T(\theta,\dot\theta_T)$ and $R_T(\theta,\dot\theta_T)$ separately. The integral of $I_T(\theta,\dot\theta_T)$ yields

$$\begin{aligned}
& \int_{V_r(\dot\theta_T)} \exp\left\{\frac{1}{2} h_T^{(2)}(\dot\theta_T)\left[\sqrt{T}(\theta-\dot\theta_T)\right]^2\right\} I_T(\theta,\dot\theta_T) d\theta \\
= & \int_{V_r(\dot\theta_T)} \exp\left\{\frac{1}{2} h_T^{(2)}(\dot\theta_T)\left[\sqrt{T}(\theta-\dot\theta_T)\right]^2\right\}\left\{T^{-\frac{1}{2}} b_T^{(1)}(\dot\theta_T)\left[\sqrt{T}(\theta-\dot\theta_T)\right]\right. \\
& \left. + \frac{T^{-\frac{1}{2}}}{6} h_T^{(3)}(\dot\theta_T) b_T(\dot\theta_T)\left[\sqrt{T}(\theta-\dot\theta_T)\right]^3 + \frac{T^{-1}}{6} h_T^{(3)}(\dot\theta_T) b_T^{(1)}(\dot\theta_T)\left[\sqrt{T}(\theta-\dot\theta_T)\right]^4\right\} d\theta \\
\overset{(a)}{=} & \int_{V_{\sqrt{T}r}(0)} \exp\left\{\frac{1}{2} h_T^{(2)}(\dot\theta_T) u^2\right\}\left[T^{-\frac{1}{2}} b_T^{(1)}(\dot\theta_T) u + \frac{T^{-\frac{1}{2}}}{6} h_T^{(3)}(\dot\theta_T) b_T(\dot\theta_T) u^3 + \frac{T^{-1}}{6} h_T^{(3)}(\dot\theta_T) b_T^{(1)}(\dot\theta_T) u^4\right] \frac{du}{T^{\frac{m}{2}}} \\
\underset{\infty}{\overset{(b)}{\sim}} & \int_{\mathbf{R}^m} \exp\left\{\frac{1}{2} h_T^{(2)}(\dot\theta_T) u^2\right\}\left[T^{-\frac{1}{2}} b_T^{(1)}(\dot\theta_T) u + \frac{T^{-\frac{1}{2}}}{6} h_T^{(3)}(\dot\theta_T) b_T(\dot\theta_T) u^3 + \frac{T^{-1}}{6} h_T^{(3)}(\dot\theta_T) b_T^{(1)}(\dot\theta_T) u^4\right] \frac{du}{T^{\frac{m}{2}}} \\
\underset{\infty}{\overset{(c)}{\sim}} & \frac{O(T^{-1})}{T^{\frac{m}{2}}}.
\end{aligned}$$

(a) Apply change of variable $\theta \mapsto \sqrt{T}(\theta-\dot\theta_T) := u$; and denote $V_{\sqrt{T}r}(0) := \left\{u \in \mathbf{R}^m : T^{-\frac{1}{2}}u + \dot\theta_T \in V_r(\dot\theta_T)\right\}$. (b) Assumption 25(b)ii); (c)ii) allows to dominate by absolute moment of Gaussian distribution scaled by a constant and then to apply the Lebesgue dominated convergence theorem. (c) The odd moments of a normal distribution are zero.

The integral of $R_T(\theta, \dot{\theta}_T)$ yields

$$
\begin{aligned}
& \left| \int_{V_r(\dot{\theta}_T)} \exp\left\{ \frac{1}{2} h_T^{(2)}(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^2 \right\} R_T(\theta, \dot{\theta}_T) d\theta \right| \\
\leqslant \quad & \int_{V_r(\dot{\theta}_T)} \left| \exp\left\{ \frac{1}{2} h_T^{(2)}(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^2 \right\} \left\{ T^{-1} \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^2 \right. \right. \\
& + \frac{T^{-\frac{3}{2}}}{6} h_T^{(3)}(\dot{\theta}_T) \overline{b}_T(\theta) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^5 \\
& + T^{-1} \overline{h}_{1,T}(\theta) b_T(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^4 + T^{-\frac{3}{2}} \overline{h}_{1,T}(\theta) b_T^{(1)}(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^5 \\
& \quad + T^{-2} \overline{h}_{1,T}(\theta) \overline{b}_T(\theta) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^6 \\
& + \frac{T^{-1}}{36} \overline{h}_{2,T}(\theta) h_T^{(3)}(\dot{\theta}_T)^2 b_T(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^6 + \frac{T^{-\frac{3}{2}}}{36} \overline{h}_{2,T}(\theta) h_T^{(3)}(\dot{\theta}_T)^2 b_T^{(1)}(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^7 \\
& \quad + \overline{h}_{2,T}(\theta) \frac{T^{-2}}{36} h_T^{(3)}(\dot{\theta}_T)^2 \overline{b}_T(\theta) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^8 \\
& + 2 \frac{T^{-\frac{5}{2}}}{6} \overline{h}_{2,T}(\theta) h_T^{(3)}(\dot{\theta}_T) \overline{h}_{1,T}(\theta) b_T(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^7 \\
& \quad + 2 \frac{T^{-3}}{6} \overline{h}_{2,T}(\theta) h_T^{(3)}(\dot{\theta}_T) \overline{h}_{1,T}(\theta) b_T^{(1)}(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^8 \\
& \quad + 2 \frac{T^{-\frac{7}{2}}}{6} \overline{h}_{2,T}(\theta) h_T^{(3)}(\dot{\theta}_T) \overline{h}_{1,T}(\theta) \overline{b}_T(\theta) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^9 \\
& + T^{-2} \overline{h}_{2,T}(\theta) \overline{h}_{1,T}(\theta)^2 b_T(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^8 + T^{-\frac{5}{2}} \overline{h}_{2,T}(\theta) \overline{h}_{1,T}(\theta)^2 b_T^{(1)}(\dot{\theta}_T) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^9 \\
& \quad \left. \left. + T^{-3} \overline{h}_{2,T}(\theta) \overline{h}_{1,T}(\theta)^2 \overline{b}_T(\theta) \left[\sqrt{T}(\theta - \dot{\theta}_T) \right]^{10} \right\} \right| d\theta \\
\overset{(a)}{\leqslant} \quad & M \int_{V_{\sqrt{T}r}(0)} \exp\left\{ \frac{1}{2} h_T^{(2)}(\dot{\theta}_T) u^2 \right\} \left\{ T^{-1} u^2 + T^{-1} u^4 + T^{-\frac{3}{2}} |u|^5 + T^{-1} u^4 + T^{-\frac{3}{2}} |u|^5 \right. \\
& \quad \left. + T^{-2} u^6 + T^{-\frac{5}{2}} |u|^7 + T^{-3} u^8 + T^{-\frac{7}{2}} |u|^9 + T^{-2} u^8 + T^{-\frac{5}{2}} |u|^9 + T^{-3} u^{10} \right\} \frac{du}{T^{\frac{m}{2}}} \\
\underset{\infty}{\overset{(b)}{\sim}} \quad & M \int_{\mathbf{R}^m} \exp\left\{ \frac{1}{2} h_T^{(2)}(\dot{\theta}_T) u^2 \right\} \left\{ T^{-1} u^2 + T^{-1} u^4 + T^{-\frac{3}{2}} |u|^5 + T^{-1} u^4 + T^{-\frac{3}{2}} |u|^5 \right. \\
& \quad \left. + T^{-2} u^6 + T^{-\frac{5}{2}} |u|^7 + T^{-3} u^8 + T^{-\frac{7}{2}} u^9 + T^{-2} u^8 + T^{-\frac{5}{2}} |u|^9 + T^{-3} u^{10} \right\} \frac{du}{T^{\frac{m}{2}}} \\
\overset{(c)}{=} \quad & \frac{O(T^{-1})}{T^{\frac{m}{2}}}.
\end{aligned}
$$

(a) By Assumption 25(b)ii);(c)ii), $\overline{h}_{1,T}(.)$, $\overline{h}_{2,T}(.)$ and $\overline{b}_T(.)$ are bounded on $V_r(\dot{\theta}_T)$. *(b)* All the moments of a Gaussian distribution are finite, thus we can apply the Lebesgue dominated convergence Theorem. *(c)* Finiteness of the moments of a Gaussian distribution.□

Lemma A.11. *Under Assumptions 1(b) and 25,*

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

Proof . We have

$$\begin{aligned}
&\int_{V_r(\dot\theta_T)} \exp\left\{\frac{T}{2}(\theta-\dot\theta_T)'\frac{\partial^2 h_T(\dot\theta_T)}{\partial\theta\partial\theta'}(\theta-\dot\theta_T)\right\}d\theta \\
\overset{(a)}{=}\;& \frac{1}{T^{\frac{m}{2}}}\frac{(2\pi)^{\frac{m}{2}}\left|\left(-\frac{\partial^2 h_T(\dot\theta_T)}{\partial\theta\partial\theta'}\right)^{-1}\right|_{det}^{\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}\left|\left(-\frac{\partial^2 h_T(\dot\theta_T)}{\partial\theta\partial\theta'}\right)^{-1}\right|_{det}^{\frac{1}{2}}}\int_{V_{\sqrt{T}r}(0)}\exp\left\{\frac{1}{2}u'\frac{\partial^2 h_T(\dot\theta_T)}{\partial\theta\partial\theta'}u\right\}du \\
\overset{(b)}{\underset{\infty}{\sim}}\;& \left(\frac{2\pi}{T}\right)^{\frac{m}{2}}\left|\left(-\frac{\partial^2 h_T(\dot\theta_T)}{\partial\theta\partial\theta'}\right)\right|_{det}^{-\frac{1}{2}}
\end{aligned}$$

(a) Change of variable $\theta \mapsto \sqrt{T}(\theta-\dot\theta_T) := u$, and denote $V_{\sqrt{T}r}(0) := \left\{u \in \mathbf{R}^m : T^{-\frac{1}{2}}u + \dot\theta_T \in V_\varepsilon(\dot\theta_T)\right\}$
(b) By Assumption 25(c), recognising the density of a normal distribution, we can apply the Lebesgue dominated convergence theorem

$$\lim_{T\to\infty}\frac{1}{(2\pi)^{\frac{m}{2}}\left|\left(-\frac{\partial^2 h_T(\dot\theta_T)}{\partial\theta\partial\theta'}\right)^{-1}\right|_{det}^{\frac{1}{2}}}\int_{V_{\sqrt{T}r}(0)}\exp\left\{-\frac{1}{2}u'\left[\left(-\frac{\partial^2 h_T(\dot\theta_T)}{\partial\theta\partial\theta'}\right)^{-1}\right]^{-1}u\right\}du = 1 \quad (27)$$

□

Proposition A.7. *Under Assumptions 1(b) and 25, 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.11. □

A.8.2 Proof of Theorems 5.1 and 5.2

Around θ_T^*: application of Laplace's approximation

Proposition A.8. *Under Assumptions 1-6, 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 $r > 0$ small enough by putting*

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

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 applies to random functions. Thus, it is now sufficient to show the above quantities satisfy Assumption 25. Corollary 1 and Assumption 3(b) provide Assumption 25(a). Assumption 6(a) with Corollary 1 provide $\psi(.) \in C^4(B_r(\theta_T^*))$ with $r > 0$ for T big enough. By the implicit function theorem and Assumption 5(a), $\tau_T(.) \in C^4(B_r(\theta_T^*))$ for T big enough. Thus, Assumption 25(b)i) (d)i) are satisfied. Assumption 6(c)(d) provide that Assumptions 25(b)ii) (d)ii). By Lemma A.13 and the definition of θ_T^*, Assumption 25(b)iii) is satisfied. By Lemma A.14, the Assumption 25(c) is also satisfied. □

Lemma A.12. *Under Assumptions 1(a)-(c), 2, 5 and 6(e), for T big enough $\mathbb{P}$-a.s.*

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

where the LHS and RHS are well-defined.

***Proof*.** To apply the implicit function theorem to the tilting equation defining $\tau_T(.)$, first show that the derivative w.r.t. τ is full rank.

$$\begin{aligned}
\lim_{T\to\infty}\left|\frac{\partial\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\right]}{\partial\tau'}\Bigg|_{\substack{\tau=\tau_T(\theta_T^*)\\ \theta=\theta_T^*}}\right|_{det} &\overset{(a)}{=} \lim_{T\to\infty}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta_T^*)'\psi_t(\theta_T^*)}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right|_{det}\\
&\overset{(b)}{=} \lim_{T\to\infty}\left|\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right|_{det}\\
&\overset{(c)}{=} \left|\mathbb{E}\left[\psi(X,\theta_0)\psi(X,\theta_0)'\right]\right|_{det} > 0
\end{aligned}$$

(a) By Assumption 2, differentiate. *(b)* By definition of $\tau_T(.)$, $\tau_T(\theta_T^*) = 0$, since θ_T^* denotes a solution to the empirical moment conditions. *(c)* By Assumption 1-2, 5(b),6(e) and Corollary 1, apply Proposition A.2 with $\mathbf{B}$ a closed ball centered at θ_0. Then use Assumption 5(a) and the definition of $\hat{\boldsymbol{\Theta}}_{\infty}^{-\eta}$.

Therefore, by Lemma A.7i) for T big enough

$$\left|\frac{\partial\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\right]}{\partial\tau'}\Bigg|_{\substack{\tau=\tau_T(\theta_T^*)\\ \theta=\theta_T^*}}\right|_{det} > 0$$

Consequently, by Assumption 2 and the implicit function theorem in a neighborhood of θ_0 for T big enough

$$\begin{aligned}
&\frac{\partial\tau_T(\theta)}{\partial\theta'}\\
=&\ -\left[\frac{\partial\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\right]}{\partial\tau'}\right]^{-1}\left[\frac{\partial\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\right]}{\partial\theta'}\right]\Bigg|_{\tau=\tau_T(\theta)}\\
=&\ -\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\frac{\partial\psi_i(\theta)}{\partial\theta'}+\psi_i(\theta)\tau'\frac{\partial\psi_i(\theta)}{\partial\theta'}\right)\right]\Bigg|_{\tau=\tau_T(\theta)}\\
=&\ -\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\left(\frac{\partial\psi_i(\theta)}{\partial\theta'}+\psi_i(\theta)\tau_T(\theta)'\frac{\partial\psi_i(\theta)}{\partial\theta'}\right)\right].
\end{aligned}$$

Put $\theta = \theta_T^*$ and note that $\tau_T(\theta_T^*) = 0_{m\times 1}$,

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

Lemma A.13. *Under the assumptions of Lemma A.12, for T big enough $\mathbb{P}$-a.s.*

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

where the LHS is well-defined.

Proof . Differentiate w.r.t. the kth component

$$\frac{\partial \ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]}{\partial\theta_k} \quad = \quad \frac{\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\left[\frac{\partial\tau_T(\theta)'}{\partial\theta_k}\psi_t(\theta)+\frac{\partial\psi_t(\theta)'}{\partial\theta_k}\tau_T(\theta)\right]}{\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}}. \qquad (28)$$

Thus, for $\theta=\theta_T^*$,

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

Stack rows to obtain the result. $\square$

Lemma A.14. *Under the assumptions of Lemma A.12 and Assumption 6(a), for T big enough $\mathbb{P}$-a.s.*

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

where the RHS and the LHS are well-defined. Thus, the RHS converges to a n-d.m.

Proof . From equation (28),

$$
\begin{aligned}
&\frac{\partial^2 \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]}{\partial\theta_k\partial\theta_l} \\
=\;& \frac{1}{\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]^2} \left\{\left\{\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} \left[\frac{\partial\tau_T(\theta)'}{\partial\theta_l}\psi_t(\theta) + \frac{\partial\psi_t(\theta)'}{\partial\theta_l}\tau_T(\theta)\right]\left[\frac{\partial\tau_T(\theta)'}{\partial\theta_k}\psi_t(\theta)\right.\right.\right. \\
&\left.\left. + \frac{\partial\psi_t(\theta)'}{\partial\theta_k}\tau_T(\theta)\right] + \left[\frac{\partial^2\tau_T(\theta)'}{\partial\theta_k\partial\theta_l}\psi_t(\theta) + \frac{\partial\psi_t(\theta)'}{\partial\theta_l}\frac{\partial\tau_T(\theta)}{\partial\theta_k}\right] + \left[\frac{\partial^2\psi_t(\theta)'}{\partial\theta_k\theta_l}\tau_T(\theta) + \frac{\partial\tau_T(\theta)'}{\partial\theta_l}\frac{\partial\psi_t(\theta)}{\partial\theta_k}\right]\right\} \\
&\times\left\{\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right\} - \left\{\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\left[\frac{\partial\tau_T(\theta)'}{\partial\theta_k}\psi_t(\theta) + \frac{\partial\psi_t(\theta)'}{\partial\theta_k}\tau_T(\theta)\right]\right\} \\
&\left.\times\left\{\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\left[\frac{\partial\tau_T(\theta)'}{\partial\theta_l}\psi_t(\theta) + \frac{\partial\psi_t(\theta)'}{\partial\theta_l}\tau_T(\theta)\right]\right\}\right\} \\
=\;& \frac{1}{\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]} \left\{\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} \left[\frac{\partial\tau_T(\theta)'}{\partial\theta_l}\psi_t(\theta) + \frac{\partial\psi_t(\theta)'}{\partial\theta_l}\tau_T(\theta)\right]\left[\frac{\partial\tau_T(\theta)'}{\partial\theta_k}\psi_t(\theta)\right.\right. \\
&\left. + \frac{\partial\psi_t(\theta)'}{\partial\theta_k}\tau_T(\theta)\right] + \left[\frac{\partial^2\tau_T(\theta)'}{\partial\theta_k\partial\theta_l}\psi_t(\theta) + \frac{\partial\psi_t(\theta)'}{\partial\theta_l}\frac{\partial\tau_T(\theta)}{\partial\theta_k}\right] + \left.\left[\frac{\partial^2\psi_t(\theta)'}{\partial\theta_k\theta_l}\tau_T(\theta) + \frac{\partial\tau_T(\theta)'}{\partial\theta_l}\frac{\partial\psi_t(\theta)}{\partial\theta_k}\right]\right\} \\
&- \frac{1}{\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]^2} \left\{\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\left[\frac{\partial\tau_T(\theta)'}{\partial\theta_k}\psi_t(\theta) + \frac{\partial\psi_t(\theta)'}{\partial\theta_k}\tau_T(\theta)\right]\right\} \\
&\times\left\{\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\left[\frac{\partial\tau_T(\theta)'}{\partial\theta_l}\psi_t(\theta) + \frac{\partial\psi_t(\theta)'}{\partial\theta_l}\tau_T(\theta)\right]\right\}.
\end{aligned}
$$

Thus,for $\theta = \theta_T^*$ and $\tau = \tau_T(\theta_T^*) = 0_{m\times 1}$

$$
\begin{aligned}
&\left.\frac{\partial^2 \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]}{\partial\theta_k\partial\theta_l}\right|_{\theta=\theta_T^*} \\
=\;& \left\{\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_l}\psi_t(\theta_T^*)\right]\left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_k}\psi_t(\theta_T^*)\right] + \left[\frac{\partial^2\tau_T(\theta_T^*)'}{\partial\theta_k\partial\theta_l}\psi_t(\theta_T^*) + \frac{\partial\psi_t(\theta_T^*)'}{\partial\theta_l}\frac{\partial\tau_T(\theta_T^*)}{\partial\theta_k}\right]\right. \\
&\left. + \left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_l}\frac{\partial\psi_t(\theta_T^*)}{\partial\theta_k}\right]\right\} - \left\{\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_k}\psi_t(\theta_T^*)\right]\right\}\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_l}\psi_t(\theta_T^*)\right]\right\} \\
=\;& \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_l}\psi_t(\theta_T^*)\right]\left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_k}\psi_t(\theta_T^*)\right] + \left[\frac{\partial\psi_t(\theta_T^*)'}{\partial\theta_l}\frac{\partial\tau_T(\theta_T^*)}{\partial\theta_k}\right] + \left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_l}\frac{\partial\psi_t(\theta_T^*)}{\partial\theta_k}\right]
\end{aligned}
$$

Stack rows and columns and use the Lemma A.12. The first term becomes

$$\begin{aligned}
&\left(\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_l}\psi_t(\theta_T^*)\right]\left[\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_k}\psi_t(\theta_T^*)\right]\right)_{(k,l)\in[\![1,m]\!]^2}\\
= \;& \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta}\psi_t(\theta_T^*)\right)\left(\psi_t(\theta_T^*)'\frac{\partial\tau_T(\theta_T^*)}{\partial\theta'}\right)\\
= \;& \frac{\partial\tau_T(\theta_T^*)'}{\partial\theta}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]\frac{\partial\tau_T(\theta_T^*)}{\partial\theta'}\\
= \;& \left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_i(\theta_T^*)}{\partial\theta'}\right]'\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]^{-1}\\
&\times\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_i(\theta_T^*)}{\partial\theta'}\right]\\
= \;& \left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_i(\theta_T^*)'}{\partial\theta}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_i(\theta_T^*)}{\partial\theta'}\right].
\end{aligned}$$

Similarly, the sum of the two other terms becomes

$$\begin{aligned}
&\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta_T^*)'}{\partial\theta_l}\frac{\partial\tau_T(\theta_T^*)}{\partial\theta_k}\right)_{(k,l)\in[\![1,m]\!]^2}+\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\tau_T(\theta_T^*)'}{\partial\theta_l}\frac{\partial\psi_t(\theta_T^*)}{\partial\theta_k}\right)_{(k,l)\in[\![1,m]\!]^2}\\
= \;& \left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_i(\theta_T^*)'}{\partial\theta}\right]\frac{\partial\tau_T(\theta_T^*)}{\partial\theta'}+\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_i(\theta_T^*)'}{\partial\theta}\right]\frac{\partial\tau_T(\theta_T^*)}{\partial\theta'}\right\}'\\
= \;& -2\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_i(\theta_T^*)'}{\partial\theta}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_i(\theta_T^*)}{\partial\theta'}\right].
\end{aligned}$$

where the last equality comes from the symmetry of each term. Add all the terms to obtain the result.□

Outside a neighborhood of θ_T^*

Proposition A.9. *Under the assumptions of Lemma A.8 and A.15 and Assumptions 1(a)-(c),2, 4(a),7(b) for all $r > 0$ small enough 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 $B_{r_1}(\dot{\theta})$ with $r_1 > 0$ and $\dot{\theta} \in \hat{\mathbf{\Theta}}_\infty^{-\eta}$. Assumptions 1(a)4(a) provide Assumption 24(a). Assumptions 1(b) provides

Assumption 24(b). By Assumptions 2 and Lemma A.8ii), for all $\theta \in \hat{\boldsymbol{\Theta}}_\infty^{-\eta}$ $\theta \mapsto \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau_\infty(\theta)'\psi_t(\theta)}$ is continuous. Then combined with Assumption 1(c) it provides Assumption 24(c). Assumptions 7(b) provides Assumption 24(d). Thus, by Proposition A.2 there exists $r_1 > 0$ such that for all $\dot{\theta} \in \hat{\boldsymbol{\Theta}}_\infty^{-\eta}$, as $T \to \infty$,

$$\sup_{\theta \in B_{r_1}(\dot{\theta})} \left\| \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau_\infty(\theta)'\psi_t(\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau_\infty(\theta)'\psi(X,\theta)}\right] \right\| \to 0.$$

Now, since $\hat{\boldsymbol{\Theta}}_\infty^{-\eta}$ is compact, there exists $\{\theta_k\}_{k=1}^K$ such that $\hat{\boldsymbol{\Theta}}_\infty^{-\eta} = \bigcup_{k=1}^K B_{r_2}(\theta_k)$ and $r_2 < r_1$. Thus, as $T \to \infty$,

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

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

$$\sup_{\theta \in \hat{\boldsymbol{\Theta}}_\infty^{-\eta}} \left\| \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau_\infty(\theta)'\psi_t(\theta)} - \mathbb{E}\left[\mathrm{e}^{\tau_\infty(\theta)'\psi(X,\theta)}\right] \right\| < \varepsilon_3.$$

Moreover, by Lemma A.15, for $\varepsilon > 0$ small enough, there exists, $r_3 > 0$ such that $\forall \theta \in \hat{\boldsymbol{\Theta}}_\infty^{-\eta} \setminus B_{r_3}(\theta_0)$, $\mathbb{E}\left[\mathrm{e}^{\tau_\infty(\theta)'\psi(X,\theta)}\right] < 1 - 2\varepsilon$, since $\theta \mapsto \mathbb{E}\left[\mathrm{e}^{\tau_\infty(\theta)'\psi(X,\theta)}\right]$ is continuous by 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 $T \geqslant \dot{T}$ $\forall \theta \in \hat{\boldsymbol{\Theta}}_\infty^{-\eta} \setminus B_{r_3}(\theta_0)$

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

since $\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} = \min_{\tau \in \mathbf{R}^m} \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}$. □

Lemma A.15. *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 p.183

Monfort, 1980), $\tau_\infty(\theta)$ is the minimum of a strictly convex function *i.e.* for all $\theta \in \hat{\mathbf{\Theta}}_\infty^{-\eta}$, for all $\tau \neq \tau_\infty(\theta)$, $\mathbb{E}\left[e^{\tau_\infty(\theta)'\psi(x,\theta)}\right] < \mathbb{E}\left[e^{\tau'\psi(X,\theta)}\right]$. Thus, for all $\tau_\infty(\theta) \neq 0_{m\times 1}$, $\mathbb{E}\left[e^{\tau_\infty(\theta)\psi(X,\theta)}\right] < \mathbb{E}\left[e^{0'_{m\times 1}\psi(X,\theta)}\right] = 1$. Now, by Assumption 4(b), for all $\theta \in \hat{\mathbf{\Theta}}_\infty^{-\eta} \setminus \{\theta_0\}$, $\tau_\infty(\theta) \neq 0_{m\times 1}$. Therefore, the result follows.□

Conclusion of the proofs

Corollary 2. *Under Assumptions 1-5, for all $r > 0$ small enough,*

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

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

Proof. *i)* By Proposition A.8, apply Proposition A.7 with Lemma A.14 so that $\mathbb{P}$-a.s.

$$\begin{aligned}\int_{B_r(\theta_T^*)} \tilde{f}_{\theta_T^*,sp}(\theta)d\theta &= \int_{B_r(\theta_T^*)} \exp\left\{T\ln\left[\frac{1}{T}\sum_{t=1}^{T} e^{\tau_T(\theta)'\psi_t(\theta)]}\right\}\left(\frac{T}{2\pi}\right)^{m/2}|\Sigma_T(\theta)|_{det}^{-\frac{1}{2}}\,d\theta \\ &\underset{\infty}{\sim} \left(\frac{2\pi}{T}\right)^{\frac{m}{2}}|\Sigma_T(\theta_T^*)^{-1}|_{det}^{-\frac{1}{2}}\left(\frac{T}{2\pi}\right)^{\frac{m}{2}}|\Sigma_T(\theta_T^*)|_{det}^{-\frac{1}{2}} \\ &\underset{\infty}{\sim} 1\end{aligned}$$

where $\Sigma_T(\theta_T^*) := \left[\frac{1}{T}\sum_{t=1}^T \frac{\partial\psi_t(\theta_T^*)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T \psi_t(\theta_T^*)\psi_t(\theta_T^*)'\right]\left[\frac{1}{T}\sum_{t=1}^T \frac{\partial\psi_t(\theta_T^*)'}{\partial\theta}\right]^{-1}$.

ii) By Proposition A.9 and Assumptions 7(a) for all $r > 0$ small enough there exists $\varepsilon > 0$ $\dot{M} \geqslant 0$ and $\dot{T} \in \mathbf{N}$ such that $\forall T \geqslant \dot{T}$ we have for all $\theta \in \hat{\mathbf{\Theta}}_T^{-\eta} \setminus B_r(\theta_0)$

$$\tilde{f}_{\theta_T^*,sp}(\theta) \leqslant \exp\{-T\varepsilon\}\left(\frac{T}{2\pi}\right)^{\frac{m}{2}}\dot{M} \quad \mathbb{P}\text{-}a.s. \tag{29}$$

By Definition 4.4 of the ESP intensity, it also holds for all $\theta \in \mathbf{\Theta} \setminus B_r(\theta_0)$. □

Conclusion of the proof of Theorem 5.1 . For all $\varphi \in C_b$ and for all $r > 0$,

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

Therefore, by the lemmas below for $r > 0$ small enough, for all $\varepsilon > 0$, $\forall T \geqslant \max(T_1, T_2, T_3)$,

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

which is the result needed. □

Lemma A.16. *For all $\varepsilon > 0$, there exists $T_1 \in \mathbf{N}$ such that for $r > 0$ small enough*

$$
\forall T \geqslant T_1, \qquad \left| \int_{B_r(\theta_T^*)} \left[\varphi(\theta) - \varphi(\theta_0) \right] \tilde{f}_{\theta_T^*,sp}(\theta) d\theta \right| \leqslant \frac{\varepsilon}{3} \quad \mathbb{P}\text{-a.s.}.
$$

Proof .

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

where M is a constant which bounds the integral by Corollary 2i).

By continuity of $\varphi(.)$, there exists $\varepsilon_3 > 0$ such that $\forall \theta \in B_{\varepsilon_3}(\theta_0)$, $|\varphi(\theta) - \varphi(\theta_0)| < \frac{\varepsilon}{3M}$. Moreover, since $\theta_T^* \to \theta_0$ by Corollary 1, $\forall \varepsilon_4 > 0$ there exists $T_4 \in \mathbf{N}$ s.t. $T \geqslant T_4$ $\|\theta_T^* - \theta_0\| \leqslant \varepsilon_4$. Thus, if r and ε_4 are such that $0 < r < \varepsilon_3 - \varepsilon_4$, for all $T \geqslant T_4$, $\theta \in B_r(\theta_T^*)$ implies

$$\begin{aligned} \|\theta - \theta_0\| &= \|\theta - \theta_T^* + \theta_T^* - \theta_0\| \\ &\leqslant \|\theta - \theta_T^*\| + \|\theta_T^* - \theta_0\| \\ &\leqslant r + \varepsilon_4 < \varepsilon; \end{aligned}$$

which in turn implies the result by putting $T_1 := T_4$ $\square$

Lemma A.17. *For all $\varepsilon > 0$, there exists $T_2 > 0$ such that*

$$\forall T \geqslant T_2, \qquad \left| \int_{\boldsymbol{\Theta} \setminus B_r(\theta_T^*)} \varphi(\theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta \right| \leqslant \frac{\varepsilon}{3} \quad \mathbb{P}\text{-}a.s..$$

Proof. Since $\varphi(.)$ is bounded, there exists $\dot{M} > 0$ such that

$$\left| \int_{\boldsymbol{\Theta} \setminus B_r(\theta_T^*)} \varphi(\theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta \right| \leqslant \dot{M} \int_{\boldsymbol{\Theta} \setminus B_r(\theta_T^*)} \tilde{f}_{\theta_T^*, sp}(\theta) d\theta$$

Thus, the result follows by Corollary 2ii). $\square$

Lemma A.18. *For all $\varepsilon > 0$, there exists $T_3 > 0$ such that*

$$\forall T \geqslant T_3, \qquad \left| \int_{B_r(\theta_T^*)} \varphi(\theta_0) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta - \varphi(\theta_0) \right| \leqslant \frac{\varepsilon}{3} \quad \mathbb{P}\text{-}a.s..$$

Proof.

$$\left| \int_{B_r(\theta_T^*)} \varphi(\theta_0) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta - \varphi(\theta_0) \right| = |\varphi(\theta_0)| \times \left| \int_{B_r(\theta_T^*)} \tilde{f}_{\theta_T^*, sp}(\theta) d\theta - 1 \right|$$

Thus the result follows by Corollary 2i). $\square$

Conclusion of the proof of Theorem 2. By Proposition A.6,

$$
\begin{aligned}
&\int_{D_T(a,\theta_T^*,b)} \tilde{f}_{\theta_T^*,sp}(\theta)d\theta \\
\underset{\infty}{\sim}\ & \int_{D_T(a,\theta_T^*,b)} \exp\left\{-\frac{T}{2}(\theta-\theta_T^*)'\Sigma_T(\theta_T^*)^{-1}(\theta-\theta_T^*)\right\}\left(\frac{T}{2\pi}\right)^{\frac{m}{2}}|\Sigma_T(\theta_T^*)|_{det}^{-\frac{1}{2}}\,d\theta \\
\overset{(a)}{\underset{\infty}{\sim}}\ & \int_{D(a,b)} \mathrm{e}^{-\frac{1}{2}z'z}\left(\frac{T}{2\pi}\right)^{\frac{m}{2}}|\Sigma_T(\theta_T^*)|_{det}^{-\frac{1}{2}}\left|T^{-1/2}\Sigma_T(\theta_T^*)^{\frac{1}{2}}\right|_{det} ds \\
\underset{\infty}{\sim}\ & \frac{1}{(2\pi)^{\frac{m}{2}}}\int_{D(a,b)} \mathrm{e}^{-\frac{1}{2}'s} ds
\end{aligned}
$$

(a) $z := \sqrt{T}\Sigma_T(\theta_T^*)^{-\frac{1}{2}}(\theta-\theta_T^*)$ □

A.8.3 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.

A.9. Proof of Proposition 7.1

i) By Proposition 4.4 and Assumption 11(a) respectively, $\tilde{f}_{\theta_T^*,sp}(.)$ and $u(.,.)$ are continuous over the compact sets $\mathbf{\Theta}$ and $\mathbf{\Theta}^2$. Thus, for all $(\theta_e,\theta)\in\mathbf{\Theta}^2$, $\|u(\theta_e,\theta)\tilde{f}_{\theta_T^*,sp}(\theta)\| \leqslant \sup_{\dot{\theta}\in\mathbf{\Theta}}\|u(\dot{\theta},\theta)\tilde{f}_{\theta_T^*,sp}(\theta)\|$ with $\int_{\mathbf{\Theta}}\sup_{\dot{\theta}\in\mathbf{\Theta}}\|u(\dot{\theta},\theta)\tilde{f}_{\theta_T^*,sp}(\theta)\|d\theta<\infty$. Thus, by the Lebesgue dominated convergence theorem, $\theta_e \mapsto \int_{\mathbf{\Theta}} u(\theta_e,\theta)\tilde{f}_{\theta_T^*,sp}(\theta)d\theta$ is continuous over $\mathbf{\Theta}$.

ii) By i) and Lemma A.19, apply Lemma 2 from Jennrich (1969).

Lemma A.19. *Under Assumption 3;11(a),* $(\omega,\theta)\mapsto\int_{\mathbf{\Theta}}u(\theta_e,\theta)\tilde{f}_{\theta_T^*,sp}(\theta)d\theta$ *is* $\mathcal{E}\otimes\mathcal{B}(\mathbf{\Theta})/\mathcal{B}(\mathbf{R})$*-measurable.*

Proof. By Assumption 3, $\tilde{f}_{\theta_T^*,sp}(.)$ is $\mathcal{E}\otimes\mathcal{B}(\mathbf{\Theta})/\mathcal{B}(\mathbf{R}_+)$-measurable. By Assumption 11(a), $u(.,.)$ is continuous. Thus, $u(.,.)\tilde{f}_{\theta_T^*,sp}(.)$ is $\mathcal{E}\otimes\mathcal{B}(\mathbf{\Theta})\otimes\mathcal{B}(\mathbf{\Theta})/\mathcal{B}(\mathbf{R}_+)$. Thus apply a standard preliminary result to the Fubini theorem (e.g. Lemma 1.26 p.14 in Kallenberg, 2001) to deduce that $\int_{\mathbf{\Theta}}u(\theta_e,\theta)\tilde{f}_{\theta_T^*,sp}(\theta)d\theta$ is $\mathcal{E}\otimes\mathcal{B}(\mathbf{\Theta})/\mathcal{B}(\mathbf{R})$-measurable. □

A.10. Proof of Proposition 7.2

i) By the Lemma A.20 below, for all $\varepsilon > 0$ there exists an open cover of $\mathbf{\Theta}$, $\bigcup_{\dot\theta\in\mathbf{\Theta}} B_{\varepsilon_{\dot\theta}}(\dot\theta)$, such that

$$\begin{cases} r_{\dot\theta} > 0 \\ \sup_{\theta_e\in B_{r_{\dot\theta}}(\dot\theta)} \left\| \int_{\mathbf{\Theta}} u(\theta_e,\theta) f_{\theta_T^*,sp}(\theta)d\theta - u(\theta_e,\theta_0)\right\| < \varepsilon \text{ for } T \text{ big enough.}\end{cases}$$

Now, any open cover of a compact set contains a finite open cover. Therefore, by Assumption 1(b), there exists $\left\{\dot\theta_k\right\}_{k=1}^K \in \mathbf{\Theta}^K$, $\{T_k\}_{k=1}^K \in \mathbf{N}^K$ and $\{r_k\}_{k=1}^K \in]0,\infty[^K$ such that

$$\begin{cases} \mathbf{\Theta} = \bigcup_{k=1}^K B_{r_k}(\dot\theta_k) \\ \sup_{\theta_e\in B_k(\dot\theta_k)} \left\| \int_{\mathbf{\Theta}} u(\theta_e,\theta) f_{\theta_T^*,sp}(\theta)d\theta - u(\theta_e,\theta_0)\right\| < \varepsilon \text{ for } T \geqslant T_k\,. \end{cases}$$

Thus for $T \geqslant \max_{k\in[\![1,K]\!]} T_k$, $\sup_{\theta_e\in\mathbf{\Theta}} \left\| \int_{\mathbf{\Theta}} u(\theta_e,\theta)\tilde f_{\theta_T^*,sp}(\theta)d\theta - u(\theta_e,\theta_0)\right\| < \varepsilon$.

ii) Since by Assumptions 11(b)(c) $u(\theta_0,\theta_0) > u(\theta_e,\theta_0)$ for all $\theta_e \in \mathbf{\Theta}$, it follows from iii). □

Lemma A.20. *Under Assumptions1-7,11(a), for all $\dot\theta_e \in \mathbf{\Theta}$ and for all $\varepsilon > 0$, there exist $r > 0$ and $\dot T \in \mathbf{N}$ such that for all $\theta_e \in B_r(\dot\theta_e)$ and for all $T \geqslant \dot T$,*

$$\left\| \int_{\mathbf{\Theta}} u(\theta_e,\theta) f_{\theta_T^*,sp}(\theta)d\theta - u(\theta_e,\theta_0)\right\| \leqslant \varepsilon.$$

Proof. For a fixed $\dot\theta_e \in \mathbf{\Theta}$, by the triangle inequality, $\forall r > 0$, $\forall \theta_e \in B_r(\dot\theta_e)$

$$\begin{aligned} &\left\| \int_{\mathbf{\Theta}} u(\theta_e,\theta) f_{\theta_T^*,sp}(\theta)d\theta - u(\theta_e,\theta_0)\right\| \\ \leqslant\ &\left\| \int_{\mathbf{\Theta}} u(\theta_e,\theta) f_{\theta_T^*,sp}(\theta)d\theta - \int_{\mathbf{\Theta}} u(\dot\theta_e,\theta) f_{\theta_T^*,sp}(\theta)d\theta\right\| + \left\| \int_{\mathbf{\Theta}} u(\dot\theta_e,\theta) f_{\theta_T^*,sp}(\theta)d\theta - u(\dot\theta_e,\theta_0)\right\| \\ &+ \left\| u(\dot\theta_e,\theta_0) - u(\theta_e,\theta_0)\right\| \end{aligned}$$

It remains to prove that for all $\varepsilon > 0$, by choosing r small enough and T big enough each of the three terms can made smaller than $\frac{\varepsilon}{3}$. For all $\dot{\varepsilon} > 0$, for r small enough

$$\begin{aligned}\left\|\int_{\mathbf{\Theta}} u(\theta_e,\theta) f_{\theta_T^*,sp}(\theta)d\theta - \int_{\mathbf{\Theta}} u(\dot{\theta}_e,\theta) f_{\theta_T^*,sp}(\theta)d\theta\right\| &= \left\|\int_{\mathbf{\Theta}} \left[u(\theta_e,\theta) - u(\dot{\theta}_e,\theta)\right] f_{\theta_T^*,sp}(\theta)d\theta\right\| \\ &\overset{(a)}{\leqslant} \dot{\varepsilon}\int_{\mathbf{\Theta}} f_{\theta_T^*,sp}(\theta)d\theta \\ &\overset{(b)}{\leqslant} \dot{\varepsilon}M \text{ with } M \in \mathbf{R}\end{aligned}$$

where *(a)* comes from the uniform continuity of $u(.,.)$ implied by Assumption 11(a) and the Heine-Cantor Theorem; and *(b)* comes from Theorem 5.1. Thus, for $\dot{\varepsilon} = \frac{\varepsilon}{3M}$, the first term is smaller than $\frac{\varepsilon}{3}$. For T big enough the second term is smaller than $\frac{\varepsilon}{3}$ by Theorem 5.1. For r small enough, the third term is smaller than $\frac{\varepsilon}{3}$. $\square$

A.11. Proof of Proposition 7.3

Notations for this proof.

$$\begin{aligned}\forall\theta_e \in \mathbf{\Theta}, \quad h_T(\theta_e) &:= \frac{1}{K_T}\int_{\mathbf{\Theta}} u(\theta_e,\theta)\tilde{f}_{\theta_T^*,sp}(\theta)d\theta \\ h(\theta_e) &:= u(\theta_e,\theta_0)\end{aligned} \tag{30}$$

i) Existence of $k_{\alpha,T}$ follows from Lemma A.21 and A.23iii) . Measurability of the ESP confidence set, $\{\theta_e : h_T(\theta_e) \geqslant k_{\alpha,T}\}$, follows from the $\mathcal{E}\otimes\mathcal{B}(\mathbf{\Theta})/\mathcal{B}(\mathbf{R})$-measurability of $h_T(.) - k_{\alpha,T}$ by Lemma A.23iv) and Lemma A.19. Thus there exists an ESP confidence set. Existence of an asymptotic ESP confidence set follows from the same arguments as in Lemma A.23iii).

ii) Proof by contradiction. Assume that $k_{\alpha,T}$ does not converge to $k_{\alpha,\infty}$ as $T \to \infty$. Then there exists $\varepsilon > 0$ and a subsequence $\left\{k_{\alpha,\beta_1(T)}\right\}_{T\geqslant 1}$ such that $\left|k_{\alpha,\beta_1(T)} - k_{\alpha,\infty}\right| > \varepsilon$. By Lemma A.22 and the Bolzano-Weierstrass theorem, there exists a converging subsequence $\left\{k_{\alpha,\beta_2\circ\beta_1(T)}\right\}_{T\geqslant 1}$ of the sequence $\left\{k_{\alpha,\beta_1(T)}\right\}_{T\geqslant 1}$. Distinguish two cases.

First case $\lim_{T\to\infty} k_{\alpha,\beta_2\circ\beta_1(T)} > k_{\alpha,\infty}$. Let $\varepsilon > 0$ s.t. $\lim_{T\to\infty} k_{\alpha,\beta_2\circ\beta_1(T)} > k_{\alpha,\infty} + \varepsilon$. Then,

$$\begin{aligned}
1-\alpha &\overset{(a)}{\leqslant} \lim_{T\to\infty} \sup \int \mathbf{1}_{\left\{\theta_e \in \mathbf{\Theta}: h_{\beta_2\circ\beta_1(T)}(\theta_e) > k_{\alpha,\beta_2\circ\beta_1(T)}\right\}}(\theta_e) h_{\beta_2\circ\beta_1(T)}(\theta_e) d\theta_e \\
&\overset{(b)}{\leqslant} \lim_{T\to\infty} \sup \int \mathbf{1}_{\left\{\theta_e \in \mathbf{\Theta}: h_{\beta_2\circ\beta_1(T)}(\theta_e) > k_{\alpha,\infty}\right\}}(\theta_e) h_{\beta_2\circ\beta_1(T)}(\theta_e) d\theta_e \\
&\overset{(c)}{\leqslant} \int \lim_{T\to\infty} \sup \mathbf{1}_{\left\{\theta_e \in \mathbf{\Theta}: h_{\beta_2\circ\beta_1(T)}(\theta_e) > k_{\alpha,\infty}\right\}}(\theta_e) h_{\beta_2\circ\beta_1(T)}(\theta_e) d\theta_e \\
&\overset{(d)}{\leqslant} \int_{\mathbf{\Theta}} \mathbf{1}_{\{\theta_e \in \mathbf{\Theta}: h(\theta_e) > k_{\alpha,\infty}+\varepsilon\}} h(\theta_e) d\theta_e. \qquad (31)
\end{aligned}$$

(a) Definition of $k_{\alpha,T}$. *(b)* By assumption, $\lim_{T\to\infty} k_{\alpha,\beta_2\circ\beta_1(T)} > k_{\alpha,\infty}$. Thus, $\mathbf{1}_{\left\{\theta_e \in \mathbf{\Theta}: h_{\beta_2\circ\beta_1(T)}(\theta_e) > k_{\alpha,\beta_2\circ\beta_1(T)}\right\}}(.) \leqslant \mathbf{1}_{\left\{\theta_e \in \mathbf{\Theta}: h_{\beta_2\circ\beta_1(T)}(\theta_e) > k_{\alpha,\infty}\right\}}(.)$. *(c)* Fatou's lemma. *(d)* Lemma A.26 and for any positive sequences $\{u_T\}_{T\geqslant 1}$ and $\{v_T\}_{T\geqslant 1}$ $\lim_{T\to\infty} \sup u_T v_T \leqslant \lim_{T\to\infty} \sup u_T \times \lim_{T\to\infty} \sup v_T$. Lebesgue dominated convergence theorem.

The inequality (31) is in contradiction with the definition of $k_{\alpha,\infty}$ since $\lim_{T\to\infty} k_{\alpha,\beta_2\circ\beta_1(T)} > k_{\alpha,\infty} + \varepsilon$.

Second case $k_{\alpha,\infty} > \lim_{T\to\infty} k_{\alpha,\beta_2\circ\beta_1(T)}$. Let $\varepsilon > 0$ s.t. $k_{\alpha,\infty} - \varepsilon > \lim_{T\to\infty} k_{\alpha,\beta_2\circ\beta_1(T)}$. Then,

$$\begin{aligned}
1-\alpha &\overset{(a)}{\leqslant} \int_{\mathbf{\Theta}} \mathbf{1}_{\{\theta_e : h(\theta_e) > k_{\alpha,\infty}\}} h(\theta_e) d\theta_e \\
&\overset{(b)}{\leqslant} \int_{\mathbf{\Theta}} \lim_{T\to\infty} \inf \mathbf{1}_{\left\{\theta_e : h_{\beta_2\circ\beta_1(T)}(\theta_e) > k_{\alpha,\infty}-\varepsilon\right\}} h_{\beta_2\circ\beta_1(T)}(\theta_e) d\theta_e \\
&\overset{(c)}{\leqslant} \lim_{T\to\infty} \inf \int_{\mathbf{\Theta}} \mathbf{1}_{\left\{\theta_e : h_{\beta_2\circ\beta_1(T)}(\theta_e) > k_{\alpha,\infty}-\varepsilon\right\}} h_{\beta_2\circ\beta_1(T)}(\theta_e) d\theta_e. \qquad (32)
\end{aligned}$$

(a) Definition of $k_{\alpha,\infty}$. *(b)* Lemma A.26 and for any positive sequences $\{u_T\}_{T\geqslant 1}$ and $\{v_T\}_{T\geqslant 1}$ $\lim_{T\to\infty} \inf u_T \times \lim_{T\to\infty} \inf v_T \leqslant \lim_{T\to\infty} \inf u_T v_T$. Lebesgue dominated convergence theorem. *(c)* Fatou's lemma.

The inequality (32) is in contradiction with the definition of $k_{\alpha,\beta_2\circ\beta_1(T)}$, since $k_{\alpha,\infty} - \varepsilon > \lim_{T\to\infty} k_{\alpha,\beta_2\circ\beta_1(T)}$.

□

Lemma A.21. *Under Assumptions 1-3;11- 12, for $\eta > 0$ small enough,*

$$K_T := \int_{\mathbf{\Theta}^2} u(\theta_e, \theta) \tilde{f}_{\theta_T^*, sp}(\theta) d\theta d\theta_e \neq 0.$$

Proof. By Assumption 11, for all $\dot{\theta} \in \mathbf{\Theta}$, there exists $r_{\dot{\theta}} > 0$ such that for all $\theta_1, \theta_2 \in B_{r_{\dot{\theta}}}(\dot{\theta}) u(\theta_1, \theta_2) > 0$. Moreover, by Proposition 4.3i), $\hat{\mathbf{\Theta}}$ is an open such that $\hat{f}_{\theta_T^*, sp}(.) > 0$. Therefore there exists an open in $\mathbf{\Theta}^2$ such that $u(.,.)f(.) > 0$. Since by Proposition 4.4i) $u(.,.)\tilde{f}_{\theta_T^*,sp}(.) \geqslant 0$ on $\mathbf{\Theta}^2$, the result follows. $\square$

Lemma A.22. *Under Assumptions 1-3;11-12,* $\mathbb{P}$*-a.s.*

i) $h_T(.) \geqslant 0$*;*

ii) $\theta_e \mapsto h_T(\theta_e)$ *is continuous;*

iii) $\int_{\mathbf{\Theta}} h_T(\theta_e)d\theta_e = 1$.

Proof. *i)* By Assumption 11(d) and Assumption 3 $u(.,)$ and $\tilde{f}_{\theta_T^*,sp}(.)$ are respectively positives.

ii) $u(.,)$ and $\tilde{f}_{\theta_T^*,sp}(.)$ are bounded as continuous functions over a compact set. Apply Lebesgue dominated convergence theorem.

iii) Note $K_T := \int_{\mathbf{\Theta}^2} u(\theta_e,\theta)\tilde{f}_{\theta_T^*,sp}(\theta)d\theta d\theta_e$. $\square$

Lemma A.23. *For a fixed* $T \in \mathbf{N}$*, define* $\mathbb{P}$ *on the probabilizable space* $(\mathbf{\Theta}, \mathcal{B}(\mathbf{\Theta}))$ *such that*

$$\forall B \in \mathcal{B}(\mathbf{\Theta}), \quad \mathbb{P}(B) := \int_B h_T(\theta_e)d\theta_e.$$

Under Assumptions 1-3;11-12, $\mathbb{P}$*-a.s.*

i) $\mathbb{P}$ *is a probability measure;*

ii) $\forall k \geqslant 0$, $k \mapsto \mathbb{P}\left(\{\theta_e \in \mathbf{\Theta} : h_T(\theta_e) \geqslant k\}\right)$ *is left-continuous decreasing function;*

iii) $k_{\alpha,T}$ *exists;*

iv) $\omega \mapsto k_{\alpha,T}(\omega)$ *is* $\mathcal{E}/\mathcal{B}(\mathbf{R})$*-measurable.*

Proof. *i)* By Lemma A.22 i) and ii), $\mathbb{P}(.) \geqslant 0$ and $\mathbb{P}(\mathbf{\Theta}) = 1$ respectively. By Fubini-Tonelli theorem, $\mathbb{P}(.)$ is countably additive.

ii) Let $\{k_n\}_{n\geqslant 1}$ be an increasing sequence converging to $\overline{k}$ Thus,

$$\begin{aligned} \mathbb{P}\left(\{\theta_e \in \mathbf{\Theta} : h_T(\theta_e) \geqslant \overline{k}\}\right) &= \mathbb{P}\left(\bigcap_{n\geqslant 1}\{\theta_e \in \mathbf{\Theta} : h_T(\theta_e) \geqslant k_n\}\right) \\ &\overset{(a)}{=} \lim_{n\to\infty} \mathbb{P}\left(\{\theta_e \in \mathbf{\Theta} : h_T(\theta_e) \geqslant k_n\}\right) \end{aligned}$$

where (a) comes from a standard continuity property of measures (e.g. Lemma 1.14 p.8 in Kallenberg, 2001).

iii) Define

$$\dot{k} := \sup_{k \in \mathbf{R}} \{k : \mathbb{P}(\{\theta_e \in \mathbf{\Theta} : h_T(\theta_3) \geqslant k\}) \geqslant 1 - \alpha\}$$

By Lemma A.22ii) $h_T(.)$ is bounded, thus by Lemma A.22iii) $\dot{k} < \infty$ exists. Therefore, by definition of a supremum, there exists an increasing sequence $\{k_n\}_{n \geqslant 1}$ converging to $\dot{k}$ such that $\forall n \geqslant 1$, $\mathbb{P}(\{\theta_e : h_T(\theta_3) \geqslant k_n\}) \geqslant 1 - \alpha$. Thus, by ii)

$$\begin{aligned} \mathbb{P}\left(\left\{\theta_e \in \mathbf{\Theta} : h_T(\theta_e) \geqslant \dot{k}\right\}\right) &= \lim_{n \to \infty} \mathbb{P}(\{\theta_e \in \mathbf{\Theta} : h_T(\theta_e) \geqslant k_n\}) \\ &\geqslant 1 - \alpha. \end{aligned}$$

iv) For the same fixed T used to define $\mathbb{P}$, define for this proof

$$\forall \omega \in \mathbf{\Omega}, \forall k \geqslant 0, \qquad g(\omega, k) := \mathbb{P}(\{\theta_e \in \mathbf{\Theta} : h_T(\theta_e) \geqslant k\})$$

By Lemma A.19 and a standard preliminary result to the Fubini theorem (e.g. Lemma 1.26 p.14 in Kallenberg, 2001) $\forall k \geqslant 0$. $g(., k)$ is $\mathcal{E}/\mathcal{B}([0, 1])$-measurable.

Follow the same idea as in Lemma 2 in Jennrich (1969) to finish the proof. $\mathbf{R}_+$ is separable. Thus, there exists an increasing sequence of finite subsets of $\mathbf{R}_+$, $\{\mathbf{R}_{+,n}\}_{n \geqslant 1}$, whose limit, $\mathbf{R}_{+,\infty}$, is dense in $\mathbf{R}_+$. W.lo.g. define $\mathbf{R}_{+,n}$ such that for all $\forall n \geqslant 1$, $0 \in \mathbf{R}_{+,n}$. Denote $\{k_1, k_2, \ldots, k_n\} := \mathbf{R}_{+,n}$

with $k_1 := 0$. For all $n \geqslant 1$, define the function $k_{n,T}(.)$ such that for all $\omega \in \mathbf{\Omega}$,

$$k_{n,T}(\omega) := \begin{cases} k_1 & \text{if } 1-\alpha > g(\omega, k_j),\ \forall j \in [\![2, n]\!] \\ k_2 & \text{if } \begin{cases} g(\omega, k_2) \geqslant 1-\alpha \\ g(\omega, k_2) \geqslant g(\omega, k_j),\ \forall j \in [\![3, n]\!] \end{cases} \\ k_3 & \text{if } \begin{cases} g(\omega, k_3) \geqslant 1-\alpha \\ g(\omega, k_3) \geqslant g(\omega, k_j),\ \forall j \in [\![4, n]\!] \\ g(\omega, k_3) > g(\omega, k_t),\ \forall i \in [\![1, 2]\!] \end{cases} \\ \vdots \\ k_q & \text{if } \begin{cases} g(\omega, k_q) \geqslant 1-\alpha \\ g(\omega, k_q) \geqslant g(\omega, k_j),\ \forall j \in [\![q+1, n]\!] \\ g(\omega, k_q) > g(\omega, k_t),\ \forall i \in [\![1, q-1]\!] \end{cases} \\ \vdots \\ k_n & \text{if } \begin{cases} g(\omega, k_n) \geqslant 1-\alpha \\ g(\omega, k_n) > g(\omega, k_t),\ \forall i \in [\![1, n-1]\!] \end{cases} \end{cases}$$

By construction of $k_{n,T}(.)$,

$$g(\omega, k_{n,T}(\omega)) = \max_{k \in \mathbf{R}_{+,n}} g(\omega, k).$$

Let $B \in \mathcal{B}(\mathbf{R}_+)$. Then

$$\begin{aligned} k_{n,T}^{-1}(B) &= \{\omega \in \mathbf{\Omega} : k_{n,T}(\omega) \in B\} \\ &= \bigcup_{q \in [\![1,n]\!] : k_q \in B} \{\omega \in \mathbf{\Omega} : k_{n,T}(\omega) = k_q\} \end{aligned}$$

Now, $\forall q \in [\![1,n]\!]$, $\{\omega \in \mathbf{\Omega} : k_{n,T}(\omega) = k_q\} \in \mathcal{E}$, since

$$
\begin{aligned}
&\{\omega \in \mathbf{\Omega} : k_{n,T}(\omega) = k_q\} \\
= &\begin{cases}
\bigcap_{j\in[\![2,n]\!]} \{\omega \in \mathbf{\Omega} : 1-\alpha > g(\omega, k_j)\} & \text{if } q=1 \\
\{\omega \in \mathbf{\Omega} : g(\omega,k_2) \geqslant 1-\alpha\} \bigcap \left[\bigcap_{j\in[\![3,n]\!]} \{\omega \in \mathbf{\Omega} : g(\omega,k_2) \geqslant g(\omega,k_j)\}\right] & \text{if } q=2 \\
\{\omega \in \mathbf{\Omega} : g(\omega,k_q) \geqslant 1-\alpha\} \bigcap \left[\bigcap_{j\in[\![q+1,n]\!]} \{\omega \in \mathbf{\Omega} : g(\omega,k_q) \geqslant g(\omega,k_j)\}\right] & \\
\qquad \bigcap \left[\bigcap_{i\in[\![1,q-1]\!]} \{\omega \in \mathbf{\Omega} : g(\omega,k_q) > g(\omega,k_t)\}\right] & \text{if } q \in [\![3,n-1]\!] \\
\{\omega \in \mathbf{\Omega} : g(\omega,k_q) \geqslant 1-\alpha\} \left[\bigcap_{i\in[\![1,n-1]\!]} \{\omega \in \mathbf{\Omega} : g(\omega,k_n) > g(\omega,k_t)\}\right] & \text{if } q=n
\end{cases}
\end{aligned}
$$

Thus, $k_{n,T}^{-1}(B) \in \mathcal{E}$, which means that $k_{n,T}(.)$ is $\mathcal{E}/\mathcal{B}(\mathbf{R}_+)$-measurable.

By the Bolzano-Weierstrass theorem, there exists a converging subsequence $\left\{k_{\beta(n),T}(\omega)\right\}_{n\geqslant 1}$. Assume that $\lim_{n\to\infty} k_{\beta(n),T}(\omega) \neq k_{\alpha,T}(\omega)$. Then, by the left-continuity of $g(\omega,.)$ there exists $\varepsilon > 0$ s.t. $\forall k \in [k_{\alpha,T}(\omega)-\varepsilon, k_{\alpha,T}(\omega)]$ $\forall n \in \mathbf{N}$ $\forall q \in [\![1,n]\!]$ $g(\omega,k_q) \leqslant g(\omega,k)$, which means that $\mathbf{R}_{+,\infty}$ is not dense in $\mathbf{R}_+$. Therefore, by contradiction $\lim_{n\to\infty} k_{\beta(n),T}(\omega) = k_{\alpha,T}(\omega)$. Consequently, $k_{\alpha,T}(.)$ is $\mathcal{E}/\mathcal{B}(\mathbf{R}_+)$-measurable, as a limit of $\mathcal{E}/\mathcal{B}(\mathbf{R}_+)$-measurable functions. □

Lemma A.24. *Under Assumptions 1-7;11-12, as $T \to \infty$*

$$\sup_{\theta_e \in \mathbf{\Theta}} \|h_T(\theta_e) - h(\theta_e)\| \to 0 \quad \mathbb{P}\text{-}a.s.$$

***Proof*.** Denote for this proof $g_T(\theta_e) := \int_{\mathbf{\Theta}} u(\theta_e,\theta)\tilde{f}_{\theta_T^*,sp}(\theta)d\theta$ and $g(\theta_e) := u(\theta_e,\theta_0)$. By the triangle inequality and then the subadditivity of supremum,

$$
\begin{aligned}
&\sup_{\theta_e\in\mathbf{\Theta}} \left\|\frac{1}{K_T}g_T(\theta_e) - \frac{1}{K_\infty}g(\theta_e)\right\| \\
\leqslant\ &\sup_{\theta_e\in\mathbf{\Theta}} \left\|\frac{1}{K_T}g_T(\theta_e) - \frac{1}{K_\infty}g_T(\theta_e)\right\| + \sup_{\theta_e\in\mathbf{\Theta}} \left\|\frac{1}{K_\infty}g_T(\theta_e) - \frac{1}{K_\infty}g(\theta_e)\right\| \\
=\ &\sup_{\theta_e\in\mathbf{\Theta}} \left\|\left[\frac{1}{K_T} - \frac{1}{K_\infty}\right]g_T(\theta)\right\| + \sup_{\theta_e\in\mathbf{\Theta}} \left\|\frac{1}{K_\infty}\left[g_T(\theta_e) - g(\theta_e)\right]\right\|.
\end{aligned}
$$

The first term of the last line can be made arbitrary small as $T \to \infty$ $\mathbb{P}$-a.s., since $K_T \to K_\infty$ by Proposition 7.1iii) and $g_T(.)$ is uniformly bounded w.r.t. T also by Proposition 7.1iii). The second term of the last line can be made arbitrary small as $T \to \infty$ $\mathbb{P}$-a.s. by Proposition 7.1iii). □

Lemma A.25. *Under Assumptions 1-7;11-12, $\{k_{\alpha,T}\}_{T\geqslant 1}$ is bounded $\mathbb{P}$-a.s.*

Proof. Proof by contradiction. Assume that $\{k_{\alpha,t}\}_{T\geqslant 1}$ is unbounded. Now, any real-valued sequence has a monotone subsequence.[40] Then there exists a subsequence $k_{\alpha,\beta(T)} \to \pm\infty$. If $k_{\alpha,\beta(T)} \to \infty$, then for T big enough $\int_{\mathbf{\Theta}} \mathbf{1}_{\{\theta_e \in \mathbf{\Theta}: h_T(\theta) \geqslant k_{\alpha,T}\}} h_T(\theta_e) d\theta_e = 0 < 1-\alpha$, since $h_T(.)$ is bounded over the compact set $\mathbf{\Theta}$ by Lemma A.22ii). If $k_{\alpha,\beta(T)} \to -\infty$, then for T big enough $k_{\alpha,\beta(T)}$ is not the highest bound s.t. $\int_{\mathbf{\Theta}} \mathbf{1}_{\{\theta_e \in \mathbf{\Theta}: h_T(\theta) \geqslant k_{\alpha,T}\}} h_T(\theta_e) d\theta_e \geqslant 1-\alpha$. □

Lemma A.26. *Under Assumptions 1-7;11-12, for all $a \geqslant 0$, for all $\varepsilon > 0$, for T big enough, for all $\dot{\theta}_e \in \mathbf{\Theta}$,*

$$\mathbf{1}_{\{\theta_e \in \mathbf{\Theta}: h_T(\theta_e) > a+\varepsilon\}}(\dot{\theta}_e) \leqslant \mathbf{1}_{\{\theta_e \in \mathbf{\Theta}: h(\theta_e) > a\}}(\dot{\theta}_e)$$

Proof. Let $\dot{\theta}_e \in \{\theta_e \in \mathbf{\Theta} : h_T(\theta_e) > a+\varepsilon\}$ *i.e.* $a+\varepsilon < h_T(\dot{\theta}_e)$. Now, by Lemma A.24, there exists $\dot{T} \in \mathbf{N}$ such that $T \geqslant \dot{T}$ implies $\forall \theta_e \in \mathbf{\Theta}$, $\|h_T(\theta_e) - h(\theta_e)\| < \varepsilon$ which in turn implies $-h(\theta_e) < \varepsilon - h_T(\theta_e)$. Thus, $a < h(\dot{\theta}_e)$ *i.e.* $\dot{\theta}_e \in \{\theta_e \in \mathbf{\Theta} : h(\theta_e) > a\}$. In other words, for $T \geqslant \dot{T}$, $\{\theta_e \in \mathbf{\Theta} : h_T(\theta_e) > a+\varepsilon\} \subset \{\theta_e \in \mathbf{\Theta} : h(\theta_e) > a\}$.□

A.12. Proof of Proposition 7.4

i) Proof by contradiction immediate.

ii) Apply the definition of Dirac distributions.

A.13. Proof of Proposition 7.5

i) By Proposition 4.4i), apply Lemma 2 from Jennrich (1969).

ii) It follows from Corollary 2.

A.14. Proof of Proposition 7.6

i) Adapt proof of Proposition 7.3i).

ii) By Corollary 2, a proof by contradiction is immediate.

[40]Let $\{u_n\}_{n\geqslant 1}$ be a real-valued sequence. Define $E := \{n \in \mathbf{N} : \forall q \geqslant n, u_q \geqslant u_n\}$. If $\#E = \infty$, any infinite subset of E corresponds to an increasing subsequence. If $\#E < \infty$, $\exists n_1 \in \mathbf{N}$ s.t. $\forall n \geqslant n_1$, $\exists q > n$ with $u_q < u_n$. Thus, we can recursively define a strictly decreasing subsequence. Consequently, in both cases the result holds.

A.15. Proof of Proposition 8.1

It is definition-chasing. By Assumption 13,

$$
\begin{aligned}
& \tilde{\mathbb{E}}[u(d_H, \theta_T^*)] \geqslant \tilde{\mathbb{E}}[u(d_A, \theta_T^*)] \\
\Leftrightarrow \quad & c_{d_H} \tilde{\mathbb{F}}_T(\mathbf{\Theta}_H) + b_{d_H} \tilde{\mathbb{F}}_T(\mathbf{\Theta}_A) \geqslant c_{d_A} \tilde{\mathbb{F}}_T(\mathbf{\Theta}_H) + b_{d_A} \tilde{\mathbb{F}}_T(\mathbf{\Theta}_A) \\
\Leftrightarrow \quad & \frac{c_{d_H} - c_{d_A}}{b_{d_A} - b_{d_H}} \tilde{\mathbb{F}}_T(\mathbf{\Theta}_H) \geqslant \tilde{\mathbb{F}}_T(\mathbf{\Theta}_A).
\end{aligned}
$$

A.16. Proof of Proposition 8.2

i) Write $d_T(.)$ with the help of indicator functions.

ii) By Corollary 2, a proof by contradiction is immediate.

A.17. Proof of Proposition 8.3

Adapt proof of Proposition 8.1.

A.18. Proof of Proposition 8.4

Adapt proof of Proposition 8.2.

A.19. Proof of Proposition 9.1

Prove the existence of $\mathbf{\Xi}$. Denote $\mathrm{co}A$ the convex hull of a set A. By Assumptions 15(e*) and triangle inequality, $\sup_{(\underline{x}_T, \phi) \in \aleph^T \times \mathbf{\Phi}} \left\| \frac{1}{T} \sum_{t=1}^T g^{(2)}(x_t, \phi) \right\| \leqslant \sup_{(x,\phi)\in\aleph\times\mathbf{\Phi}} \|g^{(2)}(x, \phi)\| < \infty$. Thus, for $\varepsilon > 0$, the set, $\left\{\xi \in \mathbf{R}^{m-q} : \|\xi - \mathbf{G^{(2)}}\| \leqslant \varepsilon\right\}$ with $\mathbf{G}^{(2)} := \left\{ \frac{1}{T} \sum_{t=1}^T g^{(2)}(x_t, \phi) : x \in \aleph \text{ and } \phi \in \mathbf{\Phi} \right\}$, is closed by continuity of the function $\|. - \mathbf{G}^{(2)}\|$ and bounded by triangle inequality. Now, the convex hull of a compact is compact (e.g., Theorem 1.4.3 on p.100 in Hiriart-Urruty and Lemaréchal, 1993). Thus, $\mathrm{co}\left\{\xi \in \mathbf{R}^{m-q} : \|\xi - \mathbf{G^{(2)}}\| \leqslant \varepsilon\right\}$ is a convex compact set, $\mathbf{\Xi}$, that includes $\left\{g^{(2)}(x, \phi) : (x, \phi) \in \aleph \times \mathbf{\Phi}\right\}$ so that for all $t \in \mathbf{N}$, $\phi \in \mathbf{\Phi}$, and $\omega \in \mathbf{\Omega}$, we can find $\dot{\xi} \in \mathrm{int}(\mathbf{\Xi})$ such that $\frac{1}{T} \sum_{t=1}^T g^{(2)}(X_t(\omega), \phi) - \dot{\xi} = 0_{(m-q)\times 1}$.

It remains to prove Assumption 1(b)-(d)

(b) By Assumption 15(b*), $\mathbf{\Phi}$ is compact. Thus, by the previous paragraph, $\mathbf{\Phi} \times \mathbf{\Xi}$ is compact.

(c) Use Assumptions 15(c*)(d*) and note that a function is measurable if and only if each of its component is measurable (e.g., Lemma 1.8 in Kallenberg, 1997).

(d) Use Assumptions 15(d*) and note that ϕ is solution to $\frac{1}{T}\sum_{t=1}^{T} g^{(1)}(x_t,\phi) = 0_{q\times 1}$ is a necessary condition for $\theta := (\phi' \quad \xi')'$ is solution to $\frac{1}{T}\sum_{t=1}^{T}\psi(x_t,\theta) = 0_{m\times 1}$. $\square$

A.20. Proof of Proposition 9.2

i) For all $(x,\theta) \in \aleph \times \boldsymbol{\Theta}$, $\psi(x,\theta) = g(x,\phi) + (0_{1\times q} \quad \xi')'$ with $\theta := (\phi' \quad \xi')'$. Thus Assumption 16 implies the result because the sum of continuously differentiable functions is continuously differentiable.

ii) By construction, $\hat{\boldsymbol{\Xi}}_T \subset \operatorname{co}\left(\left\{g_t^{(2)}(\phi)\right\}_{t=1}^{T}\right) \subset \boldsymbol{\Xi}$, which is the first result. For the second result, it is sufficient to prove $\hat{\boldsymbol{\Theta}}_T = \dot{\boldsymbol{\Phi}}_T \times \hat{\boldsymbol{\Xi}}_T$ by Lemma A.28.

By definition of $\hat{\boldsymbol{\Xi}}_T$, the first restriction on ϕ in the definition of $\dot{\boldsymbol{\Phi}}_T$ and Lemma A.27iii) is equivalent to the first restriction on θ in the definition of $\hat{\boldsymbol{\Theta}}_T$. Because a product of square matrices is invertible iff each matrix of the product is invertible, and because exponential tilting does not alter the support of the initial distribution, the second and third restriction on ϕ in the definition of $\dot{\boldsymbol{\Phi}}_T$ with Lemma A.27i) is equivalent to the first restriction on θ in the definition of $\hat{\boldsymbol{\Theta}}_T$.

iii) Because a product of square matrices is invertible iff each matrix of the product is invertible, the first restriction on $\phi \in \boldsymbol{\Phi}$ in $\check{\boldsymbol{\Phi}}_T$ combined with Lemma A.27i) and second restriction on $\phi \in \boldsymbol{\Phi}$ in the definition of $\check{\boldsymbol{\Phi}}_T$ are equivalent to the restriction on $\theta \in \boldsymbol{\Theta}$ in the definition of $\check{\boldsymbol{\Theta}}_T$. $\square$

Lemma A.27. *Under Assumptions 1-2 modified according to Assumptions 15-16, for all $\theta \in \boldsymbol{\Theta}$, for all probability measure $\mathbb{P}$,*

i) $$\mathbb{E}_{\mathbb{P}}\left[\frac{\partial\psi(X,\theta)}{\partial\theta'}\right] = \mathbb{E}_{\mathbb{P}}\left[\begin{pmatrix} \frac{\partial g^{(1)}(X,\phi)}{\partial\phi'} & 0_{q\times(m-q)} \\ \frac{\partial g^{(2)}(X,\phi)}{\partial\phi'} & I_{m-q}\end{pmatrix}\right];$$

ii) $\mathbb{V}_{\mathbb{P}}[\psi(X,\theta)]$ *full rank iff* $\mathbb{V}_{\mathbb{P}}[g(X,\phi)]$ *full rank;*

iii) for a fixed $\tau \in \mathbf{R}^m$, $\theta := (\phi' \quad \xi')'$ is solution to $\mathbb{E}_{\mathbb{P}}\left[\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\right] = 0_{m\times 1}$ iff ϕ is solution to $\mathbb{E}\left[g^{(1)}(X,\phi)\mathrm{e}^{\tau' g(X,\phi)}\right] = 0_{q\times 1}$ and $\xi = \mathbb{E}_{\mathbb{P}}\left[g^{(2)}(X,\phi)\mathrm{e}^{\tau' g(X,\phi)}\right]/\mathbb{E}\left[\mathrm{e}^{\tau' g(X,\phi)}\right]$.

Proof. *i)* Differentiate.

ii) Denote $\underline{\xi} := (0_{1\times q} \quad \xi')'$. Then, $\mathbb{V}_{\mathbb{P}}[\psi(X,\theta)] = \mathbb{V}_{\mathbb{P}}\left[g(X,\phi) - \underline{\xi}\right] = \mathbb{V}_{\mathbb{P}}[g(X,\phi)]$.

iii) Denote $\underline{\xi} := (0_{1\times q} \quad \xi')'$. Then

$$\begin{aligned}
& \mathbb{E}_{\mathbb{P}}\left[\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\right] = 0_{m\times 1} \\
\Leftrightarrow \quad & \mathrm{e}^{-\tau_\infty(\phi)'\underline{\xi}}\mathbb{E}_{\mathbb{P}}\left\{\left[g(X,\phi) - \underline{\xi}\right]\mathrm{e}^{\tau' g(X,\phi)}\right\} = 0_{m\times 1} \\
\Leftrightarrow \quad & \begin{cases} \mathbb{E}_{\mathbb{P}}\left[g^{(1)}(X,\phi)\mathrm{e}^{\tau' g(X,\phi)}\right] = 0_{q\times 1} \\ \mathbb{E}_{\mathbb{P}}\left\{\left[g^{(2)}(X,\phi) - \xi\right]\mathrm{e}^{\tau' g(X,\phi)}\right\} = 0_{q\times 1} \end{cases} \\
\Leftrightarrow \quad & \begin{cases} \mathbb{E}_{\mathbb{P}}\left[g^{(1)}(X,\phi)\mathrm{e}^{\tau' g(X,\phi)}\right] = 0_{q\times 1} \\ \xi = \mathbb{E}_{\mathbb{P}}\left[g^{(2)}(X,\phi)\mathrm{e}^{\tau' g(X,\phi)}\right] / \mathbb{E}_{\mathbb{P}}\left[\mathrm{e}^{\tau' g(X,\phi)}\right] \end{cases}
\end{aligned}$$

□

Lemma A.28. *Denote*

$$\dot{\mathbf{\Phi}}_T \quad := \quad \left\{ \phi \in \mathbf{\Phi} : \exists \tau \in \mathbf{R}^m \text{ s.t. } \begin{array}{r} \sum_{t=1}^T g_t^{(1)}(\phi)\mathrm{e}^{\tau' g_t(\phi)} = 0_{q\times 1} \\ \left|\left[\frac{1}{T}\sum_{t=1}^T \frac{\partial g_t^{(1)}(\phi)'}{\partial \phi}\right]\right|_{det} \neq 0 \\ \left|\left[\frac{1}{T}\sum_{t=1}^T g_t(\phi) g_t(\phi)'\right]\right|_{det} \neq 0 \end{array} \right\}.$$

Then, under Assumptions 1-3 modified according to Assumptions 15-17, $\dot{\mathbf{\Phi}}_T = \hat{\mathbf{\Phi}}_T$.

Proof. Denote $\mathrm{co}(A)$ the convex hull of a set A. First prove $\dot{\mathbf{\Phi}}_T \subset \hat{\mathbf{\Phi}}_T$. $\phi \in \dot{\mathbf{\Phi}}_T$ implies $0_{q\times 1} \in \mathrm{int}\left[\mathrm{co}\left(\left\{g_t^{(1)}(\phi)\right\}_{t=1}^T\right)\right]$, which in turn implies $\phi \in \hat{\mathbf{\Phi}}_T$ by duality argument. Conversely, $\hat{\mathbf{\Phi}}_T \subset \dot{\mathbf{\Phi}}_T$ because if $\phi \in \hat{\mathbf{\Phi}}_T$, then $\phi \in \dot{\mathbf{\Phi}}_T$ with $\tau = (\tau_T(\phi)' \quad 0_{(m-q)\times 1})'$.□

A.21. Proof of Proposition 9.3

Lemma A.29-A.31 below provide Assumptions 5-7. □

Lemma A.29. *Under Assumptions 1-4 modified according to Assumptions 15-18, Assumption 4 holds.*

Proof. Prove Assumption 4(b)-(e).

(b) $\mathbb{E}\left[g^{(1)}(X,\dot{\phi})\right] = 0_{m\times 1}$ is a necessary condition for $\mathbb{E}\left[\psi(X,\dot{\theta})\right] = 0_{m\times 1}$ with $\dot{\theta} = (\dot{\phi} \quad \dot{\xi})'$. Then there exists a unique solution $\theta_0 = (\phi_0 \quad \xi_0)$ to $\mathbb{E}\left[\psi(X,\theta)\right] = 0_{m\times 1}$ with $\xi_0 = \mathbb{E}\left[g^{(2)}(X,\phi_0)\right]$ and $\phi_0 \in \mathrm{int}(\mathbf{\Phi})$ by Assumption 18(b*). Thus it remains to prove $\mathbb{E}\left[g^{(2)}(X,\phi_0)\right] \in \mathrm{int}(\mathbf{\Xi})$.

Denote $y_0 := \mathbb{E}\left[g^{(2)}(X,\phi_0)\right]$. By definition of expectation and the one of $\mathbf{\Xi}$ in Proposition 9.1 on p.58, there exist $(\omega_1,\omega_2) \in \mathbf{\Omega}^2$ so that $y_1 := g^{(2)}(X(\omega_1),\phi_0) \in \mathrm{int}(\mathbf{\Xi})$, $y_2 := g^{(2)}(X(\omega_2),\phi_0) \in$

$\operatorname{int}(\boldsymbol{\Xi})$, and there is $\alpha \in [0,1]$ such that $y_0 = \alpha y_1 + (1-\alpha) y_2$. Thus there exists $r_1, r_2 > 0$ s.t. $\mathrm{B}_{r_1}(y_1) \subset \boldsymbol{\Xi}$ and $\mathrm{B}_{r_1}(y_1) \subset \boldsymbol{\Xi}$. Therefore, for all $y \in \mathrm{B}_{\min(r_1,r_2)}(y_0)$, there exists $\varepsilon \in]0, \min(r_1, r_2)]$

$$\begin{aligned} y &= y_0 + \varepsilon u = \alpha y_1 + (1-\alpha) y_2 + \varepsilon u \\ &= \alpha(y_1 + \varepsilon u) + (1-\alpha)(y_2 + \varepsilon u) \end{aligned}$$

where $u := \frac{y - y_0}{\|y - y_0\|}$. Now, $(y_1 + \varepsilon u) \in \mathrm{B}_{r_1}(y_1)$ and $(y_2 + \varepsilon u) \in \mathrm{B}_{r_2}(y_2)$. Thus, by convexity of $\boldsymbol{\Xi}$, $y \in \boldsymbol{\Xi}$, which in turn implies $y_0 \in \operatorname{int}(\boldsymbol{\Xi})$.

(c) For all $\xi \in \boldsymbol{\Xi}$, denote $\underline{\xi} := (0_{1\times q} \quad \xi')'$. Then

$$\begin{aligned} & \psi(X,\theta) = g(X,\phi) - \underline{\xi} \\ \overset{(a)}{\Rightarrow}\ & \|\psi(X,\theta)\| \leqslant \|g(X,\phi)\| + \left\|\underline{\xi}\right\| \\ \Rightarrow\ & \sup_{\theta \in \boldsymbol{\Theta}} \|\psi(X,\theta)\| \leqslant \sup_{\phi \in \boldsymbol{\Phi}} \|g(X,\phi)\| + \sup_{\xi \in \boldsymbol{\Xi}} \left\|\underline{\xi}\right\| \\ \Rightarrow\ & \mathbb{E}\left[\sup_{\theta \in \boldsymbol{\Theta}} \|\psi(X,\theta)\|\right] \leqslant \mathbb{E}\left[\sup_{\phi \in \boldsymbol{\Phi}} \|g(X,\phi)\|\right] + \sup_{\xi \in \boldsymbol{\Xi}} \|\xi\| \overset{(b)}{<} \infty \end{aligned}$$

(a) Triangle inequality. *(b)* Apply Assumption 18(c*) and use the definition of $\boldsymbol{\Xi}$ in Proposition 9.1.

(d) By Lemma A.27i) with $\mathbb{P} = \delta_X$ on p.118

$$\begin{aligned} & \left\|\frac{\partial \psi(X,\theta)}{\partial \theta'}\right\| \leqslant \left\|\frac{\partial g(X,\phi)}{\partial \phi'}\right\| + \|I_{m-q}\| \\ \Rightarrow\ & \mathbb{E}\left[\sup_{\theta \in \boldsymbol{\Theta}} \left\|\frac{\partial \psi(X,\theta)}{\partial \theta'}\right\|\right] \leqslant \mathbb{E}\left[\sup_{\phi \in \boldsymbol{\Phi}} \left\|\frac{\partial g(X,\phi)}{\partial \phi'}\right\|\right] + \|I_{m-q}\| < \infty \end{aligned}$$

where the strict inequality is implied by Assumption 18(d*).

(e) Apply Lemma A.27i).□

Lemma A.30. *Under Assumptions 1-5 modified according to Assumptions 15-19,*

i) $\hat{\boldsymbol{\Xi}}_\infty \subset \boldsymbol{\Xi}$ *and* $\hat{\boldsymbol{\Phi}}_\infty \times \hat{\boldsymbol{\Xi}}_\infty = \hat{\boldsymbol{\Theta}}_\infty$

ii) Assumption 5 holds.

Proof. *i)* By construction, $\hat{\boldsymbol{\Xi}}_\infty \subset \operatorname{co}\left(\left\{g^{(2)}(x,\phi) : (x,\phi) \in \aleph \times \boldsymbol{\Phi}\right\}\right) \subset \boldsymbol{\Xi}$, which is the first result. For the second result, it is sufficient to prove $\hat{\boldsymbol{\Theta}}_\infty = \hat{\boldsymbol{\Phi}}_\infty \times \hat{\boldsymbol{\Xi}}_\infty$.

By definition of $\hat{\boldsymbol{\Xi}}_\infty$, it is sufficient to prove that the restrictions on ϕ in the definition of $\hat{\boldsymbol{\Phi}}_\infty$ are equivalent to the restriction on θ in the definition of $\hat{\boldsymbol{\Theta}}_\infty$. By the compactness of $\boldsymbol{\Xi}$ and Assumption

19(a*), the first restrictions are equivalent because $\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X,\theta)}\right] = \mathbb{E}\left[\mathrm{e}^{\tau'g(X,\phi)}\right]\mathrm{e}^{\tau'\underline{\xi}}$. By Lemma A.27i) on p.118, the second restrictions are equivalent.

By Lemma A.27i) on p.118 the third restriction on ϕ in the definition of $\hat{\mathbf{\Phi}}_\infty$ is equivalent to $\left|\mathbb{E}\left[\mathrm{e}^{\tau'g(X,\phi)}\frac{\partial\psi(X,\theta)'}{\partial\theta}\right]\right|_{det} \neq 0$ because the determinant of block-triangular matrix equals the product of the determinant of block matrices on the diagonal. By Lemma A.27ii) on p.118, the fourth restriction on in the definition of $\hat{\mathbf{\Phi}}_\infty$ is equivalent to $\left|\mathbb{E}\left[\mathrm{e}^{\tau'g(X,\phi)}\psi(X,\phi)\psi(X,\phi)'\right]\right|_{det} \neq 0$. Now, a product of square matrices is invertible iff each matrix of the product is invertible, and exponential tilting does not alter the support of a distribution. Thus, the third and fourth restrictions in the definition of $\hat{\mathbf{\Phi}}_\infty$ are equivalent to the third restriction in definition of $\hat{\mathbf{\Theta}}_\infty$.

By Lemma A.27iii) on p.118, the fifth restriction in the definition of $\hat{\mathbf{\Phi}}_\infty$ is equivalent to the fourth restriction in definition of $\hat{\mathbf{\Theta}}_\infty$.

ii) Prove Assumption 5(a)(b)

(a) Apply i).

(b) For all $\xi \in \mathbf{\Xi}$, denote $\underline{\xi} := (0_{1\times q} \quad \xi')'$. Then

$$\begin{aligned}\sup_{\theta\in B_{r_2}(\dot{\theta})} \|\psi(X,\theta)\mathrm{e}^{\tau'\psi(X,\theta)}\| &= \sup_{(\phi,\xi)\in\mathbf{\Phi}\times\mathbf{\Xi}} \left\|\left[g(X,\phi)-\underline{\xi}\right]\mathrm{e}^{\tau'g(X,\phi)}\right\|\mathrm{e}^{\tau'\underline{\xi}} \\ &\leqslant K\left[\sup_{\phi\in\mathbf{\Phi}}\left\|g(X,\phi)\mathrm{e}^{\tau'g(X,\phi)}\right\| + \sup_{(\phi,\xi)\in\mathbf{\Phi}\times\mathbf{\Xi}}\left\|\xi\mathrm{e}^{\tau'g(X,\phi)}\right\|\right]\end{aligned}$$

where $K := \sup_{\xi\in\mathbf{\Xi}}\mathrm{e}^{\tau'\underline{\xi}} < \infty$ because $\mathbf{\Xi}$ is compact. Now, by compactness of $\mathbf{\Xi}$ and Assumption 21

$$\mathbb{E}\left[\sup_{(\phi,\xi)\in\mathbf{\Phi}\times\mathbf{\Xi}}\left\|\xi\mathrm{e}^{\tau'g(X,\phi)}\right\|\right] = \sup_{\xi\in\mathbf{\Xi}}\|\xi\|\sup_{\phi\in\mathbf{\Phi}}\mathbb{E}\left[\mathrm{e}^{\tau'g(X,\phi)}\right] < \infty.$$

Therefore, by Assumption 19(b*), Assumption 5(b) holds. □

Lemma A.31. *Under Assumptions 1-6 modified according to Assumptions 15-20 , Assumptions 6-7 hold.*

Proof. Prove Assumption 6(a)-(b) and (e).

6(a) Note that a sum of C^4 functions is C^4.

6(b) By Lemma A.27i), for $k \geqslant 1$ $D^k\psi(X,\phi)$ is not a function of ξ.

6(e) For all $\xi \in \mathbf{\Xi}$, denote $\underline{\xi} := (0_{1\times q} \quad \xi')'$. Then

$$
\begin{aligned}
\sup_{\theta\in\boldsymbol{\Theta}}\psi(X,\theta)\psi(X,\theta)' &= \sup_{(\phi,\xi)\in\boldsymbol{\Phi}\times\boldsymbol{\Xi}}[g(X,\phi)-\underline{\xi}][g(X,\phi)-\underline{\xi}]' \\
&= \sup_{(\phi,\xi)\in\boldsymbol{\Phi}\times\boldsymbol{\Xi}} g(X,\phi)g(X,\phi)' - g(X,\phi)\underline{\xi}' - \underline{\xi}g(X,\phi) + \xi\xi' \\
&\overset{(a)}{\leqslant} \sup_{\phi\in\boldsymbol{\Phi}} g(X,\phi)g(X,\phi)' + 2\left[\sup_{\xi\in\boldsymbol{\Xi}}\|\xi\|\right]\left[\sup_{\phi\in\boldsymbol{\Phi}}\|g(X,\phi)\|\right]\iota_{m\times m} \\
&\quad + \sup_{\xi\in\boldsymbol{\Xi}}\|\xi\|\,\iota_{m\times m}
\end{aligned}
$$

where $\iota_{m\times m}$ denotes a square matrix of 1. *(a)* Apply triangle inequality, use subadditivity of supremum, and note that for any matrix $A := \{a_{i,j}\}_{(i,j)\in[\![1,k]\!]\times[\![1,l]\!]}$, $\max_{(i,j)\in[\![1,k]\!]\times[\![1,l]\!]}\{a_{i,j}\} \leqslant \|A\|$ with $\|.\|$ the Euclidean norm.

Now, by Assumption 20(e*) and 18(c*), we have the result.

7(b) By compactness of $\boldsymbol{\Xi}$, Assumption 21(b*) implies Assumption 7 because

$$
\mathbb{E}\left[\sup_{(\tau,\phi,)\in B_{r_1}(\tau_\infty(\dot{\phi}))\times B_{r_2}(\dot{\phi})} \mathrm{e}^{\tau'[g(X,\phi)-\underline{\xi}]}\right] \leqslant \left[\sup_{(\tau,\xi)\in B_{r_1}(\tau_\infty(\dot{\phi}))\times\boldsymbol{\Xi}} \mathrm{e}^{\tau'\underline{\xi}}\right]\mathbb{E}\left[\sup_{(\tau,\phi)\in B_{r_1}(\tau_\infty(\dot{\phi}))\times B_{r_2}(\dot{\phi})} \mathrm{e}^{\tau' g(X,\phi)}\right].
$$

□

A.22. Proof of Proposition 9.4

Adapt proof of Proposition 8.3

A.23. Proof of Proposition 9.5

i) Adapt proof of Proposition 8.4.

ii) Convergence of a joint distribution implies convergence in distribution of marginal distributions (e.g., Theorem 4.29 in Kallenberg, 1997 after normalization of the ESP intensity). Then, adapt the proof of Proposition 8.4 to the marginal ESP intensity $\tilde{f}_{\xi_T^*,sp}(\xi) := \int_{\boldsymbol{\Phi}} \tilde{f}_{\theta_T^*,sp}(\theta)d\phi$.

A.24. Proof of Proposition 9.6

Convergence of a joint distribution implies convergence in distribution of marginal distributions (e.g., Theorem 4.29 in Kallenberg, 1997 after normalization of the ESP intensity). Then, adapt the proof of Proposition 7.6 (Appendix A.14 on p.116) to the utility-weighted marginal ESP intensity $f^{u_c}_{\xi^*_T,sp}(\xi) := \frac{1}{K^u_{T,\xi}} \int_{\mathbf{\Phi}^2} u(\phi_e,\phi)\tilde{f}_{\theta^*_T,sp}(\phi,\xi_e)d\phi d\phi_e$.

A.25. Proof of Proposition 9.7

Adapt proof of Proposition 9.6.

A.26. Lemma A.32

Lemma A.32. *Assume* $\left\{\left(\ln\left(\frac{C_{t+1}}{C_t}\right) \quad \ln\left(R_{1,t+1}\right)\right)'\right\}_{t\geqslant 1}$ *is a sequence of random vectors i.i.d. s.t.*

$$\begin{pmatrix}\ln\left(\frac{C_{t+1}}{C_t}\right)\\ \ln\left(R_{1,t+1}\right)\end{pmatrix} \hookrightarrow \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma^2 & 0\\ 0 & 9\sigma^2\end{pmatrix}\right).$$

If $\sigma = .2$, $\theta_0 = 3$, $\beta = \mathrm{e}^{-\frac{(\theta_0\sigma)^2}{2}}$ *anfd* $\log\widehat{\left(\frac{C_{t+1}}{C_t}\right)} := \log\left(\frac{C_{t+1}}{C_t}\right) + \frac{1}{3}\log\left(R_{1,t+1}\right)$, *then the moment condition*

i) $\mathbb{E}\left[\beta\exp\left(-\theta\log\widehat{\left(\frac{C_{t+1}}{C_t}\right)}\right)R_{1,t+1}-1\right] = 0$ *has only two solutions* $\theta = 3$ *and* $\theta = 0$*;*

ii) $\mathbb{E}\left\{\left[\beta\exp\left(-\theta\log\widehat{\left(\frac{C_{t+1}}{C_t}\right)}\right)R_{1,t+1}-1\right]\left(\frac{1}{3}\log\left(R_{1,t+1}\right)\right)\right\} = 0$ *has a unique solution* $\theta = 3$.

Proof. The result follows from the value of the Laplace transform for a Gaussian distribution.

i) Put $U := \log\left(\frac{C_{t+1}}{C_t}\right)$ and $V := \frac{1}{3}\log\left(R_{1,t+1}\right)$; and then rewrite the moment condition

$$\mathbb{E}\left[\beta\exp\left(-\theta\log\widehat{\left(\frac{C_{t+1}}{C_t}\right)}\right)R_{1,t+1}-1\right] = 0$$

with $\widehat{\log\left(\frac{C_{t+1}}{C_t}\right)} := \log\left(\frac{C_{t+1}}{C_t}\right) + \frac{1}{3}\log\left(R_{1,t+1}\right)$, as

$$\begin{aligned}
& \mathrm{e}^{-\frac{\theta_0^2\sigma^2}{2}}\mathbb{E}\left[\mathrm{e}^{-\theta(U+V)+3V}\right] = 1 \\
\overset{(a)}{\Leftrightarrow}\quad & \mathrm{e}^{-\frac{\theta_0^2\sigma^2}{2}}\mathbb{E}\left[\mathrm{e}^{-\theta U}\right]\mathbb{E}\left[\mathrm{e}^{(3-\theta)V}\right] = 1 \\
\overset{(b)}{\Leftrightarrow}\quad & \mathrm{e}^{-\frac{\theta_0^2\sigma^2}{2}}\mathrm{e}^{\frac{\theta^2\sigma^2}{2}}\mathrm{e}^{\frac{(3-\theta)^2\sigma^2}{2}} = 1 \\
\overset{(c)}{\Leftrightarrow}\quad & -\frac{\theta_0^2\sigma^2}{2} + \frac{\theta^2\sigma^2}{2} + \frac{(3-\theta)^2\sigma^2}{2} = 0
\end{aligned}$$

(a) Independance of U and V. *(b)* Laplace transform of a Gaussian distribution. If $Y \hookrightarrow \mathcal{N}(m_Y, \sigma_Y)$, then for all $t \in \mathbf{R}$ $\mathbb{E}\left[\mathrm{e}^{tY}\right] = \exp\left(tm_Y + \frac{t^2\sigma_Y^2}{2}\right)$. *(c)* Take logarithm.

Now, a second order polynomial can at most have two roots; and for $\theta_0 = 3$, $\theta = 3$ and $\theta = 0$ are two roots of the last equation. Thus, these are the only two solutions to the moment condition.

*ii)*Use the same notations as in i); and then rewrite the moment condition

$$\mathbb{E}\left\{\left[\beta\exp\left(-\theta\widehat{\log\left(\frac{C_{t+1}}{C_t}\right)}\right)R_{1,t+1} - 1\right]\left(\frac{1}{3}\log\left(R_{1,t+1}\right)\right)\right\} = 0$$

as

$$\begin{aligned}
& \mathrm{e}^{-\frac{\theta_0^2\sigma^2}{2}}\mathbb{E}\left[V\mathrm{e}^{-\theta(U+V)+3V} - V\right] = 0 \\
\overset{(a)}{\Leftrightarrow}\quad & \mathrm{e}^{-\frac{\theta_0^2\sigma^2}{2}}\mathbb{E}\left[\mathrm{e}^{-\theta U}\right]\mathbb{E}\left[V\mathrm{e}^{(3-\theta)V}\right] = 0 \\
\overset{(b)}{\Leftrightarrow}\quad & \mathrm{e}^{-\frac{\theta_0^2\sigma^2}{2}}\mathrm{e}^{\frac{\theta^2\sigma^2}{2}}\sigma^2(3-\theta)\mathrm{e}^{\frac{(3-\theta)^2\sigma^2}{2}} = 0
\end{aligned}$$

(a) Independance of U and V and $\mathbb{E}(V) = 0$. *(b)* Laplace transform of a Gaussian distribution and $\mathbb{E}\left[Y\mathrm{e}^{tY}\right] = \frac{\partial\mathbb{E}\left[\mathrm{e}^{tY}\right]}{\partial t} = \frac{\partial\left[\exp\left(\frac{t^2\sigma_Y^2}{2}\right)\right]}{\partial t} = \frac{\sigma^2}{2}2t\mathrm{e}^{\frac{t^2\sigma^2}{2}}$.

Since all the terms in the last equation are striclty positive except $(3-\theta)$, the only solution is $\theta = 3$.