Semi-Nonparametric Indirect Inference

Francisco Blasques*
School of Business and Economics
Maastricht University

January 29, 2011

Abstract

This paper proposes a sieve extremum estimator for semi-nonparametric models that relies on auxiliary statistics through the principle of indirect inference. The parameter space is allowed to be unbounded and infinite dimensional. The estimator is shown to be $\sqrt{T}$ consistent and asymptotically Gaussian under general regularity conditions. The data is allowed to exhibit heterogeneous and dependent behavior. Furthermore, in the tradition of indirect inference, these results apply to a large class of complex dynamic models with unobserved variables, including those yielding an estimator with no closed form algebraic representation or featuring a criterion function which is intractable or infeasible, even on appropriately chosen compact finite-dimensional sieves.

1 Introduction

Interest in conducting statistical inference on unknown parameters that are allowed to lie on infinite dimensional spaces has gained popularity often as means of avoiding the restrictiveness of parametric models and the undesirable consequences of incorrect specification. Unfortunately, conducting inference on infinite dimensional parameters has proven difficult on several occasions. Extremum estimators obtained by optimizing a criterion function defined on an infinite dimensional space may be afflicted by ill-posedness issues, inconsistency problems, or exhibit very slow convergence rates; see Chen (2007) and references therein. Such undesirable properties are generally caused by the excessive size and/or complexity (as typically measured by the covering number or entropy value) of the possibly non-compact parameter space. One possible solution to this problem involves defining an extremum

*The author is thankful to Eric Beutner, Bertrand Candelon and Jean Pierre Urbain for useful comments and suggestions. The usual disclaimer applies. Corresponding Address: Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: f.blasques@maastrichtuniversity.nl.

estimator that optimizes the criterion function over a sequence of sets (called *sieves*), of increasing size and complexity, that is dense in the entire infinite dimensional parameter space of interest.

Such estimators, called *sieve extremum estimators*, have been studied (to some extent separately) in both the statistics and econometrics literature. The theoretical properties of these estimators are often difficult to derive as they depend on both *(i)* the approximation error introduced by constraining the optimization to the sieves, and *(ii)* the convergence properties of the random estimator sequence within each sieve. Regardless of these difficulties, considerable progress has been made since the appearance in the statistics literature of the method of sieves, credited to Grenander (1981), and the introduction in econometrics of semi-nonparametric models, by the hand of Gallant (1981). See Chen (2007) for a comprehensive review of the relevant literature.

Consistency proofs have been obtained for the general sieve extremum estimator under mild regularity conditions that allow for great generality in the choice of sieves and for a variety of forms of dependence and heterogeneity to be present in the data; see e.g. Gallant (1987), White and Wooldrige (1991) and Chen (2007). A multitude of results have also been obtained for special forms of the sieve extremum estimator such as for sieve M-estimators, sieve minimum-distance estimators, and series-estimators; see Chen (2007).

General results on convergence rates and asymptotic normality have been harder to establish and are seemingly not available for the general sieve extremum estimator. Some results exist however for specific formulations of the sieve estimator that take advantage of the particular properties of the sieves and/or criterion functions being used.

Theoretical results on convergence rates are available with some generality for the special case of sieve M-estimators and series estimators. These typically rely on empirical process theory. In general, the convergence rate of sieve M-estimators will be slow if the size and/or complexity of the sieves increases slowly with T (in which case approximation errors decrease slowly and dominate) but also, when the entropy of the sieves grows too fast (in which case the estimator's convergence within sieves is typically slow and dominates). Obtaining an appropriate rate of convergence of sieve estimators thus requires an exact "rate of expansion" of the sieves that precisely balances the relation between the approximation error and the rate of convergence of the estimator within each sieve. Relevant literature on the convergence rates of the sieve M-estimator includes Van de Geer (1995), Shen and Wong (1994) and Birge and Massart (1998) and Chen and Shen (1998). Results on the sieve maximum likelihood estimator are also available in Van de Geer (1993) and Wong and Shen (1995). Finally, other results exists also for specific sieves and criterion functions; see references in Chen (2007).

Asymptotic normality results for sieve estimators are still scarce and in general apply only to either series least squares estimators or to the finite dimensional parametric part of semi-parametric models; see e.g. Andrews (1991), Gallant and Souza (1991), Newey (1994, 1997), Zhou et al. (1998) and Huang (2003) for results on series least-squares estimators, and Chen et al. (2003), Wong and Severini (1991), Gallant and Souza (1991), Shen (1997) and Chen and Shen (1998) for both two-step and simultaneous M-estimators.

This paper establishes the consistency, $\sqrt{T}$-convergence rate, and asymptotic normality of a sieve extremum estimator that relies on auxiliary statistics through the unifying principle of indirect inference introduced in Gourieroux et al. (1993) and Smith (1993). We call the resulting estimator a *Semi-NonParametric Indirect Inference* (SNPII) estimator. The properties of the SNPII estimator are shown to be preserved under mild conditions on "the rate of expansion" of the sieves and to be retained under general dependence and heterogeneity conditions on the data. These conditions seem to be weaker than those typically found in the literature. Furthermore, in the tradition of indirect inference, these results apply to a large class of models, including those yielding an extremum estimator of interest with no closed form algebraic representation or having a criterion function which is intractable or infeasible, even on appropriately chosen compact finite-dimensional sieves; see Gourieroux and Monfort (1997), e.g. Dhaene et al. (1998), Genton and de Luna (2000), Genton and Ronchetti (2003) and Dridi et al. (2007). This also seems to constitute a generalization of existing results that is likely to be of practical interest.

This paper studies also the possibility of conducting inference on continuous and/or smooth functionals defined on the parameter space. Functionals such as projection maps are likely to be of interest in applications. In particular, while the appropriately standardized SNPII estimator is shown to converge weakly to a Gaussian process, smooth functionals are shown to converge to appropriate Gaussian limits, be these finite random variables or random processes as well.

Finally, a word on notation. Throughout, we let $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$ denote the sets of natural, integer and real numbers respectively. Given a set $\mathbb{A}$, we let $\mathcal{T}_{\mathbb{A}}$ denote a topology on $\mathbb{A}$. Given a topological space $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$, we let $\mathfrak{B}(\mathbb{A})$ denote the Borel σ-algebra generated by $\mathcal{T}_{\mathbb{A}}$, and denote the closure of $\mathbb{A}$ by $\text{cl}(\mathbb{A})$. A divergence on $\mathbb{A}$ is denoted $d_{\mathbb{A}}$. A metric on $\mathbb{A}$ is denoted $\delta_{\mathbb{A}}$. When $\mathbb{A}$ is a vector space, then $\|\cdot\|_{\mathbb{A}}$ and $\langle\cdot,\cdot\rangle_{\mathbb{A}}$ denote a norm and an inner-product on $\mathbb{A}$ respectively. Given a metric space $(\mathbb{A}, \delta_{\mathbb{A}})$ we let $S_{a_0}(\epsilon)$ denote an open ball of radius $\epsilon > 0$ centered at $a_0 \in \mathbb{A}$, i.e. $S_{a_0}(\epsilon) := \{a \in \mathbb{A} \, : \, \delta_{\mathbb{A}}(a_0, a) < \epsilon\}$, and also, $S^c_{a_0}(\epsilon)$ be its complement in $\mathbb{A}$, i.e. $S^c_{a_0}(\epsilon) = \mathbb{A} \backslash S_{a_0}(\epsilon)$. For any index set $\mathbb{I}$ and a collection of sets $\mathbb{A}_i$, $i \in \mathbb{I}$, we let $\times_{i \in \mathcal{I}} \mathbb{A}_i$ denote the Cartesian product of the sets $\mathbb{A}_i$. If $\mathfrak{A}_i$ is a σ-algebra for every $i \in \mathbb{I}$, then $\otimes_{i \in \mathbb{I}} \mathfrak{A}_i$ denotes the product σ-algebra. Also, given the product space $\mathbb{A} = \times_{i \in \mathcal{I}} \mathbb{A}_i$, projections operators are denoted $\pi_i : \mathbb{A} \to \mathbb{A}_i$. Given two topological spaces $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ and $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ we let $\mathbb{C}(\mathbb{A}, \mathbb{B})$ denote the space of continuous functions mapping from $\mathbb{A}$ into $\mathbb{B}$. If $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ and $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ are topological vector spaces then $\mathbb{L}(\mathbb{A}, \mathbb{B})$ denotes the space of bounded (hence continuous) linear operators from $\mathbb{A}$ into $\mathbb{B}$. Similarly, $\mathbb{L}^2(\mathbb{A} \times \mathbb{A}, \mathbb{B})$ denotes the space of bounded bilinear operators from $\mathbb{A}$ into $\mathbb{B}$. Furthermore, given a map $f : \mathbb{A} \to \mathbb{B}$ we let $\nabla_{\mathbb{A}_0} f(a_0, a)$ denote either the Frechet or Hadamard derivative of f at $\boldsymbol{\theta}_0$, tangentially to $\mathbb{A}_0 \subseteq \mathbb{A}$, evaluated at $a \in \mathbb{A}_0$. Finally, $\stackrel{d}{\to}$, $\stackrel{p}{\to}$ and $\stackrel{a.s.}{\to}$ are used to denote convergence in distribution, probability and almost surely, respectively.

2 Preliminary Considerations

Observed data consists of a T-sequence $\mathbf{x_T}(\omega) := \{\mathbf{x}_t(\omega)\}_{t=1}^T$ of points in $\mathbb{R}^{n_x}$, $(T, n_x) \in \mathbb{N} \times \mathbb{N}$, a subset of the realized path of a n_x-variate stochastic sequence $\mathbf{x}(\omega) = \{\mathbf{x}_t(\omega), t \in \mathbb{Z}\}$ for some element ω of the event space Ω of a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F}$ denotes a σ-algebra defined on Ω and $\mathbb{P}$ a probability measure on $\mathcal{F}$.[1] The random sequence $\mathbf{x}$ is thus an $\mathcal{F}/\mathfrak{B}(\mathbb{R}^{n_x}_\infty)$-measurable mapping $\mathbf{x} : \Omega \to \mathbb{R}^{n_x}_\infty$ taking values in the Cartesian product of infinite copies of $\mathbb{R}^{n_x}$, denoted $\mathbb{R}^{n_x}_\infty := \times_{t=-\infty}^{t=\infty} \mathbb{R}^{n_x}$. The stochastic sequence $\mathbf{x}(\omega)$ lives on the space $(\mathbb{R}^{n_x}_\infty, \mathfrak{B}^{n_x}_\infty, \mathrm{D}_0)$ were the induced probability measure (p.m.) D_0 is naturally defined over the elements of the Borel σ-algebra $\mathfrak{B}^{n_x}_\infty := \mathfrak{B}(\mathbb{R}^{n_x}_\infty)$ generated by the finite dimensional product cylinders of $\mathbb{R}^{n_x}_\infty$.

The model of interest consists of a family $\mathcal{D}_\Theta$ of p.m.s $\mathrm{D}(\boldsymbol{\theta})$ defined on $\mathfrak{B}^{n_x}_\infty$. We let the elements of this family be indexed by a possibly infinite-dimensional parameter $\boldsymbol{\theta} \in \Theta$, so that $\mathcal{D}_\Theta = \{\mathrm{D}(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta\}$. The metric space (Θ, δ_Θ), called *parameter space*, is assumed to possess the following properties.

Assumption 1. *The parameter space* (Θ, δ_Θ) *is a complete, separable and measurable metric space with Borel* σ*-algebra* $\mathfrak{B}(\Theta)$ *generated by the topology* $\mathcal{T}_\Theta$ *induced by the metric* $\delta_\Theta : \Theta \times \Theta \to \mathbb{R}$ *on* Θ.[2]

Subsets of Θ, called *sieves*, will be indexed by $T \in \mathbb{N}$ and denoted $\Theta_T \subseteq \Theta\ \forall T \in \mathbb{N}$. Let $\mathcal{D}$ denote the set of all probability measures on $\mathfrak{B}^{n_x}_\infty$, then, by definition, the subset $\mathcal{D}_\Theta \subseteq \mathcal{D}$ is the image of Θ under $\mathrm{D} : \Theta \to \mathcal{D}$ with $\mathrm{D}(\boldsymbol{\theta}) : \mathfrak{B}^{n_x}_\infty \to [0, 1]$ for every $\boldsymbol{\theta} \in \Theta$. The sieves are typically designed to possess desirable features (e.g. compactness) that are especially convenient for working with extremum estimators. A mild form of correct specification shall be eventually assumed. Namely, that $\exists\, \boldsymbol{\theta}_0 \in \Theta$ such that $\mathrm{D}(\boldsymbol{\theta}_0) = \mathrm{D}_0$. However, we allow for the possibility that $\boldsymbol{\theta}_0 \notin \Theta_T\ \forall T \in \mathbb{N}$, requiring only that the sequence of sieves be increasing and dense on Θ. This is a distinct characteristic of the method of sieves.

Assumption 2. *The sieves* $\{\Theta_T\}_{T \in \mathbb{N}}$ *are non-empty compact subsets of* Θ *satisfying* $\Theta_T \subseteq \Theta_{T+1} \subseteq \Theta\ \forall T \in \mathbb{N}$ *and* $\mathrm{cl}\big(\bigcup_{T \in \mathbb{N}} \Theta_T\big) \supseteq \Theta$*. Furthermore,* $\exists \boldsymbol{\theta}_0 \in \Theta : \mathrm{D}(\boldsymbol{\theta}_0) = \mathrm{D}_0$*, i.e.* $\mathrm{D}_0 \in \mathcal{D}_\Theta$*.*

Indirect inference on $\boldsymbol{\theta}_0$ is to be conducted under the assumption that it is possible to "draw" from the distribution $\mathrm{D}(\boldsymbol{\theta})$ for every $\boldsymbol{\theta}$ lying on well chosen (possibly finite dimensional) subsets $\Theta_T \subseteq \Theta$, $\forall T \in \mathbb{N}$. In other words, it must be possible to obtain T-period subsets $\tilde{\mathbf{x}}_\mathbf{T}(\boldsymbol{\theta}, \omega) := \{\tilde{\mathbf{x}}_t(\boldsymbol{\theta}, \omega)\}_{T=1}^T$ of the realized path of the stochastic sequence $\tilde{\mathbf{x}}(\boldsymbol{\theta}) : \Omega \to \mathbb{R}^{n_x}_\infty$, living in $(\mathbb{R}^{n_x}_\infty, \mathfrak{B}^{n_x}_\infty, \mathrm{D}(\boldsymbol{\theta}))$, for every $\boldsymbol{\theta} \in \Theta_T$, $\forall T \in \mathbb{N}$. This seems hardly restrictive in practice, see e.g. Gourieroux and Monfort (1995). Note in particular,that we do not require the ability to "draw" from D_0 since it is possible that $\mathrm{D}_0 \notin \mathcal{D}_{\Theta_T}\ \forall T \in \mathbb{N}$, where $\mathcal{D}_{\Theta_T} := \{\mathrm{D}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_T\}$, even though $\mathrm{D}_0 \in \mathcal{D}_\Theta$. In this sense, statistical inference is conducted using a sequence of possibly misspecified models $\mathcal{D}_{\Theta_T} \subseteq \mathcal{D}\ \forall T \in \mathbb{N}$.

[1] Given a measurable space $(\mathcal{A}, \mathfrak{B}(\mathcal{A}))$, a measurable map $f : \Omega \to \mathcal{A}$ and some $A \in \mathfrak{B}(\mathcal{A})$, we shall often write $\mathbb{P}(A)$ instead of $\mathbb{P}(\omega \in \Omega : a(\omega) \in A)$ when there is no risk of ambiguity.

[2] Θ is thus succinctly described as a Polish metric space with Borel σ-algebra $\mathfrak{B}(\Theta)$ (see Definition 1).

When considering several draws of T-period sequences $\tilde{\mathbf{x}}_{\mathbf{T}}(\boldsymbol{\theta})$ from $\mathrm{D}(\boldsymbol{\theta})$, these shall be indexed by $s \in \{1, ..., S\}$, $S \in \mathbb{N}$ and denoted $\tilde{\mathbf{x}}^s_{\mathbf{T}}(\boldsymbol{\theta})$.

Finally, we define also a topological vector space $(\mathcal{B}, \mathcal{T}_{\mathcal{B}})$ called the *auxiliary parameter space*. Indirect inference on elements of Θ shall be conducted "through" inference on elements of $\mathcal{B}$. The auxiliary space $\mathcal{B}$ is obtained as the Cartesian product of a collection of *auxiliary factor spaces* $\mathcal{B}_L$, $L \in \mathcal{L}$, where $\mathcal{L}$ is a countable index set. We require that an appropriate topology be defined on the product space $\mathcal{B} = \times_{L\in\mathcal{L}}\mathcal{B}_L$, namely, Tychonoff's product topology. This ensures continuity of the projection maps $\pi_L : \mathcal{B} \to \mathcal{B}_L \; \forall L \in \mathcal{L}$ (Lemma 6). An immediate consequence is that convergence of a sequence $\{b_T\}_{T\in\mathbb{N}}$ on $\mathcal{B}$ implies (and is implied by) the convergence of the projection sequences $\{\pi_L(b_T)\}_{T\in\mathbb{N}}$ on $\mathcal{B}_L$ for every $L \in \mathcal{L}$ (Corollary 11). This in turn implies that continuity of operators f mapping from any topological space $\mathbb{A}$ into $\mathcal{B}$ holds if and only if $\pi_L \circ f : \mathbb{A} \to \mathcal{B}_L$ is continuous $\forall L \in \mathcal{L}$ (Lemma 7). Moreover, compactness of subsets of $\mathcal{B}^* = \times_{L\in\mathcal{L}}\mathcal{B}^*_L \subseteq \mathcal{B}$ follows from compactness of every $\mathcal{B}^*_L$ (Lemma 8). We shall often require $\mathcal{B}$ to be equipped with a metric $\delta_{\mathcal{B}}$. As such $(\mathcal{B}, \mathcal{T}_{\mathcal{B}})$ is assumed to be metrizable, and hence also Hausdorff (Lemma 2).[3] This can be obtained by having $\mathcal{B}_L$ be regular and second countable for every $L \in \mathcal{L}$ (Lemmas 1, 9 and 10). Clearly, it is imposed from the outset that the metric $\delta_{\mathcal{B}}$ on $\mathcal{B}$ be a product metric inducing the product topology $\mathcal{T}_{\mathcal{B}}$ on $\mathcal{B}$. When $\mathcal{L}$ is finite, this is not restrictive. When $\mathcal{L}$ is (countably) infinite however, then the restriction is more obvious. Product metrics inducing the desired topology on countable product spaces are,[4]

$$\delta_{\mathcal{B}}(\boldsymbol{\beta}, \boldsymbol{\beta}') = \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{\delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i}, \boldsymbol{\beta}'_{L_i})}{1 + \delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i}, \boldsymbol{\beta}'_{L_i})} \quad \text{and} \quad \delta_{\mathcal{B}}(\boldsymbol{\beta}, \boldsymbol{\beta}') = \sup_{i\in\mathbb{N}} \frac{1}{i} \frac{\delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i}, \boldsymbol{\beta}'_{L_i})}{1 + \delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i}, \boldsymbol{\beta}'_{L_i})}, \tag{1}$$

for every $(\boldsymbol{\beta}, \boldsymbol{\beta}') \in \mathcal{B} \times \mathcal{B}$ (Lemma 11) and the factor vector spaces. Finally, note that measurability statements involving $(\mathcal{B}, \mathcal{T}_{\mathcal{B}})$ are made w.r.t. the Borel σ-algebra $\mathfrak{B}(\mathcal{B})$ generated by $\mathcal{T}_{\mathcal{B}}$.

Assumption 3. *The auxiliary parameter space* $(\mathcal{B}, \delta_{\mathcal{B}})$ *is a measurable metric space, a countable Cartesian product* $\mathcal{B} := \times_{L\in\mathcal{L}}\mathcal{B}_L$ *of complete separable normed topological vector spaces* $(\mathcal{B}_L, \mathcal{T}_{\mathcal{B}_L})$ *equipped with norms* $\delta_{\mathcal{B}_L}(\boldsymbol{\beta}_L, \boldsymbol{\beta}'_L) := \|\boldsymbol{\beta}_L - \boldsymbol{\beta}'_L\|_{\mathcal{B}_L} \; \forall (\boldsymbol{\beta}_L, \boldsymbol{\beta}'_L) \in \mathcal{B}_L \times \mathcal{B}_L$ *for every* $L \in \mathcal{L}$. *The product space* $(\mathcal{B}, \delta_{\mathcal{B}})$ *is equipped with a metric* $\delta_{\mathcal{B}} : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ *inducing Tychonoff's topology* $\mathcal{T}_{\mathcal{B}}$ *on* $\mathcal{B}$ *and a Borel* σ*-algebra* $\mathfrak{B}(\mathcal{B})$ *generated by* $\mathcal{T}_{\mathcal{B}}$.

Note that Assumption 3 implies that $\mathcal{B}$ is separable (Lemma 10). Second countability of $\mathcal{B}$ is also implied by Assumption 3 (Lemma 13). Furthermore, this implies by Lemma 12 that $\mathfrak{B}(\mathcal{B}) = \otimes_{L\in\mathcal{L}}\mathfrak{B}(\mathcal{B}_L)$. These algebras are thus used interchangeably. Most importantly, the projection mappings $\pi_L : \mathcal{B} \to B_L$ are $\mathfrak{B}(\mathcal{B})/\mathfrak{B}(\mathcal{B}_L)$-measurable $\forall L \in \mathcal{L}$ (Corollary 12) and $\mathfrak{B}(\mathcal{B})/\mathfrak{B}(\mathbb{A})$-measurability of maps f from any measurable space $(\mathbb{A}, \mathfrak{B}(\mathbb{A}))$ into $(\mathcal{B}, \mathfrak{B}(\mathcal{B}))$ is implied by the $\mathfrak{B}(\mathcal{B}_L)/\mathfrak{B}(\mathbb{A})$-measurability of the projection maps $\pi_L \circ f : \mathbb{A} \to \mathcal{B}_L$ for every $L \in \mathcal{L}$ (see Corollary 13).

[3] We do not require the metric on the linear space to be a norm. The latter requires homogeneity and translation invariance which is sometimes unnecessary.

[4] When $\mathcal{L}$ is countably infinite, its elements are naturally indexed by the natural numbers $i \in \mathbb{N}$. In (1) the factor spaces $(\mathcal{B}_L, \mathcal{T}_{\mathcal{B}_L})$ are equipped with metrics $\delta_{\mathcal{B}_L} : \mathcal{B}_L \times \mathcal{B}_L \to \mathbb{R} \; \forall L \in \mathcal{L}$.

3 The SNPII Estimator

For some fixed $S \in \mathbb{N}$ (whose role is described below), define the map $\hat{\boldsymbol{\theta}}_{T,S} : \Omega \to \Theta_T$ as being such that, for every $\omega \in \Omega$, the image $\hat{\boldsymbol{\theta}}_{T,S}(\omega)$ is an element of the arg min set of a criterion function $Q_{T,S}(\omega) : \Theta \to \mathbb{R}$ over the sieve $\Theta_T \subseteq \Theta \; \forall T \in \mathbb{N}$,

$$\hat{\boldsymbol{\theta}}_{T,S} \in \arg \min_{\boldsymbol{\theta} \in \Theta_T} Q_{T,S}(\boldsymbol{\theta}). \tag{2}$$

The sequence $\{\hat{\boldsymbol{\theta}}_{T,S}(\omega)\}_{T \in \mathbb{N}}$ of points in Θ thus corresponds to a sequence of minimizers of a sequence of real-valued maps $\{Q_{T,S}(\omega)\}_{T \in \mathbb{N}}$ on a sequence of sieves $\{\Theta_T\}_{T \in \mathbb{N}}$. In what follows, conditions shall be imposed on the criterion functions $Q_{T,S} : \Theta \times \Omega \to \mathbb{R}$ and the sieves Θ_T so as to guarantee that the arg min set exists and that $Q_{T,S}$ converges in some appropriate sense to a limit deterministic criterion function $Q_\infty : \Theta \to \mathbb{R}$.[5] When such conditions are too restrictive, then the above definition can easily be relaxed to that of an approximate extremum estimator $\hat{\boldsymbol{\theta}}_{T,S}$ satisfying, for fixed $S \in \mathbb{N}$,

$$Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}) \leq \inf_{\boldsymbol{\theta} \in \Theta_T} Q_{T,S}(\boldsymbol{\theta}) + O_p(\eta_T), \tag{3}$$

With $\eta_T \to 0$ as $T \to \infty$. Clearly, setting $O_p(\eta_T) = 0 \; \forall T \in \mathbb{N}$ yields an exact sieve extremum estimator. When furthermore the arg min set exists, then the extremum estimator is given by (2) above. Now, the fundamental feature that turns the extremum estimator in either (2) or (3) into an SNPII estimator is its appropriate definition as a minimizer of a divergence defined on the auxiliary parameter space $\mathcal{B}$.

In particular, let us define the maps $\hat{\boldsymbol{\beta}}_T : \Omega \to \mathcal{B}$ and $\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) : \Omega \to \mathcal{B}$, $\forall \boldsymbol{\theta} \in \Theta$. Each of these consists of a vector of random variables (here called *auxiliary estimators* or *auxiliary statistics*) indexed by $L \in \mathcal{L}$ and taking values on the factor-spaces $\mathcal{B}_L$. The first set,

$$\hat{\boldsymbol{\beta}}_T := \{\hat{\boldsymbol{\beta}}_T^L, L \in \mathcal{L}\}$$

collects those estimators $\hat{\boldsymbol{\beta}}_T^L : \Omega \to \mathcal{B}_L$ that are functions of observed data $\mathbf{x_T}$. Immediate examples are *(i)* statistics of the type $\hat{\boldsymbol{\beta}}_T^L = \sum_{t=1}^T L(\mathbf{x}_t)$ where $L : \mathcal{X} \to \mathcal{B}_L$, and *(ii)* typical extremum estimators of the form $\hat{\boldsymbol{\beta}}_T^L = \arg\min_{\boldsymbol{\beta}_L \in \mathcal{B}_L} L(\mathbf{x_T}, \boldsymbol{\beta}_L)$ where L is some criterion function $L : \mathcal{X}_T \times \mathcal{B}_L \to \mathbb{R}$. Auxiliary estimators of interest should be simple to work with in applications and designed so as to possess desirable convergence properties. In particular, they should take values on well chosen (possibly finite dimensional compact) factor spaces $\mathcal{B}_L$ so that they do not suffer from the complications of estimation on large complex spaces. This is not restrictive, since in the spirit of indirect inference, auxiliary statistics and factor spaces $\mathcal{B}_L$ can be almost arbitrary. The second set of auxiliary estimators,

$$\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) := \left\{1/S \sum_{s=1}^{S} \tilde{\boldsymbol{\beta}}_{T,s}^L(\boldsymbol{\theta}), L \in \mathcal{L}\right\}$$

[5] There is seemingly nothing to be gained by letting $Q_{T,S}$ be defined only on the sieves Θ_T, i.e. by letting $Q_{T,S} : \Theta_T \times \Omega \to \mathbb{R}_0^+$, since an agreeing measurable extension is guaranteed to exist on Θ (see e.g. Stinchcombe and White (1992, Lemma 2.14)

collects (for any given $\boldsymbol{\theta} \in \Theta$) averages of those estimators $\tilde{\boldsymbol{\beta}}^L_T(\boldsymbol{\theta}) : \Omega \to \mathcal{B}_L$ that are functions of the "artificial" sequence of data $\tilde{\mathbf{x}}^s_{\mathbf{T}}(\boldsymbol{\theta})$ drawn from $\mathrm{D}(\boldsymbol{\theta})$. Again, immediate examples are *(i)* statistics of the type $\tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) = \sum_{t=1}^T L \circ \tilde{\mathbf{x}}^s_{\mathbf{T}}(\boldsymbol{\theta})$, and *(ii)* extremum estimators of the form $\tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\beta}_L \in \mathcal{B}_L} L(\tilde{\mathbf{x}}^s_{\mathbf{T}}(\boldsymbol{\theta}), \boldsymbol{\beta}_L)$. These estimators should have desirable properties $\forall \boldsymbol{\theta} \in \Theta_T$, $T \in \mathbb{N}$.

In particular, to be useful for indirect inference, $\hat{\boldsymbol{\beta}}_T$ should convergence in suitable manner to a limit point $\boldsymbol{\beta}^*_0$ in $\mathcal{B}$, and most importantly, the random map $\tilde{\boldsymbol{\beta}}_{T,S} : \Omega \times \Theta \to \mathcal{B}$, called *empirical binding function*, should also converge in an appropriate fashion to a limit deterministic map $\boldsymbol{\beta}^* : \Theta \to \mathcal{B}$, called the *binding function*. Under Tychonoff's topology on $\mathcal{B}$ this shall be obtained by the appropriate convergence of the projections in $\mathcal{B}_L$. The indirect inference methodology will then rely on *(i)* having an injective binding function $\boldsymbol{\beta}^*$ and *(ii)* the fact that $\boldsymbol{\beta}^*_0 = \boldsymbol{\beta}^*(\boldsymbol{\theta}_0)$, which is obtained since $\mathrm{D}_0 = \mathrm{D}(\boldsymbol{\theta}_0)$. Making use of the two sets of auxiliary estimators $\hat{\boldsymbol{\beta}}_T$ and $\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$, the SNPII estimator's criterion function, and its deterministic limit, are finally appropriately defined as the real-valued maps,

$$Q_{T,S}(\boldsymbol{\theta}) := \mu \circ \left(\hat{\boldsymbol{\beta}}_T \, , \, \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\right) \qquad \text{and} \qquad Q_\infty(\boldsymbol{\theta}) = \mu \, \left(\boldsymbol{\beta}^*_0 \, , \, \boldsymbol{\beta}^*(\boldsymbol{\theta})\right) \; , \quad \forall \; \boldsymbol{\theta} \in \Theta, \tag{4}$$

where $\mu : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ is a divergence that we shall refer to as *criterion divergence*.[6] The SNPII estimator obtained from minimizing this simple prototypical criterion function is useful in that it provides the clearest results and allows for the shortest and most comprehensible proofs. Theoretically, this is not problematic regardless of wether $\mathcal{L}$ is finite or infinite. In applications however, it is not possible to make use of an infinite number of auxiliary estimators. As such, we shall also derive results for an SNPII estimator whose criterion function is given by,

$$Q_{T,S}(\boldsymbol{\theta}) := \mu_T \circ \left(\hat{\boldsymbol{\beta}}_T \, , \, \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\right) \qquad \text{and} \qquad Q_\infty(\boldsymbol{\theta}) = \mu_\infty \, \left(\boldsymbol{\beta}^*_0 \, , \, \boldsymbol{\beta}^*(\boldsymbol{\theta})\right) \; , \quad \forall \; \boldsymbol{\theta} \in \Theta, \tag{5}$$

where $\mu_T : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ is a divergence for every $T \in \mathbb{N}$, and the sequence $\{\mu_T\}_{T \in \mathbb{N}}$ converges in a suitable manner to a limit divergence $\mu_\infty : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$.[7] In applications, when $\mathcal{L}$ is infinite, μ_T can then be appropriately chosen to be a divergence that "concentrates" on a finite subset of the auxiliary maps $\hat{\boldsymbol{\beta}}_T$ and $\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ for every $T \in \mathbb{N}$, yet converges to a divergence μ_∞ that "takes into account" the entire set of auxiliary estimators. This construction shall be made precise in Section 8.

Finally, important notational simplification is obtained by defining the *centered empirical binding function* $\Delta_{T,S}(\boldsymbol{\theta}) := \hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$, the natural estimator of the *centered binding function* $\Delta_\infty(\boldsymbol{\theta}) := b(\boldsymbol{\theta}_0) - b(\boldsymbol{\theta})$. Since $\exists \boldsymbol{\theta}_0 \in \Theta : \mathrm{D}(\boldsymbol{\theta}_0) = \mathrm{D}_0$, i.e. $\mathrm{D}_0 \in \mathcal{D}_\Theta$, the $\mathcal{B}$-valued centered binding function $\Delta_\infty : \Theta \to \mathcal{B}$ crosses the origin of $\mathcal{B}$ at $\boldsymbol{\theta}_0$, i.e. $\Delta_\infty(\boldsymbol{\theta}_0) = 0$. Its estimator $\Delta_{T,S} : \Omega \times \Theta \to \mathcal{B}$ does not necessarily cross the origin.

---

[6] A special case that might be of interest in application is $Q_{T,S}(\boldsymbol{\theta}) = \mu^p \circ \left\{\mu^p_L\left(\hat{\boldsymbol{\beta}}^L_T \, , \, \tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta})\right)\right\}_{L \in \mathcal{L}}$ and $Q_\infty(\boldsymbol{\theta}) := \mu^p \circ \left\{\mu^p_L\left(\boldsymbol{\beta}^{*,L}_0 \, , \, \tilde{\boldsymbol{\beta}}^{*,L}_0(\boldsymbol{\theta})\right)\right\}_{L \in \mathcal{L}}$ where $\mu^p_L : \mathcal{B}_L \times \mathcal{B}_L \to \mathbb{R}$ and $\mu^p : \mathbb{R}^{|\mathcal{L}|} \to \mathbb{R}$ are divergences.

[7] In applications, one might have as a special case, $Q_{T,S}(\boldsymbol{\theta}) = \mu^p_T \circ \left\{\mu^p_L\left(\hat{\boldsymbol{\beta}}^L_T \, , \, \tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta})\right)\right\}_{L \in \mathcal{L}}$ and $Q_\infty(\boldsymbol{\theta}) := \mu^p_\infty \circ \left\{\mu^p_L\left(\boldsymbol{\beta}^{*,L}_0 \, , \, \tilde{\boldsymbol{\beta}}^{*,L}_0(\boldsymbol{\theta})\right)\right\}_{L \in \mathcal{L}}$ where $\mu^p_L : \mathcal{B}_L \times \mathcal{B}_L \to \mathbb{R}$ and $\mu^p_T : \mathbb{R}^{|\mathcal{L}|} \to \mathbb{R}$ and $\mu^p_\infty : \mathbb{R}^{|\mathcal{L}|} \to \mathbb{R}^+_0$ are divergences.

4 Existence and Measurability

As we shall now see, measurability of the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S} : \Omega \to \Theta_T$ follows almost immediately from a measurability result for sieve extremum estimators established by Theorem 2.2 in White and Wooldrige (1991), the measurability of the auxiliary maps $\hat{\boldsymbol{\beta}}_T : \Omega \to \mathcal{B}$ and $\tilde{\boldsymbol{\beta}}_{T,S} : \Omega \times \Theta \to \mathcal{B}$, which in turn, is directly obtained from the measurability of the individual auxiliary estimators under the Borel σ-algebra $\mathfrak{B}(\mathcal{B})$ generated by the product topology $\mathcal{T}_{\mathcal{B}}$.

Assumption 4. (i) $\hat{\boldsymbol{\beta}}_T^L : \Omega \to \mathcal{B}_L$ *is* $\mathcal{F}/\mathfrak{B}(\mathcal{B}_L)$*-measurable* $\forall (T, L) \in \mathbb{N} \times \mathcal{L}$
(ii) $\tilde{\boldsymbol{\beta}}_{T,s}^L(\cdot, \boldsymbol{\theta}) : \Omega \to \mathcal{B}_L$ *is* $\mathcal{F}/\mathfrak{B}(\mathcal{B}_L)$*-measurable* $\forall (\boldsymbol{\theta}, T, s, L) \in \Theta \times \mathbb{N} \times \{1, ..., S\} \times \mathcal{L}$.

We also impose the following continuity assumptions.

Assumption 5. $\tilde{\boldsymbol{\beta}}_{T,s}^L(\omega, \cdot) : \Theta \to \mathcal{B}_L$ *is continuous on* Θ $\forall (\omega, T, s, L) \in \Omega \times \mathbb{N} \times \{1, ..., S\} \times \mathcal{L}$.

Assumption 6. $\mu : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ *is continuous on* $\mathcal{B} \times \mathcal{B}$.

The following establishes the desired existence result for the approximate SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (3) with fixed criterion divergence as defined in (4).

Theorem 1. (Existence of SNPII Estimator) *Let Assumptions 1-6 hold, then there exists a map* $\hat{\boldsymbol{\theta}}_{T,S} : \Omega \to \Theta_T$ *satisfying (3) and (4) for every* $T \in \mathbb{N}$ *and* $S \in \mathbb{N}$ *that is* $\mathcal{F}/\mathfrak{B}(\Theta_T)$*-measurable.*

The same measurability result applies immediately to the exact estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (2).

Corollary 1. (Existence of SNPII Estimator) *Let Assumptions 1-6 hold, then there exists a map* $\hat{\boldsymbol{\theta}}_{T,S} : \Omega \to \Theta_T$ *satisfying (2) and (4) for every* $T \in \mathbb{N}$ *and* $S \in \mathbb{N}$ *that is* $\mathcal{F}/\mathfrak{B}(\Theta_T)$*-measurable.*

Theorem 1 and Corollary 1 can be naturally extended to the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ with criterion divergence indexed by T, defined in (7), by extending the continuity Assumption in 6 to every divergence in the sequence $\{\mu_T\}_{T \in \mathbb{N}}$.

Assumption 7. $\mu_T : \mathcal{B} \times \mathcal{B} \to \mathbb{R}_0^+$ *is continuous on* $\mathcal{B} \times \mathcal{B}\, \forall T \in \mathbb{N}$.

Theorem 2. (Existence of SNPII Estimator) *Let Assumptions 1-5 and 7 hold, then there exists* $\hat{\boldsymbol{\theta}}_{T,S} : \Omega \to \Theta_T$ *satisfying (3) and (7) for every* $T \in \mathbb{N}$ *and* $S \in \mathbb{N}$ *that is* $\mathcal{F}/\mathfrak{B}(\Theta_T)$*-measurable.*

As before, the same measurability result applies immediately to the exact estimator defined in (2).

Corollary 2. (Existence of SNPII Estimator) *Let Assumptions 1-5 and 7 hold, then there exists a map* $\hat{\boldsymbol{\theta}}_{T,S} : \Omega \to \Theta_T$ *satisfying (2) and (5) for every* $T \in \mathbb{N}$ *and* $S \in \mathbb{N}$ *that is* $\mathcal{F}/\mathfrak{B}(\Theta_T)$*-measurable.*

We thus proceed under the established result that $\hat{\boldsymbol{\theta}}_{T,S}$ is a random element taking values in subsets of Θ for every $T \in \mathbb{N}$. Statements involving convergence in law, in probability or almost surely of $\hat{\boldsymbol{\theta}}_{T,S}$ are from now on considered sound under the set of Assumptions 1-7.

5 Consistency

This section establishes the consistency of the SNPII estimator. In particular, the convergence in probability (and almost surely) of $\hat{\boldsymbol{\theta}}_{T,S}$, as defined in either (2) or (3), to the parameter $\boldsymbol{\theta}_0 \in \Theta$. Note that in Section 2 we have not been precise as to which metric $\delta_{\mathcal{B}}$ is defined on $\mathcal{B}$, requiring only that it induces Tychonoff's topology on the set. This requirement should be enough to obtain meaningful convergence results since, given a set of metrics $\{\delta_{\mathcal{B}_L}\}_{L\in\mathcal{L}}$ on the auxiliary factor spaces $\{\mathcal{B}_L\}_{L\in\mathcal{L}}$, any pair of product metrics inducing Tychonoff's topology on $\mathcal{B}$ is, by definition, topologically equivalent, and convergence in one implies convergence in the other (see Definition 9 and Remark 2)). Nonetheless, to obtain simpler proofs for the theorems that follow, we shall further restrict the class of metrics that are allowed to equip $\mathcal{B}$. The following assumption provides an explicit restriction on this class. In particular, it imposes the seemingly mild regularity condition (satisfied e.g. by both metrics in (1); see Proposition 1) that the product metric $\delta_{\mathcal{B}}$ be *Lipschitz weaker* than the *uniform product metric* (Definitions 10 and 11).

Assumption 8. $\exists\, k \in \mathbb{R}^+$ *such that* $\delta_{\mathcal{B}}(\boldsymbol{\beta}, \boldsymbol{\beta}') \;\leq\; k \cdot \sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}(\boldsymbol{\beta}_L, \boldsymbol{\beta}'_L)\;\; \forall\, (\boldsymbol{\beta}, \boldsymbol{\beta}') \in \mathcal{B}\times\mathcal{B}$.

As we shall now see, under Assumption 8 and appropriate regularity conditions, consistency of the SNPII estimator follows from ensuring *(i)* the uniform convergence in probability [a.s.] of the criterion function $Q_{T,S}$ across the sieves $\Theta_T\ \forall T \in \mathbb{N}$, and *(ii)* the identifiable uniqueness of $\boldsymbol{\theta}_0 \in \Theta$, i.e. that $\boldsymbol{\theta}_0$ is a well separated minimizer of the limiting criterion function Q_∞.

Appropriate uniform convergence of $Q_{T,S}$ shall be obtained from the convergence in probability [a.s.] of auxiliary estimators uniformly over $\mathcal{L}$ and across the sieves $\Theta_T\ \forall T \in \mathbb{N}$. Such convergence is easily satisfied in applications since, in general, auxiliary estimators can be chosen so as to retain a $\sqrt{T}$ convergence rate for every $L \in \mathcal{L}$ and $\boldsymbol{\theta} \in \Theta$ (see Section 8 for simple primitive conditions).[8]

Assumption 9. (i) $\sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}\Big(\hat{\boldsymbol{\beta}}^L_T\,,\,\boldsymbol{\beta}^*_L(\boldsymbol{\theta}_0)\Big) \xrightarrow{p} 0$ *[a.s.] as* $T\to\infty$*;*
(ii) $\sup_{\boldsymbol{\theta}\in\Theta_T} \sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}\Big(\tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta})\,,\,\boldsymbol{\beta}^*_L(\boldsymbol{\theta})\Big) \xrightarrow{p} 0$ *[a.s.] as* $T\to\infty\ \forall\, s\in\{1,...,S\}$.

In the context of indirect inference, identification of $\boldsymbol{\theta}_0$ requires the fundamental condition that the product binding function $\boldsymbol{\beta}^*$ be injective. This is ensured by having, for every pair $(\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta\times\Theta$, at least one $L\in\mathcal{L}$ such that the limit $\boldsymbol{\beta}^*_L$ of the auxiliary estimator $\tilde{\boldsymbol{\beta}}^L_{T,s}$ satisfies $\boldsymbol{\beta}^*_L(\boldsymbol{\theta}) \neq \boldsymbol{\beta}^*_L(\boldsymbol{\theta}')$. Furthermore, to ensure the "transfer" of some topological structure from Θ to the factor spaces $\mathcal{B}_L$ (and ultimately to $\mathcal{B}$), we shall assume that the factor binding function $\boldsymbol{\beta}^*_L$ is an open map $\;\forall L\in\mathcal{L}$. Finally, to guarantee the continuity of the limit criterion function Q_∞ we also impose that $\boldsymbol{\beta}^*$ be continuous on $\Theta\;\;\forall L\in\mathcal{L}$. Together, these conditions can also be shown to imply that the product binding function $\boldsymbol{\beta}^*$ is a homeomorphism on its range (see proof of Theorem 3). The parameter space Θ is thus homeomorphic (topologically equivalent) to a subset of $\mathcal{B}$. This conveys a natural sense in which inference on Θ can be conducted through inference on $\mathcal{B}$.

[8] In Assumption 9 the following definition is used $\boldsymbol{\beta}^*_L(\boldsymbol{\theta}) := \pi_L \circ \boldsymbol{\beta}^*(\boldsymbol{\theta}) \in \mathcal{B}_L\ \forall\, (\boldsymbol{\theta}, L) \in \Theta\times\mathcal{L}$.

Assumption 10. $\boldsymbol{\beta}_L^* : \Theta \to \mathcal{B}_L$ *is* (i) *an open map* $\forall L \in \mathcal{L}$*;* (ii) *continuous on* $\Theta\ \forall L \in \mathcal{L}$*; and* (iii) *for every* $(\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta \times \Theta$, $\exists\, L \in \mathcal{L}\ :\ \boldsymbol{\beta}_L^*(\boldsymbol{\theta}) = \boldsymbol{\beta}_L^*(\boldsymbol{\theta}')$.

Finally, as we shall see, given Assumption 10, a sufficient condition for $\boldsymbol{\theta}_0$ to be an identifiably unique minimizer (Definition 14) of the limit criterion function Q_∞, is that $\boldsymbol{\beta}_0$ be itself a well-separated minimizer of the criterion divergence μ. In applications, most divergences of interest satisfy this condition (see Section 8 for examples).

Assumption 11. $\inf_{\boldsymbol{\beta} \in S^c_{\boldsymbol{\beta}_0}(\epsilon) \subset \mathcal{B}} \left|\mu(\boldsymbol{\beta}) - \mu(\boldsymbol{\beta}_0)\right| > 0\ \ \forall \epsilon > 0.$

We are now ready to provide the first consistency result. The following theorem establishes the convergence in probability [a.s.] of the SNPII estimator.

Theorem 3. (Consistency of SNPII Estimator) *Let Assumptions 1-6 and 8-11 hold. Then, the approximate SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (3) and (4) satisfies* $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ *[a.s.] as* $T \to \infty$.

Since the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ in (2) is a special case of the one in (3), the following result is obtained immediately as a corollary.

Corollary 3. (Consistency of SNPII Estimator) *Let Assumptions 1-6 and 8-11 hold. Then, the exact SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (2) and (4) satisfies* $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ *[a.s.] as* $T \to \infty$.

In general, it is easy to verify that the convergence of auxiliary estimators, uniformly in $L \in \mathcal{L}$ and across the sequence of sieves $\{\Theta_T\}_{T\in\mathbb{N}}$ (Assumption 9), holds true by noting that every auxiliary estimator retains a $\sqrt{T}$-convergence in $\mathcal{B}_L$ for every $\boldsymbol{\theta} \in \Theta_T\ \forall T \in \mathbb{N}$, i.e. for every $\mathrm{D}(\boldsymbol{\theta}) \in \mathcal{D}_{\Theta_T}\ \forall T \in \mathbb{N}$ (see Section 8). Still, the consistency result established in Theorem 3 above applies also naturally to auxiliary estimators $\tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta})$ exhibiting varying convergence rates across elements of $\mathcal{D}_{\Theta_T}\ \forall T \in \mathbb{N}$. In applications where auxiliary estimators exhibit varying convergence rates across Θ and $\mathcal{L}$ (except possibly at $\boldsymbol{\theta}_0$), it is desirable to dispose of alternative ways of verifying that Assumption 9 is satisfied.

To obtain an analogue of Theorem 3 while substituting the uniform convergence postulated in Assumption 12 by pointwise convergence (over $\mathcal{L}$ and Θ), we do however require the added influence of conditions that might be more or less restrictive depending on the application. In particular, we make use of *(i)* a uniform generalized stochastic Lipschitz condition on $\tilde{\boldsymbol{\beta}}^L_{T,s}$, *(ii)* a generalized Lipschitz condition on the criterion divergence μ, and *(iii)* the total boundedness of the parameter space Θ.

Assumption 12. (i) $\delta_{\mathcal{B}}\Big(\hat{\boldsymbol{\beta}}^L_T\, , \boldsymbol{\beta}^*_L(\boldsymbol{\theta}_0)\Big) \xrightarrow{p} 0$ *[a.s.] as* $T \to \infty\ \ \forall L \in \mathcal{L}$*;*
(ii) $\delta_{\mathcal{B}_L}\Big(\tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta})\, , \boldsymbol{\beta}^*_L(\boldsymbol{\theta})\Big) \xrightarrow{p} 0$ *[a.s.] as* $T \to \infty\ \forall (\boldsymbol{\theta}, s, L) \in \Theta \times \{1, ..., S\} \times \mathcal{L}$.

Assumption 13. $\exists\, T^* \in \mathbb{N}$ *such that* $\delta_{\mathcal{B}_L}\Big(\tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) - \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}')\Big) \leq \zeta_T \xi\big(\delta_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}')\big)$ *holds a.s.* $\forall (\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta_T \times \Theta_T$ *and every* $T > T^*$*, where* ξ *is a nonstochastic function satisfying* $\lim_{x\to 0} \xi(x) = 0$ *and* ζ_T *is a stochastic sequence satisfying either* (i) $\zeta_T = O_p(1)$ *or* (ii) $\limsup_{T\in\mathbb{N}} \zeta_T < \infty$ *a.s.*.

Assumption 14. *$\mu(\boldsymbol{\beta},\boldsymbol{\beta}') = \mu^*(\boldsymbol{\beta}-\boldsymbol{\beta}')\ \forall(\boldsymbol{\beta},\boldsymbol{\beta}') \in \mathcal{B}\times\mathcal{B}$ and $\left|\mu^*(\boldsymbol{\beta}) - \mu^*(\boldsymbol{\beta}')\right| \leq \xi_\mu\Big(\delta_{\mathcal{B}}(\boldsymbol{\beta},\boldsymbol{\beta}')\Big)$ $\forall(\boldsymbol{\beta},\boldsymbol{\beta}') \in \mathcal{B}\times\mathcal{B}$ where $\xi : \mathbb{R}\to\mathbb{R}$ is ζ_μ-homogeneous (see Definition 15) and satisfies $\lim_{x\to 0}\xi(x) = 0$.*

Assumption 15. *(Θ, δ_Θ) is a totally bounded metric space.*

The generalized Lipschitz continuity postulated in Assumption 14 allows e.g. for the criterion divergence μ to be Lipschitz or Holder continuous. The total boundedness of the parameter space imposed in Assumption 15 might be too restrictive from a sieve estimation perspective when interest lies in having a sequence of sieves $\{\Theta_T\}$ that is "increasing in size" and not necessarily one that is only "increasing in dimension". Note however that the restrictiveness of total boundedness on separable metric spaces such as Θ might be only apparent (Lemma 27). In any case, total boundedness of Θ is not a necessary condition and can be relaxed as long as some mild structure is imposed on the rate of increase in the entropy of the sieves. We shall provide such a consistency result, but only in Section 6, as we must then make use of conditions (introduced only in that section) that allow us to effectively translate convergence rate properties from auxiliary estimators in the factor spaces $\mathcal{B}_L$ to the empirical binding function in the product space $\mathcal{B}$, and ultimately, to the SNPII criterion function $Q_{T,S}$. For now, we impose the total boundedness of Θ (and hence a finite ϵ-entropy on Θ $\forall\,\epsilon > 0$) which effectively reduces the consistency proof to that commonly found in standard extremum estimation problems. Finally, it is important to note that Assumption 15 does not make the use of sieves pointless since various desirable properties for the SNPII estimator can be obtained by choosing appropriate subsets of a possibly infinite dimensional (yet totally bounded) parameter space Θ. Also, from a practical perspective, there is a clear impossibility of conducting indirect inference directly on the infinite dimensional Θ, again even if it is totally bounded.

Theorem 4. (Consistency of SNPII Estimator) *Let Assumptions 1-6, 8 and 10-15 hold. Then, the approximate SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (3) and (4) satisfies $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ [a.s.] as $T\to\infty$.*

Again, consistency of the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ in (2) follows immediately as a corollary.

Corollary 4. (Consistency of SNPII Estimator) *Let Assumptions 1-6, 8 and 10-15 hold. Then, the exact SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (2) and (4) satisfies $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ [a.s.] as $T\to\infty$.*

Finally, we provide analogues of both Theorems 3 and 4 for the SNPII estimator described in (5) whose criterion divergence μ_T is indexed by T. The only changes in the set of assumptions pertain exactly to the new criterion divergence. In particular, we now require the uniform convergence of the deterministic sequence of criterion divergences $\{\mu_T\}_{T\in\mathbb{N}}$ to a limit criterion divergence μ_∞ that satisfies an identifiable uniqueness w.r.t. $\boldsymbol{\beta}_0 \in \mathcal{B}$.

Assumption 16. *$\{\mu_T\}_{T\in\mathbb{N}}$ satisfies $\sup_{\boldsymbol{\beta}\in\mathcal{B}}|\mu_T(\boldsymbol{\beta}) - \mu_\infty(\boldsymbol{\beta})|\to 0$ for some continuous $\mu_\infty : \mathbb{B}\to\mathbb{R}$.*

Assumption 17. $\inf_{\boldsymbol{\beta}\in S^c_{\boldsymbol{\beta}_0}(\epsilon)\subset\mathcal{B}}\left|\mu_\infty(\boldsymbol{\beta}) - \mu_\infty(\boldsymbol{\beta}_0)\right| > 0\ \ \forall\epsilon > 0.$

Section 8 provides examples and discusses the verification of Assumptions 16 and 17 in applications.

Theorem 5. (Consistency of SNPII Estimator) *Let Assumptions 1-5, 7-10, 16 and 17 hold. Then, the approximate SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (3) and (5) satisfies* $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ *[a.s.] as* $T \to \infty$.

The usual Corollary establishing the consistency of the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (2) and (5) follows immediately.

Corollary 5. (Consistency of SNPII Estimator) *Let Assumptions 1-5, 7-10, 16 and 17 hold. Then, the exact SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (2) and (5) satisfies* $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ *[a.s.] as* $T \to \infty$.

An analogue of Theorem 4 that follows from pointwise convergence of auxiliary estimators (over $\mathcal{L}$ and Θ) can also be derived by appropriately substituting Assumption 14 for the following equicontinuity condition for the sequence $\{\mu_T\}_{T\in\mathbb{N}}$.

Assumption 18. $\mu_T(\boldsymbol{\beta}, \boldsymbol{\beta}') = \mu_T^*(\boldsymbol{\beta} - \boldsymbol{\beta}')\ \forall (\boldsymbol{\beta}, \boldsymbol{\beta}') \in \mathcal{B} \times \mathcal{B}$ *and* $\left|\mu_T^*(\boldsymbol{\beta}) - \mu_T^*(\boldsymbol{\beta}')\right| \leq \xi_\mu\Big(\delta_{\mathcal{B}}(\boldsymbol{\beta}, \boldsymbol{\beta}')\Big)$ $\forall (\boldsymbol{\beta}, \boldsymbol{\beta}') \in \mathcal{B} \times \mathcal{B}$ *and every* $T \in \mathbb{N}$ *where* $\xi : \mathbb{R} \to \mathbb{R}$ *is* ζ_μ*-homogeneous (see Definition 15) and satisfies* $\lim_{x\to 0} \xi(x) = 0$.

By Arzela-Ascolli's Theorem (Lemma 31), under the additional assumption that $\mathcal{B}$ is compact (which is implied by compactness of $\mathcal{B}_L\ \forall L \in \mathcal{L}$ in the product topology Tychonoff's Theorem, Lemma 8), the a.s. continuity of the limit criterion μ_∞ postulated in Assumption 16 implies (and is implied by) the equicontinuity in Assumption 18 and the uniform convergence in Assumption 16.

Theorem 6. (Consistency of SNPII Estimator) *Let Assumptions 1-5, 7-8, 10, 12, 13, 15-17 and 18 hold. Then, the approximate SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (3) and (5) satisfies* $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ *[a.s.] as* $T \to \infty$.

Finally, a corollary of Theorem 6.

Corollary 6. (Consistency of SNPII Estimator) *Let Assumptions 1-5, 7-8, 10, 12, 13, 15-17 and 18 hold. Then, the exact SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (2) and (5) satisfies* $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ *[a.s.] as* $T \to \infty$.

Before moving on to the next section, a note is due also on the convergence of continuous functionals of $\hat{\boldsymbol{\theta}}_{T,S}$. In applications, researchers might be interested in the study of such quantities as $\big(\phi_1(\boldsymbol{\theta}_0), ..., \phi_{n_\phi}(\boldsymbol{\theta}_0)\big)$ where $\phi_i : \Theta \to \Phi$ is some continuous functional defined on Θ, $i = 1, ..., n_\phi$, $n_\phi \in \mathbb{N}$, be this a map to finite or infinite dimensional spaces. Examples of interest are likely to include finite dimensional objects such as a set of derivatives (when elements $\boldsymbol{\theta}$ are functions) or projections to finite dimensional subsets of Θ (see Andrews (1991) for more examples). Wether elements $\boldsymbol{\theta}$ are infinite vectors, functions, or collections of functions, interest in conducting inference on projections is almost certain to occur in applications. In any case it is the continuity of the projection maps w.r.t. Tychonoff's topology that enables the immediate deduction of convergence of the possible finite dimensional subsets of $\hat{\boldsymbol{\theta}}_{T,S}$ to the corresponding finite dimensional subsets of $\boldsymbol{\theta}_0$.

Corollary 7. *Let the conditions of any of the above Theorems 3-6 be satisfied. Let $\hat{\boldsymbol{\theta}}_{T,S}$ denote the corresponding SNPII estimator. Let $\phi : \Theta \to \Phi$ denote a continuous functional (possibly a projection). Then, we have that $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ [a.s.] as $T \to \infty$ and by the Continuous Mapping Theorem it follows that $\delta_\Phi(\phi(\hat{\boldsymbol{\theta}}_{T,S}), \phi(\boldsymbol{\theta}_0)) \xrightarrow{p} 0$ [a.s.] as $T \to \infty$.*

6 Convergence Rate

This section establishes a $\sqrt{T}$-convergence rate for the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$, as defined by alternative combinations of (2), (3), (4) and (5). In essence, the rate of convergence is derived, under appropriate regularity conditions, from the $\sqrt{T}$-convergence rate of the individual auxiliary estimators $\hat{\boldsymbol{\beta}}^L_T$ and $\tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta}_0)\ \forall\, L \in \mathcal{L}$. The set of regularity conditions is however more restrictive than that employed in Section 5 to obtain consistency. First of all, Θ is now required to be a normed vector space. Together with Assumption 1 this implies that Θ is now a separable Banach space. The linear space structure is required for us to make use of differentiability and linearity concepts. The vector space assumptions thus turns out to be unavoidable as linear operators and differentiable maps are an integral part of the theory that follows, and it is on linear spaces that such maps are naturally defined. Second, the auxiliary factor spaces $\mathcal{B}_L$ are assumed to be compact normed vector spaces $\forall\, L \in \mathcal{L}$. Compactness of auxiliary factor spaces $\mathcal{B}_L$, and ultimately of the product auxiliary space $\mathcal{B}$, not only simplifies proofs, it enables the use of several well-established results of weak convergence on compact sets. It should be stressed however that, while the first assumption constitutes a visible restriction on the generality of the SNPII methodology, the second assumption does not really carry any serious practical implication.

Assumption 19. $(\Theta, \|\cdot\|_\Theta)$ *is a normed vector space.*[9]

Assumption 20. $(\mathcal{B}_L, \|\cdot\|_{\mathcal{B}_L})$ *is a compact normed vector space* $\forall\, L \in \mathcal{L}$.

Besides controlling for the convergence rate of auxiliary estimators and establishing a primitive characterization of the relevant parameter spaces, a number of regularity conditions must also be imposed on *(i)* the rate at which the "error" η_T of approximate estimation in (3) declines to zero, *(ii)* the smoothness and invertibility of auxiliary estimators and their limits, *(iii)* the smoothness and invertibility of the criterion divergence maps μ on $\mathcal{B}$, and finally, *(iv)* the rate of "expansion" of the sieves $\{\Theta_T\}_{T\in\mathbb{N}}$. Some conditions might not always help in keeping with clarity and intuition. Nonetheless, some light can be shed on the structure of the convergence rate of the SNPII estimator by decomposing, for every $(\omega, T, S) \in \Omega \times \mathbb{N} \times \mathbb{N}$, the distance $\|\hat{\boldsymbol{\theta}}_{T,S}(\omega) - \boldsymbol{\theta}_0\|_\Theta$ into parts that explicitly

[9] The metric δ_Θ in Assumption 1 is thus assumed to be the one induced by the norm $\|\cdot\|_\Theta$ according to $\delta_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}') := \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_\Theta\ \forall\, (\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta \times \Theta$. The denseness of the sieves postulated in Assumption 2 thus hold also w.r.t. $\|\cdot\|_\Theta$.

identify the minimizers of the criterion function $Q_{T,S}(\omega)$ and its limit Q_∞ on Θ_T. In particular, let $\boldsymbol{\theta}^*_{T,S}(\omega)$ and $\boldsymbol{\theta}^0_T$ denote elements of the arg min set of $Q_{T,S}(\omega)$ and Q_∞ respectively over the sieve Θ_T,

$$\boldsymbol{\theta}^*_{T,S}(\omega) \in \arg\min_{\boldsymbol{\theta}\in\Theta_T} Q_{T,S}(\omega,\boldsymbol{\theta}) \quad \forall(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N} \quad \text{and} \quad \boldsymbol{\theta}^0_T \in \arg\min_{\boldsymbol{\theta}\in\Theta_T} Q_\infty(\boldsymbol{\theta}) \quad \forall T\in\mathbb{N}.$$

Then, for every $(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N}$ it holds true that,

$$\|\hat{\boldsymbol{\theta}}_{T,S}(\omega)-\boldsymbol{\theta}_0\|_\Theta \;\leq\; \|\hat{\boldsymbol{\theta}}_{T,S}(\omega)-\boldsymbol{\theta}^*_{T,S}(\omega)\|_\Theta \;+\; \|\boldsymbol{\theta}^*_{T,S}(\omega)-\boldsymbol{\theta}^0_T\|_\Theta \;+\; \|\boldsymbol{\theta}^0_T-\boldsymbol{\theta}_0\|_\Theta. \tag{6}$$

This decomposition allows us to separate the study of the convergence rate of $\|\hat{\boldsymbol{\theta}}_{T,S}(\omega)-\boldsymbol{\theta}_0\|_\Theta$ into *(i)* the convergence rate of the "error" $\|\hat{\boldsymbol{\theta}}_{T,S}(\omega)-\boldsymbol{\theta}^*_{T,S}(\omega)\|_\Theta$ introduced by having $\hat{\boldsymbol{\theta}}_{T,S}$ be an *approximate* extremum estimator; *(ii)* the convergence rate of the *exact* SNPII estimator "within sieves" $\|\boldsymbol{\theta}^*_{T,S}(\omega)-\boldsymbol{\theta}^0_T\|_\Theta$; and *(iii)* the "sieve approximation error" $\|\boldsymbol{\theta}^0_T-\boldsymbol{\theta}_0\|_\Theta$ introduced by restricting the optimization to the sequence of sieves $\{\Theta_T\}_{T\in\mathbb{N}}$. In what follows, a $\sqrt{T}$-convergence rate for the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ is thus obtained essentially by showing that each of these terms is $O_p(T^{-1/2})$ [a.s.].

Clearly, if $\hat{\boldsymbol{\theta}}_{T,S}$ is an exact extremum estimator (as defined in (2)) then it holds true that

$$\hat{\boldsymbol{\theta}}_{T,S}(\omega)=\boldsymbol{\theta}^*_{T,S}(\omega) \quad \forall(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N}$$

and the term $\|\hat{\boldsymbol{\theta}}_{T,S}(\omega)-\boldsymbol{\theta}^*_{T,S}(\omega)\|_\Theta$ vanishes. If $\hat{\boldsymbol{\theta}}_{T,S}$ is an approximate extremum estimator (as defined in (3)) then an appropriate rate of convergence for $\|\hat{\boldsymbol{\theta}}_{T,S}-\boldsymbol{\theta}^*_{T,S}(\omega)\|_\Theta$ can be obtained under the suitable behavior (postulated in Assumption 21) of the error term η_T in (3) and the smoothness conditions maintained by Assumption 22 below (see proof of Theorem 7). The $\sqrt{T}$ rate of convergence of the "within sieves" term $\|\boldsymbol{\theta}^*_{T,S}(\omega)-\boldsymbol{\theta}^0_T\|_\Theta$ is essentially derived from the $\sqrt{T}$-convergence rate of the individual auxiliary estimators (Assumption 24), the smoothness conditions of Assumption 22 and the invertibility conditions of Assumption 23 (see proof of Theorem 7). Finally, the rate of convergence of the "sieve approximation error" term $\|\boldsymbol{\theta}^0_T-\boldsymbol{\theta}_0\|_\Theta$ is likewise deduced from the smoothness conditions in Assumption 22 and by establishing an appropriate rate of "expansion" for the sequence of sieves $\{\Theta_T\}_{T\in\mathbb{N}}$ (Assumption 21) as characterized by the convergence of the sequence of δ_Θ-metric projections $\pi_{\Theta_T}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ onto Θ_T $\forall T\in\mathbb{N}$ (also in proof of Theorem 7).[10]

Note here that, by definition, each $\boldsymbol{\theta}^0_T\in\{\boldsymbol{\theta}^0_T\}_{T\in\mathbb{N}}\subset\Theta$ corresponds to an element of the projection set (w.r.t. the divergence Q_∞) of $\boldsymbol{\theta}_0$ onto Θ_T.[11] It is important to note however that, in applications, while the convergence rate of $\|\pi_{\Theta_T}(\boldsymbol{\theta}_0)-\boldsymbol{\theta}_0\|_\Theta$ is typically known from the outset, the convergence

[10]The fundamental denseness property of the method of sieves (postulated in Assumption 2) implies by construction that there exists a sequence $\{\pi_{\Theta_T}(\boldsymbol{\theta}_0)\}_{T\in\mathbb{N}}$ with $\pi_{\Theta_T}(\boldsymbol{\theta}_0)\in\Theta_T$ $\forall T\in\mathbb{N}$ such that $\|\pi_{\Theta_T}(\boldsymbol{\theta}_0)-\boldsymbol{\theta}_0\|_\Theta\to 0$ as $T\to\infty$.

[11]That Q_∞ is a divergence on Θ w.r.t. $\boldsymbol{\theta}_0$ follows from its definition in terms of the divergence μ (or μ_∞) and the injective nature of the product binding function $\boldsymbol{\beta}^*$ derived in Proposition 21 from Assumption 10. Existence of the arg min set follows immediately, by Weierstrass's *Extreme Value Theorem* (Lemma 32), from the compactness of each Θ_T (Assumption 2) and continuity of Q_∞ (derived in Theorems 3 and 4 under Assumptions 1, 1, 10 and either 6 or 7). Uniqueness of $\boldsymbol{\theta}^0_T$ (i.e. reduction of the arg min set to a singleton) follows for norm-divergences μ (or μ_∞) by the strict convexity of μ (or μ_∞ respectively); see Theorem 5 in Blasques (2010)

rate of $\|\boldsymbol{\theta}_T^0 - \boldsymbol{\theta}_0\|_\Theta$ is most likely not known unless further regularity conditions are satisfied. In particular, in applications, the choice of norm $\|\cdot\|_\Theta$ is likely to be dictated by the availability of results establishing *(i)* the denseness of the sequence of sieves $\{\Theta_T\}_{T\in\mathbb{N}}$ on Θ w.r.t. $\|\cdot\|_\Theta$ and *(ii)* the properties of the sequence $\{\Theta_T\}_{T\in\mathbb{N}}$ in terms of convergence rates of $\|\pi_{\Theta_T}(\boldsymbol{\theta}_0) - \boldsymbol{\theta}_0\|_\Theta$. No similar results are necessarily available for the sequence $\{\boldsymbol{\theta}_T^0\}_{T\in\mathbb{N}}$ of projections w.r.t. the (probably very complex) divergence Q_∞. Control over the convergence rate of the "sieve approximation error" term $\|\boldsymbol{\theta}_T^0 - \boldsymbol{\theta}_0\|_\Theta$ should thus be exerted indirectly by imposing a primitive "rate of expansion" of the sieves (as described by the convergence rate of $\|\pi_{\Theta_T}(\boldsymbol{\theta}_0) - \boldsymbol{\theta}_0\|_\Theta$). Assumption 21 provides the appropriate convergence rate for $\|\pi_{\Theta_T}(\boldsymbol{\theta}_0) - \boldsymbol{\theta}_0\|_\Theta$ which, given the regularity conditions imposed below, turns out also to be $\sqrt{T}$.

Assumption 21. $\|\pi_{\Theta_T}(\boldsymbol{\theta}_0) - \boldsymbol{\theta}_0\|_\Theta = O(T^{-1/2})$

Typical sieve estimation problems require also an upper bound on the "rate of expansion" of the sieves. Strictly speaking, there is no such requirement in SNPII estimation. The SNPII estimator is, to the author's knowledge, unique among nontrivial sieve extremum estimators in not requiring such an upper bound. This somewhat surprising property occurs because the rate of convergence of $\hat{\boldsymbol{\theta}}_{T,S}$ is derived from the rate of convergence of a set of well behaved auxiliary estimators whose convergence rate is to a large extent independent of the size and/or complexity of the sequence of sieves $\{\Theta_T\}_{T\in\mathbb{N}}$ defined on Θ.[12]

The smoothness requirements mentioned above will involve Frechet or Hadamard differentiability concepts (see Definition 22). These forms of differentiability shall be imposed on *(i)* the divergence criterion μ and its derivative $\nabla\mu$, *(ii)* the auxiliary estimators $\tilde{\boldsymbol{\beta}}_{T,S}^L$, *(iii)* their limits $\boldsymbol{\beta}_L^*$ and respective derivatives $\nabla\boldsymbol{\beta}_L^*$ for every $L \in \mathcal{L}$. Since auxiliary estimators of interest in applications are likely to be finite dimensional (i.e. likely to take values on some finite dimensional $\mathcal{B}_L$) the concepts of Frechet and Hadamard differentiability are equivalent and correspond to the usual notion of differentiability. Infinite dimensionality of auxiliary estimators is nonetheless allowed for and it thus becomes relevant to distinguish between such concepts. Careful distinction is also important in what concerns the definition of *derivative function* and consequently of *second-order differentiability*; see Ren and Sen (2001, Remark 1) and references therein. Alternative definitions exist in the literature. This makes it hard to interpret what is meant by any of these concepts outside a specific context. To avoid ambiguity, the following notational convention applies throughout. Unless explicitly stated otherwise, following Definition 22, given a pair of topological vector spaces $(\mathbb{A}, \mathcal{T}_\mathbb{A})$ and $(\mathbb{B}, \mathcal{T}_\mathbb{B})$ and a map $f : \mathbb{A}_f \subseteq \mathbb{A} \to \mathbb{B}$, the map $\nabla_{\mathbb{A}_0} f : \mathbb{A}_\nabla \to \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ denotes the *derivative function* of f tangentially to $\mathbb{A}_0$, a map

[12] In practice however, an upper bound could be required. As explained in Section 8 the need for an upper bound is related to the fact that, in applications, the preservation of the injective nature of the binding function (Assumption 10) is likely to require the divergence μ_T to take into account a number of auxiliary estimators that grows faster (in some specific sense) than the complexity of the sieves Θ_T. This however might imply an explosive behavior for the sequence $\mu_T(\Delta_{T,S}(\boldsymbol{\theta}))$ depending on the weighting scheme that μ_T is chosen to establish over the elements of $\Delta_{T,S}$.

from the set $\mathbb{A}_\nabla \subseteq \mathbb{A}_f$ of points at which f is differentiable into the *conjugate space* or *dual space* of $\mathbb{A}_0$ over the field $\mathbb{B}$, the space of bounded linear functionals $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$. It is easy to verify that the dual space $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$ becomes a vector space once the operations of addition and multiplication by scalars are defined in the usual way.[13] Clearly, $\nabla_{\mathbb{A}_0} f(a_\nabla) \in \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ denotes a bounded linear operator, an element of $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$, for any $a_\nabla \in \mathbb{A}_\nabla$. For conciseness, $\nabla_{\mathbb{A}_0} f(a_\nabla)$ might also be denoted $\nabla_{\mathbb{A}_0} f_{a_\nabla}$, so that $\nabla_{\mathbb{A}_0} f_{a_\nabla}(a_0)$ is thus a point in $\mathbb{B}$ for every $a_0 \in \mathbb{A}_0$, because the differentiation point a_∇ has been fixed. When convenient, the more extensive form $\nabla_{\mathbb{A}_0} f(a_\nabla, a_0)$ will be used. If the tangent set $\mathbb{A}_0$ coincides with the entire set $\mathbb{A}$, then it is omitted from the notation. Thus $\nabla f(a_\nabla) \in \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ denotes the derivative of f at a_∇ tangentially to the entire $\mathbb{A}_f$. Now, the usual notion of *second-order differentiability* of f is one requiring the differentiability of both $f : \mathbb{A}_f \to \mathbb{B}$ and $\nabla_{\mathbb{A}_0} f : \mathbb{A}_\nabla \to \mathbb{L}(\mathbb{A}_0, \mathbb{B})$.[14] Yet, in some occasions $\nabla_{\mathbb{A}_0} f$ shall be defined in alternative ways such as $\nabla_{\mathbb{A}_0} f : \mathbb{A} \times \mathbb{A}_0 \to \mathbb{B}$ depending on our specific needs, in which case *second-order differentiability* takes a different meaning. Indeed, note that alternative definitions of $\nabla_{\mathbb{A}_0} f$ carry important implications since differentiability in one does not necessarily imply differentiability in the other.[15] Such deviations from standard notation shall however be made explicit. Most importantly, in applications, the distinction is irrelevant as long as the alternatives are equally verifiable. Finally, regardless of these considerations, *second-order derivative functions* $\nabla^2_{\mathbb{A}_0} f : \mathbb{A}_\nabla \to \mathbb{L}^2(\mathbb{A}_0 \times \mathbb{A}_0, \mathbb{B})$ are always understood as a map for the from the set $\mathbb{A}_\nabla$ of points at which f is differentiable to the space of bounded bilinear operators $\mathbb{L}^2(\mathbb{A}_0 \times \mathbb{A}_0, \mathbb{B})$.[16] One should thus note that while $\nabla^2_{\mathbb{A}_0} f(a_\nabla) \in \mathbb{L}^2(\mathbb{A}_0 \times \mathbb{A}_0, \mathbb{B})$ denotes a bilinear operator for every $a_\nabla \in \mathbb{A}_\nabla$, we have that $\nabla^2_{\mathbb{A}_0} f(a_\nabla, a_0) \in \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ denotes a linear operator, an element of $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$, for every $(a_\nabla, a_0) \in \mathbb{A}_\nabla \times \mathbb{A}_0$.

Assumption 22. (i) $\tilde{\boldsymbol{\beta}}^L_{T,S} : \Omega \times \Theta \to \mathcal{B}_L$ *is a.s. Frechet [Hadamard] continuously differentiable in* $\boldsymbol{\theta} \in \Theta$ *tangentially to* $\Theta_T \forall (T, L) \in \mathbb{N} \times \mathcal{L}$*;* (ii) $\boldsymbol{\beta}^*_L : \Theta \to \mathcal{B}_L$ *and* $\nabla_\Theta \boldsymbol{\beta}^*_L : \Theta \to \mathbb{L}(\Theta, \mathcal{B}_L)$ *are both continuously Frechet [Hadamard] differentiable in* $\boldsymbol{\theta} \in \Theta \; \forall (T, L) \in \mathbb{N} \times \mathcal{L}$*;* (iii) $\mu : \mathcal{B} \to \mathbb{R}$ *is continuously Frechet [Hadamard] differentiable in* $\boldsymbol{\beta} \in \mathcal{B}$ *and* $\nabla \mu : \mathcal{B} \times \mathbb{L}(\Theta_T, \mathcal{B}) \to \mathbb{L}(\Theta, \mathbb{R})$ *is continuously Frechet [Hadamard] differentiable at every point of its domain* $\mathcal{B} \times \mathbb{L}(\Theta, \mathcal{B})$*.*

From the conditions stated above, we shall obtain the differentiability of $Q_{T,S}$ and consequently allow for a (approximate) Z-estimator formulation of the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ (see e.g. the proof of Theorem 7). In particular, as noted in Van der Vaart (1995) and Van der Vaart and Wellner (1996, ch. 3.3), the derivative function $\nabla_{\Theta_T} Q_{T,S} : \Theta \to \mathbb{L}(\Theta_T, \mathbb{R})$ can be seen in this context as an infinite

[13] Also, if $(\mathbb{A}_0, \| \cdot \|_{\mathbb{A}})$ is a normed vector space over $\mathbb{R}$, then $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$ with operator norm (Definition 24) is a Banach space $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$ (Lemma 36)

[14] Various other definitions of second-order differentiability have however been proposed that do not follow this reasoning; see e.g. Ren and Sen (2001, Remark 1) and references therein.

[15] Any equivalence relation between these concepts would depend first of all on the norm (or at least the topology) defined on $\mathbb{L}(\mathbb{A}, \mathbb{B})$.

[16] The more immediate definition is $\nabla^2_{\mathbb{A}_0} f : \mathbb{A}_\nabla \to \mathbb{L}(\mathbb{A}_0, \mathbb{L}(\mathbb{A}_0, \mathbb{B}))$. However, following Denkowski et al. (2003, Proposition 5.1.17, p.525) we note that $\mathbb{L}(\mathbb{A}_0, \mathbb{L}(\mathbb{A}_0, \mathbb{B}))$ can be identified with $\mathbb{L}^2(\mathbb{A}_0 \times \mathbb{A}_0, \mathbb{B})$.

set of *estimating equations* of the form $\nabla_{\Theta_T} Q_{T,S}(\cdot, \boldsymbol{\theta}) \in \mathbb{C}(\Theta_T, \mathbb{R})$ [a.s.] indexed by $\boldsymbol{\theta} \in \Theta_T$ that are (approximately) set to zero by the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$. Here, continuity of $\nabla_{\Theta_T} Q_{T,S}(\cdot, \boldsymbol{\theta})$ follows by the a.s. continuous differentiability of $Q_{T,S}$ on Θ tangentially to Θ_T. Clearly, from a statistical point of view, there are however too many equations present in this set $\nabla Q_{T,S}(\cdot, \Theta) \subset \mathbb{C}(\Theta, \mathbb{R})$. Indeed, it is not only unpractical to have uncountably many equations for estimation, no statistical information is in effect being added by a large number these equations. Note for instance that statistical inference based on a subset of $\nabla Q_{T,S}(\cdot, \Theta)$ containing the pair of estimating equations $\nabla Q_{T,S}(\cdot, \boldsymbol{\theta})$ and $\nabla Q_{T,S}(\cdot, \boldsymbol{\theta}')$ will be *informationally equivalent* to that which includes a third equation $\nabla Q_{T,S}(\cdot, \boldsymbol{\theta}'')$ that is a linear combination of the former. A statistically more sensible set of equations is thus one that excludes such elements. An example of such set (if is exists) is the *Schauder basis* (see Definition 23 and Corollary 16) of the vector space $\nabla Q_{T,S}(\cdot, \Theta)$.[17] From an optimization point of view, there is a similar natural "reduction" of the system of estimating equations $\nabla Q_{T,S}(\cdot, \Theta)$ since setting to zero the derivatives of $Q_{T,S}$ in the direction of elements of the Schauder basis Θ^{SB} of Θ provides us with the first-order optimality condition (Proposition 5). This implies "reduction" to the subset $\nabla Q_{T,S}(\cdot, \Theta^{\text{SB}})$.[18] In practice, it is enough that $Q_{T,S}$ be differentiable tangentially to the sieve Θ_T for every $T \in \mathbb{N}$ and to appropriately set to zero a smaller but increasing selection of estimating equations that are obtained as derivatives in the direction of elements of the sieves. Differentiating $Q_{T,S}$ only in the direction of elements of the Schauder basis of Θ_T provides further reduction of the set of equations to a set that is manageable in applications. In particular, if the sieves Θ_T are finite dimensional, the finite Schauder basis (and correspondingly the set of estimating equations) is finite.[19] From a theoretical perspective, details on the construction of such reduced systems are however immaterial as long as the present set of assumptions are satisfied by $\nabla Q_{T,S} : \Theta \to \mathbb{L}(\Theta, \mathbb{R})$ or some of the above mentioned reductions of the system. We thus proceed without further comments on this matter.

We also make use of appropriate primitive invertibility assumptions designed to ensure that $\nabla^2_{\Theta_T} Q^0_\infty(\boldsymbol{\theta}^0_T, \cdot) : \text{lin}(\Theta_T) \to \mathbb{R}$ is continuously invertible for every $T \in \mathbb{N} : \boldsymbol{\theta}^0_T \in S_{\boldsymbol{\theta}_0}(\epsilon)$ for some $\epsilon > 0$, and also, that $\nabla^2_\Theta Q^0_\infty(\boldsymbol{\theta}_0, \cdot) : \text{lin}(\Theta) \to \mathbb{R}$ is continuously invertible.[20]

Assumption 23. (i) $\nabla_{\text{lin}(\Theta_T)} \boldsymbol{\beta}_\Delta(\boldsymbol{\theta}, \cdot) : \text{lin}(\Theta_T) \to \mathcal{B}_\Delta$ *is continuously invertible for every* $\boldsymbol{\theta} \in S_{\boldsymbol{\theta}_0}(\epsilon)$ *for some* $\epsilon > 0$. (ii) $\nabla^2_{\mathcal{B}_\Delta} \mu(\boldsymbol{\beta}_\Delta(\boldsymbol{\theta}_0), \cdot) : \mathcal{B}_\Delta \to \mathbb{R}$ *is continuously invertible.* (iii) $\nabla_{\text{lin}(\Theta)} \boldsymbol{\beta}_\Delta(\boldsymbol{\theta}_0, \cdot) : \text{lin}(\Theta) \to \mathcal{B}_\Delta$ *is continuously invertible.*

[17] Existence of a Schauder basis for $\nabla Q_{T,S}(\cdot, \Theta)$ is not guaranteed if Θ is non-compact.

[18] Also here the assumption that Θ admits a Schauder basis would be lacking. Indeed, Assumptions 1 and 19 by themselves, although guaranteeing that Θ is a complete separable Banach space, do not ensure that a Schauder basis exists for Θ as shown by Enflo's negative answer in Enflo (1973) to Banach's question on this matter.

[19] On a finite dimensional space, the Schauder basis coincides with the usual basis for a vector space, the *Hamel basis*.

[20] Given the linearity of $\nabla_{\text{lin}(\Theta_T)} \boldsymbol{\beta}_\Delta(\boldsymbol{\theta}, \cdot)$ in Assumption 23, sufficient conditions for Assumption 23 could be stated in terms of the map's Kernel and range. Namely, for *(i)* $\text{Ker}\Big(\nabla_{\text{lin}(\Theta_T)} \boldsymbol{\beta}_\Delta(\boldsymbol{\theta}, \cdot)\Big) = \{0\}\ \forall \boldsymbol{\theta} \in S_{\boldsymbol{\theta}_0}(\epsilon)$ and $\nabla_{\text{lin}(\Theta_T)} \boldsymbol{\beta}_\Delta(\boldsymbol{\theta}, \text{lin}(\Theta_T)) = \mathcal{B}_\Delta\ \forall \boldsymbol{\theta} \in S_{\boldsymbol{\theta}_0}(\epsilon)$, for *(ii)* $\text{Ker}\Big(\nabla^2_{\mathcal{B}_\Delta} \mu(\boldsymbol{\beta}_\Delta(\boldsymbol{\theta}_0), \cdot)\Big) = \{0\}$ and $\nabla^2_{\mathcal{B}_\Delta} \mu(\boldsymbol{\beta}_\Delta(\boldsymbol{\theta}_0), \mathcal{B}_\Delta) = \mathbb{R}$, and for *(iii)* $\text{Ker}\Big(\nabla_{\text{lin}(\Theta)} \boldsymbol{\beta}_\Delta(\boldsymbol{\theta}_0, \cdot)\Big) = \{0\}$ and $\nabla_{\text{lin}(\Theta)} \boldsymbol{\beta}_\Delta(\boldsymbol{\theta}_0, \text{lin}(\Theta)) = \mathcal{B}$.

Finally, to ensure that the derivative of the criterion function have a convergence rate no slower that $\sqrt{T}$ when evaluated at the sequence $\{\boldsymbol{\theta}_T^0\}_{T\in\mathbb{N}}$,

$$\sqrt{T}\Big[\nabla_{\Theta_T} Q^0_{T,S}(\boldsymbol{\theta}_T^0) - \nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}_T^0)\Big] = O_p(1) \text{ [a.s.]}$$

we also impose the fundamental condition of $\sqrt{T}$-convergence of the auxiliary estimators $\hat{\boldsymbol{\beta}}_T^L$ and $\tilde{\boldsymbol{\beta}}_{T,S}^L(\boldsymbol{\theta}_0)$, and also, the partial Frechet/Hadamard differentiability uniformly on a parameter. The latter concept is introduced in Definition 27 and its generality discussed in Remark 8.

Assumption 24. (i) $\sqrt{T}\delta_{\mathcal{B}_L}\big(\hat{\boldsymbol{\beta}}_T^L, \boldsymbol{\beta}_L^*(\boldsymbol{\theta}_0)\big) = O_p(1)$ *[a.s.]* $\forall\, L \in \mathcal{L}$;
(ii) $\sqrt{T}\delta_{\mathcal{B}_L}\big(\tilde{\boldsymbol{\beta}}_{T,S}^L(\boldsymbol{\theta}_T^0), \boldsymbol{\beta}_L^*(\boldsymbol{\theta}_0)\big) = O_p(1)$ *[a.s.]* $\forall\,(S,L) \in \mathbb{N}\times\mathcal{L}$.

Assumption 25. $\nabla_{\mathcal{B}}\mu(\cdot,\boldsymbol{\beta}') : \mathcal{B} \to \mathbb{R}$ *is Frechet [Hadamard] differentiable at the origin tangentially to* $\mathcal{B}$ *uniformly in* $\boldsymbol{\beta}' \in \mathcal{B}$.

The following theorem establishes a $\sqrt{T}$-convergence rate for the approximate SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ with fixed divergence criterion defined in (3) and (4). The equivalent result for the exact SNPII estimator defined in (2) and (4) follows immediately after.

Theorem 7. *Let Assumptions 1-6, 8-11 and 19-25 hold. Then, the approximate SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (3) and (4) satisfies* $\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\|_\Theta = O_p(1)$ *[a.s.].*

Corollary 8. *Let Assumptions 1-6, 8-11 and 19-25 hold. Then, the approximate SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (2) and (4) satisfies* $\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\|_\Theta = O_p(1)$ *[a.s.].*

Counterparts of Theorem 7 and Corollary 8 for SNPII estimators obtained by minimizing a sequence of criterion divergences $\{\mu_T\}_{T\in\mathbb{N}}$ can be easily obtained by imposing appropriate smoothness and convergence properties on the deterministic sequence $\{\mu_T\}_{T\in\mathbb{N}}$.

[to be completed]

7 Asymptotic Normality

Given the established results of existence, consistency and $\sqrt{T}$-convergence rate of the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$, asymptotic normality can now be derived with little effort. In particular, the weak convergence of $\sqrt{T}(\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0)$ to a well-defined limit Gaussian process shall be derived from the convergence in distribution of the appropriately standardized sequence of individual auxiliary estimators $\hat{\boldsymbol{\beta}}_T^L$ and $\tilde{\boldsymbol{\beta}}_{T,S}^L(\boldsymbol{\theta}_T^0)$, and the convergence in probability [a.s.] of their derivatives $\nabla_{\Theta_T}\tilde{\boldsymbol{\beta}}_{T,S}^L(\boldsymbol{\theta}_T^0, 0)$ to $\nabla_\Theta \boldsymbol{\beta}_L^*(\boldsymbol{\theta}_0, 0) \in \mathcal{B}_L$ for every $L \in \mathcal{L}$.

Assumption 26. (i) $\sqrt{T}(\hat{\boldsymbol{\beta}}_T^L - \boldsymbol{\beta}_L^*(\boldsymbol{\theta}_0)) \xrightarrow{d} \mathcal{N}(0, \Sigma_0^L)\ \ \forall L \in \mathcal{L}$; (ii) $\sqrt{T}(\tilde{\boldsymbol{\beta}}_{T,S}^L(\boldsymbol{\theta}_T^0) - \boldsymbol{\beta}_L^*(\boldsymbol{\theta}_T^0)) \xrightarrow{d} \mathcal{N}(0, \Sigma(\boldsymbol{\theta}_0)^L)\ \ \forall L \in \mathcal{L}$; *and* (iii) $\nabla_{\Theta_T}\tilde{\boldsymbol{\beta}}_{T,S}^L(\boldsymbol{\theta}_T^0, 0) \xrightarrow{p} \nabla_\Theta \boldsymbol{\beta}_L^*(\boldsymbol{\theta}_0, 0)\ \forall (S,L) \in \mathbb{N} \times \mathcal{L}$.

Under certain regularity conditions already incorporated in the set of assumptions postulated in the preceding sections, Assumption 26 is shown to be sufficient to obtain the convergence in distribution of the derivatives of the criterion function evaluated at the sequence $\{\boldsymbol{\theta}_T^0\}_{T\in\mathbb{N}}$ of points in Θ,

$$\sqrt{T}\Big[\nabla_{\Theta_T} Q_{T,S}^0(\boldsymbol{\theta}_T^0) - \nabla_{\Theta_T} Q_\infty^0(\boldsymbol{\theta}_T^0)\Big] \xrightarrow{d} \mathbb{G}_0 := \mathcal{N}\big(0, \sigma_0^2\big) \quad \text{as } \ T \to \infty.$$

The following theorem draws from Van der Vaart (1995) and Van der Vaart and Wellner (1996) and establishes the asymptotic Gaussianity of the appropriately standardized sequence of SNPII estimators $\hat{\boldsymbol{\theta}}_{T,S}$ with fixed divergence criterion defined in (3) and (4). The equivalent result for the exact SNPII estimator defined in (2) and (4) follows immediately after.

Theorem 8. *Let Assumptions 1-6, 8-11 and 19-26 hold. Then, the approximate SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (3) and (4) satisfies* $\sqrt{T}\big(\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\big) \xrightarrow{d} -\Big(\nabla_\Theta^2 Q_\infty^0(\boldsymbol{\theta}_0, \cdot)\Big)^{-1}\big(\mathbb{G}_0\big)$ *as* $T \to \infty$.

Corollary 9. *Let Assumptions 1-6, 8-11 and 19-26 hold. Then, the approximate SNPII estimator* $\hat{\boldsymbol{\theta}}_{T,S}$ *defined in (2) and (4) satisfies* $\sqrt{T}\big(\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\big) \xrightarrow{d} -\Big(\nabla_\Theta^2 Q_\infty^0(\boldsymbol{\theta}_0, \cdot)\Big)^{-1}\big(\mathbb{G}_0\big)$ *as* $T \to \infty$.

The following remark describes the covariance structure of the Gaussian process $\Big(\nabla_\Theta^2 Q_\infty^0(\boldsymbol{\theta}_0, \cdot)\Big)^{-1}\big(\mathbb{G}_0\big)$ determined by both the variance σ_0^2 of the real-valued random variable $\mathbb{G}_0$ and the properties of the inverse of $\nabla_\Theta^2 Q_\infty^0(\boldsymbol{\theta}_0, \cdot) : \Theta \to \mathbb{R}$.

[to be completed]

Counterparts of Theorem 8 and Corollary 9 for SNPII estimators obtained by minimizing a sequence of criterion divergences $\{\mu_T\}_{T\in\mathbb{N}}$ are obtained by imposing further smoothness and convergence properties on the deterministic sequence $\{\mu_T\}_{T\in\mathbb{N}}$.

[to be completed]

8 Some Applications of SNPII

[to be completed]

9 Final Remarks

[to be completed]

A Propositions, Lemmas and Definitions

Definition 1. (Polish Space) *A topological space is said to be a Polish space if it separable and there exists a metric that generates the topology for which the space is complete. Any separable complete metric space is thus a Polish space.*

Lemma 1. [Klambauer 1973, Proposition 31, p.257] (Urysohn-Tychonoff Theorem) *Every regular second-countable topological space* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *is metrizable.*

Definition 2. (Metrizable Space) *A topological space* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *is said to be metrizable if and only if there exists a metric* $\delta_{\mathbb{A}}$ *that induces* $\mathcal{T}_{\mathbb{A}}$ *on* $\mathbb{A}$*, i.e. such that sets of* $\mathcal{T}_{\mathbb{A}}$ *are open w.r.t.* $\delta_{\mathbb{A}}$*.*

Definition 3. (Regular Space) *A topological space* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *is called regular if for every point* $a \notin \mathbb{A}_0 \subset \mathbb{A}$ *there are disjoint open sets* $\mathbb{A}_1$ *and* $\mathbb{A}_2$ *with* $a \in \mathbb{A}_1$ *and* $\mathbb{A}_0 \subset \mathbb{A}_2$*.*

Definition 4. (Second Countable Space) *A topological space* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *is said to be second countable if* $\mathcal{T}_{\mathbb{A}}$ *has a countable base.*

Definition 5. (Base for a Topology) *A base for a topology* $\mathcal{T}$ *is any collection* $\mathcal{T}_0 \subset \mathcal{T}$ *such that for every* $\mathcal{T}_1 \subset \mathcal{T}$*, we have* $\mathcal{T}_1 = \bigcup \left\{\mathcal{T}_0' \in \mathcal{T}_0 : \mathcal{T}_0' \subset \mathcal{T}_1\right\}$*.*

Lemma 2. [Sutherland 2009, Proposition 11.4, p.110] (Metrizable-Hausdorff Space) *Every metrizable space* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *is Hausdorff.*

Definition 6. (Hausdorff Space) *A topological space* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *is Hausdorff if and only if* $\forall (a_1, a_2) \in \mathbb{A} \times \mathbb{A}$ *there exists open sets* $\mathbb{A}_1 \subset \mathbb{A}$ *and* $\mathbb{A}_2 \subset \mathbb{A}$ *such that* $a_1 \in \mathbb{A}_1$*,* $a_2 \in \mathbb{A}_2$ *and* $A_1 \cap \mathbb{A}_2 = \emptyset$*.*

Definition 7. (Measurable Map) *Let* $(\mathbb{A}, \mathfrak{A})$ *and* $(\mathbb{B}, \mathfrak{B})$ *be measurable spaces. A map* $f : \mathbb{A} \to \mathbb{B}$ *is* $\mathcal{B}/\mathcal{A}$*-measurable if* $f^{-1}(B) \in \mathfrak{A}$ *for every* $B \in \mathfrak{B}$*.*

Lemma 3. [Billingsley (1995, Theorem 13.1, p.182)] (Measurable Map) *Let* $(\mathbb{A}, \mathfrak{A})$ *and* $(\mathbb{B}, \mathfrak{B})$ *be measurable spaces. Let* $f : \mathbb{A} \to \mathbb{B}$ *be such that* $f^{-1}(B) \in \mathfrak{A}$ *for every* $B \in \mathcal{B}_0$ *and let* $\mathfrak{B}$ *be generated by* $\mathcal{B}_0$*, then* f *is* $\mathcal{B}/\mathcal{A}$*-measurable.*

Lemma 4. [Klambauer 1973, Proposition 4, p.234] (Inverse of Continuous Operator) *Let* $(\mathbb{A}, \mathfrak{A})$ *and* $(\mathbb{B}, \mathfrak{B})$ *be topological spaces. A map* $f : \mathbb{A} \to \mathbb{B}$ *is continuous map if and only if its inverse is an open map.*

The following is an immediate Corollary of Lemmas 3 and 4.

Corollary 10. (Continuous Borel Map) *Let* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *and* $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ *be topological spaces with Borel* σ*-algebra* $\mathfrak{B}(\mathbb{A})$ *and* $\mathfrak{B}(\mathbb{B})$ *generated by* $\mathcal{T}_{\mathbb{A}}$ *and* $\mathcal{T}_{\mathbb{B}}$ *respectively. Then a continuous map* $f : \mathbb{A} \to \mathbb{B}$ *is* $\mathfrak{B}(\mathbb{B})/\mathfrak{B}(\mathbb{A})$*-measurable.*

Lemma 5. [Billingsley 1995, Theorem 13.1, p.182] (Measurable Composition) *Let* $(\mathbb{A},\mathfrak{A})$*,* $(\mathbb{B},\mathfrak{B})$ *and* $(\mathbb{C},\mathfrak{C})$ *be measurable spaces. Let* $f:\mathbb{A}\to\mathbb{B}$ *be* $\mathcal{B}/\mathcal{A}$*-measurable and* $g:\mathbb{B}\to\mathbb{C}$ *be* $\mathfrak{C}/\mathfrak{B}$*-measurable. Then* $g\circ f:\mathbb{A}\to\mathbb{C}$ *is* $\mathfrak{C}/\mathfrak{A}$*-measurable.*

Lemma 6. [Gamelin and Greene 1999, Theorem 12.1, p.101] (Continuous Projections) *Let* $(\mathbb{A}_i,\mathcal{T}_{\mathbb{A}_i})$ *be topological spaces for all* i *in some set* $\mathbb{I}$ *and let* $\mathbb{A}=\times_{i\in\mathbb{I}}\mathbb{A}_i$*. The product topology is the smallest topology making the coordinate projections* $\pi_i:\mathbb{A}\to\mathbb{A}_i$ *continuous* $\forall\, i\in\mathbb{I}$*.*

The following result thus follows as a Corollary of Lemma 6.

Corollary 11. [James 1987, Corollary 2.12, p.33](Product Topology Convergent Sequences) *Let* $(\mathbb{A}_i,\mathcal{T}_{\mathbb{A}_i})$ *be topological spaces for all* i *in some set* $\mathbb{I}$ *and let* $\mathbb{A}=\times_{i\in\mathbb{I}}\mathbb{A}_i$*. Let* $(\mathbb{A},\mathcal{T}_{\mathbb{A}})$ *be the product space with product topology* $\mathcal{T}_{\mathbb{A}}$*. A sequence* $\{a_n\}_{n\in\mathbb{N}}\subset\mathbb{A}$ *satisfies* $a_n\to a\in\mathbb{A}$ *if and only if* $\pi_i(a_n)\to\pi_i(a)\in\mathbb{A}_i\ \forall\, i\in\mathbb{I}$*.*

Also, a Corollary of Lemmas 10 and 6 is as follows.

Corollary 12. (Measurable Projections) *Let* $(\mathbb{A}_i,\mathcal{T}_{\mathbb{A}_i})$ *be topological spaces with Borel* σ*-algebra* $\mathfrak{B}(\mathbb{A}_i)$ *generated by* $\mathcal{T}_{\mathbb{A}_i}$ *for all* i *in some set* $\mathbb{I}$*. Let* $(\mathbb{A},\mathcal{T}_{\mathbb{A}})$ *be the product space* $\mathbb{A}=\times_{i\in\mathbb{I}}\mathbb{A}_i$ *with product topology* $\mathcal{T}_{\mathbb{A}}$ *and Borel* σ*-algebra* $\mathfrak{B}(\mathbb{A})$ *generated by* $\mathcal{T}_{\mathbb{A}}$*. Then the projection maps* $\pi_i:\mathbb{A}\to\mathbb{A}_i$ *are* $\mathfrak{B}(\mathbb{A}_i)/\mathfrak{B}(\mathbb{A})$*-measurable* $\forall\, i\in\mathbb{I}$*.*

Lemma 7. [Gamelin and Greene 1999, Theorem 12.2, p.101] (Continuous Map into Product Spaces) *Let* $(\mathbb{A}_i,\mathcal{T}_{\mathbb{A}_i})$ *be topological spaces for all* i *in some set* $\mathbb{I}$ *and let* $\mathbb{A}=\times_{i\in\mathbb{I}}\mathbb{A}_i$*. Let* $(\mathbb{A},\mathcal{T}_{\mathbb{A}})$ *be the product space with product topology* $\mathcal{T}_{\mathbb{A}}$ *and* $(\mathbb{B},\mathcal{T}_{\mathbb{B}})$ *be some topological space. An operator* $f:\mathbb{B}\to\mathbb{A}$ *is continuous at* $b\in\mathbb{B}$ *if and only if* $\pi_i\circ f:\mathbb{B}\to\mathbb{A}_i$ *is continuous at* b *for every* $i\in\mathbb{I}$*.*

Lemma 8. [Dudley 2002, Theorem 2.2.8, p.39] (Tychonoff's Theorem) *Let* $(\mathbb{A}_i,\mathcal{T}_i)$ *be compact topological spaces for each* i *in a set* $\mathbb{I}$*. Then the Cartesian product* $\times_{i\in\mathbb{I}}\mathbb{A}_i$ *with product topology is compact.*

Lemma 9. [Munkres 2000, Theorem 31.2, p.196] (Subsets and Countable Products of Regular Spaces) *Any subspace of a regular space is regular. Any product of regular spaces is regular.*

Lemma 10. [Davidson 1994, Theorem 6.16, p.103] (Countable Products of Separable Spaces) *Let* $(\mathbb{A}_i,\mathcal{T}_{\mathbb{A}_i})$ *be topological spaces for all* i *in some countable set* $\mathbb{I}$ *and let* $\mathbb{A}=\times_{i\in\mathbb{I}}\mathbb{A}_i$*. Let* $(\mathbb{A},\mathcal{T}_{\mathbb{A}})$ *be the product space with product topology* $\mathcal{T}_{\mathbb{A}}$*. Then* $\mathbb{A}$ *is separable if and only if* $\mathbb{A}_i$ *is separable for every* $i\in\mathbb{I}$*.*

Lemma 11. [Dudley 2002, Proposition 2.4.4, p.50] (Metrization of Product Topology) *For every sequence of metric spaces* $\big\{(\mathbb{A}_i,\delta_{\mathbb{A}_i})\big\}_{i\in\mathbb{N}}$*, the topological product space* $(\times_{i\in\mathbb{N}}\mathbb{A}_i,\mathcal{T}_{\mathbb{A}})$ *with product topology* $\mathcal{T}_{\mathbb{A}}$ *is metrizable by the product-metric,*

$$\delta_{\mathbb{A}}(a,a'):=\sum_{i\in\mathbb{N}}\frac{1}{2^i}\frac{\delta_{\mathbb{A}_i}(a_i,a'_i)}{1+\delta_{\mathbb{A}_i}(a_i,a'_i)}\qquad\forall\ (a,a')=\big(\{a_i\}_{i\in\mathbb{N}},\{a'_i\}_{i\in\mathbb{N}}\big)\in\mathbb{A}\times\mathbb{A}.$$

Remark 1. *Uncountable product spaces with product topology are not metrizable.*

Lemma 12. [Dudley 2002, Proposition 4.1.7, p.119] (Algebra on Product Spaces) *Let* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *and* $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ *be any two topological spaces. Let* $(\mathbb{A} \times \mathbb{B}, \mathcal{T}_{\mathbb{A}\times\mathbb{B}})$ *be the product space with product Tychonoff's topology* $\mathcal{T}_{\mathbb{A}\times\mathbb{B}}$ *and let* $\mathfrak{B}(\mathbb{A} \times \mathbb{B})$ *denote the Borel* σ*-algebra generated by the product topology* $\mathcal{T}_{\mathbb{A}\times\mathbb{B}}$ *on* $\mathbb{A} \times \mathbb{B}$. *Then* $\mathfrak{B}(\mathbb{A} \times \mathbb{B})$ *includes the product* σ*-algebra* $\mathfrak{B}(\mathbb{A}) \otimes \mathfrak{B}(\mathbb{B})$. *If both* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *and* $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ *are second-countable then the two* σ*-algebras on* $\mathbb{A} \times \mathbb{B}$ *are equal.*

Lemma 13. (Separability and Second-Countability) *A metric space* $(\mathbb{A}, \delta_{\mathbb{A}})$ *is separable if and only if it is second-countable.*

Lemma 14. [Foland 2009, p.24] (Measurable Maps and Product σ-Algebra) *Let* $(\mathbb{A}, \mathfrak{A})$ *and* $(\mathbb{B}_i, \mathfrak{B}_i)$ *be measurable spaces for all* i *in some set* $\mathbb{I}$. *Let* $(\mathbb{B}, \mathfrak{B})$ *be the product space* $\mathbb{B} = \times_{i\in\mathbb{I}}\mathbb{B}_i$ *with product* σ*-algebra* $\mathfrak{B} = \otimes_{i\in\mathbb{I}}$. *Then the map* $f : \mathbb{A} \to \mathbb{B}$ *is* $\mathfrak{B}/\mathfrak{A}$*-measurable if and only if the projection maps* $\pi_i \circ f : \mathbb{A} \to \mathbb{B}_i$ *are* $\mathfrak{B}_i/\mathfrak{A}$*-measurable* $\forall\, i \in \mathbb{I}$.

The following is obtained as a Corollary of Lemmas 12 and 14.

Corollary 13. (Measurable Maps into Product Spaces) *Let* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *and* $(\mathbb{B}_i, \mathcal{T}_{\mathbb{B}_i})$ *be topological spaces with Borel* σ*-algebra* $\mathfrak{B}(\mathbb{A})$ *and* $\mathfrak{B}(\mathbb{B}_i)$ *generated by* $\mathcal{T}_{\mathbb{A}}$ *and* $\mathcal{T}_{\mathbb{B}_i}$ *respectively for all* i *in some set* $\mathbb{I}$. *Let* $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ *be the product space* $\mathbb{B} = \times_{i\in\mathbb{I}}\mathbb{B}_i$ *with product topology* $\mathcal{T}_{\mathbb{B}}$ *and Borel* σ*-algebra* $\mathfrak{B}(\mathbb{B})$ *generated by* $\mathcal{T}_{\mathbb{B}}$. *Then the map* $f : \mathbb{A} \to \mathbb{B}$ *is* $\mathfrak{B}(\mathcal{T}_{\mathbb{B}})/\mathfrak{B}(\mathcal{T}_{\mathbb{A}})$*-measurable if the projection maps* $\pi_i \circ f : \mathbb{A} \to \mathbb{B}_i$ *are* $\mathfrak{B}(\mathbb{B}_i)/\mathfrak{B}(\mathbb{A})$*-measurable* $\forall\, i \in \mathbb{I}$.

Definition 8. (Topological Vector Space) *A topological vector space* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *is a vector space* $\mathbb{A}$ *endowed with a topology* $\mathcal{T}_{\mathbb{A}}$ *such that vector addition and scalar multiplication are continuous functions.*

Lemma 15. [Sutherland 2009, Proposition 8.4, p.84] (Continuous Composition) *Let* $(\mathbb{A}, \delta_{\mathbb{A}})$, $(\mathbb{B}, \delta_{\mathbb{B}})$ *and* $(\mathbb{C}, \delta_{\mathbb{C}})$ *be topological spaces and* $f : \mathbb{A} \to \mathbb{B}$ *and* $g : \mathbb{B} \to \mathbb{C}$ *be continuous at* $a \in \mathbb{A}$ *and* $b \in \mathbb{B}$ *respectively. Then* $g \circ f : \mathbb{A} \to \mathbb{C}$ *is continuous at* $a \in \mathbb{A}$.

Lemma 16. [Klein and Thompson 1984, Lemma 13.2.3,p.154] (Measurable Maps) *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* (Θ, δ_Θ) *be a separable metric space. If* $Q(\omega, \cdot) : \Theta \to \mathbb{R}_0^+$ *is continuous in* Θ *for every* $\omega \in \Omega$ *and* $Q(\cdot, \boldsymbol{\theta}) : \Omega \to \mathbb{R}_0^+$ *is measurable for every* $\boldsymbol{\theta} \in \Theta$, *then* $Q : \Omega \times \Theta \to \mathbb{R}_0^+$ *is* $\mathcal{F} \times \mathfrak{B}(\Theta)$*-measurable.*

Lemma 17. [Debreu 1967, Theorem 4.5] (Measurability) *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $(\mathbb{A}, \delta_{\mathbb{A}})$ *be a complete separable metric space with Borel* σ*-algebra* $\mathfrak{B}(\mathbb{A})$. *Let the random sieve-correspondence* $\mathbb{A}_T : \Omega \to \mathbb{A}$ *have a measurable graph* $\mathfrak{gr}(\mathbb{A}_T) \in \mathcal{F} \otimes \mathfrak{B}(\mathbb{A})$ *and the sieves* $\mathbb{A}_T(\omega) \subset \mathbb{A}$ *are non-empty and compact for every* $\omega \in \Omega$. *Finally, let the criterion mapping* $f_T : \mathfrak{gr}(\mathbb{A}_T) \to \mathbb{R}_0^+$ *be* $\mathcal{F} \otimes \mathfrak{B}(\mathbb{A})$*-measurable and* $f_T(\omega) : \mathbb{A} \to \mathbb{R}_0^+$ *be continuous on* $\mathbb{A}$. *Then* $f_T^{\inf} : \Omega \to \mathbb{R}_0^+$ *is* $\mathcal{F}^P/\mathfrak{B}(\mathbb{R}_0^+)$*-measurable and the minimizer set* $\widehat{\mathbb{A}\Omega} \in \mathbb{A} \times \Omega$ *defined as* $\hat{\mathbb{A}} : \Omega \to \Theta$ *satisfying* $\hat{\mathbb{A}}(\omega) := \{a \in \mathbb{A}_T(\omega) : f_T(\omega, a) = \inf_{\boldsymbol{\theta}\in\mathbb{A}_T(\omega)} f_T(\omega, a)\}$ *for every* $\omega \in \Omega$ *belongs to* $\mathcal{F}^P \otimes \mathfrak{B}(\Theta)$.

Lemma 18. [Hildenbrand 1974, p.55] (Measurable Selection) *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $\mathbb{A}$ *a complete separable metric space with its Borel* σ*-field* $\mathfrak{B}(\mathbb{A})$ *and* $f_T^{\sup} : \Omega \to 2^{\mathbb{A}}$ *a closed valued correspondence s.t.* $\{\omega \in \Omega : f_T^{\sup}(\omega) \cap \mathbb{A}^*\} \in \mathcal{F}$ *for every closed subset* $\mathbb{A}^* \subset \mathbb{A}$*. Then* $f_T^{\sup} : \Omega \to 2^{\mathbb{A}}$ *admits a measurable selector, i.e. there exists a map* $\hat{a}_T : \Omega \to 2^{\mathbb{A}}$ *that is measurable and for every* $\omega \in \Omega$ *it satisfies* $\hat{a}_T(\omega) \in f_T^{\sup}(\omega)$*.*

Corollary 14. [White and Wooldrige 1991, Theorem 2.2, p.646] (Measurable Extrema) *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a complete probability space and* $(\mathbb{A}, \delta_{\mathbb{A}})$ *be a complete separable metric space. Let* $\{\mathbb{A}_n\}_{n\in\mathbb{N}}$ *be a sequence of compact subsets of* $\mathbb{A}$*. Let* $f_n : \Omega \times \mathbb{A}_n \to \mathbb{R}$ *be* $\mathcal{F} \otimes \mathfrak{B}(\mathbb{A}_n)/\mathfrak{B}(\mathbb{R})$*-measurable for every* $n \in \mathbb{N}$ *and* $f_n(\omega, \cdot) : \mathbb{A} \to \mathbb{R}$ *be continuous on* $\mathbb{A}_n$ *for every* $(\omega, n) \in \Omega \times \mathbb{N}$*. Then there exists an* $\mathcal{F}/\mathfrak{B}(\mathbb{A}_n)$*-measurable map* $\hat{a}_n : \Omega \to \mathbb{A}_n$ *satisfying* $f_n(\omega, \hat{a}_n(\omega)) = \inf_{a\in\mathbb{A}_n} f_n(\omega, a)$ *for every* $\omega \in \Omega$ *and every* $n \in \mathbb{N}$*.*

Definition 9. (Metric Equivalence) *Let* $\mathbb{A}$ *be a set. Two metrics,* $\delta^1_{\mathbb{A}} : \mathbb{A}\times\mathbb{A} \to \mathbb{R}$ *and* $\delta^2_{\mathbb{A}} : \mathbb{A}\times\mathbb{A} \to \mathbb{R}$ *are said to be topologically equivalent if they define the same open sets, i.e. if they induce the same topology* $\mathcal{T}_{\mathbb{A}}$ *on* $\mathbb{A}$*.*

Remark 2. *Let* $\delta^1_{\mathbb{A}}$ *and* $\delta^2_{\mathbb{A}}$ *be any two topologically equivalent metrics on the set* $\mathbb{A}$*. If a sequence in* $\mathbb{A}$ *is* $\delta^1_{\mathbb{A}}$*-convergent then it is also* $\delta^2_{\mathbb{A}}$*-convergent.*

Definition 10. (Lipschitz Stronger/Weaker Metric) *Given a pair of metrics* $\delta_{\mathbb{A}}$ *and* $\delta'_{\mathbb{A}}$ *defined on the product* $\mathbb{A}\times\mathbb{A}$ *of some set* $\mathbb{A}$*, the metric* $\delta_{\mathbb{A}}$ *is said to be Lispchitz weaker than* $\delta'_{\mathbb{A}}$ *if* $\exists\, k \in \mathbb{R}^+$ *such that* $\delta_{\mathbb{A}}(a, a') \le k \cdot \delta'_{\mathbb{A}}(a, a')\ \forall\, (a, a') \in \mathbb{A} \times \mathbb{A}$*. The metric* $\delta'_{\mathbb{A}}$ *is also said to be Lispchitz stronger than* $\delta_{\mathbb{A}}$*. Furthermore, if* $\exists\, (k, k') \in \mathbb{R}^+ \times \mathbb{R}^+$ *such that* $k \cdot \delta'_{\mathbb{A}}(a, a') \le \delta_{\mathbb{A}}(a, a') \le k' \cdot \delta'_{\mathbb{A}}(a, a')\ \forall\, (a, a') \in \mathbb{A} \times \mathbb{A}$ *then* $\delta_{\mathbb{A}}$ *and* $\delta'_{\mathbb{A}}$ *are said to be Lipschitz equivalent.*

Lemma 19. [Sutherland 2009, Proposition 6.34, p.70] (Lispchitz Topological Equivalence) *A pair of Lipschitz equivalent metrics* $\delta_{\mathbb{A}}$ *and* $\delta'_{\mathbb{A}}$ *defined on the product* $\mathbb{A}\times\mathbb{A}$ *of some set* $\mathbb{A}$ *is also topologically equivalent.*

Definition 11. (Uniform Product Metric) *Given metric spaces* $(\mathbb{A}_i, \delta_{\mathbb{A}_i})$*,* $i \in \mathbb{I}$ *where* $\mathbb{I}$ *is a countable index set and a product space* $\mathbb{A} := \times_{i\in\mathbb{I}}\mathbb{A}_i$*. The product metric* $\delta_{\mathbb{A}}(a, a') := \sup_{i\in\mathbb{I}} \delta_{\mathbb{A}_i}(a_i, a'_i)\ \forall\, (a, a') \in \mathbb{A} \times \mathbb{A}$ *is called the uniform product metric on* $\mathbb{A}$*.*

Proposition 1. (Lipschitz Weaker Metrics) *Both product metrics in (1) are Lipschitz weaker than the uniform product metric.*

Proof. Immediate from the definitions in (1) since,

$$\begin{aligned}\delta_{\mathcal{B}}(\boldsymbol{\beta}, \boldsymbol{\beta}') &= \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{\delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i}, \boldsymbol{\beta}'_{L_i})}{1 + \delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i}, \boldsymbol{\beta}'_{L_i})} \le \sum_{i=1}^{\infty} \frac{1}{2^i} \delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i}, \boldsymbol{\beta}'_{L_i}) \\ &\le \sum_{i=1}^{\infty} \frac{1}{2^i} \sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}(\boldsymbol{\beta}_L, \boldsymbol{\beta}'_L) = \sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}(\boldsymbol{\beta}_L, \boldsymbol{\beta}'_L).\end{aligned}$$

and also,

$$\delta_{\mathcal{B}}(\boldsymbol{\beta},\boldsymbol{\beta}') = \sup_{i\in\mathbb{N}} \frac{1}{i}\frac{\delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i},\boldsymbol{\beta}'_{L_i})}{1+\delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i},\boldsymbol{\beta}'_{L_i})} \leq \sup_{i\in\mathbb{N}} \frac{1}{i}\delta_{\mathcal{B}_{L_i}}(\boldsymbol{\beta}_{L_i},\boldsymbol{\beta}'_{L_i}) \leq \sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}(\boldsymbol{\beta}_L,\boldsymbol{\beta}'_L), \tag{7}$$

□

Lemma 20. [Sutherland 2009, Proposition 3.18, p.13] (Inverse Bijection) *Let $\mathbb{A}$ and $\mathbb{B}$ be two sets and $f : \mathbb{A} \to \mathbb{B}$. The map f is invertible if and only if it is bijective.*

Definition 12. (Homeomorphism) *Let $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ and $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ be topological spaces. A map $f : \mathbb{A} \to \mathbb{B}$ is said to be a Homeomorphism iff it is continuous, bijective, and has continuous inverse f^{-1}.*

Lemma 21. [Lee 2000, Proposition 3.13, p.51, James 1987, p.31] (Product Homeomorphisms) *Let $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ and $\{(\mathbb{B}_i, \mathcal{T}_{\mathbb{B}})\}_{i\in\mathbb{I}}$ be topological spaces and $\mathbb{I}$ be an arbitrary set. Let $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ denote the product space $\mathbb{B} := \times_{i\in\mathbb{I}}\mathbb{B}_i$ with Tychonoff's topology $\mathcal{T}_{\mathbb{B}}$. A map $f : \mathbb{A} \to \mathbb{B}$ is a homeomorphism if every projection map $\pi_i f : \mathbb{A} \to \mathbb{B}_i$ is a homeomorphism for every $i \in \mathbb{I}$.*

Lemma 22. [Basener 1973, p.13](Open Sets in Product Topology) *Let $(\mathbb{A}_i, \mathcal{T}_{\mathbb{A}_i})$ be topological spaces for every i in some set $\mathbb{I}$ and $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ be the product space $\mathbb{A} = \times_{i\in\mathbb{I}}\mathbb{A}_i$ with product topology $\mathcal{T}_{\mathbb{A}}$. Then a subset $\mathcal{O} \subseteq \mathbb{A}$ is open if and only if $\pi_i(\mathbb{O}) \subseteq \mathbb{A}_i$ is open for every $i \in \mathbb{I}$.*

Proposition 2. (Homeomorphisms on Product Spaces) *Let $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ and $\{(\mathbb{B}_i, \mathcal{T}_{\mathbb{B}})\}_{i\in\mathbb{I}}$ be topological spaces and $\mathbb{I}$ be a countable set. Let $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ denote the product space $\mathbb{B} := \times_{i\in\mathbb{I}}\mathbb{B}_i$ with Tychonoff's topology $\mathcal{T}_{\mathbb{B}}$. Let $\{f_i\}_{i\in\mathbb{I}}$ denote a collection of maps $f_i : \mathbb{A} \to \mathbb{B}_i$ such that* (i) *f_i is continuous on $\mathbb{A}\ \forall\, i \in \mathbb{I}$;* (ii) *$f_i$ is open $\forall\, i \in \mathbb{I}$; and* (iii) *for every pair $(a, a') \in \mathbb{A}\times\mathbb{A}$, $\exists\, i \in \mathbb{I}\ :\ f_i(a) \neq f_i(a')$. Then the product map $f : \mathbb{A} \to \mathbb{B}$ satisfying $f(a) = (f_{i_1}(a), f_{i_2}(a), ...)$ is a homeomorphism on its range.*

Proof. Continuity of the product map f follows by continuity of each projection map f_i by Lemma 7. Openess of f follows by noting that the image $f(\mathbb{A}_{\mathcal{T}})$ of an open set $\mathbb{A}_{\mathcal{T}} \in \mathcal{T}_{\mathbb{A}}$ must be an open subset $f(\mathbb{A}_{\mathcal{T}}) \subseteq \mathbb{B}$ by Lemma 22 since $f_i(\mathbb{A}_{\mathcal{T}}) \subseteq \mathbb{B}_i$ is an open set (i.e. $f_i(\mathbb{A}_{\mathcal{T}}) \in \mathcal{T}_{\mathbb{B}_i}$) for every $i \in \mathbb{I}$. The injective nature follows easily since, by contradiction, if $\exists\,(a, a') \in \mathbb{A}\times\mathbb{A}$ such that $f(a) = f(a')$, then by construction is must be that $f_i(a) = f_i(a')\ \forall\, i \in \mathbb{I}$, but this contradicts the assumption that for every pair $(a, a') \in \mathbb{A}\times\mathbb{A}$, $\exists\, i \in \mathbb{I}\ :\ f_i(a) \neq f_i(a')$. □

Definition 13. (Divergence) *Let $\mathbb{A}$ be a non-empty set and $f : \mathbb{A}\times\mathbb{A} \to \mathbb{R}$. The real-valued map f is said to be a divergence on $\mathbb{A}$ if and only if it satisfies* (i) *non-negativity $f(a, a') \geq 0\ \forall\,(a, a') \in \mathbb{A}\times\mathbb{A}$, and* (ii) *identity of indiscernibles $f(a) = f(a')$ iff $a = a'$, $\forall\,(a, a') \in \mathbb{A}\times\mathbb{A}$.*

Definition 14. (Identifiably Unique Minimizer) *Let $(\mathbb{A}, \delta_{\mathbb{A}})$ be a metric space and $f : \mathbb{A} \to \mathbb{R}$ be some real-valued map. Then $a_0 \in \mathbb{A}$ is called an identifiably unique minimizer of f if and only if $\inf_{a\in S^c_{a_0}(\epsilon)} |f(a) - f(a_0)| > 0$ for every $\epsilon > 0$.*

Definition 15. (g-Homogeneous Function) *Let $\mathbb{A}$ be a vector space. A function $f : \mathbb{A} \to \mathbb{R}$ is called g-homogeneous if and only if there exists a function $g : \mathbb{A} \to \mathbb{R}$ satisfying $g(a_n) = O_p(1)$ for every sequence $\{a_n\}_{n\in\mathbb{N}}$ satisfying $a_n = O_p(1)$ and $\limsup_{n\in\mathbb{N}} g(a_n) < \infty$ a.s. for every sequence $\{a_n\}_{n\in\mathbb{N}}$ satisfying $\limsup_{n\in\mathbb{N}} a_n < \infty$ a.s. , such that $f(a \cdot a') = g(a) \cdot f(a')$.*

Lemma 23. *[Van der Vaart and Wellner 1996, Theorem 1.11.1, p.67](Extended Continuous Mapping Theorem) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $(\mathbb{A}, \mathfrak{B}(\mathbb{A}))$ and $(\mathbb{B}, \mathfrak{B}(\mathbb{B}))$ be measurable spaces with Borel σ-algebras $\mathfrak{B}(\mathbb{A})$ and $\mathfrak{B}(\mathbb{B})$ respectively. Let $f_T : \mathbb{A}_T \to \mathbb{B}$ be measurable maps defined on subsets $\mathbb{A}_T \subset \mathbb{A}\ \forall T \in \mathbb{N}$ satisfying $f_T(a_T) \to f(a)$ for every $a_T \to a$ with $a_T \in \mathbb{A}_T\ \forall T \in \mathbb{N}$, $a \in \mathbb{A}_0$ and some measurable $f : \mathbb{A}_0 \to \mathbb{B}$ with $\mathbb{A}_0 \subset \mathbb{A}$. Let $X_T : \Omega \to \mathbb{A}_T$ be $\mathcal{F}/\mathfrak{B}(\mathbb{A}_T)$-measurable maps taking values in $\mathbb{A}_T$ and X be $\mathcal{F}/\mathfrak{B}(\mathbb{A})$-measurable and separable and take values in $\mathbb{A}_0$. Then, (i) $X_T \xrightarrow{d} X$ implies $f_T(X_T) \xrightarrow{d} f(X)$, (ii) $X_T \xrightarrow{p} X$ implies $f_T(X_T) \xrightarrow{p} f(X)$, and (iii) $X_T \xrightarrow{a.s.} X$ implies $f_T(X_T) \xrightarrow{a.s.} f(X)$.*

Corollary 15. (Continuous Mapping Theorem) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $(\mathbb{A}, \mathfrak{B}(\mathbb{A}))$ and $(\mathbb{B}, \mathfrak{B}(\mathbb{B}))$ be measurable spaces with Borel σ-algebras $\mathfrak{B}(\mathbb{A})$ and $\mathfrak{B}(\mathbb{B})$ respectively. Let $g : \mathbb{A} \to \mathbb{B}$ be continuous. Let $X : \Omega \to \mathbb{A}$ be $\mathcal{F}/\mathfrak{B}(\mathbb{A})$-measurable and separable and take values in $\mathbb{A}$. Then, (i) $X_T \xrightarrow{d} X$ implies $f(X_T) \xrightarrow{d} f(X)$, (ii) $X_T \xrightarrow{p} X$ implies $f(X_T) \xrightarrow{p} f(X)$, and (iii) $X_T \xrightarrow{a.s.} X$ implies $f(X_T) \xrightarrow{a.s.} f(X)$.*

Lemma 24. [Davidson and Donsig 2009, Theorem 2.4.6, p.17] (Squeeze Theorem) *Let $\{a_n\}_{n\in\mathbb{N}}$, $\{b_n\}_{n\in\mathbb{N}}$ and $\{c_n\}_{n\in\mathbb{N}}$ be sequences satisfying $a_n \leq b_n \leq c_n\ \forall n \in \mathbb{N}$ and $\lim_{n\to\infty} a_n = b$ and $\lim_{n\to\infty} c_n = b$. Then $\lim_{n\to\infty} b_n = b$.*

Proposition 3. (Uniform Continuity Preserves Uniform Convergence) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathbb{A}, \delta_{\mathbb{A}})$, $(\mathbb{B}, \delta_{\mathbb{B}})$ and $(\mathbb{C}, \delta_{\mathbb{C}})$ be measurable metric spaces with Borel σ-algebras $\mathfrak{B}(\mathbb{A})$, $\mathfrak{B}(\mathbb{B})$ and $\mathfrak{B}(\mathbb{C})$ respectively. Let $g : \mathbb{B} \to \mathbb{C}$ be a uniformly continuous map on $\mathbb{B}$. If $\{f_T\}_{T\in\mathbb{N}}$ are measurable maps $f_T : \Omega \times \mathbb{A} \to \mathbb{B}$ satisfying $\sup_{a\in\mathbb{A}} \delta_{\mathbb{B}}(f_T(a), f_\infty(a)) \xrightarrow{p} 0$ for some measurable $f_\infty : \mathbb{A} \to \mathbb{B}$, then $\sup_{a\in\mathbb{A}} \delta_{\mathbb{C}}(g \circ f_T(a), g \circ f_\infty(a)) \xrightarrow{p} 0$, and if $\sup_{a\in\mathbb{A}} \delta_{\mathbb{B}}(f_T(a), f_\infty(a)) \xrightarrow{a.s.} 0$ then $\sup_{a\in\mathbb{A}} \delta_{\mathbb{C}}(g \circ f_T(a), g \circ f_\infty(a)) \xrightarrow{a.s.} 0$.*

Proof. By uniform continuity of g on Y we have that for every $(\omega, T) \in \Omega \times \mathbb{N}$ and every $\epsilon > 0$, $\exists\ \epsilon' > 0$ such that having

$$\delta_Y(f_T(\omega, a), f_\infty(a)) < \epsilon' \quad \text{implies} \quad \delta_Z(g \circ f_T(\omega, a), g \circ f_\infty(\omega, a)) < \epsilon. \tag{8}$$

Now, convergence in probability follows since for every $T \in \mathbb{N}$ it holds true that

$$\mathbb{P}(\sup_{a\in\mathbb{A}} \delta_{\mathbb{C}}(g \circ f_T(a), g \circ f_\infty(a)) < \epsilon) \geq \mathbb{P}(\sup_{a\in\mathbb{A}} \delta_Y(f_T(a), f_\infty(a)) < \epsilon')$$

because the second implies the first $\forall \omega \in \Omega$. Hence, since pointwise convergence in probability $\lim_{T\to\infty} \mathbb{P}(\sup_{a\in\mathbb{A}} \delta_Y(f_T(a), f_\infty(a)) < \epsilon') = 1\ \forall \epsilon' > 0$ holds by assumption, it follows that

$$\lim_{T\to\infty} \mathbb{P}(\sup_{a\in\mathbb{A}} \delta_Z(g \circ f_T(a), g \circ f_\infty(a)) < \epsilon) = 1\ \forall \epsilon > 0.$$

Convergence a.s. follows since $\forall T \in \mathbb{N}$ it holds true, by (8) and Lemma 24 that for every $\omega \in \Omega$, $\lim_{T\to\infty} \delta_Y(f_T(\omega,a), f_\infty(a)) < \epsilon'$ implies $\lim_{T\to\infty} \delta_Z(g \circ f_T(\omega,a), g \circ f_\infty(\omega,a)) < \epsilon$, and hence that,

$$\mathbb{P}(\lim_{T\to\infty} \sup_{a\in\mathbb{A}} \delta_{\mathbb{C}}(g \circ f_T(a), g \circ f_\infty(a)) < \epsilon) \geq \mathbb{P}(\lim_{T\to\infty} \sup_{a\in\mathbb{A}} \delta_Y(f_T(a), f_\infty(a)) < \epsilon')$$

because the second implies the first $\forall \omega \in \Omega$. Hence, since pointwise a.s. convergence holds by assumption, i.e. $\mathbb{P}(\lim_{T\to\infty} \sup_{a\in\mathbb{A}} \delta_Y(f_T(a), f_\infty(a)) < \epsilon') = 1 \; \forall \epsilon' > 0$, it follows that

$$\mathbb{P}(\lim_{T\to\infty} \sup_{a\in\mathbb{A}} \delta_Z(g \circ f_T(a), g \circ f_\infty(a)) < \epsilon) = 1.$$

□

Lemma 25. [Davidson 1994, Theorem 2.19, p.28] (Heine-Cantor Theorem) *Let* $(\mathbb{A}, \delta_{\mathbb{A}})$ *and* $(\mathbb{B}, \delta_{\mathbb{B}})$ *be metric spaces and* $f : \mathbb{A} \to \mathbb{B}$ *be a continuous map at every* $a \in \mathbb{A}$. *Then, if* $\mathbb{A}$ *is compact,* f *is uniformly continuous on* $\mathbb{A}$.

Lemma 26. [Chen 2007, Theorem 3.1 and Remark 3.2 and White and Wooldrige 1991, Proposition 2.4 and Corollary 2.6] (Convergence of Sieve Estimators) *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a complete probability space,* $(\mathbb{A}, \delta_{\mathbb{A}})$ *be a metric space with Borel* σ*-algebra* $\mathfrak{B}(\mathbb{A})$ *and* $\{\mathbb{A}_T\}_{T\in\mathbb{N}}$ *be a sequence of compact subsets of* $\mathbb{A}$ *such that* $\text{cl}\big(\bigcup_{T\in\mathbb{N}} \mathbb{A}_T\big) \supseteq \mathbb{A}$. *Suppose that the sequence* $\{f_T\}_{T\in\mathbb{N}}$ *of functions* $f : \Omega \times \mathbb{A} \to \mathbb{R}$, *continuous of* $\mathbb{A}$ $\forall T \in \mathbb{N}$ *and such that,* $\lim_{T\to\infty} \mathbb{P}\Big(\sup_{a\in\mathbb{A}} |f_T(a) - f(a)| > \epsilon\Big) = 0 \;\; \forall \epsilon > 0$, *for some continuous deterministic function* $f : \mathbb{A} \to \mathbb{R}$ *satisfying,* $f(a_0) = 0$ *and* $\inf_{a\in S^c_{a_0}(\epsilon)} |f(a_0) - f(a)| > 0 \;\; \forall \epsilon > 0$. *Let* $\hat{a}_T : \Omega \to \mathbb{A}$ *be an* $\mathcal{F}/\mathfrak{B}(\mathbb{A})$*-measurable map such that,* $f_T(\hat{a}_T) \leq \inf_{a\in\mathbb{A}_T} f_T(a) + O_p(\eta_T)$ *with* $\eta_T \to 0$ *as* $T \to \infty$. *Then,* $\lim_{T\to\infty} \big(\delta_{\mathbb{A}}(\hat{a}_T, a_0) > \epsilon\big) = 0 \; \forall \epsilon > 0$.

Lemma 27. [Dudley 2002, Theorem 2.8.2, p.72] (Compactification) *Any separable metric space* $(\mathbb{A}, \delta_{\mathbb{A}})$ *has a totally bounded metrization, i.e. there exists a metric* $\delta'_{\mathbb{A}}$ *on* $\mathbb{A}$ *inducing the same topology as* $\delta_{\mathbb{A}}$ *on* $\mathbb{A}$ *such that* $(\mathbb{A}, \delta'_{\mathbb{A}})$ *is totally bounded, so that the completion for* $\delta'_{\mathbb{A}}$ *is a compact metric space and a compactification of* $\mathbb{A}$.

Lemma 28. [Van der Vaart and Wellner 1996, Theorem 1.4.8, p. 32] (Weak Convergence on Product Spaces) *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a complete probability space,* $(\mathbb{A}_i, \mathcal{T}_{\mathbb{A}_i})$ *be a topological space for every* i *on a countable set* $\mathbb{I}$ *and* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *be the product space* $\mathbb{A} = \times_{i\in\mathbb{I}} \mathbb{A}_i$ *with product topology* $\mathcal{T}_{\mathbb{A}_i}$. *Let* $\{X_T(\omega)\}_{T\in\mathbb{N}}$ *with* $X_T : \Omega \to \mathbb{A}$ *be a sequence in* $\mathbb{A}$ *for every* $\omega \in \Omega$ *and* $X : \Omega \to \mathbb{A}$ *be a separable random element. Then* X_T *converges weakly to* X *if and only if* $(X_{T,i_1}, ..., X_{T,i_n})$ *converges weakly to* $(X_{i_1}, ..., X_{i_n})$ *for every* $n \in \mathbb{N}$.

Definition 16. (Separable Process) *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a complete probability space,* $\mathbb{A}$ *be a separable set,* $(\mathbb{B}, \mathfrak{B})$ *be a measurable space, and* $X : \Omega \times \mathbb{A} \to \mathbb{B}$ *be an* $\mathcal{F}/\mathfrak{B}$*-measurable element, a stochastic process.* X *is said to be separable with respect to* $\mathbb{A}'$ *if* $\mathbb{A}'$ *is a countable dense subset of* $\mathbb{A}$, *and there is a measure-zero set* $\Omega^* \subset \Omega$, $\mathbb{P}(\Omega^*) = 0$, *such that for every* $\omega \notin \Omega^*$, $X(\omega, \cdot)$ *is almost surely* $\mathbb{A}'$*-separable.*

Definition 17. (Separable Map) *Let* $(\mathbb{A}, \delta_{\mathbb{A}})$ *and* $(\mathbb{B}, \delta_{\mathbb{B}})$ *be metric spaces, and* $\mathbb{A}$ *be separable. Let* $\mathbb{A}'$ *be a countable, dense subset of* $\mathbb{A}$. *A function* $f : \mathbb{A} \to \mathbb{B}$ *is* $\mathbb{A}'$*-separable, or separable with respect to* $\mathbb{A}'$, *if* $\forall\, a \in \mathbb{A}$, *there exists a sequence* $a_i \in \mathbb{A}'$ *such that* $a_i \to \mathbb{A}$ *and* $f(a_i) \to f(a)$.

Remark 3. *We cannot easily guarantee that a process is separable. We can however "turn" a non-separable process, into a separable process with the same finite-dimensional distributions (Lemma 29).*

Lemma 29. [Gusak et al. 2010, Theorem 3.2, p.22] (Separable Modification) *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a complete probability space,* $(\mathbb{A}, \delta_{\mathbb{A}})$ *be a separable metric space,* $(\mathbb{B}, \delta_{\mathbb{B}})$ *a compact metric space, and* $X : \Omega \times \mathbb{A} \to \mathbb{B}$ *a stochastic process. Then there exists a separable version* $\tilde{X} : \Omega \times \mathbb{A} \to \mathbb{B}$ *of* X. *This is called a separable modification of* X.

Definition 18. (Stochastic Process Versioning) *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a complete probability space,* $\mathbb{A}$ *be a set and* $(\mathbb{B}, \mathfrak{B})$ *be a measurable space. Two stochastic processes* $X : \Omega \times \mathbb{A} \to \mathbb{B}$ *and* $Y : \Omega \times \mathbb{A} \to \mathbb{B}$ *are said to be versions of one another if* $\forall\, a \in \mathbb{A}$, $\mathbb{P}(\omega : X(\omega, a) = Y(\omega, a)) = 1$.

Remark 4. *If stochastic processes* X *and* Y *are versions of one another, they have the same finite-dimensional distributions.*

Remark 5. *In Lemma 29, if* $\mathbb{B}$ *is not compact, there still exists a separable version of* X *in some compactification* $\tilde{\mathbb{B}}$ *of* $\mathbb{B}$.

Definition 19. (Compact Topological Space) *A set* $\mathbb{A}'$ *in a topological space* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *is compact if every covering of* $\mathbb{A}'$ *by open sets contains a finite sub-cover.* $\mathbb{A}$ *is a compact space if it is itself a compact set.*

Definition 20. (Covering Number) *Let* $(\mathbb{A}, \delta_{\mathbb{A}})$ *be a metric space. The* ϵ*-covering number of* $\mathbb{A}' \subseteq \mathbb{A}$ *is the smallest number of open balls of radius* ϵ *required to cover* $\mathbb{A}'$.

Definition 21. (Totally Bounded Metric Space) *A metric space* $(\mathbb{A}, \delta_{\mathbb{A}})$ *is called totally bounded if and only if for every* $\epsilon > 0$ *there is a finite set* $\mathbb{A}_F \subseteq \mathbb{A}$ *such that for every* $a \in \mathbb{A}$ *there exists some* $a_F \in \mathbb{A}_F$ *such that* $\delta_{\mathbb{A}}(a, a_F) < \epsilon$.

Lemma 30. [Van der Vaart and Wellner 1996, Lemma 1.3.2, p.17] (Tightness and Separability) *Let* $(\mathbb{A}, \delta_{\mathbb{A}})$ *be a metric space. A Borel probability measure on* $(\mathbb{A}, \delta_{\mathbb{A}})$ *is pre-tight if and only if it is separable. Let* $(\mathbb{A}, \delta_{\mathbb{A}})$ *be a complete metric space. For a Borel probability measure on* $(\mathbb{A}, \delta_{\mathbb{A}})$, *separability, pre-tightness and tightness are equivalent. Any Polish Borel probability measure is tight.*

Proposition 4. (Degenerate Weak Convergence Implies Convergence in Probability) *Let* $(\Omega, \mathfrak{F}, \mathbb{P})$ *be a complete probability space,* $(\mathbb{A}, \delta_{\mathbb{A}})$ *be a measurable metric space,* $\{X_T\}_{T \in \mathbb{N}}$ *be a sequence of* $\mathbb{A}$*-valued random variables* $X_T : \Omega \to \mathbb{A}$, *and* $X : \Omega \to \mathbb{A}$ *be some* $\mathbb{A}$*-valued random variable. Then* $X_T \xrightarrow{p} X$ *if* $X_T \xrightarrow{d} X$ *and* X *is degenerate.*

Proof. Immediate extension of the common result for real-valued random variables that can be found e.g. in Davidson (1994, Theorem 22.5, p.349) and Potscher and Prucha (2001, Theorem 10, p.209). □

Lemma 31. [Dudley 2002, Theorem 2.4.7, p.52] (Arzelà-Ascoli Theorem) *Let* $(\mathbb{A}, \delta_{\mathbb{A}})$ *be a compact metric space,* $(\mathbb{C}(\mathbb{A}), \delta_{\mathbb{C}}^{\sup})$ *be the space of rel-valued continuous functions defined on* $\mathbb{A}$ *with* sup*-norm. A subset* $\mathbb{C}' \subset \mathbb{C}(\mathbb{A})$ *is totally bounded if and only if it is uniformly equicontinuous.*

Lemma 32. Munkres 2000, Theorem 27.4, p.174(Weierstrass's Extreme Value Theorem) *Let* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *and* $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ *be topological spaces with* $\mathbb{Y}$ *an ordered set in the order topology and* $f : \mathbb{A} \to \mathbb{B}$ *be a continuous map. If* X *is compact, then there exists points* $(a', a'') in \mathbb{A} \times \mathbb{A}$ *such that* $f(a') \leq f(a) \leq f(a'')$ *for every* $a \in \mathbb{A}$.

Definition 22. (Frechet and Hadamard Derivatives) *Let* $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ *and* $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ *be topological vector spaces. Let map* $f : \mathbb{A}_f \subseteq \mathbb{A} \to \mathbb{B}$ *and consider for some* $(a_\nabla, a_0) \in \mathbb{A}_f \times \mathbb{A}$ *the limit of the sequence in* $\mathbb{B}$*,*

$$\nabla f(a_\nabla, a_0) := \lim_{t \to 0} \frac{f(a_\nabla + t a_0) + f(a_\nabla)}{t}.$$

If this limit exists, then it is called the first variation of f *at* $a_\nabla \in \mathbb{A}_f$ *in the direction of* $a_0 \in \mathbb{A}$*. If* $\nabla f(a_0, a)$ *exists for every direction* $a \in \mathbb{A}$ *and* $\nabla f(a_\nabla) : \mathbb{A} \to \mathbb{B}$ *is a linear operator, then* $\nabla f(a_\nabla) \in \mathbb{L}(\mathbb{A}, \mathbb{B})$ *is called the Gateaux derivative of* $f : \mathbb{A}_f \to \mathbb{B}$ *at* $a_\nabla \in \mathbb{A}_f$ *and* f *is said to be Gateaux differentiable at* a_∇*. In this case,* $\nabla f(a_\nabla, a_0)$ *is called the Gateaux derivative of* $f : \mathbb{A}_f \to \mathbb{B}$ *at* $a_\nabla \in \mathbb{A}_f$ *in the direction of* $a_0 \in \mathbb{A}$*. If the limit,*

$$\nabla f(a_\nabla, a_0) := \lim_{t_n \to 0} \frac{f(a_\nabla + t_n a_n) + f(a_\nabla)}{t_n}$$

exists and $\nabla f(a_0) \in \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ *for every sequence* $t_n \to 0$ *and* $a_n \to a_0$ *with* $a_\nabla + t a_n \in \mathbb{A}_f$*, then* $\nabla f(a_0)$ *is called the Hadamard derivative of* f *at* $\mathbb{A}_\nabla$ *and* f *is said to be Hadamard differentiable at* a_∇*. For a normed vector space* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$*, Hadamard differentiability can be shown to be equivalent to compact differentiability. In particular, a map* $f : \mathbb{A}_f \subseteq \mathbb{A} \to \mathbb{B}$ *is said to be Compactly differentiable at* $a_\nabla \in \mathbb{A}_f$ *if the limit,*

$$\nabla f(a_\nabla, a_0) := \lim_{t_n \to 0} \sup_{a_0 \in \mathbb{A}_0, a_\nabla + t_n a_0 \in \mathbb{A}_f} \frac{f(a_\nabla + t_n a_0) + f(a_\nabla)}{t_n}$$

exists uniformly over all compact subsets $\mathbb{A}_0$ *of* $\mathbb{A}$ *and* $\nabla f(a_0) : \mathbb{A} \to \mathbb{B}$ *is a unique (Lemma 33) bounded linear operator. If the same holds over all bounded subsets* $\mathbb{A}_0$ *of* $\mathbb{A}$*, then the map said to be Frechet differentiable at* a_∇ *and* $\nabla f(a_0)$ *is called the Frechet derivative of* f *at* a_∇*.*

Remark 6. *The following notational convention applies throughout. Unless explicitly stated otherwise,* $\nabla_{\mathbb{A}_0} f$ *denotes always the "derivative function of* f *tangentially to* $\mathbb{A}_0$*", a map* $\nabla_{\mathbb{A}_0} f : \mathbb{A}_\nabla \to \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ *from the set* $\mathbb{A}_\nabla$ *of points at which* f *is differentiable into the space of bounded linear functionals* $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$*. So* $\nabla_{\mathbb{A}_0} f(a_\nabla) \in \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ *denotes a bounded linear operator, an element of* $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$*,* *for any $a_\nabla \in \mathbb{A}_\nabla$. For conciseness $\nabla_{\mathbb{A}_0} f(a_\nabla)$ shall often be denoted $\nabla_{\mathbb{A}_0} f_{a_\nabla}$, so that $\nabla_{\mathbb{A}_0} f_{a_\nabla}(a_0)$ is thus a point in $\mathbb{B}$ for every $a_0 \in \mathbb{A}_0$, because the differentiation point a_∇ has been fixed. When appropriate, this map is also alternatively denoted in a more extensive form as $\nabla_{\mathbb{A}_0} f(a_\nabla, a_0)$. If the tangent set $\mathbb{A}_0$ coincides with the entire set $\mathbb{A}$, then it is omitted from the notation. Thus the derivative $\nabla f(a_\nabla)$, an element of $\mathbb{L}(\mathbb{A}, \mathbb{B})$, corresponds to the derivative of f at a_∇ tangentially to the entire A.*

Second-order derivative functions $\nabla^2_{\mathbb{A}_0} f : \mathbb{A}_\nabla \to \mathbb{L}^2(\mathbb{A}_0 \times \mathbb{A}_0, \mathbb{B})$ are always understood as a map for the from the set $\mathbb{A}_\nabla$ of points at which f is differentiable to the space of bounded bilinear operators $\mathbb{L}^2(\mathbb{A}_0 \times \mathbb{A}_0, \mathbb{B})$. One should thus note that while $\nabla^2_{\mathbb{A}_0} f(a_\nabla) \in \mathbb{L}^2(\mathbb{A}_0 \times \mathbb{A}_0, \mathbb{B})$ denotes a bilinear operator for every $a_\nabla \in \mathbb{A}_\nabla$, we have that $\nabla^2_{\mathbb{A}_0} f(a_\nabla, a_0) \in \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ denotes a linear operator, an element of $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$, for every $(a_\nabla, a_0) \in \mathbb{A}_\nabla \times \mathbb{A}_0$.

Lemma 33. [Luenberger 1997, Proposition 2, p. 173] (Derivative Uniqueness) *Let $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ and $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ be normed vector spaces and $f : \mathbb{A}_f \subseteq \mathbb{A} \to \mathbb{B}$ be Frechet (Hadamard) differentiable at $a_\nabla \in \mathbb{A}_f$ tangentially to $\mathbb{A}_0 \subseteq \mathbb{A}$. Then there exists a unique continuous linear map $\nabla_{\mathbb{A}_0} f(a_\nabla) : \mathbb{A}_0 \to \mathbb{B}$ satisfying the definition of Frechet (Hadamard) derivative above.*

Lemma 34. [Denkowski et al. 2003, Proposition 5.1.17, p.525](Identification of $\mathbb{L}$) *Let $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ and $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ be topological vector spaces and let the spaces $\mathbb{L}(\mathbb{A}, \mathbb{B})$ of bounded linear operators from $\mathbb{A}$ into $\mathbb{B}$ be equipped with the uniform norm. Then, the space $\mathbb{L}(\mathbb{A}, \mathbb{L}(\mathbb{A}, \mathbb{B}))$ is isometrically isomorphic to the space $\mathbb{L}(\mathbb{A} \times \mathbb{A}, \mathbb{B})$ of bounded bilinear operators from $\mathbb{A} \times \mathbb{A}$ into $\mathbb{B}$ with uniform norm.*

Lemma 35. (Continuity and Differentiability) [Cheney 2001, Theorem 3, P.117] *If f is differentiable at x then it is continuous at x. (Cheney, Theorem 3, Analysis for applied mathematics, p 117.)*

Definition 23. (Schauder Basis) *Let $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ be a Banach space. A sequence $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{A}$ is a Schauder basis of $\mathbb{A}$ if for every $a \in \mathbb{A}$ there is a unique sequence of scalars $\{r_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ such that $a = \lim_{N\to\infty} \sum_n^N r_n a_n$.*

The following is thus an immediate corollary of Definition 23

Corollary 16. (Linear Independence of Schauder Basis) *Any finite collection of elements of the Schauder basis of a vector space consists of a set of linearly independent vectors.*

Remark 7. *Examples of spaces with Schauder basis are: The standard bases of C^0 and L_p for $1 \le p < \infty$ are Schauder bases. Every orthonormal basis in a separable Hilbert space is a Schauder basis. The Haar system is an example of a basis for $L_p(0,1)$ with $1 \le p < \infty$. The Banach space $C([0,1])$ of continuous functions on the interval $[0,1]$, with the supremum norm, admits a Schauder basis. A Banach space with a Schauder basis is necessarily separable, but the converse is false, as described below. Every Banach space with a Schauder basis has the approximation property.*

A question of Banach asked whether every separable Banach space has a Schauder basis; this was negatively answered by Per Enflo who constructed a separable Banach space without a Schauder basis.

However, a theorem of Mazur asserts that every infinite-dimensional Banach space has an infinite-dimensional subspace that has a Schauder basis.

Definition 24. (Operator Norm) *Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ *and* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *be normed vector spaces and* $\mathbb{L}(\mathbb{A},\mathbb{B})$ *denote the space of bounded linear operators form* $\mathbb{A}$ *into* $\mathbb{B}$*. The operator norm* $\|\cdot\|_{\mathbb{L}_{\mathbb{A}}^{\mathbb{B}}}$ *on* $\mathbb{L}(\mathbb{A},\mathbb{B})$ *is defined alternatively as*

$$\|f\|_{\mathbb{L}_{\mathbb{A}}^{\mathbb{B}}} := \sup_{a\in\mathbb{A}} \frac{\|f(a)\|_{\mathbb{B}}}{\|a\|_{\mathbb{A}}} \quad \text{or} \quad \|f\|_{\mathbb{L}_{\mathbb{A}}^{\mathbb{B}}} := \sup_{a\in\mathbb{A}:\|a\|_{\mathbb{A}}\leq 1} \|f(a)\|_{\mathbb{B}} \quad \text{for every} \quad f \in \mathbb{L}(\mathbb{A},\mathbb{B}).$$

Lemma 36. [Dudley 2002, Theorem 6.1.3, p.191] (Complete Dual Normed Vector Space) *For any normed vector space* $(\mathbb{A}, \|\cdot\|)$ *over* $\mathbb{R}$ *the dual space with operator norm* $(\mathbb{L}(\mathbb{A},\mathbb{B}), \|\cdot\|_{\mathbb{L}_{\mathbb{A}}^{\mathbb{B}}})$ *is a Banach space.*

Proposition 5. (Generalized Fermat's Stationary Points Theorem) *Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ *and* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *be normed vector spaces and let* $f : \mathbb{A} \to \mathbb{B}$ *be Hadamard differentiable at* $a_0 \in \text{int}(\mathbb{A})$ *tangentially to* $\mathbb{A}_0 \subseteq \mathbb{A}$*, with continuous linear Hadamard derivative denoted* $\nabla_{\mathbb{A}_0} f_{a_0} : \mathbb{A}_0 \to \mathbb{B}$*. Furthermore, suppose that* $\mathbb{B}$ *is a totally order set and let* a_0 *be a local minimizer of* f *on* $\mathbb{A}$*, i.e. there exists an open ball of radius* $\epsilon > 0$ *centered in* a_0*, denoted* $S_{a_0}(\epsilon) \subset \mathbb{A}$ *such that* $f(a_0) \leq f(a)\; \forall a \in S_{a_0}(\epsilon)$*. Then* $\nabla_{\mathbb{A}_0} f_{a_0}(a) = 0\; \forall a \in \mathcal{A}_0$*.*

Proof. Note first that, by definition $f(a) - f(a_0) \geq 0$ holds for every $a \in S_{a_0}(\epsilon)$. As a result, it is also true that,

$$\frac{f(a_0 + ta) - f(a_0)}{t} \geq 0 \;\; , \; \forall t : (a_0 + ta) \in S_{a_0}(\epsilon).$$

Now, by the definition of compact set and the Heine-Borel Theorem we have that $a_n = O(1)$ holds for every sequence $\{a_n\}_{n\in\mathbb{N}}$ on compact subsets of $\mathbb{A}_0$. Hence, any sequence $t_n \to 0$ satisfies $t_n a_n = o(1) \cdot O(1) = o(1)$. As a result, by continuity of norms and the continuous mapping theorem, we have that, for every $\epsilon > 0$, $\exists\, n \in \mathbb{N}$ such that $\|t_n a_n\|_{\mathbb{A}} < \epsilon$. Thus, for any $\epsilon > 0$, $\exists\, n \in \mathbb{N}$ such that $(a_0 + t_n a_n) \in S_{a_0}(\epsilon)$. This implies immediately that,

$$\nabla_{\mathbb{A}_0} f_{a_0}(a) = \lim_{t_n \to 0} \frac{f(a_0 + t_n a_n) - f(a_0)}{t} \geq 0$$

holds for every sequence $a_n \to a \in \mathbb{A}_0$ on compact subsets of $\mathbb{A}_0$. Finally, linearity of $\nabla_{\mathbb{A}_0} f_{a_0} : \mathbb{A}_0 \to \mathbb{B}$ implies that $\nabla_{\mathbb{A}_0} f_{a_0}(a) \geq 0$ and that $\nabla_{\mathbb{A}_0} f_{a_0}(-a) = -\nabla_{\mathbb{A}_0} f_{a_0}(a) \geq 0 \;\; \forall a \in \mathbb{A}_0$. This is only possible if $\nabla_{\mathbb{A}_0} f_{a_0}(a) = 0\; \forall a \in \mathbb{A}_0$. ☐

Proposition 6. (Linear Coordinate Projections) *Let* $\mathbb{B}_i$ *be a vector space for every* i *in some countable set* $\mathbb{I}$ *and* $\mathbb{B} := \times_{i\in\mathbb{I}} \mathbb{B}_i$ *be the associated product space. Then the coordinate projections* $\pi_i : \mathbb{B} \to \mathbb{B}_i$ *are linear for every* $i \in \mathbb{I}$*.*

Proof. Immediate since given a scalar $c \in \mathbb{R}$ and a vectors $b = (b_1, b_2, ...) \in \mathbb{B}$, the ith projection π_i satisfies $\pi_i(c \cdot b) = \pi_i((c \cdot b_1, c \cdot b_2, ...)) = c \cdot b_i$ and $c \cdot \pi_i(b) = c\dot{b}_i$ and thus $\pi_i(c \cdot b) = c \cdot \pi_i(b) \; \forall (c, b, i) \in \mathbb{R} \times \mathbb{B} \times \mathbb{I}$. Furthermore, given a pair of vectors $(b, b') \in \mathbb{B} \times \mathbb{B}$, $\pi_i(b+b') = \pi_i((b_1+b'_1, b_2+b'_2, ...)) = b_i + b'_i$ and $\pi_i(b) + \pi_i(b') = b_i + b'_i$ and thus $\pi_i(b + b') = \pi_i(b) + \pi_i(b') \; \forall (b, b', i) \in \mathbb{B} \times \mathbb{B} \times \mathbb{I}$. ☐

Lemma 37. [Winitzki (2010, Statement 2, p.28)](Linear Composition) *Let $\mathbb{A}$, $\mathbb{B}$ and $\mathbb{C}$ be vector spaces and $f : \mathbb{A} \to \mathbb{B}$ and $g : \mathbb{B} \to \mathbb{C}$ be linear maps. Then the composition map $h := g \circ f : \mathbb{A} \to \mathbb{C}$ is linear.*

Proposition 7. (Differentiability with Product Topology) *Let $(\mathbb{A}, \mathcal{T}_{\mathbb{A}})$ and $(\mathbb{B}_i, \mathcal{T}_{\mathbb{B}_i})$ be topological vector spaces for every i in some countable index set $\mathbb{I}$ and $(\mathbb{B}, \mathcal{T}_{\mathbb{B}})$ be the product space $\mathbb{B} = \times_{i \in \mathbb{I}} \mathbb{B}_i$ with product topology $\mathcal{T}_{\mathbb{B}}$. Then, a map $f : \mathbb{A}_f \subseteq \mathbb{A} \to \mathbb{B}$ is Frechet (Hadamard) differentiable at a point $a_\nabla \in \mathbb{A}_f$ tangentially to $\mathbb{A}_0 \subseteq \mathbb{A}$ if and only if the coordinate projection $\pi_i f : \mathbb{A}_f \to \mathbb{B}_i$ is also Frechet (Hadamard) differentiable at $a_\nabla \in \mathbb{A}_f$ tangentially to $\mathbb{A}_0 \in \mathbb{A}$ for every $i \in \mathbb{I}$.*

Proof. By definition, f is Frechet (Hadamard) differentiable at $a_\nabla \in \mathbb{A}_f$ tangentially to $\mathbb{A}_0 \subseteq \mathbb{A}$ if and only if there exists a continuous linear functional $\nabla_{\mathbb{A}_0} f(a_\nabla) : \mathbb{A}_0 \to \mathbb{B}$ such that, every sequence $\big\{b_n(t_n, a_n)\big\}_{T \in \mathbb{N}} \subset \mathbb{B}$ defined as,

$$b_n(t_n, a_n) := \frac{f(a_\nabla + t_n a_n) - f(a_\nabla) - t_n \nabla_{\mathbb{A}_0} f(a_\nabla, a_0)}{t_n}$$

converges to zero, for every $t_n \to 0$ and $a_n \to a_0 \in \mathbb{A}_0$ with $a_\nabla + t_n a_n \in \mathbb{A}_f \; \forall n \in \mathbb{N}$. Now, by Corollary 11 given the product topology $\mathcal{T}_{\mathbb{B}}$ on $\mathbb{B}$, convergence of the sequence $b_n(t_n, a_n) \to 0$ on the product space $\mathbb{B}$ occurs if and only if its coordinate projections $\pi_i b_n(t_n, a_n)$ also converge $\pi_i b_n(t_n, a_n) \to 0$ in $\mathbb{B}_i$ for every $i \in \mathbb{I}$. By linearity of the coordinate projection (Proposition 6) we then have that $b_n(t_n, a_n) \to 0$ if and only if,

$$\begin{aligned} \pi_i b_n(t_n, a_n) : &= \pi_i \left(\frac{f(a_\nabla + t_n a_n) - f(a_\nabla) - t_n \nabla_{\mathbb{A}_0} f(a_\nabla, a_0)}{t_n} \right) \\ &= \frac{\pi_i f(a_\nabla + t_n a_n) - \pi_i f(a_\nabla) - t_n \pi_i \nabla_{\mathbb{A}_0} f(a_\nabla, a_0)}{t_n} \to 0 \text{ for every } i \in \mathbb{I}. \end{aligned} \tag{9}$$

Finally, since by Lemma 37 a composition of linear maps is linear, and by lemma 15 a composition of continuous maps is continuous, $\pi_i \circ \nabla_{\mathbb{A}_0} f(a_\nabla)$ is a continuous linear map on $\mathbb{A}_0$. This implies, by definition, that the convergence in (9) above holds if and only if $\pi_i f$ is Frechet (Hadamard) differentiable for every $i \in \mathbb{I}$. Hence, the complete argument goes as follows: *(i)* f is Frechet (Hadamard) at a_∇ if and only if every sequence $b_n(t_n, a_n) \to 0$; *(ii)* every sequence $b_n(t_n, a_n) \to 0$ if and only if every sequence $\pi_i b_n(t_n, a_n) \to 0 \; \forall i \in \mathbb{I}$, and; *(iii)* every $\pi_i b_n(t_n, a_n) \to 0 \; \forall i \in \mathbb{I}$ if and only if every $\pi_i f$ is Frechet (Hadamard) at a_∇. We thus conclude that f is Frechet (Hadamard) at a_∇ if and only if $\pi_i f$ is Frechet (Hadamard) at a_∇. ☐

The following corollary follows immediately by continuity of continuous compositions (Lemma 15), Proposition 7 above, and the fact that $\pi_i \nabla_{\mathbb{A}_0} f(a_\nabla, a_0) = \nabla_{\mathbb{A}_0} \pi_i f(a_\nabla, a_0)$.

Corollary 17. (Continuous Differentiability with Product Topology) *Let $(\mathbb{A}, \mathcal{T}_\mathbb{A})$ and $(\mathbb{B}_i, \mathcal{T}_{\mathbb{B}_i})$ be topological vector spaces for every i in some countable index set $\mathbb{I}$ and $(\mathbb{B}, \mathcal{T}_\mathbb{B})$ be the product space $\mathbb{B} = \times_{i \in \mathbb{I}} \mathbb{B}_i$ with product topology $\mathcal{T}_\mathbb{B}$. Then, a map $f : \mathbb{A}_f \subseteq \mathbb{A} \to \mathbb{B}$ is continuously Frechet (Hadamard) at every point of $\mathbb{A}_\nabla \subseteq \mathbb{A}_f \subseteq \mathbb{A}$ tangentially to $\mathbb{A}_0 \subseteq \mathbb{A}$ if and only if the coordinate projection $\pi_i f : \mathbb{A}_f \to \mathbb{B}_i$ is also continuously Frechet (Hadamard) differentiable at every point of $\mathbb{A}_\nabla$ tangentially to $\mathbb{A}_0$ for every $i \in \mathbb{I}$.*

Lemma 38. (Chain Rule) *Let $(\mathbb{A}, \mathcal{T}_\mathbb{A})$, $(\mathbb{B}, \mathcal{T}_\mathbb{B})$ and $(\mathbb{C}, \mathcal{T}_\mathbb{C})$ be topological vector spaces. Let $f : \mathbb{A}_f \subseteq \mathbb{A} \to \mathbb{B}$ be Frechet (Hadamard) differentiable at every point of $\mathbb{A}_\nabla \subseteq \mathbb{A}_f$ tangentially to $\mathbb{A}_0$ and let $g : \mathbb{B}_g \subseteq \mathbb{B} \to \mathbb{C}$ be Frechet (Hadamard) differentiable at $\mathbb{B}_\nabla := f(\mathbb{A}_\nabla) \subseteq \mathbb{B}_g$ tangentially to $\mathbb{B}_0 := \nabla_{\mathbb{A}_0} f(\mathbb{A}_\nabla, \mathbb{A}_0)$. Then, $g \circ f : \mathbb{A}_f \to \mathbb{C}$ is differentiable at every point of $\mathbb{A}_\nabla$ tangentially to $\mathbb{A}_0$ with derivative $\nabla_{\mathbb{B}_0} g\big(f(a_\nabla), \nabla_{\mathbb{A}_0} f(a_\nabla)\big)$.*

Proposition 8. (Twice Differentiable Compositions) *Let $(\mathbb{A}, \mathcal{T}_\mathbb{A})$, $(\mathbb{B}, \mathcal{T}_\mathbb{B})$, $(\mathbb{C}, \mathcal{T}_\mathbb{C})$, $\big(\mathbb{L}(\mathbb{A}_0, \mathbb{B}), \mathcal{T}_{\mathbb{L}^{\mathbb{B}}_{\mathbb{A}_0}}\big)$ and $\big(\mathbb{L}(\mathbb{B}_0, \mathbb{C}), \mathcal{T}_{\mathbb{L}^{\mathbb{C}}_{\mathbb{B}_0}}\big)$ be topological vector spaces where $\mathbb{L}(\mathbb{A}_0, \mathbb{B})$ and $\mathbb{L}(\mathbb{B}_0, \mathbb{C})$ denote the spaces of bounded linear operators from $\mathbb{A}_0 \subseteq \mathbb{A}$ into $\mathbb{B}$ and $\mathbb{B}_0 \subseteq \mathbb{B}$ into $\mathbb{C}$ respectively. Let $f : \mathbb{A}_f \subseteq \mathbb{A} \to \mathbb{B}$ be Frechet (Hadamard) differentiable at every point of $\mathbb{A}_\nabla \subseteq \mathbb{A}_f$ tangentially to $\mathbb{A}_0$, with derivative at $a_\nabla \in \mathbb{A}_\nabla$ in the direction of $a \in \mathbb{A}_0$ denoted $\nabla_{\mathbb{A}_0} f(a_\nabla, a)$, and $g : \mathbb{B}_g \subseteq \mathbb{B} \to \mathbb{C}$ be Frechet (Hadamard) differentiable at every point of $\mathbb{B}_\nabla := f(\mathbb{A}_\nabla)$ tangentially to $\mathbb{B}_0 := \nabla_{\mathbb{A}_0} f(\mathbb{A}_\nabla, \mathbb{A}_0)$. Then, the composition map $h := g \circ f : \mathbb{A}_f \to \mathbb{C}$ is Frechet (Hadamard) differentiable at every point of $\mathbb{A}_\nabla$ tangentially to $\mathbb{A}_0$. If furthermore, the derivative function $\nabla_{\mathbb{A}_0} f : \mathbb{A}_\nabla \to \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ is Frechet (Hadamard) differentiable at every point of its domain $\mathbb{A}_\nabla$ tangentially to $\mathbb{A}_0$, and the map $\nabla_{\mathbb{B}_0} g : \mathbb{B}_\nabla \times \mathbb{L}_\nabla(\mathbb{A}_0, \mathbb{B}) \to \mathbb{L}(\mathbb{A}_0, \mathbb{C})$ is Frechet (Hadamard) differentiable at every point of its domain $\mathbb{B}_\nabla \times \mathbb{L}_\nabla(\mathbb{A}_0, \mathbb{B}) := f(\mathbb{A}_\nabla) \times \nabla_{\mathbb{A}_0} f(\mathbb{A}_\nabla) \subseteq \mathbb{B} \times \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ tangentially to $\mathbb{B}_0 \times \mathbb{L}_0(\mathbb{A}_0, \mathbb{B}) := \nabla_{\mathbb{A}_0} f(\mathbb{A}_\nabla, \mathbb{A}_0) \times \nabla^2_{\mathbb{A}_0} f(\mathbb{A}_\nabla, \mathbb{A}_0) \subseteq \mathbb{B} \times \mathbb{L}(\mathbb{A}_0, \mathbb{B})$, then the derivative function $\nabla_{\mathbb{A}_0} h := \nabla_{\mathbb{B}_0} g(f, \nabla_{\mathbb{A}_0} f) : \mathbb{A}_f \to \mathbb{L}(\mathbb{A}_0, \mathbb{C})$ is also Frechet (Hadamard) differentiable at every point of $\mathbb{A}_\nabla$ tangentially to $\mathbb{A}_0$.*

Proof. Differentiability of $h : \mathbb{A} \to \mathbb{C}$ on $\mathbb{A}_\nabla$ tangentially to $\mathbb{A}_0$ follows immediately from the chain-rule Lemma 38. For completeness, the proof goes as follows. For every $a_\nabla \in \mathbb{A}_\nabla$ and every sequence $t_n \to 0$ and $a_n \to a_0 \in \mathbb{A}_0$ as $n \to \infty$ with $(a_0 + a_n t_n) \in \mathbb{A}_f$ $\forall n \in \mathbb{N}$ it holds true that,

$$\frac{h(a_\nabla + t_n a_n) - h(a_0)}{t_n} = \frac{g\big(f(a_\nabla + t_n a_n)\big) - g\big(f(a_\nabla)\big)}{t_n} = \frac{g\big(b_\nabla + b_n t_n)\big) - g\big(b_\nabla\big)}{t_n} \tag{10}$$

where

$$b_\nabla = f(a_\nabla) \quad \text{and} \quad b_n := \frac{f(a_\nabla + t_n a_n) - f(a_\nabla)}{t_n} \to b_0 := \nabla_{\mathbb{A}_0} f(a_\nabla, a_0) \quad \forall t_n \to 0 \text{ and } a_n \to a_0 \in \mathbb{A}_0, \tag{11}$$

with the convergence $b_n \to b_0 \in \mathbb{B}_0$ being implied by differentiability of $f : \mathbb{A}_f \to \mathbb{B}$ at $a_\nabla \in \mathbb{A}_\nabla$ tangentially to $\mathbb{A}_0$. As a result, by differentiability of g at $b_\nabla \in \mathbb{B}_\nabla$ tangentially to $\mathbb{B}_0$,

$$\frac{g\big(b_\nabla + t_n b_n)\big) - g\big(b_\nabla\big)}{t_n} \to \nabla_{\mathbb{B}_0} g(b_\nabla, b_0) = \nabla_{\mathbb{B}_0} g\big(f(a_\nabla), \nabla_{\mathbb{A}_0} f(a_\nabla, a_0)\big) \quad \forall t_n \to 0 \text{ and } b_n \to b_0 \in \mathbb{B}_0$$

By (10) this implies the desired result that,

$$\frac{h(a_\nabla + t_n a_n) - h(a_\nabla)}{t_n} \to \nabla_{\mathbb{A}_0} h(a_\nabla, a_0) \quad \forall t_n \to 0 \text{ and } a_n \to a_0 \in \mathbb{A}_0.$$

Now, differentiability of the derivative function $\nabla_{\mathbb{A}_0} h := \nabla_{\mathbb{B}_0} g(f, \nabla_{\mathbb{A}_0} f) : \mathbb{A}_f \to \mathbb{L}(\mathbb{A}_0, \mathbb{C})$ at every point of $\mathbb{A}_\nabla$ tangentially to $\mathbb{A}_0$ follows by a similar argument. In particular, differentiability of $\nabla_{\mathbb{A}_0} f : \mathbb{A}_\nabla \to \mathbb{L}(\mathbb{A}_0, \mathbb{B})$ at every point of $\mathbb{A}_\nabla$ tangentially to $\mathbb{A}_0$ implies convergence of the sequence $\{L_n\}_{n\in\mathbb{N}} \subset \mathbb{L}(\mathbb{A}_0, \mathbb{B})$, defined below, to a point $L_0 \in \mathbb{L}_0(\mathbb{A}_0, \mathbb{B}) := \nabla^2_{\mathbb{A}_0} f(\mathbb{A}_\nabla, \mathbb{A}_0) \subseteq \mathbb{L}(\mathbb{A}_0, \mathbb{B})$,

$$L_n := \frac{\nabla_{\mathbb{A}_0} f(a_\nabla + t_n a_n) - \nabla_{\mathbb{A}_0} f(a_\nabla)}{t_n} \to L_0 := \nabla^2_{\mathbb{A}_0} f(a_\nabla, a_0) \quad \forall\ t_n \to 0 \text{ and } a_n \to a_0 \in \mathbb{A}_0.$$

Furthermore, differentiability of $\nabla_{\mathbb{B}_0} g : \mathbb{B}_\nabla \times \mathbb{L}_\nabla(\mathbb{A}_0, \mathbb{B}) \to \mathbb{L}(\mathbb{A}_0, \mathbb{C})$ at every point of its domain $\mathbb{B}_\nabla \times \mathbb{L}_\nabla(\mathbb{A}_0, \mathbb{B}) := f(\mathbb{A}_\nabla) \times \nabla_{\mathbb{A}_0} f(\mathbb{A}_\nabla)$ tangentially to $\mathbb{B}_0 \times \mathbb{L}_0(\mathbb{A}_0, \mathbb{B}) := \nabla_{\mathbb{A}_0} f(\mathbb{A}_\nabla, \mathbb{A}_0) \times \nabla^2_{\mathbb{A}_0} f(\mathbb{A}_\nabla, \mathbb{A}_0)$ implies that,

$$\frac{\nabla_{\mathbb{B}_0} g\Big((b_\nabla, L_\nabla) + t_n(b_n, L_n)\Big) - \nabla_{\mathbb{B}_0} g\Big((b_\nabla, L_\nabla)\Big)}{t_n} \to \nabla^2_{\mathbb{B}_0} g\Big((b_\nabla, L_\nabla)\,,\,(b_0, L_0)\Big)$$

for every sequence $t_n \to 0$ and $(b_n, L_n) \to (b_0, L_0) \in \mathbb{B}_0 \times \mathbb{L}_0(\mathbb{A}_0, \mathbb{B})$ with $(b_\nabla, L_\nabla) + t_n(b_n, L_n) \in \mathbb{B}_\nabla \times \mathbb{L}_\nabla(\mathbb{A}_0, \mathbb{B})\ \forall n \in \mathbb{N}$. The desired result now follows by noting precisely that,

$$\begin{aligned}
&\frac{\nabla_{\mathbb{A}_0} h(a_\nabla + t_n a_n) - \nabla_{\mathbb{A}_0} h(a_\nabla, \cdot)}{t_n} \\
&\qquad = \frac{\nabla_{\mathbb{B}_0} g\Big(f(a_\nabla + t_n a_n), \nabla_{\mathbb{A}_0} f(a_\nabla + t_n a_n)\Big) - \nabla_{\mathbb{B}_0} g\Big(f(a_\nabla), \nabla_{\mathbb{A}_0} f(a_\nabla)\Big)}{t_n} \\
&\qquad\qquad = \frac{\nabla_{\mathbb{B}_0} g\Big((b_\nabla, L_\nabla) + t_n(b_n, L_n)\Big) - \nabla_{\mathbb{B}_0} g\Big((b_\nabla, L_\nabla)\Big)}{t_n},
\end{aligned}$$

so that $\nabla_{\mathbb{A}_0} h := \nabla_{\mathbb{B}_0} g(f, \nabla_{\mathbb{A}_0} f) : \mathbb{A}_f \to \mathbb{L}(\mathbb{A}_0, \mathbb{C})$ is differentiable with derivative,

$$\nabla^2_{\mathbb{A}_0} h = \nabla^2_{\mathbb{B}_0} g\Big(\big(f, \nabla_{\mathbb{A}_0} f\big)\ ,\ \big(\nabla_{\mathbb{A}_0} f, \nabla^2_{\mathbb{A}_0} f\big)\Big) : \mathbb{A}_f \to \mathbb{L}^2(\mathbb{A}_0 \times \mathbb{A}_0, \mathbb{C}).$$

□

Lemma 39. [Luenberger 1997, Theorem 1, p.240] (Generalized Inverse Function Theorem) *Let* $(\mathbb{A}, \|\cdot\|_\mathbb{A})$ *and* $(\mathbb{B}, \|\cdot\|_\mathbb{B})$ *be Banach spaces and* $f : \mathbb{A}_f \to \mathbb{B}$ *be continuously Frechet differentiable on an open set* $\mathbb{A}_\nabla \subset \mathbb{A}_f$. *Let* $a_\nabla \in \mathbb{A}_\nabla$ *be a regular point of* f, *then* f *is invertible with derivative* f^{-1} *defined on an* ϵ*-ball* $S(b_\nabla, \epsilon)$ *centered at* $b_\nabla = f(a_\nabla)$ *for some* $\epsilon > 0$ *and there exists* $c > 0$ *such that* $\|a - a_\nabla\|_\mathbb{A} \le c\|f(a) - f(a_\nabla)\|$

In what follows, Corrolary 18 below states an immediate implication of the definition of Hadamard differentiability encountered above. Next, Proposition 9 establishes trivially that $\nabla_{\mathbb{A}_0} f(a_0, \cdot)$ is simply a restriction of $\nabla_{\mathbb{A}_0} f(a_0, \cdot)$ onto $\mathbb{A}_0$.

Corollary 18. (Differentiability Sub-Tangentially) *Let $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ and $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ be normed vector spaces. Let $f : \mathbb{A} \to \mathbb{B}$ be Hadamard differentiable at $a_0 \in \mathbb{A}$ tangentially to $\mathbb{A}_0 \subseteq \mathbb{A}$. Then, f is also Hadamard differentiable at $a_0 \in \mathbb{A}$ tangentially to $\mathbb{A}_1 \subseteq \mathbb{A}_0$.*

Proposition 9. (Equivalence of Sub-Tangential Derivative) *Let $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ and $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ be normed vector spaces and $f : \mathbb{A} \to \mathbb{B}$ be Hadamard differentiable at $a_0 \in \mathbb{A}$ tangentially to $\mathbb{A}_0 \subseteq \mathbb{A}$. Let $\nabla_{\mathbb{A}_0} f(a_0, \cdot) : \mathbb{A}_0 \to \mathbb{B}$ and $\nabla_{\mathbb{A}_1} f(a_1, \cdot) : \mathbb{A}_1 \to \mathbb{B}$ denote the Hadamard derivatives of $f : \mathbb{A} \to \mathbb{B}$ at $a_0 \in \mathbb{A}$ and $a_1 \in \mathbb{A}$ tangentially to $\mathbb{A}_0$ and $\mathbb{A}_1$ respectively with $\mathbb{A}_1 \subseteq \mathbb{A}_0 \subseteq \mathbb{A}$. Then $a_1 = a_0$ implies $\nabla_{\mathbb{A}_1} f(a_1, a) = \nabla_{\mathbb{A}_0} f(a_0, a)\ \forall a \in \mathbb{A}_1$ and furthermore, there exists a continuous linear extension of $\nabla_{\mathbb{A}_1} f(a_1, \cdot)$ denoted $\nabla^{\mathbb{A}_0}_{\mathbb{A}_1} f(a_1, \cdot) : \mathbb{A}_0 \to \mathbb{B}$ such that $a_1 = a_0$ implies $\nabla^{\mathbb{A}_0}_{\mathbb{A}_1} f(a_1, a) = \nabla_{\mathbb{A}_0} f(a_0, a)\ \forall a \in \mathbb{A}_0$.*

Proof. The proof is trivial. Existence of $\nabla_{\mathbb{A}_1} f(a_1, \cdot) : \mathbb{A}_1 \to \mathbb{B}$ is guaranteed by Proposition 18. The first claim, that $a_1 = a_0$ implies $\nabla_{\mathbb{A}_1} f(a_1, a) = \nabla_{\mathbb{A}_0} f(a_0, a)\ \forall a \in \mathbb{A}_1$, follows immediately from Lemma 33. The second claim, that there exists an extension of $\nabla_{\mathbb{A}_1} f(a_1, \cdot)$ denoted $\nabla^{\mathbb{A}_0}_{\mathbb{A}_1} f(a_1, \cdot) : \mathbb{A}_0 \to \mathbb{B}$ such that $a_1 = a_0$ implies $\nabla^{\mathbb{A}_0}_{\mathbb{A}_1} f(a_1, a) = \nabla_{\mathbb{A}_0} f(a_0, a)\ \forall a \in \mathbb{A}_0$, follows from the first claim and by setting $\nabla^{\mathbb{A}_0}_{\mathbb{A}_1} f(a_0, a) = \nabla_{\mathbb{A}_0} f(a_0, a)\ \forall a \in \mathbb{A}_0 \backslash \mathbb{A}_1$. Finally, the implicit claim that $\nabla^{\mathbb{A}_0}_{\mathbb{A}_1} f(a_1, \cdot) : \mathbb{A}_0 \to \mathbb{B}$ is a continuous linear map on $\mathbb{A}_0$ is a straightforward implication of the fact that $\nabla^{\mathbb{A}_0}_{\mathbb{A}_1} f(a_0, a) = \nabla_{\mathbb{A}_0} f(a_0, a)\ \forall a \in \mathbb{A}_0$ and that $\nabla_{\mathbb{A}_0} f(a_0, \cdot)$ is linear and continuous by construction. □

Since a_0 and a_1 above were arbitrary elements of $\mathbb{A}$. It follows also immediately that $\nabla_{\mathbb{A}_1} f$ is simply a restriction of $\nabla_{\mathbb{A}_0} f$ onto $\mathbb{A} \times \mathbb{A}_0$. For the sake of completeness, this trivial implication is recorded as a corollary below.

Corollary 19. (Equivalence of Sub-Tangential Derivative) *Let $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ and $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ be normed vector spaces and $f : \mathbb{A} \to \mathbb{B}$ be Hadamard differentiable at $a_0 \in \mathbb{A}$ tangentially to $\mathbb{A}_0 \subseteq \mathbb{A}$. Then the maps $\nabla_{\mathbb{A}_0} f : \mathbb{A} \times \mathbb{A}_0 \to \mathbb{B}$ and $\nabla_{\mathbb{A}_1} f : \mathbb{A} \times \mathbb{A}_1 \to \mathbb{B}$ satisfy $\nabla_{\mathbb{A}_1} f(a, a') = \nabla_{\mathbb{A}_0} f(a, a')\ \forall (a, a') \in \mathbb{A} \times \mathbb{A}_1$ and furthermore, there exists an extension of $\nabla_{\mathbb{A}_1} f$ denoted $\nabla^{\mathbb{A}_0}_{\mathbb{A}_1} f : \mathbb{A} \times \mathbb{A}_0 \to \mathbb{B}$ such that $\nabla^{\mathbb{A}_0}_{\mathbb{A}_1} f(a, a') = \nabla_{\mathbb{A}_0} f(a, a')\ \forall (a, a') \in \mathbb{A} \times \mathbb{A}_0$.*

Proposition 10. (Convergence of Hadamard Derivatives) *Let $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ and $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ be normed vector spaces and let $\{\mathbb{A}_T\}_{T \in \mathbb{N}}$ be a sequence of compact subsets of $\mathbb{A}$ satisfying $\mathbb{A}_T \subseteq \mathbb{A}_{T+1} \subseteq \mathbb{A}_0 \subseteq \mathbb{A}$ with $\mathbb{A}_\infty := \bigcup_{T \in \mathbb{N}} \mathbb{A}_T$ such that $\mathrm{cl}(\mathbb{A}_\infty) \supseteq \mathbb{A}_0$. Let $f : \mathbb{A} \to \mathbb{B}$ be continuously Hamadamard differentiable on an open ball of radius $\epsilon > 0$ centered at a_0 and denoted $S_{a_0}(\epsilon) \subseteq \mathbb{A}$ tangentially to $\mathbb{A}_0$. Finally, let $\{a^0_T\}_{T \in \mathbb{N}}$ be a sequence in $\mathbb{A}$ satisfying $a_T \in \mathbb{A}_T\ \forall T \in \mathbb{N}$ and $a^0_T \to a_0$. Then $\nabla_{\mathbb{A}_T} f(a^0_T, a_T) \to \nabla_{\mathbb{A}_0} f(a_0, a)$ for every sequence $\{a_T\}_{T \in \mathbb{N}}$ satisfying $a_T \in \mathbb{A}_T\ \forall T \in \mathbb{N}$ and $a_T \to a \in \mathbb{A}_0$.*

Proof. Choose T^* large enough so that $a_T^0 \in S_{a_0}(\epsilon)\ \forall T > T^*$. Then, by assumption, f is Hadamard differentibale at a_T^0 tagentially to $\mathbb{A}_0$ for every $T > T^*$. By Corrolary 18, since $\mathbb{A}_T \subseteq \mathbb{A}_0\ \forall T \inf \mathbb{N}$, it follows that f is also Hadamard differentibale at $\boldsymbol{\theta}_T^0$ tagentially to $\mathbb{A}_T$ for every $T > T^*$, and hence $\nabla_{\mathbb{A}_T} f$ is a well defined for every $T > T^*$. Now,

$$\begin{aligned}\left|\nabla_{\mathbb{A}_T} f(a_T^0, a_T) - \nabla_{\mathbb{A}_0} f(a_0, a)\right| \leq & \left|\nabla_{\mathbb{A}_T} f(a_T^0, a_T) - \nabla_{\mathbb{A}_0} f(a_T^0, a_T)\right| \\ & + \left|\nabla_{\mathbb{A}_0} f(a_T^0, a_T) - \nabla_{\mathbb{A}_0} f(a_0, a_T)\right| \\ & + \left|\nabla_{\mathbb{A}_0} f(a_0, a_T) - \nabla_{\mathbb{A}_0} f(a_0, a)\right|\end{aligned}$$

and hence $\left|\nabla_{\mathbb{A}_T} f(a_T^0, a_T) - \nabla_{\mathbb{A}_0} f(a_0, a)\right| \to 0$ is deduced from the fact that *(i)* by Corrolary 19, $\left|\nabla_{\mathbb{A}_T} f(a_T^0, a_T) - \nabla_{\mathbb{A}_0} f(a_T^0, a_T)\right| = 0\ \forall (a_T^0, a_T) \in \mathbb{A}_T$ and every $T > T^*$; *(ii)* $\left|\nabla_{\mathbb{A}_0} f(a_0, a_T) - \nabla_{\mathbb{A}_0} f(a_0, a)\right| \to 0$ by continuity of $\nabla_{\mathbb{A}_0} f(a_0, a_T)$, convergence $a_T \to a$ and the continuous mapping theorem; and finally *(iii)* $\left|\nabla_{\mathbb{A}_0} f(a_T^0, a_T) - \nabla_{\mathbb{A}_0} f(a_0, a_T)\right| \to 0$ by uniform convergence of $\nabla_{\mathbb{A}_0} f(a_T^0, \cdot)$ to $\nabla_{\mathbb{A}_0} f(a_0, \cdot)$ on $\mathbb{A}_0$ which follows from the following argument. Define $g_T(a) = \nabla_{\mathbb{A}_0} f(a_T^0, a)\ \forall a \in \mathbb{A}_0$ and $g(a) = \nabla_{\mathbb{A}_0} f(a_0, a)\ \forall a \in \mathbb{A}_0$. Note that $\{g_T\}$ is thus a sequence of Bounded linear operators on $\mathbb{A}_0$. By the continuous mapping theorem, $g_T(a) \to g(a)$ pointwise for every $a \in \mathbb{A}_0$. Finally, since $\mathbb{A}_T$ compact for every $T \in \mathbb{N}$, by Proposition ... $g_T(a) \to g(a)$ uniformly on $\mathbb{A}_0$ and the desired result is obtained. ☐

Lemma 40. (Injective Linear Operator) *Let $T : V \to W$ be a linear map. Then T is injective if and only if* $\mathrm{Ker}(T) = \{0\}$.

Lemma 41. (Bounded Inverse) [Kolmogorov and Fomin 1975, Theorem 2, p.229] *Let f be an invertible bounded linear operator mapping from a Banach space $\mathbb{A}$ onto another Banach space $\mathbb{B}$. Then the inverse operator f^{-1} is itself bounded.*

Lemma 42. (Bounded Linear Operator) [Sviridyuk and Fedorov 2003, Theorem 1.1.1, p.3] *Let an operator $f : \mathbb{A} \to \mathbb{B}$ be linear. Then the following statements are equivalent:* (i) *the operator f is continuous at one point;* (ii) *the operator f is continuous;* (iii) *the operator f is bounded.*

Lemma 43. (Inverse of Linear Operator) [Kolmogorov and Fomin 1975, Theorem 1, p. 228, Luenberger 1997, Proposition 1, p.174] *The inverse of a linear operator between topological vector spaces is itself linear.*

The following Corrolary follows immediately from Lemmas 41, 42 and 43.

Corollary 20. (Continuous Inverse) *Let f be an invertible continuous linear operator. Then f^{-1} is continuous.*

Definition 25. (Continuously Invertible Operator) *Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ *and* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *be normed vector spaces. A bounded linear operator* $f \in \mathbb{L}(\mathbb{A}, \mathbb{B})$ *is said to be* continuously invertible *if its inverse is an operator* $f^{-1} \in \mathbb{L}(\mathbb{B}, \mathbb{A})$*, i.e. if it is defined and bounded on the range of* f*. In other words, let* $f \in \mathbb{L}(\mathbb{A}, \mathbb{B})$*,* $f(\mathbb{A}) = \mathbb{B}$ *and* f *be invertible. Then, the operator* f *is continuously invertible.*

Lemma 44. (Continuous Invertibility and Bounded Inverse) [Sviridyuk and Fedorov 2003] *Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ *and* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *be Banach spaces and* $f : \mathbb{A}_f \subseteq \mathbb{A} \to \mathbb{B}$ *be a bounded linear operator from* $\mathbb{A}_f$ *into* $\mathbb{B}$*, i.e.* $f \in \mathbb{L}(\mathbb{A}_f, \mathbb{B})$*. Then, the inverse operator* $f^{-1} : f(\mathbb{A}_f) \to \mathbb{A}_f$ *exists and is bounded on* $f(\mathbb{A}_f)$ *if and only if there exists* $m \in \mathbb{R}_+$ *such that* $\|f(a)\|_{\mathbb{B}} \geq m\|a\|_{\mathbb{A}} \;\; \forall\, a \in \mathbb{A}_f$*.*

Definition 26. (Uniformly Continuously Invertible Operator) *Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ *and* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *be Banach spaces and* $f_i \in \mathbb{L}(\mathbb{A}, \mathbb{B})$ *for every* $i \in \mathbb{I}$*. Building on Lemma 44, the family* $f : \mathbb{I} \times \mathbb{A} \to \mathbb{B}$ *is said to be continuously invertible uniformly in* $i \in \mathbb{I}$ *if and only if* $f_i : \mathbb{L}(\mathbb{A}, \mathbb{B})$ *is continuously invertible for every* $i \in \mathbb{I}$ *and there exists* $m \in \mathbb{R}_+$ *such that* $\|f_i(a)\|_{\mathbb{B}} \geq m\|a\|_{\mathbb{A}} \;\; \forall\,(a, i) \in \mathbb{A} \times \mathbb{I}$*.*

Proposition 11. (Continuously Invertible Composition)) *Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$*,* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *and* $(\mathbb{C}, \|\cdot\|_{\mathbb{C}})$ *be Banach spaces and* $f \in \mathbb{L}(\mathbb{A}, \mathbb{B})$ *and* $g \in \mathbb{L}(\mathbb{B}, \mathbb{C})$ *be continuously invertible bounded linear operators. Then* $h := g \circ f : \mathbb{A} \to \mathbb{C}$ *is also a continuously invertible bounded linear operator.*

Proof. By Lemma 37, $h := g \circ f$ is a linear map. By Lemma 44, $\exists\,(m_f, m_g) \in \mathbb{R}_+ \times \mathbb{R}_+$ such that $\|f(a)\|_{\mathbb{B}} \geq m_f\|a\|_{\mathbb{A}} \;\; \forall\, a \in \mathbb{A}$ and $\|g(b)\|_{\mathbb{C}} \geq m_g\|b\|_{\mathbb{B}} \;\; \forall\, b \in \mathbb{B}$. It thus follows that $\|h(a)\|_{\mathbb{C}} = \|g(f(a))\|_{\mathbb{C}} \geq m_g\|f(a)\|_{\mathbb{B}} \geq m_g m_f\|a\|_{\mathbb{A}} = m_h\|a\|_{\mathbb{A}} \;\; \forall\, a \in \mathbb{A}$ and again by Lemma 44 that h is continuously invertible. □

Proposition 12. (Uniformly Continuously Invertible Composition)) *Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$*,* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *and* $(\mathbb{C}, \|\cdot\|_{\mathbb{C}})$ *be Banach spaces. Furthermore, let* $f_i \in \mathbb{L}(\mathbb{A}, \mathbb{B})$ *be continuously invertible uniformly in* $i \in \mathbb{I}$ *and* $g \in \mathbb{L}(\mathbb{B}, \mathbb{C})$ *be continuously invertible. Then* $h_i := g \circ f_i : \mathbb{A} \to \mathbb{C}$ *is a bounded linear operator continuously invertible in* $i \in \mathbb{I}$*.*

Proof. By Lemma 37, $h_i := g \circ f_i$ is a linear map for every $i \in \mathbb{I}$. By Lemma 44 and Definition 26, $\exists\,(m_f, m_g) \in \mathbb{R}_+ \times \mathbb{R}_+$ such that $\|f_i(a)\|_{\mathbb{B}} \geq m_f\|a\|_{\mathbb{A}} \;\; \forall\,(a, i) \in \mathbb{A} \times \mathbb{I}$ and $\|g(b)\|_{\mathbb{C}} \geq m_g\|b\|_{\mathbb{B}} \;\; \forall\, b \in \mathbb{B}$. It thus follows that $\|h_i(a)\|_{\mathbb{C}} = \|g(f_i(a))\|_{\mathbb{C}} \geq m_g\|f_i(a)\|_{\mathbb{B}} \geq m_g m_f\|a\|_{\mathbb{A}} = m_h\|a\|_{\mathbb{A}} \;\; \forall\,(a, i) \in \mathbb{A} \times \mathbb{I}$ and again by Lemma 44 that h_i is continuously invertible uniformly in $i \in \mathbb{I}$. □

Lemma 45. (Banach Space of Bounded Linear Operators) [Sviridyuk and Fedorov 2003] *Let* $\mathbb{B}$ *be a Banach space. Then the space* $\mathbb{L}(\mathbb{A}, \mathbb{B})$ *is a Banach space when equipped with the supremum norm.*

Lemma 46. (Linearity of Pointwise Limit of Sequence of Linear Functions) [Denkowski et al. 2003, Proposition 3.2.3, p.267] *If* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ *is a Banach space,* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *is a normed vector space,* $\{f_n\}_{n\in\mathbb{N}} \subseteq \mathbb{L}(\mathbb{A}, \mathbb{B})$ *and for every* $a \in \mathbb{A}$*,* $f(a) = \lim f_n(a)$ *exists in* $\mathbb{B}$*, then* $f \in \mathbb{L}(\mathbb{A}, \mathbb{B})$*.*

Proposition 13. *(Uniform Bound on Bounded Linear Operators) Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ *a normed vector space and* $\{\mathbb{A}_T\}_{T\in\mathbb{N}}$ *be a sequence of compact subsets of* $\mathbb{A}$ *satisfying* $\mathbb{A}_T \subseteq \mathbb{A}_{T+1} \subseteq \mathbb{A}\ \forall T \in \mathbb{N}$. *Furthermore, let* $f_T : \mathbb{A}_T \to \mathbb{R}$ *be a continuous linear map satisfying* $|f_T(a)| \geq c_T\|a\|_{\mathbb{A}}\ \forall a \in \mathbb{A}_T$ *for every* $T \in \mathbb{N}$, *and let* $f : \mathbb{A} \to \mathbb{R}$ *be also a continuous linear map satisfying* $f(a) \geq c\|a\|_{\mathbb{A}} \forall a \in \mathbb{A}$. *Finally, suppose that* $f_T(a_T) \to f(a)$ *for every sequence* $a_T \to a \in \mathbb{A}$ *with* $a_T \in \mathbb{A}_T\ \forall T \in \mathbb{N}$. *Then* $\exists\, c^* > 0$ *and* $T^* \in \mathbb{N}$ *such that, for every* $T > T^*$, $|f_T(a_T)| \geq c^*\|a_T\|_{\mathbb{A}}$ *for every sequence* $a_T \to a \in \mathbb{A}$ *with* $a_T \in \mathbb{A}_T\ \forall T \in \mathbb{N}$.

Proof. Set $c^* = c/2$. Suppose by contradiction that $\nexists\, T^* \in \mathbb{N}$ such that $|f_T(a_T)| \geq c^*\|a_T\|_{\mathbb{A}}\ \forall T > T^*$ and every $a_T \to a$. Then for every $T^* \in \mathbb{N}$ we can always find a sequence $\{a_T\}_{T\in\mathbb{N}}$ and some $T > T^*$ such that $|f_T(a_T)| < c^*\|a_T\|_{\mathbb{A}}$. Together with the fact that $f(a) \geq c\|a\|_{\mathbb{A}} \forall a \in \mathbb{A}$, this implies that, for every $T^* \in \mathbb{N}$ there exists a sequence $\{a_T\}_{T\in\mathbb{N}}$ and some $T > T^*$ such that,

$$\begin{aligned} |f_T(a_T) - f(a_T)| &\geq |c^*\|a_T\|_{\mathbb{A}} - c\|a_T\|_{\mathbb{A}}| \\ &= |(c/2)\|a_T\|_{\mathbb{A}} - c\|a_T\|_{\mathbb{A}}| \\ &= (c/2)\|a_T\|_{\mathbb{A}}. \end{aligned}$$

But this contradicts the Assumption that $f_T(a_T) \to f(a)$ for every sequence $a_T \to a \in \mathbb{A}$ with $a_T \in \mathbb{A}_T\ \forall T \in \mathbb{N}$ except when $a_T \to 0$. □

Definition 27. *(Partial Frechet/Hadamard Differentiability Uniformly on a Parameter) Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$, $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *and* $(\mathbb{C}, \|\cdot\|_{\mathbb{C}})$ *be normed vector spaces. A map* $f : \mathbb{A} \times \mathbb{B} \to \mathbb{C}$ *is said to be Hadamard differentiable at* $a_0 \in \mathbb{A}$ *tangentially to* $\mathbb{A}_0 \subseteq \mathbb{A}$ *uniformly in* $b \in \mathbb{B}$ *if there exists a continuous map* $\nabla_{\mathbb{A}_0} f_{a_0} : \mathbb{A}_0 \times \mathbb{B} \to \mathbb{C}$, *linear in the first argument, such that,*

$$\frac{f(a_0 + t_n h_n, b_n) - f(a_0, b_0)}{t_n} \to \nabla_{\mathbb{A}_0} f_{a_0}(h_0, b_0)$$

holds for every sequence $t_n \to 0$, $a_n \to a_0$ *and* $b_n \to b_0$ *with* $a_0 + t_n h_n \in \mathbb{A}$ *and* $b_n \in \mathbb{B}$ *for all* $n \in \mathbb{N}$.

Remark 8. *In general, the requirement of existence of a map* $\nabla_{\mathbb{A}_0} f_{a_0}$ *satisfying the conditions in Definition 27 is more or less restrictive depending on the convergence rate of the sequence* $\{b_n\}_{n\in\mathbb{N}}$ *to its limit* b_0. *If* $\{b_n\}_{n\in\mathbb{N}}$ *converges at a fast rate then there is a large class of functions satisfying the property of partial Frechet/Hadamard differentiability uniformly on* $\mathbb{B}$. *If* $\{b_n\}_{n\in\mathbb{N}}$ *converges slowly then fewer maps satisfy this definition. Some characterization of the family of maps satisfying this property is thus useful. Proposition 14 below shows that bilinear maps are partially differentiable uniformly on a parameter regardless of the converge rate of* $\{b_n\}_{n\in\mathbb{N}}$.

Proposition 14. (Partial Differentiability of Continuous Bilinear Operator) *Let* $\mathbb{A}_1$, $\mathbb{A}_2$ *and* $\mathbb{B}$ *be vector spaces and* $f : \mathbb{A}_1 \times \mathbb{A}_2 \to \mathbb{B}$ *be a continuous bilinear map. Then* $f(\cdot, a)$ *is partially Hadamard differentiable at* $a_0 \in \mathbb{A}$ *uniformly in* $a \in \mathbb{A}$. *In particular,*

$$\frac{f(a_0 + t_n a_n, \bar{a}_n) - f(a_0, \bar{a}_n)}{t_n} \to \nabla f_{a_0}(a, \bar{a})$$

for every $t_n \to 0$, $a_n \to a \in \mathbb{A}_1$ *and* $\bar{a}_n \to \bar{a} \in \mathbb{A}_2$ *with* $(t_n + a_n)$ *in compact subsets of* $\mathbb{A}_1$.

Proof. Immediate by setting $\nabla f_{a_0}(a,\bar{a}) = f(a,\bar{a})$, in which case,

$$\begin{aligned}\frac{f(a_0 + t_n a_n, \bar{a}_n) - f(a_0, \bar{a}_n) - t_n f(a,\bar{a})}{t_n} &= \frac{f(a_0,\bar{a}_n) + f(t_n a_n, \bar{a}_n) - f(a_0,\bar{a}_n) - t_n f(a,\bar{a})}{t_n} \\ &= \frac{t_n f(a_n,\bar{a}_n) - t_n f(a,\bar{a})}{t_n} \\ &= f(a_n,\bar{a}_n) - f(a,\bar{a}) \to 0\end{aligned}$$

where the first and second equalities involve linearity of f in the first argument and the last step follows by continuity and the continuous mapping theorem. □

Lemma 47. (Delta Method) [Van der Vaart and Wellner 1996, Theorem 3.9.4, p.374] *Let* $\mathbb{A}$ *and* $\mathbb{B}$ *be metrizable topological vector spaces. Let* $f : \mathbb{A}_f \subset \mathbb{A} \to \mathbb{B}$ *be Hadamard differentiable at* a_0 *tangentially to* $\mathbb{A}_0$*. Let* $X_n : \Omega_n \to \mathbb{A}_f$ *be maps with* $r_n(X_n - a_0) \xrightarrow{d} X$ *for some constants* $r_n \to \infty$*, where* X *is separable and takes its values in* $\mathbb{A}_0$*. Then* $r_n(f(X_n) - f(a_0)) \xrightarrow{d} f'_{a_0}(X)$*. If* f'_{a_0} *is defined and continuous on the whole of* $\mathbb{A}$ *then the sequence* $r_n(f(X_n) - f(a_0)) - f'_\theta(r_n(X_n - a_0))$ *converges to zero in outer probability.*

Lemma 48. (Banach-Steinhaus Theorem) [Dudely, Theorem 6.5.1, p.212] *Let* $(\mathbb{A}, \|\cdot\|_{\mathbb{A}})$ *be a Banach space and* $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ *a normed vector space. Let* $f_T : \mathbb{A} \to \mathbb{B}$ *be a bounded linear operator for every* $T \in \mathbb{N}$*. If* $\sup_{T\in\mathbb{N}} \|f_T(a)\|_{\mathbb{B}} < \infty \; \forall a \in \mathbb{A}$ *then* $\sup_{T\in\mathbb{N}} \|f_T\| < \infty$ *in operator norm* $\|\cdot\|$.

B Proofs

B.1 Proof of Theorem 1

Proof. Clearly, the $\mathcal{F}/\mathfrak{B}(\mathcal{B}_L)$-measurability of each auxiliary estimators $\tilde{\boldsymbol{\beta}}^L_{T,s}(\cdot,\boldsymbol{\theta}) : \Omega \to \mathcal{B}_L$ for every $(\boldsymbol{\theta}, T, s, L) \in \Theta \times \mathbb{N} \times \{1, ..., S\} \times \mathcal{L}$, postulated in Assumption 4, implies the $\mathcal{F}/\mathfrak{B}(\mathcal{B}_L)$-measurability of the average $\tilde{\boldsymbol{\beta}}^L_{T,S}(\cdot,\boldsymbol{\theta}) : \Omega \to \mathcal{B}_L$ obtained as $\tilde{\boldsymbol{\beta}}^L_{T,S}(\cdot,\boldsymbol{\theta}) = 1/S \sum_{s=1}^{S} \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta})$, for every $(\boldsymbol{\theta}, T, S, L) \in \Theta \times \mathbb{N} \times \mathbb{N} \times \mathcal{L}$ by the continuity of vector addition and scalar multiplication under Assumption 3 (see Definition 8) and measurability of continuous functions (Lemma 3 and Corollary 10). This implies in turn that, given Assumption 3, $\tilde{\boldsymbol{\beta}}_{T,S}(\cdot,\boldsymbol{\theta}) : \Omega \to \mathcal{B}$ is $\mathcal{F}/\mathfrak{B}(\mathcal{B})$-measurable $\forall\, (\boldsymbol{\theta}, T, S) \in \Theta \times \mathbb{N} \times \mathbb{N}$ (Lemma 6 and Corollary 13). By the same argument, the $\mathcal{F}/\mathfrak{B}(\mathcal{B}_L)$-measurability of the auxiliary estimators $\hat{\boldsymbol{\beta}}^L_T : \Omega \to \mathcal{B}_L \;\forall\, (T, L) \in \mathbb{N}\times\mathcal{L}$ implies the $\mathcal{F}/\mathfrak{B}(\mathcal{B})$-measurability of $\hat{\boldsymbol{\beta}}_T : \Omega \to \mathcal{B}_L \;\forall\, (T, L) \in \mathbb{N}$. Furthermore, Assumption 6 implies that $\mu : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ is $\mathfrak{B}(\mathcal{B})/\mathfrak{B}(\mathbb{R})$-measurable (Lemma 3 and Corollary 10), and hence, together with the measurability of $\tilde{\boldsymbol{\beta}}_{T,S}(\cdot,\boldsymbol{\theta}) : \Omega \to \mathcal{B} \;\forall\, (\boldsymbol{\theta}, T, S) \in \Theta\times\mathbb{N}\times\mathbb{N}$ and $\hat{\boldsymbol{\beta}}_T \;\forall T \in \mathbb{N}$, we have that $Q_{T,S}(\boldsymbol{\theta}) := \mu \circ \Big(\hat{\boldsymbol{\beta}}_T \,,\, \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\Big) : \Omega \to \mathbb{R}$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$-measurable for every $(\boldsymbol{\theta}, T, S) \in \Theta \times \mathbb{N} \times \mathbb{N}$ by measurability of measurable compositions (Lemma 5).

Now, Assumption 5 implies immediately the continuity the average map $\tilde{\boldsymbol{\beta}}^{L}_{T,S}(\omega,\cdot):\Theta\to\mathcal{B}_L$ on $\Theta\ \forall(\omega,T,S,L)\in\Omega\times\mathbb{N}\times\mathbb{N}\times\mathcal{L}$ (under Assumption 3, Definition 8 and Lemma 15). This in turn implies that (under Assumption 3) the continuity of $\tilde{\boldsymbol{\beta}}_{T,S}(\omega,\cdot):\Theta\to\mathcal{B}_L$ on $\Theta\ \forall(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N}$ (Lemma 7). Together with the continuity of μ postulated in Assumption 5 this implies the continuity of $Q_{T,S}(\omega,\cdot):=\mu\circ\Big(\hat{\boldsymbol{\beta}}_T(\omega)\ ,\ \tilde{\boldsymbol{\beta}}_{T,S}(\omega,\cdot)\Big):\Theta\to\mathbb{R}$ on Θ for every $(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N}$ (Lemma 15).

Finally, $\mathcal{F}/\mathfrak{B}(\mathbb{R})$-measurability of $Q_{T,S}(\boldsymbol{\theta}):\Omega\to\mathbb{R}$ for every $(\boldsymbol{\theta},T,S)\in\Theta\times\mathbb{N}\times\mathbb{N}$ and continuity of $Q_{T,S}(\omega,\cdot):\Theta\to\mathbb{R}$ on Θ for every $(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N}$ implies by Lemma 16 that $Q_{T,S}:\Omega\times\Theta\to\mathbb{R}$ is $\mathcal{F}\otimes\mathfrak{B}(\Theta)/\mathfrak{B}(\mathbb{R})$-measurable. Together with Assumptions 1 and 2 the desired result follows from Lemmas 17, 18 and Corollary 14 adapted from Debreu (1967, Theorem 4.5), Hildenbrand (1974, p.55) and White and Wooldrige (1991, Theorem 2.2, p.646), i.e. that there exists a $\hat{\boldsymbol{\theta}}_{T,S}:\Omega\to\Theta_T$ satisfying (3) for every $T\in\mathbb{N}$ and $S\in\mathbb{N}$ that is $\mathcal{F}/\mathfrak{B}(\Theta_T)$-measurable.[21] □

B.2 Proof of Theorem 2

Proof. The $\mathcal{F}/\mathfrak{B}(\mathcal{B})$-measurability of $\tilde{\boldsymbol{\beta}}_{T,S}(\cdot,\boldsymbol{\theta}):\Omega\to\mathcal{B}\ \forall(\boldsymbol{\theta},T,S)\in\Theta\times\mathbb{N}\times\mathbb{N}$ and $\hat{\boldsymbol{\beta}}_T:\Omega\to\mathcal{B}_L\ \forall(T,L)\in\mathbb{N}$ is obtained by the same argument as in the proof of Theorem 1. Assumption 7 implies that $\mu_T:\mathcal{B}\times\mathcal{B}\to\mathbb{R}$ is $\mathfrak{B}(\mathcal{B})/\mathfrak{B}(\mathbb{R})$-measurable (Corollary 10) for every $T\in\mathbb{N}$, and hence, together with the measurability of $\tilde{\boldsymbol{\beta}}_{T,S}(\cdot,\boldsymbol{\theta}):\Omega\to\mathcal{B}\ \forall(\boldsymbol{\theta},T,S)\in\Theta\times\mathbb{N}\times\mathbb{N}$ and $\hat{\boldsymbol{\beta}}_T\ \forall T\in\mathbb{N}$, we have that $Q_{T,S}(\boldsymbol{\theta}):=\mu_T\circ\Big(\hat{\boldsymbol{\beta}}_T\ ,\ \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\Big):\Omega\to\mathbb{R}$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$-measurable for every $(\boldsymbol{\theta},T,S)\in\Theta\times\mathbb{N}\times\mathbb{N}$ by measurability of measurable compositions (Lemma 5).

Now, the continuity of $\tilde{\boldsymbol{\beta}}_{T,S}(\omega,\cdot):\Theta\to\mathcal{B}_L$ on $\Theta\ \forall(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N}$ is again obtained by following the argument in the proof of Theorem 1. Together with the continuity of μ_T postulated in Assumption 7 this implies the continuity of $Q_{T,S}(\omega,\cdot):=\mu_T\circ\Big(\hat{\boldsymbol{\beta}}_T(\omega)\ ,\ \tilde{\boldsymbol{\beta}}_{T,S}(\omega,\cdot)\Big):\Theta\to\mathbb{R}$ on Θ for every $(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N}$.

Finally, $\mathcal{F}\otimes\mathfrak{B}(\Theta)/\mathfrak{B}(\mathbb{R})$-measurability of $Q_{T,S}:\Omega\times\Theta\to\mathbb{R}$ is by the same argument as in the proof of Theorem 1, and once more, together with Assumptions 1 and 2 the desired result follows, by Corollary 14, that there exists a $\hat{\boldsymbol{\theta}}_{T,S}:\Omega\to\Theta_T$ satisfying (3) for every $T\in\mathbb{N}$ and $S\in\mathbb{N}$ that is $\mathcal{F}/\mathfrak{B}(\Theta_T)$-measurable.

□

B.3 Proof of Theorem 3

Proof. First, note that given Assumptions 1, 3 and 10 the product binding function $\boldsymbol{\beta}^*:\Theta\to\mathcal{B}$ is a homeomorphism and thus injective, continuous and open (Proposition 2). By injectivity $\boldsymbol{\beta}^*_0=\boldsymbol{\beta}^*(\boldsymbol{\theta}_0)$

[21] Note how Lemmas 17 and 18 allow for random sieves to be considered. Note also that, in what near-measurability is concerned, completeness and separability of Θ could be weakened to the requirement that Θ be a Souslin measurable space; see Stinchcombe and White (1992).

and $\boldsymbol{\beta}_0^* \neq \boldsymbol{\beta}^*(\boldsymbol{\theta})\ \forall \boldsymbol{\theta} \in \Theta \backslash \{\boldsymbol{\theta}_0\}$, which implies, by the properties of divergences (Definition 13) that,

$$Q_\infty(\boldsymbol{\theta}_0) := \mu\Big(\boldsymbol{\beta}_0^*\,,\, \boldsymbol{\beta}^*(\boldsymbol{\theta}_0)\Big) = 0 \text{ and } Q_\infty(\boldsymbol{\theta}) := \mu\Big(\boldsymbol{\beta}_0^*\,,\, \boldsymbol{\beta}^*(\boldsymbol{\theta})\Big) > 0\ \forall \boldsymbol{\theta} \in \Theta \backslash \{\boldsymbol{\theta}_0\}. \tag{12}$$

By openness, $\boldsymbol{\beta}^*\Big(S_{\boldsymbol{\theta}_0}(\epsilon)\Big) \subset \mathcal{B}$ is open for every open ball $S_{\boldsymbol{\theta}_0}(\epsilon) \subset \Theta$, $\epsilon > 0$, and thus there exist an open ball $S_{\boldsymbol{\beta}_0}(\epsilon') \subset \mathcal{B}$ centered at $\boldsymbol{\beta}_0$ with radius $\epsilon' > 0$ such that,

$$S_{\boldsymbol{\beta}_0}(\epsilon') \subset \boldsymbol{\beta}^*\Big(S_{\boldsymbol{\theta}_0}(\epsilon)\Big) \subset \mathcal{B}.$$

It thus follows immediately that $\boldsymbol{\theta}_0$ is identifiably unique since, for every $\epsilon > 0$,

$$\begin{aligned}
\inf_{\boldsymbol{\theta} \in S^c_{\boldsymbol{\theta}_0}(\epsilon)} \Big| Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) \Big| &= \inf_{\boldsymbol{\theta} \in S^c_{\boldsymbol{\theta}_0}(\epsilon)} \Big| \mu\Big(\boldsymbol{\beta}_0^*\,,\, \boldsymbol{\beta}^*(\boldsymbol{\theta})\Big) - \mu\Big(\boldsymbol{\beta}_0^*\,,\, \boldsymbol{\beta}^*(\boldsymbol{\theta}_0)\Big) \Big| \\
&= \inf_{\boldsymbol{\theta} \in S^c_{\boldsymbol{\theta}_0}(\epsilon)} \Big| \mu\Big(\boldsymbol{\beta}_0^*\,,\, \boldsymbol{\beta}^*(\boldsymbol{\theta})\Big) \Big| = \inf_{\boldsymbol{\beta} \in \boldsymbol{\beta}^*\left(S^c_{\boldsymbol{\theta}_0}(\epsilon)\right)} \Big| \mu\big(\boldsymbol{\beta}_0^*\,,\, \boldsymbol{\beta}\big) \Big| \\
&\geq \inf_{\boldsymbol{\beta} \in S_{\boldsymbol{\beta}_0}(\epsilon') \subset \boldsymbol{\beta}^*\left(S^c_{\boldsymbol{\theta}_0}(\epsilon)\right)} \Big| \mu\big(\boldsymbol{\beta}_0^*\,,\, \boldsymbol{\beta}\big) \Big| > 0\ \ \forall \epsilon'\ :\ S_{\boldsymbol{\beta}_0}(\epsilon') \subseteq \boldsymbol{\beta}^*\big(S^c_{\boldsymbol{\theta}_0}(\epsilon)\big),
\end{aligned} \tag{13}$$

where the second equality follows by identity of inderscernibles of divergences (Definition 13) and the last inequality by Assumption 11.

Now, given the Lipschitz weakness of $\delta_{\mathcal{B}}$ postulated in Assumption 8, the uniform convergence of $\tilde{\boldsymbol{\beta}}^L_{T,s}$ over Θ (Assumption 9) implies the uniform convergence of the product empirical binding function $\boldsymbol{\beta}^*$ on Θ. Indeed, for every $\epsilon > 0$, it holds true that,

$$\begin{aligned}
\mathbb{P}\Big(\sup_{\boldsymbol{\theta} \in \Theta_T} \delta_{\mathcal{B}}\Big(\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\,,\, \boldsymbol{\beta}^*(\boldsymbol{\theta})\Big) > \epsilon \Big) &\leq \mathbb{P}\Big(\sup_{\boldsymbol{\theta} \in \Theta_T} k \cdot \sup_{L \in \mathcal{L}} \delta_{\mathcal{B}_L}\Big(\tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta})\,,\, \boldsymbol{\beta}^*_L(\boldsymbol{\theta})\Big) > \epsilon \Big) \\
&= \mathbb{P}\Big(k \cdot \sup_{\boldsymbol{\theta} \in \Theta_T} \sup_{L \in \mathcal{L}} \Big\| \tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta}) - \boldsymbol{\beta}^*_L(\boldsymbol{\theta}) \Big\|_{\mathcal{B}_L} > \epsilon \Big) \\
&= \mathbb{P}\Big(k \cdot \sup_{\boldsymbol{\theta} \in \Theta_T} \sup_{L \in \mathcal{L}} \Big\| 1/S \sum_{s=1}^{S} \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) - 1/S \sum_{s=1}^{S} \boldsymbol{\beta}^*_L(\boldsymbol{\theta}) \Big\|_{\mathcal{B}_L} > \epsilon \Big) \\
&\leq \mathbb{P}\Big(k \cdot \sup_{\boldsymbol{\theta} \in \Theta_T} \sup_{L \in \mathcal{L}} 1/S \sum_{s=1}^{S} \Big\| \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) - \boldsymbol{\beta}^*_L(\boldsymbol{\theta}) \Big\|_{\mathcal{B}_L} > \epsilon \Big) \\
&\leq \mathbb{P}\Big(k/S \sum_{s=1}^{S} \sup_{\boldsymbol{\theta} \in \Theta_T} \sup_{L \in \mathcal{L}} \Big\| \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) - \boldsymbol{\beta}^*_L(\boldsymbol{\theta}) \Big\|_{\mathcal{B}_L} > \epsilon \Big)
\end{aligned} \tag{14}$$

as $T \to \infty$ for every $S \in \mathbb{N}$ and some $k \in \mathbb{R}^+$, and where the first inequality follows by Assumption 8, the second by norm sub-additivity, and the third by supremum sub-additivity. Hence, by Assumption 3 the continuous mapping Theorem (Corollary 15, see also Definition 8 and note that a degenerate random variable is separable) and part *(ii)* of Assumption 9 we have that, for every $\epsilon > 0$,

$$\lim_{T \to \infty} \mathbb{P}\Big(k/S \sum_{s=1}^{S} \sup_{\boldsymbol{\theta} \in \Theta_T} \sup_{L \in \mathcal{L}} \Big\| \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) - \boldsymbol{\beta}^*_L(\boldsymbol{\theta}) \Big\|_{\mathcal{B}_L} > \epsilon \Big) = 0.$$

This implies by (14) and Lemma 24 that,

$$\lim_{T\to\infty} \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\Big(\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\,,\,\boldsymbol{\beta}^*(\boldsymbol{\theta})\Big) > \epsilon\Big) = 0 \ \ \forall \epsilon > 0. \tag{15}$$

For almost sure uniform convergence simply note that following the argument in (14) and Lemma 24,

$$\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\Big(\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\,,\,\boldsymbol{\beta}^*(\boldsymbol{\theta})\Big) \ \le\ \lim_{T\to\infty} k/S\sum_{s=1}^{S} \sup_{\boldsymbol{\theta}\in\Theta_T} \sup_{L\in\mathcal{L}} \Big\|\tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) - \boldsymbol{\beta}^*_L(\boldsymbol{\theta})\Big\|_{\mathcal{B}_L},$$

and hence, since by Assumption 3 the continuous mapping Theorem (Corollary 15) and part *(ii)* of Assumption 9,

$$\mathbb{P}\Big(\lim_{T\to\infty} k/S\sum_{s=1}^{S} \sup_{\boldsymbol{\theta}\in\Theta_T} \sup_{L\in\mathcal{L}} \Big\|\tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) - \boldsymbol{\beta}^*_L(\boldsymbol{\theta})\Big\|_{\mathcal{B}_L} > \epsilon\Big) = 0 \,\forall \epsilon > 0,$$

we have by Lemma 24,[22]

$$\mathbb{P}\Big(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\Big(\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\,,\,\boldsymbol{\beta}^*(\boldsymbol{\theta})\Big) > \epsilon\Big) = 0 \,\forall \epsilon > 0. \tag{16}$$

Convergence in probability and a.s. of $\hat{\boldsymbol{\beta}}_T$ is implied by Assumption 8, part *(i)* of Assumption 9 and Lemma 24 since it follows immediately from $\delta_{\mathcal{B}}\big(\hat{\boldsymbol{\beta}}_T\,,\,\boldsymbol{\beta}^*_0\big) \le k\cdot \sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}\big(\hat{\boldsymbol{\beta}}^L_T\,,\,\boldsymbol{\beta}^*_L(\boldsymbol{\theta}_0)\big)$ that,

$$\lim_{T\to\infty} \mathbb{P}\Big(\delta_{\mathcal{B}}\big(\hat{\boldsymbol{\beta}}_T\,,\,\boldsymbol{\beta}^*_0\big) > \epsilon\Big) \le \lim_{T\to\infty} \mathbb{P}\Big(k\cdot \sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}\big(\hat{\boldsymbol{\beta}}^L_T\,,\,\boldsymbol{\beta}^*_L(\boldsymbol{\theta}_0)\big) > \epsilon\Big) = 0 \ \ \forall \epsilon > 0, \tag{17}$$

and also that,

$$\mathbb{P}\Big(\lim_{T\to\infty} \delta_{\mathcal{B}}\big(\hat{\boldsymbol{\beta}}_T\,,\,\boldsymbol{\beta}^*_0\big) > \epsilon\Big) \le \mathbb{P}\Big(\lim_{T\to\infty} k\cdot \sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}\big(\hat{\boldsymbol{\beta}}^L_T\,,\,\boldsymbol{\beta}^*_L(\boldsymbol{\theta}_0)\big) > \epsilon\Big) = 0 \ \ \forall \epsilon > 0. \tag{18}$$

Uniform convergence in probability of the centered empirical binding function $\Delta_{T,S}(\boldsymbol{\theta}) := \hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ to $\Delta_\infty(\boldsymbol{\theta}) := \boldsymbol{\beta}^*(\boldsymbol{\theta}_0) - \boldsymbol{\beta}^*(\boldsymbol{\theta})$ across the sequence of sieves $\{\Theta_T\}_{T\in\mathbb{N}}$ now follows immediately

[22] It is also clear that under appropriate regularity conditions, the almost sure convergence of the product binding function $\boldsymbol{\beta}^*$ uniformly on Θ is also obtained directly by convergence of the projection maps under Assumption 3 and Corollary 11 without the need for Assumption 8.

from (15) and (17) since it holds true that,

$$\begin{aligned}
\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta})\,,\,\Delta_\infty(\boldsymbol{\theta})\big)>\epsilon\Big) &= \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\big\|\Delta_{T,S}(\boldsymbol{\theta})-\Delta_\infty(\boldsymbol{\theta})\big\|_{\mathcal{B}}>\epsilon\Big)\\
&= \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\big\|\hat{\boldsymbol{\beta}}_T-\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})-\boldsymbol{\beta}^*(\boldsymbol{\theta}_0)+\boldsymbol{\beta}^*(\boldsymbol{\theta})\big\|_{\mathcal{B}}>\epsilon\Big)\\
&\le \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\big\|\hat{\boldsymbol{\beta}}_T-\boldsymbol{\beta}^*(\boldsymbol{\theta}_0)\big\|_{\mathcal{B}}+\big\|\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})-\boldsymbol{\beta}^*(\boldsymbol{\theta})\big\|_{\mathcal{B}}>\epsilon\Big)\\
&\le \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\big\|\hat{\boldsymbol{\beta}}_T-\boldsymbol{\beta}^*(\boldsymbol{\theta}_0)\big\|_{\mathcal{B}}+\sup_{\boldsymbol{\theta}\in\Theta_T}\big\|\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})-\boldsymbol{\beta}^*(\boldsymbol{\theta})\big\|_{\mathcal{B}}>\epsilon\Big)\\
&= \mathbb{P}\Big(\big\|\hat{\boldsymbol{\beta}}_T-\boldsymbol{\beta}^*(\boldsymbol{\theta}_0)\big\|_{\mathcal{B}}+\sup_{\boldsymbol{\theta}\in\Theta_T}\big\|\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})-\boldsymbol{\beta}^*(\boldsymbol{\theta})\big\|_{\mathcal{B}}>\epsilon\Big)\\
&\le \mathbb{P}\Big(\big\|\hat{\boldsymbol{\beta}}_T-\boldsymbol{\beta}^*(\boldsymbol{\theta}_0)\big\|_{\mathcal{B}}>\epsilon/2\Big)\\
&\qquad+\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\big\|\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})-\boldsymbol{\beta}^*(\boldsymbol{\theta})\big\|_{\mathcal{B}}>\epsilon/2\Big)
\end{aligned}\tag{19}$$

where the first inequality follows from norm sub-additivity, the second by sub-additivity of the supremum, and the third by the fact that $\{a+b>\epsilon\}\subseteq\{a>\epsilon/2\}\cup\{b>\epsilon/2\}$ and that, for random events, this implies $\mathbb{P}(a+b>\epsilon)\le\mathbb{P}(a>\epsilon/2)+\mathbb{P}(b>\epsilon/2)$. Finally, by the convergence results obtained in 15 and 17, the last two terms converge to zero which implies by Lemma 24 the convergence of the centered empirical binding function,

$$\lim_{T\to\infty}\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta})\,,\,\Delta_\infty(\boldsymbol{\theta})\big)>\epsilon\Big)=0\;\;\forall\epsilon>0.\tag{20}$$

The almost sure counterpart of this result is obtained by following the same argument as in (19) to conclude that, for every $\epsilon>0$,

$$\begin{aligned}
\mathbb{P}\Big(\lim_{T\to\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta})\,,\,\Delta_\infty(\boldsymbol{\theta})\big)>\epsilon\Big) &\le \mathbb{P}\Big(\lim_{T\to\infty}\big\|\hat{\boldsymbol{\beta}}_T-\boldsymbol{\beta}^*(\boldsymbol{\theta}_0)\big\|_{\mathcal{B}}>\epsilon/2\Big)\\
&\qquad+\mathbb{P}\Big(\lim_{T\to\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\big\|\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})-\boldsymbol{\beta}^*(\boldsymbol{\theta})\big\|_{\mathcal{B}}>\epsilon/2\Big)
\end{aligned}$$

and thus obtain by the a.s. convergence results in (16) and (18),

$$\mathbb{P}\Big(\lim_{T\to\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta})\,,\,\Delta_\infty(\boldsymbol{\theta})\big)>\epsilon\Big)=0\;\;\forall\epsilon>0.\tag{21}$$

Now, uniform convergence across $\{\Theta_T\}_{T\in\mathbb{N}}$ of the composition criterion $Q_{T,S}(\boldsymbol{\theta}):=\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big)$ can be obtained from (20) under the uniform continuity of μ on $\mathcal{B}$ by making use of the fact that uniform continuous compositions preserve uniform convergence in probability and almost surely (Proposition 3).[23] If $\mu:\mathcal{B}\to\mathbb{R}$ is uniformly continuous on $\mathcal{B}$ then, $\forall\,\epsilon>0$, $\exists\,\epsilon'>0$ such that, $\forall\,(\boldsymbol{\beta},\boldsymbol{\beta}')\in\mathcal{B}\times\mathcal{B}$

[23]Given the continuity of μ postulated in Assumption 6, uniform continuity of μ on $\mathcal{B}$ can be obtained e.g. by compactness of $\mathcal{B}$ (compactness of every $\mathcal{B}_L$ under Tychonoff's topology) or Holder continuity of μ on a possibly non-compact $\mathcal{B}$.

satisfying $\delta_{\mathcal{B}}(\boldsymbol{\beta},\boldsymbol{\beta}')<\epsilon'$ we have that $|\mu(\boldsymbol{\beta})-\mu(\boldsymbol{\beta}')|<\epsilon$, and this implies that for every pair of maps $\boldsymbol{\beta}:\Theta\rightarrow\mathcal{B}$ and $\boldsymbol{\beta}':\Theta\rightarrow\mathcal{B}$ and every $\epsilon>0$, $\exists\epsilon'>0$ such that,

$$\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\Big(\boldsymbol{\beta}(\boldsymbol{\theta})-\boldsymbol{\beta}'(\boldsymbol{\theta})\Big)<\epsilon' \quad\Rightarrow\quad \sup_{\boldsymbol{\theta}\in\Theta_T}\Big|\mu\big(\boldsymbol{\beta}(\boldsymbol{\theta})\big)-\mu\big(\boldsymbol{\beta}'(\boldsymbol{\theta})\big)\Big|<\epsilon.$$

Now, since for every $(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N}$ and every $\forall\epsilon>0$, $\exists\epsilon'>0$ such that,

$$\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\big(\Delta_{T,S}(\omega,\boldsymbol{\theta})-\Delta_\infty(\boldsymbol{\theta})\big)<\epsilon' \;\Rightarrow\; \sup_{\boldsymbol{\theta}\in\Theta_T}\Big|\mu\big(\Delta_{T,S}(\omega,\boldsymbol{\theta})\big)-\mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big|<\epsilon, \tag{22}$$

it follows immediately that, $\forall(T,S)\in\mathbb{N}\times\mathbb{N}$, $\forall\epsilon>0$, $\exists\epsilon'>0$ such that,

$$\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta})-\Delta_\infty(\boldsymbol{\theta})\big)<\epsilon'\Big)\leq\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big)-\mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big|<\epsilon\Big).$$

because the first implies the second $\forall(\omega,T,S)\in\Omega\times\mathbb{N}\times\mathbb{N}$. Finally, since by (20),

$$\lim_{T\rightarrow\infty}\mathbb{P}(\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\Big(\Delta_{T,S}(\boldsymbol{\theta})-\Delta_\infty(\boldsymbol{\theta}))<\epsilon'\Big)=1\;\;\forall\epsilon'>0,$$

it follows immediately that $\lim_{T\rightarrow\infty}\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big)-\mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big|<\epsilon\Big)=1\,\forall\epsilon>0$, and hence, for every $S\in\mathbb{N}$,

$$\begin{aligned}\lim_{T\rightarrow\infty}\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|Q_{T,S}(\boldsymbol{\theta})-Q_\infty(\boldsymbol{\theta})\Big|>\epsilon\Big)&=1-\lim_{T\rightarrow\infty}\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|Q_{T,S}(\boldsymbol{\theta})-Q_\infty(\boldsymbol{\theta})\Big|<\epsilon\Big)\\&=1-\lim_{T\rightarrow\infty}\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big)-\mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big|<\epsilon\Big)=0\,\forall\epsilon>0.\end{aligned} \tag{23}$$

The almost sure counterpart of this result follows by the same argument. In particular by (22) for every $(\omega,S)\in\Omega\times\mathbb{N}$ and every $\epsilon>0$, $\exists\epsilon'>0$ such that,

$$\lim_{T\rightarrow\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\big(\Delta_{T,S}(\omega,\boldsymbol{\theta})-\Delta_\infty(\boldsymbol{\theta})\big)<\epsilon' \;\Rightarrow\; \lim_{T\rightarrow\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|\mu\big(\Delta_{T,S}(\omega,\boldsymbol{\theta})\big)-\mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big|<\epsilon,$$

and hence, for every $\epsilon>0$, $\exists\epsilon'>0$ such that,

$$\mathbb{P}\Big(\lim_{T\rightarrow\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta})-\Delta_\infty(\boldsymbol{\theta})\big)<\epsilon'\Big)\leq\mathbb{P}\Big(\lim_{T\rightarrow\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big)-\mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big|<\epsilon\Big).$$

because the first implies the second $\forall\omega\in\Omega$. Finally, since by (21),

$$\mathbb{P}(\lim_{T\rightarrow\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\Big(\Delta_{T,S}(\boldsymbol{\theta})-\Delta_\infty(\boldsymbol{\theta}))<\epsilon'\Big)=1\;\;\forall\epsilon'>0,$$

it follows immediately that $\mathbb{P}\Big(\lim_{T\rightarrow\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big)-\mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big|<\epsilon\Big)=1\,\forall\epsilon>0$, and hence,

$$\begin{aligned}\mathbb{P}\Big(\lim_{T\rightarrow\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|Q_{T,S}(\boldsymbol{\theta})-Q_\infty(\boldsymbol{\theta})\Big|>\epsilon\Big)&=1-\mathbb{P}\Big(\lim_{T\rightarrow\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|Q_{T,S}(\boldsymbol{\theta})-Q_\infty(\boldsymbol{\theta})\Big|<\epsilon\Big)\\&=1-\mathbb{P}\Big(\lim_{T\rightarrow\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big)-\mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big|<\epsilon\Big)=0\,\forall\epsilon>0.\end{aligned} \tag{24}$$

To see that the same result holds given only the continuity of μ on $\mathcal{B}$ (Assumption 6) and the uniform convergence of the centered empirical binding function $\sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\Big(\Delta_{T,S}(\boldsymbol{\theta}) - \Delta_\infty(\boldsymbol{\theta})\Big) \xrightarrow{p} 0$ obtained in probability in (20) and almost surely in (21), let $\mathcal{B}_0(\epsilon') \subset \mathcal{B}$ denote a closed ball of radius ϵ' centered at the zero element of $\mathcal{B}$, i.e. $\mathcal{B}_0(\epsilon') := \{\boldsymbol{\beta} \in \mathcal{B} : \delta_{\mathcal{B}}(\boldsymbol{\beta}) \leq \epsilon'\}$. By Assumption 6, the compactness of $\mathcal{B}_0(\epsilon')$ and the well known Heine–Cantor Theorem (Lemma 25), for every $(\omega, T, S) \in \Omega \times \mathbb{N} \times \mathbb{N}$ and every $\epsilon > 0$, $\exists \epsilon' > 0$ such that,

$$\Big\{\Delta_{T,S}(\omega, \Theta_T) - \Delta_\infty(\Theta_T)\Big\} \subset B_0(\epsilon') \quad \Rightarrow \quad \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu\big(\Delta_{T,S}(\omega, \boldsymbol{\theta})\big) - \mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| < \epsilon$$

This in turn implies that $\forall (T, S) \in \mathbb{N} \times \mathbb{N}$ and every $\epsilon > 0$, there exists $\epsilon' > 0$, such that,

$$\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big) - \mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| < \epsilon\Big) \geq \mathbb{P}\left(\Big\{\Delta_{T,S}(\Theta_T) - \Delta_\infty(\Theta_T)\Big\} \subset B_0(\epsilon')\right).$$

Now, for every $(\omega, T, S) \in \Omega \times \mathbb{N} \times \mathbb{N}$, having $\sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\big(\Delta_{T,S}(\omega, \boldsymbol{\theta}), \Delta_\infty(\boldsymbol{\theta})\big) < 2\epsilon'$ implies by construction that $\big\{\Delta_{T,S}(\omega, \Theta_T) - \Delta_\infty(\Theta_T)\big\} \subseteq \mathcal{B}_0(\epsilon')$. Hence, $\forall T \in \mathbb{N}$ we have that,

$$\mathbb{P}\Big(\Big\{\Delta_{T,S}(\Theta_T) - \Delta_\infty(\Theta_T)\Big\} \subset \mathcal{B}_0(\epsilon')\Big) \geq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\Big(\Delta_{T,S}(\boldsymbol{\theta}) - \Delta_\infty(\boldsymbol{\theta})\Big) < 2\epsilon'\right).$$

The two previous inequalities can now be used to conclude that, for every $T \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| < \epsilon\Big) &= \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big) - \mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| < \epsilon\Big) \\ &\geq \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta}), \Delta_\infty(\boldsymbol{\theta})\big) < 2\epsilon'\Big)\end{aligned}$$

and hence, by (20) and Lemma 24 we finally obtain the uniform convergence in probability (25) of the sequence of SNPII criterion functions $\{Q_{T,S}\}_{T\in\mathbb{N}}$ across the sieves $\{\Theta_T\}_{T\in\mathbb{N}}$ for every $S \in \mathbb{N}$,

$$\lim_{T\to\infty} \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| > \epsilon\Big) = 0 \;\forall \epsilon > 0. \tag{25}$$

The almost sure counterpart of this result is obtained in a similar way by noting that, for every $(\omega, S) \in \Omega \times \mathbb{N}$ and every $\epsilon > 0$, $\exists \epsilon' > 0$ such that,

$$\lim_{T\to\infty} \Big\{\Delta_{T,S}(\omega, \Theta_T) - \Delta_\infty(\Theta_T)\Big\} \subset B_0(\epsilon') \quad \Rightarrow \quad \lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu\big(\Delta_{T,S}(\omega, \boldsymbol{\theta})\big) - \mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| < \epsilon$$

which implied immediately,

$$\mathbb{P}\Big(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu\big(\Delta_{T,S}(\boldsymbol{\theta})\big) - \mu\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| < \epsilon\Big) \geq \mathbb{P}\left(\lim_{T\to\infty} \Big\{\Delta_{T,S}(\Theta_T) - \Delta_\infty(\Theta_T)\Big\} \subset B_0(\epsilon')\right).$$

Now, for every $(\omega, T, S) \in \Omega \times \mathbb{N} \times \mathbb{N}$, having $\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\big(\Delta_{T,S}(\omega, \boldsymbol{\theta}), \Delta_\infty(\boldsymbol{\theta})\big) < 2\epsilon'$ implies by construction that $\lim_{T\to\infty} \big\{\Delta_{T,S}(\omega, \Theta_T) - \Delta_\infty(\Theta_T)\big\} \subseteq \mathcal{B}_0(\epsilon')$. Hence, $\forall T \in \mathbb{N}$ we have that,

$$\mathbb{P}\Big(\lim_{T\to\infty} \Big\{\Delta_{T,S}(\Theta_T) - \Delta_\infty(\Theta_T)\Big\} \subset \mathcal{B}_0(\epsilon')\Big) \geq \mathbb{P}\left(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\Big(\Delta_{T,S}(\boldsymbol{\theta}) - \Delta_\infty(\boldsymbol{\theta})\Big) < 2\epsilon'\right).$$

The two previous inequalities can now be used to conclude that, for every $T \in \mathbb{N}$,

$$\mathbb{P}\Big(\lim_{T\to\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| < \epsilon\Big) \geq \mathbb{P}\Big(\lim_{T\to\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta}), \Delta_\infty(\boldsymbol{\theta})\big) < 2\epsilon'\Big)$$

and hence that, by (21) and Lemma 24, for every $S \in \mathbb{N}$,

$$\mathbb{P}\Big(\lim_{T\to\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| > \epsilon\Big) = 0\ \forall\,\epsilon > 0. \tag{26}$$

Continuity of the limit criterion function Q_∞ on Θ follows by *(i)* the continuity of the product binding function $\boldsymbol{\beta}^*$ on Θ which is implied by Assumptions 1, 3, 10 and Lemma 7 or Proposition 2; *(ii)* the continuity of the divergence criterion μ on $\mathcal{B}\times\mathcal{B}$, postulated in Assumption 5, and *(iii)* the continuity of continuous compositions (see Lemma 15).

$$Q_\infty(\cdot) := \mu\Big(\boldsymbol{\beta}_0^*, \boldsymbol{\beta}^*(\cdot)\Big) : \Theta \to \mathbb{R} \text{ is continuous in } \boldsymbol{\theta}\in\Theta. \tag{27}$$

Finally, recall from Theorem 1 that measurability of $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (3) and (4) follows from 1-6. Now, given Assumptions 1-6 and 8-11 and the intermediate results of *(i)* identifiable uniqueness of $\boldsymbol{\theta}_0$ obtained in (13), and *(ii)* continuity of the limit criterion function Q_∞ derived in (27); the desired conclusion that the approximate SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (3) and (4) satisfies $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \overset{p}{\to} 0$ follows by uniform convergence in probability of the criterion function $Q_{T,S}$ established in (25) and Lemma 26 adapted from Theorem 3.1 in Chen (2007) (see also Proposition 2.4 and Corollary 2.6 in White and Wooldrige (1991)). The convergence $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \overset{a.s.}{\to} 0$ follows by the uniform a.s. convergence of the criterion function $Q_{T,S}$ established in (26) and Lemma 26 (see Theorem 3.1 and Remark 3.2 in Chen (2007)). □

B.4 Proof of Theorem 4

Proof. Note first that the identifiable uniqueness of $\boldsymbol{\theta}_0$ obtained in (13) and the continuity of the limit criterion function Q_∞ derived in (27) in the proof of Theorem 3 still apply here. Uniform convergence of the criterion function $Q_{T,S}$ must however be derived from the new set of conditions. In particular, by Assumptions 1, 3 and convergence in probability in part *(ii)* of Assumption 12 we obtain by the Continuous Mapping Theorem (Corollary 15),

$$\lim_{T\to\infty}\mathbb{P}\Big(\delta_{\mathcal{B}_L}\big(\tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta}), \boldsymbol{\beta}^*_L(\boldsymbol{\theta})\big) > \epsilon\Big) = 0\ \forall\,\epsilon > 0 \text{ and every } (\boldsymbol{\theta}, L) \in \Theta\times\mathcal{L}; \tag{28}$$

The a.s. counterpart of (28) follows also from the same set of Assumptions 1, 3, the a.s. convergence in part *(ii)* of Assumption 12, and Corollary 15,

$$\mathbb{P}\Big(\lim_{T\to\infty}\delta_{\mathcal{B}_L}\big(\tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta}), \boldsymbol{\beta}^*_L(\boldsymbol{\theta})\big) > \epsilon\Big) = 0\ \forall\,\epsilon > 0 \text{ and every } (\boldsymbol{\theta}, L) \in \Theta\times\mathcal{L}; \tag{29}$$

The pointwise convergence in probability of the empirical binding function $\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ then follows from (28), Assumptions 1, 3, Lemma 28 and Proposition 4 (see also Remark 3 and Lemmas 29 and 5),

$$\lim_{T\to\infty} \mathbb{P}\Big(\delta_{\mathcal{B}}\big(\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}), \boldsymbol{\beta}^*_L(\boldsymbol{\theta})\big) > \epsilon\Big) = 0 \; \forall \epsilon > 0 \; \text{ and every } \; (\boldsymbol{\theta}) \in \Theta; \tag{30}$$

Almost sure convergence of $\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ can be easily deduced, given Assumption 3, from (29), Lemma 6 and Corollary 11,

$$\mathbb{P}\Big(\lim_{T\to\infty} \delta_{\mathcal{B}}\big(\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}), \boldsymbol{\beta}^*_L(\boldsymbol{\theta})\big) > \epsilon\Big) = 0 \; \forall \epsilon > 0 \; \text{ and every } \; (\boldsymbol{\theta}) \in \Theta; \tag{31}$$

The pointwise convergence in probability of $\hat{\boldsymbol{\beta}}_T$ follows naturally from Assumption 3, part *(i)* of Assumption 12, Lemma 28 and Proposition 4,

$$\lim_{T\to\infty} \mathbb{P}\Big(\delta_{\mathcal{B}}\big(\hat{\boldsymbol{\beta}}_T, \boldsymbol{\beta}^*_0\big) > \epsilon\Big) = 0 \; \forall \epsilon > 0. \tag{32}$$

Likewise its a.s. counterpart follows from Assumption 3, the a.s. convergence in part *(i)* of Assumption 12 and Corollary 11,

$$\mathbb{P}\Big(\lim_{T\to\infty} \delta_{\mathcal{B}}\big(\hat{\boldsymbol{\beta}}_T, \boldsymbol{\beta}^*_0\big) > \epsilon\Big) = 0 \; \forall \epsilon > 0. \tag{33}$$

The pointwise convergence of the centered product binding function $\Delta_{T,S}(\boldsymbol{\theta}) := \hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})$ (in probability and almost surely) to the limit centered binding function $\Delta_\infty(\boldsymbol{\theta}) := \boldsymbol{\beta}^*_0 - \boldsymbol{\beta}^*_0(\boldsymbol{\theta})$, by (30), (32) and the Continuous Mapping Theorem (Corollary 15),

$$\lim_{T\to\infty} \mathbb{P}\Big(\delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta}), \Delta_\infty(\boldsymbol{\theta})\big) > \epsilon\Big) = 0 \; \forall \epsilon > 0 \; \text{ and every } \; (\boldsymbol{\theta}) \in \Theta; \tag{34}$$

$$\mathbb{P}\Big(\lim_{T\to\infty} \delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta}), \Delta_\infty(\boldsymbol{\theta})\big) > \epsilon\Big) = 0 \; \forall \epsilon > 0 \; \text{ and every } \; (\boldsymbol{\theta}) \in \Theta; \tag{35}$$

Ultimately, we obtain the pointwise convergence in probability of the criterion function $Q_{T,S}$ to its limit Q_∞ by (34), Assumption 14 and the Continuous Mapping Theorem,

$$\lim_{T\to\infty} \mathbb{P}\Big(\Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| > \epsilon\Big) = 0 \; \forall \epsilon > 0 \; \text{ and every } \; \boldsymbol{\theta} \in \Theta. \tag{36}$$

The same applies to a.s. convergence by (35), Assumption 14 and the Continuous Mapping Theorem,

$$\mathbb{P}\Big(\lim_{T\to\infty} \Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| > \epsilon\Big) = 0 \; \forall \epsilon > 0 \; \text{ and every } \; \boldsymbol{\theta} \in \Theta. \tag{37}$$

Furthermore, under the weak Lipschitz product metric condition on $\delta_{\mathcal{B}}$ (Assumption 8), the uniform generalized stochastic Lipschitz condition on $\tilde{\boldsymbol{\beta}}^L_{T,s}$ (Assumption 13) is naturally inherited by *(i)* the average empirical product binding function $\tilde{\boldsymbol{\beta}}_{T,S}$; *(ii)* the centered empirical product binding function

$\Delta_{T,S}$; and ultimately *(iii)* the SNPII criterion function $Q_{T,S}$. Indeed, note first that for every $T > T^*$,

$$\begin{aligned}
\delta_{\mathcal{B}}\Big(\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\,,\,\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}')\Big) &\leq k \cdot \sup_{L\in\mathcal{L}} \delta_{\mathcal{B}_L}\left(\tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta})\,,\,\tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta}')\right) \\
&= k \cdot \sup_{L\in\mathcal{L}} \Big\| 1/S \sum_{s=1}^{S} \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) - 1/S \sum_{s=1}^{S} \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}') \Big\|_{\mathcal{B}_L} \\
&\leq k \cdot 1/S \sum_{s=1}^{S} \sup_{L\in\mathcal{L}} \Big\| \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}) - \tilde{\boldsymbol{\beta}}^L_{T,s}(\boldsymbol{\theta}') \Big\|_{\mathcal{B}_L} \\
&\leq k \cdot 1/S \sum_{s=1}^{S} \zeta_T \xi\Big(\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')\Big) = k \cdot \zeta_T \xi\Big(\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')\Big) \;\text{ a.s. }\; \forall (\boldsymbol{\theta},\boldsymbol{\theta}') \in \Theta_T \times \Theta_T,
\end{aligned} \tag{38}$$

where the first inequality follows by Assumption 8, the second by the sub-additivity of $\|\cdot\|_{\mathcal{B}_L}$ and $\sup_{L\in\mathcal{L}}$, and the third by the uniform generalized stochastic Lipschitz condition postulated in Assumption 13. Now, $\Delta_{T,S}$ inherits immediately the same form of stochastic smoothness from $\tilde{\boldsymbol{\beta}}_{T,S}$ since for every $T > T^*$,

$$\begin{aligned}
\delta_{\mathcal{B}}\big(\Delta_{T,S}(\boldsymbol{\theta})\,,\,\Delta_{T,S}(\boldsymbol{\theta}')\big) &= \Big\| \big(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta})\big) - \big(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}')\big) \Big\|_{\mathcal{B}} \\
&= \Big\| \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) - \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}') \Big\|_{\mathcal{B}} = \delta_{\mathcal{B}}\Big(\tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}) - \tilde{\boldsymbol{\beta}}_{T,S}(\boldsymbol{\theta}')\Big) \\
&\leq k \cdot \zeta_T \xi\Big(\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')\Big) \;\text{ a.s. }\; \forall (\boldsymbol{\theta},\boldsymbol{\theta}') \in \Theta_T \times \Theta_T,
\end{aligned} \tag{39}$$

where the inequality was obtained in (38). Finally, under Assumption 14, the SNPII criterion $Q_{T,S}$ is also generalized Lipschitz continuous. In particular, it holds for every $T > T^*$ that,

$$\begin{aligned}
\Big| Q_{T,S}(\boldsymbol{\theta}) - Q_{T,S}(\boldsymbol{\theta}') \Big| &= \Big| \mu(\Delta_{T,S}(\boldsymbol{\theta})) - \mu(\Delta_{T,S}(\boldsymbol{\theta}')) \Big| \\
&\leq \xi_\mu\Big(\delta_{\mathcal{B}}(\Delta_{T,S}(\boldsymbol{\theta}), \Delta_{T,S}(\boldsymbol{\theta}'))\Big) \;\text{ surely }\; \forall T \in \mathbb{N} \\
&\leq \xi_\mu\Big(k \cdot \zeta_T \xi\Big(\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')\Big)\Big) \;\text{ a.s. }\; \forall T > T^* \\
&= \zeta_\mu\Big(k \cdot \zeta_T\Big) \xi_\mu\Big(\xi\Big(\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')\Big)\Big) = \zeta^Q_T \xi^Q\Big(\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')\Big) \;\text{ a.s. }\; \forall (\boldsymbol{\theta},\boldsymbol{\theta}') \in \Theta_T \times \Theta_T,
\end{aligned} \tag{40}$$

where the first inequality follows from generalized Lipschitz smoothness of μ on $\mathcal{B}$ (Assumption 14), the second from (39), and the last two equalities by the ζ_μ-homogeneity of ξ_μ (Assumption 14) and by defining $\zeta^Q_T := \zeta_\mu\Big(k \cdot \zeta_T\Big)$ and $\xi^Q(x) := \xi_\mu \circ \xi(x)\ \forall x \in \mathbb{R}$. Clearly, by condition *(i)* in Assumption 13 and Definition 15 on obtains $\zeta^Q_T = O_p(1)$, whereas by condition *(ii)* in Assumption 13 on obtains $\limsup_{t\in\mathbb{N}} \zeta^Q_T < \infty$ a.s.. Also, by Assumptions 13 and 14 we have that $\lim_{x\to 0} \xi^Q(x) = 0$. Now, a straightforward adaptation of Lemma 1 and Theorem 1 in Andrews (1992) (see also Davidson Davidson (1994, Theorem 21.9 and Theorem 21.10, p.337,339) for a textbook treatment) yields the desired uniform convergence if the SNPII criterion across sieves. First, we obtain that the sequence $\{Q_{T,S}-Q_\infty\}_{T\in\mathbb{N}}$ is Asymptotically Uniformly Stochastically Equicontinuous (AUSE) across the sieves $\{\Theta\}_{T\in\mathbb{N}}$. In particular, by (40) it holds true that, for every $\epsilon'>0$,

$$\sup_{\boldsymbol{\theta}\in\Theta_T}\ \sup_{\boldsymbol{\theta}'\in\Theta_T:\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')\leq\epsilon'} |Q_{T,S}(\boldsymbol{\theta})-Q_{T,S}(\boldsymbol{\theta}')| \leq \zeta_T\xi(\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')) \leq \zeta_T\xi(\epsilon') \ \text{ a.s. } \ \forall T>T^*, \tag{41}$$

and hence, using *(i)* in Assumption 13 for every $\epsilon>0$, there exists $\epsilon'>0$ such that,

$$\limsup_{T\to\infty}\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T}\ \sup_{\boldsymbol{\theta}'\in\Theta_T:\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')\leq\epsilon'} |Q_{T,S}(\boldsymbol{\theta})-Q_{T,S}(\boldsymbol{\theta}')|>\epsilon\Big) \leq \limsup_{T\to\infty}\mathbb{P}\Big(\zeta_T>\epsilon/\xi(\epsilon')\Big)<\epsilon \tag{42}$$

where the last inequality follows since $\zeta_T=O_p(1)$ and $\epsilon/\xi(\epsilon')\to\infty$ as $\epsilon'\to 0$. Alternatively, using *(ii)* in Assumption 13 it follows in a similar fashion that for every $\epsilon>0$, there exists $\epsilon'>0$ such that,

$$\mathbb{P}\Big(\limsup_{T\to\infty}\sup_{\boldsymbol{\theta}\in\Theta_T}\ \sup_{\boldsymbol{\theta}'\in\Theta_T:\delta_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')\leq\epsilon'} |Q_{T,S}(\boldsymbol{\theta})-Q_{T,S}(\boldsymbol{\theta}')|>\epsilon\Big) \leq \mathbb{P}\Big(\limsup_{T\to\infty}\zeta_T>\epsilon/\xi(\epsilon')\Big)<\epsilon \tag{43}$$

since $\limsup_{n\in\mathbb{N}}\zeta_T<\infty$ a.s. and $\epsilon/\xi(\epsilon')\to\infty$ as $\epsilon'\to 0$. Second, define $\Delta Q_{T,S}(\boldsymbol{\theta})=Q_{T,S}(\boldsymbol{\theta})-Q_\infty(\boldsymbol{\theta})\ \forall\boldsymbol{\theta}\in\Theta$. By Assumption 1, the sieves $\Theta_T\subset\Theta$ are compact and therefore totally bounded for every $T\in\mathbb{N}$. Let $S_T:=\{S(\boldsymbol{\theta}_i,\epsilon'),i=1,...,I_T(\epsilon')\}$ be a finite cover for Θ_T for every $T\in\mathbb{N}$. Then, for every $\epsilon>0$,

$$\begin{aligned}
\mathbb{P}\big(\sup_{\boldsymbol{\theta}\in\Theta_T}|\Delta Q_{T,S}(\boldsymbol{\theta})|>2\epsilon\big) &= \mathbb{P}\big(\max_{i\leq I_T(\epsilon')}\sup_{\boldsymbol{\theta}\in S(\boldsymbol{\theta}_i,\epsilon')}|\Delta Q_{T,S}(\boldsymbol{\theta})|>2\epsilon\big)\\
&\leq \mathbb{P}\big(\max_{i\leq I_T(\epsilon')}\sup_{\boldsymbol{\theta}\in S(\boldsymbol{\theta}_i,\epsilon')}\Big[|\Delta Q_{T,S}(\boldsymbol{\theta})-\Delta Q_{T,S}(\boldsymbol{\theta}_i)|+|\Delta Q_{T,S}(\boldsymbol{\theta}_i)|\Big]>2\epsilon\big)\\
&\leq \mathbb{P}\big(\max_{i\leq I_T(\epsilon')}\sup_{\boldsymbol{\theta}\in S(\boldsymbol{\theta}_i,\epsilon')}|\Delta Q_{T,S}(\boldsymbol{\theta})-\Delta Q_{T,S}(\boldsymbol{\theta}_i)|>\epsilon\big)\\
&\quad+\mathbb{P}\big(\max_{i\leq I_T(\epsilon')}|\Delta Q_{T,S}(\boldsymbol{\theta}_i)|>\epsilon\big)
\end{aligned} \tag{44}$$

where the first inequality is obtained by adding and subtracting $\Delta Q_{T,S}(\boldsymbol{\theta}_i)$ and by norm sub-additivity. Now, since Q_∞ is a fixed limit deterministic function, (42) implies immediately that $\{\Delta Q_{T,S}\}_{T=1}^\infty$ is AUSE, and thus, for every $\epsilon>0$, there exists $\epsilon'>0$ such that,

$$\begin{aligned}
&\limsup_{T\to\infty}\mathbb{P}\big(\max_{i\leq I_T(\epsilon')}\sup_{\boldsymbol{\theta}\in S(\boldsymbol{\theta}_i,\epsilon')}|\Delta Q_{T,S}(\boldsymbol{\theta})-\Delta Q_{T,S}(\boldsymbol{\theta}_i)|>\epsilon\big)\\
&\qquad=\limsup_{T\to\infty}\mathbb{P}\big(\sup_{\boldsymbol{\theta}_i\in\Theta_T}\sup_{\boldsymbol{\theta}\in S(\boldsymbol{\theta}_i,\epsilon')}|\Delta Q_{T,S}(\boldsymbol{\theta})-\Delta Q_{T,S}(\boldsymbol{\theta}_i)|>\epsilon\big)<\epsilon.
\end{aligned}$$

Furthermore, by Assumption 15, $\exists\,\{I_T(\epsilon')\}_{T\in\mathbb{N}}\ :\ I_T(\epsilon')=O(1)\ \forall\,\epsilon'>0$ (Definition 21). This is satisfied in particular by the sequence of covering numbers of $\{\Theta_T\}_{T\in\mathbb{N}}$ (Definition 20)) Hence, for every $\epsilon'>0$, $\exists\,I(\epsilon')\in\mathbb{N}$ such that $\sup_{T\in\mathbb{N}}I_T(\epsilon')<I(\epsilon')$. It thus follows by (34) and the Continuous

Mapping Theorem that,

$$\limsup_{T\to\infty} \mathbb{P}\big(\max_{i\le I_T(\epsilon')} |\Delta Q_{T,S}(\boldsymbol{\theta}_i)| > \epsilon\big) \le \limsup_{T\to\infty} \mathbb{P}\big(\bigcup_{i\le I_T(\epsilon')} |\Delta Q_{T,S}(\boldsymbol{\theta}_i)| > \epsilon\big)$$
$$= \limsup_{T\to\infty} \sum_{i\le I_T(\epsilon')} \mathbb{P}\big(|\Delta Q_{T,S}(\boldsymbol{\theta}_i)| > \epsilon\big)$$
$$\le \limsup_{T\to\infty} \sum_{i\le I(\epsilon')} \mathbb{P}\big(|\Delta Q_{T,S}(\boldsymbol{\theta}_i)| > \epsilon\big) < \epsilon \;\forall \epsilon' > 0.$$

Thus, taking $\limsup_{T\to\infty}$ on both sides of (44) yields, $\limsup_{T\to\infty} \mathbb{P}\big(\sup_{\boldsymbol{\theta}\in\Theta_T} |\Delta Q_{T,S}(\boldsymbol{\theta})| > 2\epsilon\big) \le \epsilon$ which implies,

$$\lim_{T\to\infty} \mathbb{P}\big(\sup_{\boldsymbol{\theta}\in\Theta_T} |Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})| > \epsilon\big) = 0 \;\forall \epsilon > 0.$$

Finally, the desired conclusion that the approximate SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (3) and (4) satisfies $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{p} 0$ follows by Theorem 4.

For almost sure convergence we note that the a.s. counterpart of (44) is,

$$\begin{aligned}\mathbb{P}\big(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} |\Delta Q_{T,S}(\boldsymbol{\theta})| > 2\epsilon\big) &\le \mathbb{P}\big(\lim_{T\to\infty} \max_{i\le I_T(\epsilon')} \sup_{\boldsymbol{\theta}\in S(\boldsymbol{\theta}_i,\epsilon')} |\Delta Q_{T,S}(\boldsymbol{\theta}) - \Delta Q_{T,S}(\boldsymbol{\theta}_i)| > \epsilon\big) \\ &\quad + \mathbb{P}\big(\lim_{T\to\infty} \max_{i\le I_T(\epsilon')} |\Delta Q_{T,S}(\boldsymbol{\theta}_i)| > \epsilon\big)\end{aligned} \tag{45}$$

and that by (43) for every $\epsilon > 0$, there exists $\epsilon' > 0$ such that,

$$\mathbb{P}\big(\limsup_{T\to\infty} \max_{i\le I_T(\epsilon')} \sup_{\boldsymbol{\theta}\in S(\boldsymbol{\theta}_i,\epsilon')} |\Delta Q_{T,S}(\boldsymbol{\theta}) - \Delta Q_{T,S}(\boldsymbol{\theta}_i)| > \epsilon\big) < \epsilon.$$

Also, by the same argument there exists $\sup_{T\in\mathbb{N}} I_T(\epsilon') < I(\epsilon') \;\forall \epsilon' > 0$ and hence, by (35) it follows that,

$$\mathbb{P}\big(\limsup_{T\to\infty} \max_{i\le I_T(\epsilon')} |\Delta Q_{T,S}(\boldsymbol{\theta}_i)| > \epsilon\big) \le \sum_{i\le I(\epsilon')} \mathbb{P}\big(\limsup_{T\to\infty} |\Delta Q_{T,S}(\boldsymbol{\theta}_i)| > \epsilon\big) < \epsilon \;\forall \epsilon' > 0.$$

and hence,

$$\mathbb{P}\big(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} |Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})| > \epsilon\big) = 0 \;\forall \epsilon > 0.$$

Finally, the desired conclusion that the approximate SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (3) and (4) satisfies $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$ follows by Theorem 4. □

B.5 Proof of Theorem 5

Proof. Identifiable uniqueness of $\boldsymbol{\theta}_0 \in \Theta$ w.r.t. the limit criterion divergence μ_∞ follows by applying the same argument as that encountered in the proof of Theorem 3, only this time to μ_∞. In particular, given Assumptions 1, 3 and 10 the product binding function $\boldsymbol{\beta}^* : \Theta \to \mathcal{B}$ is a homeomorphism and thus injective, continuous and open (Proposition 2). By injectivity $\boldsymbol{\beta}_0^* = \boldsymbol{\beta}^*(\boldsymbol{\theta}_0)$ and $\boldsymbol{\beta}_0^* \neq \boldsymbol{\beta}^*(\boldsymbol{\theta})\ \forall \boldsymbol{\theta} \in \Theta \backslash \{\boldsymbol{\theta}_0\}$, which implies,

$$Q_\infty(\boldsymbol{\theta}_0) := \mu_\infty\Big(\boldsymbol{\beta}_0^*\,,\, \boldsymbol{\beta}^*(\boldsymbol{\theta}_0)\Big) = 0 \ \text{ and } \ Q_\infty(\boldsymbol{\theta}) := \mu_\infty\Big(\boldsymbol{\beta}_0^*\,,\, \boldsymbol{\beta}^*(\boldsymbol{\theta})\Big) > 0\ \forall \boldsymbol{\theta} \in \Theta\backslash\{\boldsymbol{\theta}_0\} \tag{46}$$

and also, by openness, $\boldsymbol{\beta}^*\Big(S_{\boldsymbol{\theta}_0}(\epsilon)\Big)$ is open for every open ball $S_{\boldsymbol{\theta}_0}(\epsilon) \subset \Theta$, $\epsilon > 0$, and thus there exist an open ball $S_{\boldsymbol{\beta}_0}(\epsilon')$, $\epsilon' > 0$ such that,

$$S_{\boldsymbol{\beta}_0}(\epsilon') \subset \boldsymbol{\beta}^*\Big(S_{\boldsymbol{\theta}_0}(\epsilon)\Big) \subset \mathcal{B}.$$

Together with Assumption 17 it thus follows immediately that $\boldsymbol{\theta}_0$ is identifiably unique since,

$$\begin{aligned}\inf_{\boldsymbol{\theta}\in S^c_{\boldsymbol{\theta}_0}(\epsilon)} \Big|Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0)\Big| &= \inf_{\boldsymbol{\theta}\in S^c_{\boldsymbol{\theta}_0}(\epsilon)} \Big|\mu_\infty\Big(\boldsymbol{\beta}_0^*\,,\,\boldsymbol{\beta}^*(\boldsymbol{\theta})\Big) - \mu_\infty\Big(\boldsymbol{\beta}_0^*\,,\,\boldsymbol{\beta}^*(\boldsymbol{\theta}_0)\Big)\Big| \\ &= \inf_{\boldsymbol{\theta}\in S^c_{\boldsymbol{\theta}_0}(\epsilon)} \Big|\mu_\infty\Big(\boldsymbol{\beta}_0^*\,,\,\boldsymbol{\beta}^*(\boldsymbol{\theta})\Big)\Big| = \inf_{\boldsymbol{\beta}\in\boldsymbol{\beta}^*\left(S^c_{\boldsymbol{\theta}_0}(\epsilon)\right)} \Big|\mu_\infty\big(\boldsymbol{\beta}_0^*\,,\,\boldsymbol{\beta}\big)\Big| \\ &\leq \inf_{\boldsymbol{\beta}\in S_{\boldsymbol{\beta}_0}(\epsilon')\subset\boldsymbol{\beta}^*\left(S^c_{\boldsymbol{\theta}_0}(\epsilon)\right)} \Big|\mu_\infty\big(\boldsymbol{\beta}_0^*\,,\,\boldsymbol{\beta}\big)\Big| > 0 \qquad \forall\ \epsilon' > 0.\end{aligned} \tag{47}$$

The uniform convergence of the centered empirical binding function derived in the proof of Theorem 3 in (20) follows exactly by the same argument and hence requires no further explanation.

Now, the uniform convergence across $\{\Theta_T\}_{T\in\mathbb{N}}$ of the composition criterion $Q_{T,S}(\boldsymbol{\theta}) := \mu_T\big(\Delta_{T,S}(\boldsymbol{\theta})\big)$ is obtained by noting that, for every $T \in \mathbb{N}$ and every $\epsilon > 0$, it holds true that,

$$\begin{aligned}\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| > \epsilon\Big) &= \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_T\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| > \epsilon\Big) \\ &\leq \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_T\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big)\Big| \\ &\qquad + \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| > \epsilon\Big) \\ &\leq \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_T\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big)\Big| > \epsilon/2\Big) \\ &\qquad + P\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| > \epsilon/2\Big)\end{aligned} \tag{48}$$

where the first inequality is obtained by simply adding and subtracting $\mu_\infty(\Delta_T(\boldsymbol{\theta}))$ and by norm sub-additivity of the absolute value and supremum functions. Now,

$$\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_T\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big)\Big| > \epsilon/2\Big) \to 0 \tag{49}$$

holds true by the sure uniform convergence of μ_T (Assumption 16), and

$$\mathbb{P}(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| > \epsilon/2\Big) \to 0 \tag{50}$$

is implied by the uniform convergence of the centered empirical binding function in (20) and by uniform continuity of μ_∞ on compact sets (Assumption 16 and Lemma 25), namely on $\mathcal{B}_0(\epsilon') \subset \mathcal{B}$, a closed ball of radius ϵ' centered at the zero element of $\mathcal{B}$, i.e. $\mathcal{B}_0(\epsilon') := \{\boldsymbol{\beta} \in \mathcal{B} : \delta_{\mathcal{B}}(\boldsymbol{\beta}) \leq \epsilon'\}$. In particular, we have as before that, for every $\omega \in \Omega$ and every $\epsilon > 0$, $\exists \epsilon' > 0$ such that, $\Big\{\Delta_T(\omega, \Theta_T) - \Delta_\infty(\omega, \Theta_T)\Big\} \subset B_0(\epsilon')$ implies that $\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_\infty\big(\Delta_T(\omega, \boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_\infty(\omega, \boldsymbol{\theta})\big)\Big| < \epsilon$ and hence it follows immediately that,

$$\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| < \epsilon\Big) \geq \mathbb{P}\left(\Big\{\Delta_T(\Theta_T) - \Delta_\infty(\Theta_T)\Big\} \subset B_0(\epsilon')\right).$$

We note also that, for every $\omega \in \Omega$ and every $T \in \mathbb{N}$, having $\sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\big(\Delta_T(\omega, \boldsymbol{\theta}), \Delta_\infty(\omega, \boldsymbol{\theta})\big) < 2\epsilon'$ implies by construction that $\Delta_T(\omega, \Theta_T) - \Delta_\infty(\omega, \Theta_T) \subseteq \mathcal{B}_0$ Hence, $\forall T \in \mathbb{N}$ we have that,

$$\mathbb{P}\Big(\Big\{\Delta_T(\Theta_T) - \Delta_\infty(\Theta_T)\Big\} \subset \mathcal{B}_0(\epsilon')\Big) \geq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\Big(\Delta_T(\boldsymbol{\theta}) - \Delta_\infty(\boldsymbol{\theta})\Big) < 2\epsilon'\right).$$

The two previous inequalities can now be used to conclude that, for every $T \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| < \epsilon\Big) &\geq \mathbb{P}\Big(\Delta_T(\Theta_T) - \Delta_\infty(\Theta_T) \subset \mathcal{B}_0(\epsilon')\Big) \\ &\geq \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \delta_{\mathcal{B}}\big(\Delta_T(\boldsymbol{\theta}), \Delta_\infty(\boldsymbol{\theta})\big) < 2\epsilon'\Big).\end{aligned}$$

As a result, (20) and Assumption 16 implies (50). Finally, (49) and (50) imply by (48) that

$$\lim_{T\to\infty} \mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\Theta_T} \Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| > \epsilon\Big) = 0. \tag{51}$$

The almost sure counterpart of this result is obtained by the same argument. In particular, similarly to (48),

$$\begin{aligned}\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| \leq& \lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_T\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big)\Big| \\ &+ \lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big|\end{aligned} \tag{52}$$

and then,

$$\mathbb{P}\Big(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_T\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big)\Big| > \epsilon/2\Big) = 0 \;\forall \epsilon > 0 \tag{53}$$

holds also true by the sure uniform convergence of μ_T (Assumption 16), and

$$\mathbb{P}(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|\mu_\infty\big(\Delta_T(\boldsymbol{\theta})\big) - \mu_\infty\big(\Delta_\infty(\boldsymbol{\theta})\big)\Big| > \epsilon/2\Big) = 0 \;\forall \epsilon > 0 \tag{54}$$

is implied by the uniform convergence in (20) and uniform continuity of μ_∞ on compact sets by applying the well known result that uniform continuity preserves uniform convergence of deterministic sequences to every $\omega \in \Omega^*$ where $\mathbb{P}(\Omega^*) = 1$. As a result, (21) and Assumption 16 implies (54). Finally, (53) and (54) imply by (52) that

$$\mathbb{P}\Big(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} \Big|Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})\Big| > \epsilon\Big) = 0. \tag{55}$$

Continuity of the limit criterion function Q_∞ on Θ follows by applying the same argument encountered in the proof of Theorem 3 to the limit divergence μ_∞. In particular, continuity of Q_∞ follows from *(i)* the continuity of the product binding function $\boldsymbol{\beta}^*$ on Θ (implied by Assumptions 1, 3, 10 and Lemma 21 since a homeomorphism is continuous by definition), *(ii)* the continuity of μ_∞ on $\mathcal{B} \times \mathcal{B}$ (postulated in Assumption 16), and *(iii)* the continuity of continuous compositions (Lemma 15).

$$Q_\infty(\cdot) := \mu_\infty\Big(\boldsymbol{\beta}_0^*\,, \boldsymbol{\beta}^*(\cdot)\Big) : \Theta \to \mathbb{R} \text{ is continuous in } \boldsymbol{\theta} \in \Theta. \tag{56}$$

Finally, obtain measurability of $\hat{\boldsymbol{\theta}}_{T,S}$ follows from 1-6 and Theorem 1. Given Assumptions 1-6 and 8-11 and the intermediate results of *(i)* identifiable uniqueness of $\boldsymbol{\theta}_0$ obtained in (47), *(ii)* uniform convergence of the criterion function $Q_{T,S}$ established in (51) and *(iii)* the continuity of the limit criterion function Q_∞ derived in (56), the desired conclusion that the approximate SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (3) and (4) satisfies $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \overset{p}{\to} 0$ [a.s.] follows by Lemma 26 adapted from Theorem 3.1 in Chen (2007) (see also Proposition 2.4 and Corollary 2.6 in White and Wooldrige (1991)). □

B.6 Proof of Theorem 6

Proof. Identifiable uniqueness of $\boldsymbol{\theta}_0 \in \Theta$ is obtained in Theorem 5 in (47) from Assumptions 1, 3, 10 and 17. The pointwise convergence in probability (and almost surely) of the SNPII criterion function $Q_{T,S}$ is derived in Theorem 4 in (34) and (35) from Assumptions 1, 3, 12 and 14. Now, given the stochastic smoothness of the empirical binding function $\Delta_{T,S}$ derived in Theorem 4 in (39) under Assumptions 8, 13, 13 and 14, the uniform convergence of $Q_{T,S}$ across $\{\Theta_T\}_{T\in\mathbb{N}}$ follows from a simple adaptation of the argument found in (40) in Theorem 4. Indeed, note first that for every $T > T^*$,

$$\begin{aligned}
\Big|Q_{T,S}(\boldsymbol{\theta}) - Q_{T,S}(\boldsymbol{\theta}')\Big| &= \Big|\mu_T(\Delta_{T,S}(\boldsymbol{\theta})) - \mu_T(\Delta_{T,S}(\boldsymbol{\theta}'))\Big| \\
&\leq \xi_\mu\Big(\delta_{\mathcal{B}}(\Delta_{T,S}(\boldsymbol{\theta}), \Delta_{T,S}(\boldsymbol{\theta}'))\Big) \;\text{ surely } \;\; \forall T \in \mathbb{N} \\
&\leq \xi_\mu\Big(k \cdot \zeta_T \xi\Big(\delta_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}')\Big)\Big) \;\text{ a.s. } \;\; \forall T > T^* \\
&= \zeta_\mu\Big(k \cdot \zeta_T\Big)\xi_\mu\Big(\xi\Big(\delta_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}')\Big)\Big) = \zeta_T^Q \xi^Q\Big(\delta_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}')\Big) \;\text{ a.s. } \; \forall\, (\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta_T \times \Theta_T,
\end{aligned} \tag{57}$$

where the first inequality follows from generalized Lipschitz smoothness of μ_T $\forall T \in \mathbb{N}$ on $\mathcal{B}$ (Assumption 14), the second from (39), and the last two equalities by the ζ_μ-homogeneity of ξ_μ (Assumption

14) and by defining $\zeta_T^Q := \zeta_\mu\Big(k \cdot \zeta_T\Big)$ and $\xi^Q(x) := \xi_\mu \circ \xi(x)\ \forall x \in \mathbb{R}$. Now, following the argument in Theorem 4, we obtain the desired result that,

$$\lim_{T\to\infty} \mathbb{P}\big(\sup_{\boldsymbol{\theta}\in\Theta_T} |Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})| > \epsilon\big) = 0 \text{ and } \mathbb{P}\big(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_T} |Q_{T,S}(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})| > \epsilon\big) = 0\ \forall \epsilon > 0.$$

Also, continuity of Q_∞ is obtained as in Theorem 5 in (56) under Assumptions 1, 3, 10, 16. Finally, measurability of $\hat{\boldsymbol{\theta}}_{T,S}$ and consistency $\delta_\Theta(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}_0) \overset{p}{\to} 0$ [a.s.] of the SNPII estimator $\hat{\boldsymbol{\theta}}_{T,S}$ defined in (3) and (4) follows by the same argument as in Theorem 5. □

B.7 Proof of Theorem 7

Proof. Obtain first the a.s. continuous Frechet (Hadamard) differentiability of the criterion function $Q_{T,S}$ on Θ tangentially to Θ_T for every $(T, S) \in \mathbb{N} \times \mathbb{N}$. In particular, given the product topology on $\mathcal{B}$ (Assumption 3) and the a.s. continuously Frechet (Hadamard) differentiability of $\tilde{\boldsymbol{\beta}}_{T,S}^L : \Omega \times \Theta \to \mathcal{B}$ on Θ tangentially to Θ_T, $\forall (T, S, L) \in \mathbb{N} \times \mathbb{N} \times \mathcal{L}$ (part *(i)* of Assumption 22), it follows immediately by Proposition 7 and Corollary 17 that the empirical binding function $\tilde{\boldsymbol{\beta}}_{T,S} : \Omega \times \Theta \to \mathcal{B}$ is likewise a.s. continuously Frechet (Hadamard) differentiable on Θ tangentially to Θ_T, $\forall (T, S) \in \mathbb{N}\times\mathbb{N}$. Trivial algebra shows that the same holds for the centered empirical binding function $\Delta_{T,S} : \Omega \times \Theta \to \mathcal{B}$. Finally, by the continuous Frechet (Hadamard) differentiability of $\mu : \mathcal{B} \to \mathbb{R}$ (part (iii) of Assumption 22), the chain rule (Lemma 38), and the continuity of continuous compositions (Lemma 15) we obtain the desired result that $Q_{T,S}$ is a.s. continuously differentiable on Θ tangentially to Θ_T with derivative,

$$\nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}, \cdot) = \nabla_{B_T} \mu\Big(\Delta_{T,S}(\boldsymbol{\theta}), \nabla_{\Theta_T} \Delta_{T,S}(\boldsymbol{\theta}, \cdot)\Big)$$

for every $\boldsymbol{\theta} \in \Theta$ where $B_T := \Delta_{T,S}(\Theta_T)$. We thus state for future reference that,

$$\begin{gathered} Q_{T,S} : \Omega \times \Theta \to \mathbb{R} \text{ is a.s. continuously Frechet (Hadamard) differentiable} \\ \text{in } \boldsymbol{\theta} \in \Theta \text{ tangentially to } \Theta_T\ \forall (T, S) \in \mathbb{N} \times \mathbb{N}. \end{gathered} \tag{58}$$

By a similar argument, equivalent smoothness results are also available for the limit criterion function Q_∞ and its derivative function $\nabla_\Theta Q_\infty : \Theta \to \mathbb{L}(\Theta_T, \mathbb{R})$. In particular, by Proposition 7 and Corollary 17, the continuous Frechet (Hadamard) differentiability of $\boldsymbol{\beta}_L^* : \Theta \to \mathcal{B}_L\ \forall L \in \mathcal{L}$ on Θ implies the same property for $\boldsymbol{\beta}^* : \Theta \to \mathcal{B}$ and, by trivial algebra, the same holds for $\Delta_\infty : \Theta \to \mathcal{B}$. By the continuous partial Frechet (Hadamard) differentiability of $\mu : \mathcal{B} \to \mathbb{R}$ (part *(iii)* of Assumption 22), the chain rule (Lemma 38) and continuity of continuous compositions (Lemma 15) we then have that Q_∞ is differentiable with derivative given by,

$$\nabla_\Theta Q_\infty(\boldsymbol{\theta}, \cdot) = \nabla_B \mu\Big(\Delta_\infty(\boldsymbol{\theta}), \nabla_\Theta \Delta_\infty(\boldsymbol{\theta}, \cdot)\Big)$$

which is continuous on Θ, and hence the desired result follows. For reference we state,

$$Q_\infty : \Theta \to \mathbb{R} \text{ is continuously Frechet (Hadamard) differentiable in } \boldsymbol{\theta} \in S_{\boldsymbol{\theta}_0}(\epsilon) \text{ for some } \epsilon > 0. \tag{59}$$

An analogous result can be derived for the criterion's derivative function $\nabla_{\Theta_T} Q_\infty : \Theta \to \mathbb{L}(\Theta_T, \mathbb{R})$ from the continuous Frechet (Hadamard) differentiability of $\nabla_{\Theta_T} \boldsymbol{\beta}^*_L : \Theta \to \mathbb{L}(\Theta_T, \mathcal{B}_L) \, \forall L \in \mathcal{L}$ on Θ tangentially to Θ_T (part (ii) of Assumption 22) and the continuous Frechet (Hadamard) differentiability of $\mu : \mathcal{B} \times \mathbb{L}(\Theta_T, \mathcal{B}) \to \mathbb{L}(\mathcal{B}_T, \mathbb{R})$ (part (iii) of Assumption 22). In particular, given the product topology on $\mathcal{B}$ (Assumption 3), it follows immediately by Proposition 7 and Corollary 17 that $\nabla_{\Theta_T} \boldsymbol{\beta}^* : \Theta \to \mathbb{L}(\Theta_T, \mathcal{B}_L)$ is continuously Frechet (Hadamard) differentiable on Θ tangentially to Θ_T, $\forall T \in \mathbb{N}$. The same holds for $\nabla_{\Theta_T} \Delta_\infty : \Theta \to \mathbb{L}(\Theta_T, \mathcal{B})$ since $\nabla_{\Theta_T} \Delta_\infty = \nabla_{\Theta_T} \boldsymbol{\beta}^*$ trivially on Θ. Finally, by the continuous Frechet (Hadamard) differentiability of $\nabla_{\mathcal{B}} \mu : \times \mathbb{L}(\Theta_T, \mathcal{B}) \to \mathbb{L}(\mathcal{B}_T, \mathbb{R})$ (part (iii) of Assumption 22), Proposition 8, and the continuity of continuous compositions (Lemma 15) we obtain the desired result that $\nabla_{\Theta_T} Q_\infty : \Theta \to \mathbb{L}(\Theta_T, \mathbb{R})$ is continuously differentiable on Θ tangentially to Θ_T.

$$\nabla_{\Theta_T} Q_\infty : \Theta \to \mathbb{L}(\Theta_T, \mathbb{R}) \text{ is continuously Frechet (Hadamard) differentiable on } \Theta \text{ tangentially to } \Theta_T. \tag{60}$$

Now, using (59), (60) and the $\sqrt{T}$ sieves' expansion rate $\|\pi_T(\boldsymbol{\theta}_0) - \boldsymbol{\theta}_0\|_\Theta$ (Assumption 21) it is easy to derive an identical $\sqrt{T}$ convergence rate for the "sieve approximation error" term $\|\boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0\|_\Theta$. In particular, note that by Lemma 39 from the differentiability of $\nabla_{\Theta_T} Q_\infty$ derived in (60) and the regularity of $\boldsymbol{\theta}_0 \in \Theta$ w.r.t. $\nabla_{\Theta_T} Q_\infty$, it follows that, for every $T \in \mathbb{N}$, $\exists \, c_t > 0$ such that,[24]

$$\|\boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0\|_\Theta \leq c_t \Big| \nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}^0_T, \boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0) - \nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}_0, \boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0) \Big| \leq \bar{c} \Big| \nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}^0_T, \boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0) \Big|, \tag{61}$$

were the second inequality follows from having $\nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}_0, \boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0) = 0 \, \forall T \in \mathbb{N}$ (Proposition 5) and also because $\nabla_{\Theta_T} Q_\infty$ is a restriction of ∇Q_∞ (Proposition 9), by convergence of tangential derivatives (Proposition 10), the fact that ∇Q_∞ also satisfies Lemma 39 and hence that bounded convergence holds (Proposition 13). As a result, it follows that,

$$\begin{aligned}
\|\boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0\|_\Theta &\leq \bar{c} \Big| Q_\infty(\boldsymbol{\theta}^0_T) - Q_\infty(\boldsymbol{\theta}_0) \Big| + o(\|\boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0\|_\Theta) \\
&\leq \bar{c} \Big| Q_\infty(\pi_T \boldsymbol{\theta}_0) - Q_\infty(\boldsymbol{\theta}_0) \Big| + o(\|\boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0\|_\Theta) \\
&\leq o(\|\pi_T \boldsymbol{\theta}_0 - \boldsymbol{\theta}_0\|_\Theta) + o(\|\boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0\|_\Theta)
\end{aligned}$$

where the first inequality follows from (61) and by differentiability of Q_∞ at $\boldsymbol{\theta}^0_T \in \Theta$ derived in (59) and the implied tangential differentiability (Corollary 18), the second inequality is obtained since $Q_\infty(\pi_T \boldsymbol{\theta}_0) > Q_\infty(\boldsymbol{\theta}^0_T)$ by construction, and the last inequality follows again simply by applying the definition of differentiable operator. Finally, since $o(\|\pi_T \boldsymbol{\theta}_0 - \boldsymbol{\theta}_0\|_\Theta) = O(T^{-1/2})$ (Assumption 21), we obtain,

$$\begin{aligned}
&\|\boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0\|_\Theta (1 + o(1)) = o(\|\pi_T \boldsymbol{\theta}_0 - \boldsymbol{\theta}_0\|_\Theta) \\
&\Leftrightarrow \|\boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0\|_\Theta = \frac{1}{(1 + o(1))} O(T^{-1/2}) = O(T^{-1/2}).
\end{aligned} \tag{62}$$

[24] Regularity of $\boldsymbol{\theta}_0$ is obtained from the identifiable uniqueness of $\boldsymbol{\theta}_0$ derived in (13) from Assumptions 1, 3, 10 and 11.

From this result it follows easily that $\left|\nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}_T^0, \boldsymbol{\theta})\right|$ converges to zero at an appropriate rate for every $\boldsymbol{\theta} \in \Theta_T$. In particular,

$$\begin{aligned}
\left|\nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}_T^0, \boldsymbol{\theta})\right| &= \left|\nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}_T^0, \boldsymbol{\theta}) - \nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}_0, \boldsymbol{\theta})\right| \quad \forall \boldsymbol{\theta} \in \Theta_T \\
&\leq c_{t,\boldsymbol{\theta}} \|\boldsymbol{\theta}_T - \boldsymbol{\theta}_0\| = o(\|\boldsymbol{\theta}_T - \boldsymbol{\theta}_0\|) \quad \forall \boldsymbol{\theta} \in \Theta_T \\
&= o(O(^{-1/2})) = o(T^{-1/2}) \quad \forall \boldsymbol{\theta} \in \Theta_T.
\end{aligned} \tag{63}$$

where the first equality follows from the fact that $\nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = 0 \, \forall \boldsymbol{\theta} \in \Theta_T$ (Proposition 5) and the inequality inequality by the differentiability of $\nabla_{\Theta_T} Q_\infty$ derived in (60).

Finally, a similar result completes an approximate Z-estimator formulation of $\hat{\boldsymbol{\theta}}_{T,S}$ as follows. First, by the a.s. Frechet (Hadamard) differentiability of $\nabla_{\Theta_T} Q_{T,S}$ and the regularity of $\boldsymbol{\theta}^*_{T,S}$, it follows by Proposition 5 that,

$$\begin{aligned}
\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S}\|_\Theta &\leq c_t \left|\nabla_{\Theta_T} Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}, \hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S}) - \nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^*_{T,S}, \hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S})\right| \\
&\leq c_t \left|\nabla_{\Theta_T} Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}, \hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S})\right| + \left|\nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^*_{T,S}, \hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S})\right| \\
&\leq \left|Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}) - Q_{T,S}(\boldsymbol{\theta}^*_{T,S})\right| + o(\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S}\|_\Theta) \\
&\qquad + \left|Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}) - Q_{T,S}(\boldsymbol{\theta}^*_{T,S})\right| + o(\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S}\|)
\end{aligned}$$

where the second inequality follows by norm sub-additivity and the third by applying the definition of Frechet (Hadamard) differentiable map. This implies since $\eta_T = \left|Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}) - Q_{T,S}(\boldsymbol{\theta}^*_{T,S})\right| = O_p(T^{-1/2})$ that,

$$\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S}\|_\Theta (1 + o(1)) \leq 2\eta_T \Leftrightarrow \|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S}\|_\Theta \leq \frac{2O_p(T^{-1/2})}{1 + o(1)} = O_p(T^{-1/2}).$$

Finally, the desired result follows by noting that,

$$\begin{aligned}
\left|\nabla_{\Theta_T} Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta})\right| &= \left|\nabla_{\Theta_T} Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}) - \nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^{**}_{T,S}, \boldsymbol{\theta})\right| \\
&\leq \left|\nabla_{\Theta_T} Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}) - \nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^*_{T,S}, \boldsymbol{\theta})\right| \\
&\qquad + \left|\nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^*_{T,S}, \boldsymbol{\theta}) - \nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^{**}_{T,S}, \boldsymbol{\theta})\right| \\
&= o(\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^*_{T,S}\|_\Theta) + o(\|\boldsymbol{\theta}^*_{T,S} - \boldsymbol{\theta}^{**}_{T,S}\|_\Theta) \, \forall \boldsymbol{\theta} \in \Theta_T,
\end{aligned}$$

where the first inequality follows by adding and subtracting $\nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^*_{T,S}, \boldsymbol{\theta})$ and by norm sub-additivity, and the second inequality follows by differentiability of $\nabla_{\Theta_T} Q_{T,S}$. Now since, $\nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^*_{T,S}$ is Frechet (Hadamard) differentiable and $\boldsymbol{\theta}^*_{T,S}$ is a regular point, it follows that,

$$\|\boldsymbol{\theta}^*_{T,S} - \boldsymbol{\theta}^{**}_{T,S}\|_\Theta \leq c_t \left|\nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^*_{T,S}, \boldsymbol{\theta}^*_{T,S} - \boldsymbol{\theta}^{**}_{T,S}) - \nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^{**}_{T,S}, \boldsymbol{\theta}^*_{T,S} - \boldsymbol{\theta}^{**}_{T,S})\right|$$

which implies that,

$$
\begin{aligned}
\|\boldsymbol{\theta}^*_{T,S} - \boldsymbol{\theta}^{**}_{T,S}\|_\Theta &\leq \left|\nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^*_{T,S}, \boldsymbol{\theta}^*_{T,S} - \boldsymbol{\theta}^{**}_{T,S})\right| \\
&\leq \left|Q_{T,S}(\boldsymbol{\theta}^*_{T,S}) - Q_{T,S}(\boldsymbol{\theta}^{**}_{T,S})\right| + o(\|\boldsymbol{\theta}^*_{T,S} - \boldsymbol{\theta}^{**}_{T,S}\|_\Theta) \\
&\leq \left|Q_{T,S}(\pi_T(\boldsymbol{\theta}^{**}_{T,S})) - Q_{T,S}(\boldsymbol{\theta}^{**}_{T,S})\right| + o(\|\boldsymbol{\theta}^*_{T,S} - \boldsymbol{\theta}^{**}_{T,S}\|_\Theta) \\
&\leq o(\|\pi_T(\boldsymbol{\theta}^{**}_{T,S}) - \boldsymbol{\theta}^{**}_{T,S}\|_\Theta) + o(\|\boldsymbol{\theta}^*_{T,S} - \boldsymbol{\theta}^{**}_{T,S}\|_\Theta)
\end{aligned}
$$

where the first inequality follows immediately by (B.7) and the fact that $\nabla_{\Theta_T} Q_{T,S}(\boldsymbol{\theta}^{**}_{T,S}, \boldsymbol{\theta}) = 0\ \forall \boldsymbol{\theta} \in \Theta_T$, the second inequality by applying the definition of Frechet (Hadamard) differentiability, the third inequality is obtained since $Q_{T,S}(\pi_T(\boldsymbol{\theta}^{**}_{T,S})) \leq Q_{T,S}(\boldsymbol{\theta}^*_{T,S}) - Q_{T,S}(\boldsymbol{\theta}^{**}_{T,S})\Big| + o(\|\boldsymbol{\theta}^*_{T,S}$ and the last inequality by differentiability of $Q_{T,S}$. It thus follows by (B.7) that

$$\nabla_{\Theta_T} Q_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}, \boldsymbol{\theta}) = o_p(T^{-1/2})\ \forall \boldsymbol{\theta} \in \Theta_T. \tag{64}$$

An appropriate characterization of the convergence of the convergence rate of the limit criterion function evaluated at the sequence of projection points $\{\boldsymbol{\theta}^T_0\}_{T\in\mathbb{N}}$ is also In particular, ... It thus follows that

$$\nabla_{\Theta_T} Q_\infty(\boldsymbol{\theta}^0_T, \boldsymbol{\theta}) = o(T^{-1/2})\ \forall \boldsymbol{\theta} \in \Theta_T \quad \text{and} \quad \nabla_\Theta Q_\infty(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = 0\ \forall (\boldsymbol{\theta}, T) \in \Theta_T \times \mathbb{N}. \tag{65}$$

Given the $\sqrt{T}$ sieve expansion rate (Assumption 21) and suitable generalized Lipschitz smoothness conditions on the auxiliary estimators, their limits and derivatives below, we shall obtain an important intermediate condition, namely,

$$\sqrt{T}\Big[\big(\nabla_{\Theta_T} Q^0_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}) - \nabla_{\Theta_T} Q^0_\infty(\hat{\boldsymbol{\theta}}_{T,S})\big) - \big(\nabla_{\Theta_T} Q^0_{T,S}(\boldsymbol{\theta}^0_T) - \nabla_{\Theta_T} Q^0_\infty(\boldsymbol{\theta}^0_T)\big)\Big] = o_p\big(\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^0_T\|_\Theta\big).$$

The conditions *(i)* $\left\|\nabla_{\Theta_T}\tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta}', 0) - \nabla_{\Theta_T}\tilde{\boldsymbol{\beta}}^L_{T,S}(\boldsymbol{\theta}'', 0)\right\|_{\mathcal{B}} \leq \zeta_{\nabla,T}\xi_\nabla\big(\|\boldsymbol{\theta}' - \boldsymbol{\theta}''\|_\Theta\big)$; *(ii)* $\left\|\tilde{\boldsymbol{\beta}}^L_\infty(\boldsymbol{\theta}') - \tilde{\boldsymbol{\beta}}^L_\infty(\boldsymbol{\theta}'')\right\|_{\mathcal{B}} \leq \xi_\infty\big(\|\boldsymbol{\theta}' - \boldsymbol{\theta}''\|_\Theta\big)$; and *(iii)* $\left\|\nabla_{\Theta_T}\tilde{\boldsymbol{\beta}}^L_\infty(\boldsymbol{\theta}', 0) - \nabla_{\Theta_T}\tilde{\boldsymbol{\beta}}^L_\infty(\boldsymbol{\theta}'', 0)\right\|_{\mathcal{B}} \leq \xi^\infty_\nabla\big(\|\boldsymbol{\theta}' - \boldsymbol{\theta}''\|_\Theta\big)$; hold a.s. $\forall (\boldsymbol{\theta}', \boldsymbol{\theta}'') \in S_{\boldsymbol{\theta}_0}(\delta) \times S_{\boldsymbol{\theta}_0}(\delta) \subset \Theta \times \Theta$ for some $\delta > 0$ and every $L \in \mathcal{L}$ and $T > T^* \in \mathbb{N}$ with $\zeta^\Delta_{\nabla,T} = O_p(1)$ and ξ_∇, ξ_∞ and ξ^∞_∇ satisfying $\xi_\nabla(x) \to 0$, $\xi_\infty(x) \to 0$ and $\xi^\infty_\nabla(x) \to 0$ as $x \to 0$.

By Assumption AC1, there exists $T^* \in \mathbb{N}$ such that $\boldsymbol{\theta}^0_T \in S_{\boldsymbol{\theta}_0}(\epsilon)\ \forall T \geq T^*$. Assumption AC8 then ensures that $\nabla^2_\Theta Q^0_\infty(\boldsymbol{\theta}^0_T, \cdot)$ is defined for every $T > T^*$. Furthermore, by Corolarry 18, the tangential derivative $\nabla^2_{\Theta_T} Q^0_\infty(\boldsymbol{\theta}^0_T, \cdot)$ is equally defined for all $T > T^*$. Now, by Assumption AC9, $\nabla^2_{\Theta_T} Q^0_\infty(\boldsymbol{\theta}^0_T, \cdot)$ has an inverse defined on the range of $\nabla^2_{\Theta_T} Q^0_\infty(\boldsymbol{\theta}^0_T, \cdot)$ for every $T > T^*$. Since the inverse of a continuous linear map is itself both linear and continuous, then that same inverse is also bounded. Finally, it holds true that, for every $T > T^*$,

$$\exists\, c_T \in \mathbb{R}^+_0 \text{ such that } \left|\nabla^2_{\Theta_T} Q^0_\infty(\boldsymbol{\theta}^0_T, \boldsymbol{\theta} - \boldsymbol{\theta}^0_T)\right| \geq c_T\|\boldsymbol{\theta} - \boldsymbol{\theta}^0_T\|_\Theta \quad \forall \boldsymbol{\theta}\ :\ (\boldsymbol{\theta} - \boldsymbol{\theta}^0_T) \in \text{lin}(\Theta_T).$$

Furthermore, by the same argument, under Assumption AC10,

$$\exists\, \bar{c} \in \mathbb{R}_0^+ \text{ such that } \left|\nabla^2_{\Theta} Q^0_{\infty}(\boldsymbol{\theta}_0, \boldsymbol{\theta} - \boldsymbol{\theta}_0)\right| \geq \bar{c}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_{\Theta} \quad \forall \boldsymbol{\theta} \;:\; (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \in \text{lin}(\Theta).$$

Also, Assumptions AC1 and AC8 together imply that,

$$\left|\nabla^2_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T, \boldsymbol{\theta}_T) - \nabla^2_{\Theta} Q^0_{\infty}(\boldsymbol{\theta}_0, \boldsymbol{\theta})\right| = o(\|\boldsymbol{\theta}^0_T - \boldsymbol{\theta}_0\|_{\mathbb{A}}), \tag{66}$$

and in particular by Proposition 10, $\left|\nabla^2_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T, \boldsymbol{\theta}_T) - \nabla^2_{\Theta} Q^0_{\infty}(\boldsymbol{\theta}_0, \boldsymbol{\theta})\right| \to 0$ as $T \to \infty$ for every sequence $\boldsymbol{\theta}_T \to \boldsymbol{\theta} \in \Theta$ with $\boldsymbol{\theta}_T \in \text{lin}(\Theta_T)\ \forall T > T*$. This in turn implies by Proposition 13 that $\exists\, c > 0$ such that,

$$\left|\nabla^2_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T, \boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T)\right| \geq c\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta} \tag{67}$$

holds for every sequence $\{\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\}_{T\in\mathbb{N}}$ such that $(\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T) \in \text{lin}(\Theta_T)\ \forall T > T^*$.

Now, the continuous Frechet differentiability of $\nabla_{\Theta} Q^0_{\infty} : \Theta \to \mathbb{R}$ on $S_{\boldsymbol{\theta}_0}(\epsilon)$ postulated in Assumption AC8 implies that $\nabla_{\Theta} Q^0_{\infty}$ is uniformly Frechet differentiable along every sequence $\boldsymbol{\theta}^0_T \to \boldsymbol{\theta}_0$. In particular,[25]

$$\left|\nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}_T) - \nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T) - \nabla^2_{\Theta} Q^0_{\infty}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T)\right| = o(\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta}) \tag{68}$$

holds for every sequence $\{\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\}_{T\in\mathbb{N}}$ such that $(\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T) \in \text{lin}(\Theta_T)\ \forall T > T^*$. Hence, using (67), it follows immediately that,

$$\begin{aligned}
\left|\nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}_T) - \nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T)\right| - \left|\nabla^2_{\Theta} Q^0_{\infty}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T)\right| \;&=\; o(\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta}) \\
\Leftrightarrow \left|\nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}_T) - \nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T)\right| \;&=\; \left|\nabla^2_{\Theta} Q^0_{\infty}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T)\right| - c\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta} \\
&\qquad + c\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta} + o(\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta}) \\
\Leftrightarrow \left|\nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}_T) - \nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T)\right| \;&\geq\; c\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta} + o(\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta})
\end{aligned}$$

holds also for every sequence $\{\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\}_{T\in\mathbb{N}}$ such that $(\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T) \in \text{lin}(\Theta_T)\ \forall T > T^*$, where the first equality follows from (68) by norm sub-additivity, the first equivalence by adding and subtracting $c\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta}$, and the second by (67). Finally, we can conclude that,

$$\begin{aligned}
&\left|\nabla_{\Theta_T} Q^0_{\infty}(\hat{\boldsymbol{\theta}}_{T,S}) - \nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T)\right| \geq c\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^0_T\|_{\Theta} + o(\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^0_T\|_{\Theta}) \\
&\Leftrightarrow \|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^0_T\|_{\Theta}\big(c + o(1)\big) \leq \left|\nabla_{\Theta_T} Q^0_{\infty}(\hat{\boldsymbol{\theta}}_{T,S}) - \nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T)\right| \\
&\Leftrightarrow \sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^0_T\|_{\Theta}\big(c + o(1)\big) \leq \sqrt{T}\left|\nabla_{\Theta_T} Q^0_{T,S}(\boldsymbol{\theta}^0_T) - \nabla_{\Theta_T} Q^0_{\infty}(\boldsymbol{\theta}^0_T)\right| + o_p(\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^0_T\|_{\Theta}) \\
&\Leftrightarrow \sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^0_T\|_{\Theta}\big(c + o(1) - o_p(1)\big) \leq O_p(1) \\
&\Leftrightarrow \sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}^0_T\|_{\Theta} \leq \frac{O_p(1)}{c + o_p(1)} = O_p(1)
\end{aligned} \tag{69}$$

[25] Here continuous differentiability is not really required since for asymptotically tight sequences $\|\boldsymbol{\theta}_T - \boldsymbol{\theta}^0_T\|_{\Theta}$, deferentiability at the limit point $\boldsymbol{\theta}_0$ is sufficient. However, continuous differentiability is required above for independent reasons and hence we make arbitrary use of it.

holds again for every sequence $\{\hat{\boldsymbol{\theta}}_{T,S}\}_{T\in\mathbb{N}}$ and $\{\boldsymbol{\theta}_T^0\}_{T\in\mathbb{N}}$ such that $(\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0) \in \Theta_T \ \forall T \in \mathbb{N}$. Here, the first equivalence in (69) is obtained simply by rewriting $c\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta - o(\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta)$ as $\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta\big(c + o(1)\big)$. The second equivalence is obtained by multiplying both sides by $\sqrt{T}$ and substituting $\sqrt{T}\Big|\nabla_{\Theta_T} Q_\infty^0(\hat{\boldsymbol{\theta}}_{T,S}) - \nabla_{\Theta_T} Q_\infty^0(\boldsymbol{\theta}_T^0)\Big|$ by $\sqrt{T}\Big|\nabla_{\Theta_T} Q_{T,S}^0(\boldsymbol{\theta}_T^0) - \nabla_{\Theta_T} Q_\infty^0(\boldsymbol{\theta}_T^0)\Big| + o_p(\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta)$, which follows from the fact that,

$$\begin{aligned}
\sqrt{T}\Big[\nabla_{\Theta_T} Q_\infty^0(\hat{\boldsymbol{\theta}}_{T,S}) - \nabla_{\Theta_T} Q_\infty^0(\boldsymbol{\theta}_T^0)\Big] &= \sqrt{T}\nabla_{\Theta_T} Q_\infty^0(\hat{\boldsymbol{\theta}}_{T,S}) \\
&= \sqrt{T}\nabla_{\Theta_T} Q_\infty^0(\hat{\boldsymbol{\theta}}_{T,S}) - \sqrt{T}\nabla_{\Theta_T} Q_{T,S}^0(\hat{\boldsymbol{\theta}}_{T,S}) + o_p(1) \\
&= -\sqrt{T}\big(\nabla_{\Theta_T} Q_{T,S}^0(\boldsymbol{\theta}_T^0) - \nabla_{\Theta_T} Q_\infty^0(\boldsymbol{\theta}_T^0)\big) \\
&\qquad + o_p(\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta)
\end{aligned} \tag{70}$$

where the first equality in (70) holds by definition (since $\nabla_{\Theta_T} Q_\infty^0(\boldsymbol{\theta}_T^0) = 0$ in Assumption AC5), the second by adding and subtracting $\sqrt{T}\nabla_{\Theta_T} Q_{T,S}^0(\hat{\boldsymbol{\theta}}_{T,S})$ and noting that $\sqrt{T}\nabla_{\Theta_T} Q_{T,S}^0(\hat{\boldsymbol{\theta}}_{T,S}) = o_p(1)$ (since $\nabla_{\Theta_T} Q_{T,S}^0(\hat{\boldsymbol{\theta}}_{T,S}) = o_p(T^{-1/2})$ in Assumption AC5), and the third follows immediately from Assumption AC6. Finally, the third equivalence in (69) is obtained by taking the term $o_p(\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta)$ to the left-hand-side and rewriting $\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta\big(c + o(1)\big) - o_p(\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta)$ as $\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta\big(c + o(1) - o_p(1)\big)$, and also, by noting that $\sqrt{T}\big(\nabla Q_{T,S}^0(\boldsymbol{\theta}_T^0) - \nabla_{\Theta_T} Q_\infty^0(\boldsymbol{\theta}_T^0)\big) = O_p(1)$ (Assumption AC7).

Finally, $\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta = O_p(1)$ implies by Assumption AC1 that,

$$\begin{aligned}
\sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\|_\Theta &= \sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0 + \boldsymbol{\theta}_T^0 - \boldsymbol{\theta}_0\|_\Theta \\
&\le \sqrt{T}\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_T^0\|_\Theta + \sqrt{T}\|\boldsymbol{\theta}_T^0 - \boldsymbol{\theta}_0\|_\Theta = O_p(1) + o_p(1) = O_p(1).
\end{aligned}$$

□

B.8 Proof of Theorem 8

[to be completed]

References

Andrews, D. W. (1992). Generic uniform convergence. *Econometric Theory*, 8:241–257.

Andrews, D. W. K. (1991). Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica*, 59(2):307–45.

Basener, W. F. (1973). *Topology and its Applications*. Pure and Applied Mathematics: A Wiley-Interscience. John Wiley and Sons, Inc.

Billingsley, P. (1995). *Probability and Measure*. Wiley-Interscience.

Birge, L. and Massart, P. (1998). Minimum constrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli*, (4):329–375.

Blasques, F. (2010). Identifiable uniqueness conditions for a large class of extremum estimators. Meteor research memorandum, Maastricht University.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6 of *Handbook of Econometrics*, chapter 76. Elsevier.

Chen, X., Linton, O., and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, (71):1591–1608.

Chen, X. and Shen, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, (66):289–314.

Cheney, W. (2001). *Analysis for Applied Mathematics*. Graduate Texts in Mathematics. Springer-Verlag, New York Inc.

Davidson, J. (1994). *Stochastic Limit Theory*. Advanced Texts in Econometrics. Oxford University Press.

Davidson, K. R. and Donsig, A. P. (2009). *Real Analysis and Applications: Theory in Practice*. Springer, 1 edition.

Debreu (1967). Integration of correspondences. *Proceeding of the Fifth Berkley Symposium on Mathematical Statistics and Probability*, 2(1).

Denkowski, Z., Migorski, S., and Papageorgiou, N. S. (2003). *An Introduction to Nonlinear Analysis: Theory*. Kluwer Academic/Plenum Publishers, New York.

Dhaene, G., Gourieroux, C., and Scaillet, O. (1998). Instrumental models and indirect encompassing. *Econometrica*, 66(3):673–688.

Dridi, R., Guay, A., and Renault, E. (2007). Indirect inference and calibration of dynamic stochastic general equilibrium models. *Journal of Econometrics*, 136(2):397–430.

Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Enflo, P. (1973). A counterexample to the approximation problem in banach spaces. *Acta Mathematica*, 130:309–317. 10.1007/BF02392270.

Foland, G. B. (2009). *A Guide to Advanced Real Analysis*. Dolciani Mathematical Expositions. Cambridge University Press.

Gallant, A. R. (1981). On the bias in flexible functional forms and an essentially unbiased form: The fourier flexible form. *Journal of Econometrics*, 15:211–245.

Gallant, A. R. (1987). Identification and consistency in seminonparametric regression. *Bewley, Truman F., ed. (1987), Advances in Econometrics Fifth World Congress*, 1:145–170.

Gallant, A. R. and Souza, G. (1991). On the asymptotic normality of fourier flexible form estimates. *Journal of Econometrics*, 50(3):329–353.

Gamelin, T. W. and Greene, R. E. (1999). *Introduction to Topology.* Dover Publications.

Genton, M. G. and de Luna, X. (2000). Robust simulation-based estimation. *Statistics & Probability Letters*, 48(3):253–259.

Genton, M. G. and Ronchetti, E. (2003). Robust indirect inference. *Journal of the American Statistical Association*, 98(461):67–76.

Gourieroux, C. and Monfort, A. (1997). *Simulation-based Econometric Methods.* Oxford Scholarship Online Monographs.

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:85–118.

Grenander, U. (1981). *Abstract inference / Ulf Grenander.* Wiley, New York :.

Gusak, D., Kukush, A., Kulik, A., Mishura, Y., and Pilipenko, A. (2010). *Theory of stochastic processes. With applications to financial mathematics and risk theory.* Problem Books in Mathematics. New York, NY: Springer. xii, 375 p. EUR 64.15 .

Hildenbrand, W. (1974). *Core and Equilibria of a Large Economy.* Princeton University Press.

Huang, J. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics*, (28):960–999.

James, I. M. (1987). *Topological and Uniform Spaces.* Springer-Verlag, New York Inc.

Klambauer, G. (1973). *Real analysis.* North-Holland, New York, NY.

Klein, E. and Thompson, A. (1984). *Theory of correspondences: including applications to mathematical economics.* Canadian Mathematical Society series of monographs and advanced texts. Wiley.

Kolmogorov, A. N. and Fomin, S. V. (1975). *Introductory Real Analysis.* Dover Publications.

Lee, J. M. (2000). *Introduction to Topological Manifolds.* Springer.

Luenberger, D. G. (1997). *Optimization by Vector Space Methods (Series in Decision and Control)*. Wiley-Interscience, 1969 edition.

Munkres, J. (2000). *Topology*. Prentice Hall, 2 edition.

Newey, W. K. (1994). Series estimation of regression functionals. *Econometric Theory*, (10):1–28.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, (79):147–168.

Potscher, B. M. and Prucha, I. R. (2001). *A Companion to Theoretical Econometrics*, chapter Basic Elements of Asymptotic Theory. Blackwell Publishing.

Ren, J.-J. and Sen, P. K. (2001). Second order hadamard differentiability in statistical applications. *Journal of Multivariate Analysis*, 77(2):187 – 228.

Shen, X. (1997). On methods of sieves and penalization. *The Annals of Statistics*, (25):2555–2591.

Shen, X. and Wong, W. (1994). Converge rate of sieve estimates. *The Annals of Statistics*, (22):580–615.

Smith, A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8:563–584.

Stinchcombe, M. B. and White, H. (1992). Some measurability results for extrema of random functions over random sets. *Review of Economic Studies*, 59(3):495–514.

Sutherland, W. A. (2009). *Introduction to metric and topological spaces*. Oxford University Press, second edition edition.

Sviridyuk, G. A. and Fedorov, V. E. (2003). *Linear Sobolev type equations and degenerate semigroups of operators*. Inverse and Ill-Posed Problems Series. VSP.

Van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *The Annals of Statistics*, 21:14–44.

Van de Geer, S. (1995). The method of sieves and minimum contrast estimators. *Mathematical Methods of Statistics*, (4):20–38.

Van der Vaart, A. W. (1995). Efficiency of infinite dimensional m-estimators. *Statistica Neerlandica*, 49(1).

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer-Verlag, New York.

White, H. and Wooldrige, J. M. (1991). *Some Results on Sieve Estimation with Dependent Observations*, chapter 18. Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press.

Winitzki, S. (2010). *Linear Algebra via Exterior Products*. lulu.com, GNU Free Documentation License.

Wong, W. and Severini, T. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces. *The Annals of Statistics*, (19):603–632.

Wong, W. H. and Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates for sieve mle's. *The Annals of Statistics*, 23:339–362.

Zhou, S., Shen, X., and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, (26):1760–1782.