INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

Nonparametric minimum-distance estimation of simulation models

Mario Martinoli [a]

[a] Institute of Economics and L'EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy

2025/06 **March 2025**

ISSN(ONLINE): 2284-0400
DOI: 10.57838/sssa/zcvj-6e29

Nonparametric minimum-distance estimation of simulation models

Mario Martinoli*

Abstract

We propose a novel nonparametric minimum-distance estimator for the estimation of simulation models. Our approach leverages a nonparametric smoothing step to approximate the distance between real-world observations and data simulated from a model, allowing for the estimation of model parameters without relying on specific auxiliary models or moment selection. By employing sieve estimation techniques, we approximate the objective function using a series of basis functions, ensuring consistency and providing nonparametric rates of convergence. We investigate the asymptotic properties of our estimator and demonstrate its performance through Monte Carlo experiments and an empirical application to financial market data. Our method addresses the limitations of traditional simulation-based estimation techniques, particularly in cases where the stochastic equicontinuity condition is violated, and offers a robust framework for estimating parameters in heterogeneous agents models and other complex systems.

Keywords: Simulated minimum-distance; sieve estimation; stochastic equicontinuity.
JEL classification: C15; C52; C63

1 Introduction

Simulation-based econometrics has become a central tool to estimate the parameters of complex heterogeneous models in which the likelihood function is intractable or difficult to compute. Among the available methods, simulated minimum-distance (SMD) techniques provide flexible approaches to estimation by minimizing a discrepancy measure between simulated and observed data. Rooted in the simulated method of moments (SMM) pioneered by McFadden (1989) and Pakes and Pollard (1989), SMD has evolved into several strategies. One of the most exploited techniques is indirect inference (II, Gouriéroux, Monfort, and Renault, 1993; Smith Jr., 1993), which aims at connecting the theoretical (simulated) model to real-world data through a simplified auxiliary model. The parameters of the auxiliary model are estimated on both simulated and real-world data and the distance between the estimated parameters in the two cases is minimized.

*Institute of Economics & L'EMbeDS, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, Pisa (Italy). Email: *m.martinoli@santannapisa.it*
The author acknowledges support from the project How good is your model? Empirical evaluation and validation of quantitative models in economics, PRIN grant no. 20177FX2A7. I am grateful to Raffaello Seri, Samuele Centorrino, Alessio Moneta and Fulvio Corsi for insightful comments.

SMD holds several advantages with respect to SMM or II: it does not rely on the choice of a specific auxiliary model and it allows to bypass the choice of moments to match (Gallant and Tauchen, 1996). Moreover, SMD is computationally cheaper than performing a full simulated maximum likelihood (SML, Lee, 1992; Kristensen and Shin, 2012) estimation as it avoids to evaluate the model likelihood function for each combination of parameters. Rather, the alignment between model output and empirical evidence is done by minimizing a weighted distance between selected distributional features.

Although these approaches have been widely applied in macroeconomics (e.g., Christiano, Eichenbaum, and Evans, 2005), finance (e.g., Duffie and Singleton, 1993; Nikolov and Whited, 2014), labor (e.g., Voena, 2015), and industrial organization (e.g., Goettler and Gordon, 2011; Aguirregabiria and Magesan, 2020), there are still some rooms of improvement. In fact, a further drawback hitting the estimation of complex heterogeneous models, and that has not been fully recognized in the previous literature, is that several simulation models do not respect some of the conditions for the identification of the parameter $\boldsymbol{\theta}$ in simulation-based estimation methods. These issues mainly relate to the violation of the stochastic equicontinuity condition (see, e.g., McFadden, 1989, Pakes and Pollard, 1989 and Newey and McFadden, 1994, pp. 2136-2137 for a definition). To provide an intuition of the problem, let us define the unknown function $f(\boldsymbol{\theta}) \coloneqq f\left(\widehat{\mathbb{P}}_{\mathbf{y}}, \mathbb{P}_{\mathbf{z}(\boldsymbol{\theta})}\right)$ measuring the distance between $\widehat{\mathbb{P}}_{\mathbf{y}}$, the estimator of the probability measure $\mathbb{P}_{\mathbf{y}}$ based on real data $\{y_n\}_{n=1}^{N}$, and $\mathbb{P}_{\mathbf{z}(\boldsymbol{\theta})}$, the probability measure of the data simulated from the model $\{z_m(\boldsymbol{\theta})\}_{m=1}^{M}$. $\mathbb{P}_{\mathbf{z}(\boldsymbol{\theta})}$ cannot be computed but can be approximated through $\widehat{\mathbb{P}}_{\mathbf{z}(\boldsymbol{\theta})}$, the empirical probability measure of a sample extracted from $\mathbb{P}_{\mathbf{z}(\boldsymbol{\theta})}$. This approximation allows to define the function $\widehat{f}(\boldsymbol{\theta}) \coloneqq f\left(\widehat{\mathbb{P}}_{\mathbf{y}}, \widehat{\mathbb{P}}_{\mathbf{z}(\boldsymbol{\theta})}\right)$ used in simulation-based estimation techniques. Stochastic equicontinuity condition guarantees that the mapping $\boldsymbol{\theta} \mapsto f\left(\widehat{\mathbb{P}}_{\mathbf{y}}, \widehat{\mathbb{P}}_{\mathbf{z}(\boldsymbol{\theta})}\right)$ satisfies suitable continuity properties and, therefore, model parameters can be consistently estimated. However, this assumption cannot be verified for heterogeneous agents models and network models. To face this situation, we propose to approximate $f(\boldsymbol{\theta})$ through a smooth function $f_K(\boldsymbol{\theta})$ given by an expansion in a series of basis functions. A more detailed treatment of the dependence of the simulated data on $\boldsymbol{\theta}$ and the consequences of the violation of the stochastic equicontinuity condition are exposed in Section 2 (see also Martinoli, 2021, Chapter 2, for a related discussion).

We go beyond a purely parametric moment-based approach. Rather, in the same vein of Corradi and Swanson (2007) who develop an evaluation method for Dynamic Stochastic General Equilibrium (DSGE) models based on distributional comparisons, we put forward an estimator that combines SMD with a nonparametric smoothing step. We call it simulated nonparametric minimum-distance (SNPMD) estimator. As pointed out by Seri, Martinoli, et al. (2021), several metrics can be used to quantify the dissimilarity between real-world and model output. When dealing with ergodic time series in which a single instance of a time series is sufficient to estimate probabilities of events, as in our case, a good choice concerns distances between the probability measures that generated the data. For this reason, we adopt the distributional metric developed by Gray (1988) (see also D. Ryabko and B. Ryabko, 2010). Let y_n be a real-world time series extracted from a process distribution ρ^y and let $z_m(\boldsymbol{\theta})$ be a simulated time series extracted from a process distribution $\rho^z_{\boldsymbol{\theta}}$, with $\boldsymbol{\theta} \subseteq \boldsymbol{\Theta} \in \mathbb{R}^d$, we consider the metric $D(\rho^y, \rho^z_{\boldsymbol{\theta}})$ and we represent it as a function of $\boldsymbol{\theta}$, which we call $\widehat{f}(\boldsymbol{\theta})$ (a

more precise way to indicate it would be $\widehat{f}_{N,M}(\boldsymbol{\theta})$, but we use $\widehat{f}(\boldsymbol{\theta})$ for ease notation). Precisely, $\widehat{f}(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is the sampling error of $\widehat{f}(\boldsymbol{\theta})$ around $f(\boldsymbol{\theta})$, and the function f is estimated through the estimator $\widehat{f}_K$ using K sieve spaces. The estimator $\widehat{\boldsymbol{\theta}}$ is then obtained by solving the following minimization problem:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \widehat{f}_K(\boldsymbol{\theta}).$$

Sieve estimation is a nonparametric technique that approximates the unknown function of interest using a finite linear combination of some known basis function (e.g., power series, regression splines, trigonometric polynomials). Function approximation is progressively refined as the sample size grows, ensuring that the estimator becomes increasingly accurate. This framework is particularly useful in settings where the functional form of the data generating process (DGP) is left unspecified or highly complex. As such, it provides a powerful tool for handling high-dimensional and nonparametric components. Key contributions to the theory of sieve estimation include the seminal book of Grenander (1981) and the works of Newey (1997), de Jong (2002), Chen (2007), Belloni et al. (2015), and Chen and Christensen (2015), among others. Other results on the consistency and the rate of convergence of the estimators, as well as asymptotic normality of functional of the parameters, can be found in Chen and Pouzo (2012) and Chen and Pouzo (2015).

Our approach is in line with recent methodological advancements that have refined SMD estimation in multiple directions (see, e.g., Altissimo and Mele, 2009; Hnatkovska, Marmer, and Tang, 2012; Dominicy and Veredas, 2013; Gospodinov, Komunjer, and Ng, 2017). To the best of our knowledge, the first contributions merging sieve estimation and SMD can be attributed to Newey (2001) and Bierens and Song (2012), which propose semiparametric estimators for some static models. The authors show that, under some specific assumptions, their simulated estimators attain parametric rate of convergence for the estimator and the functional of the parameters. However, their outcomes are too stringent for our framework as we want to model the function linking the distribution and the parameters fully nonparametric. A notable exception is the work of Forneron (2023), in which the author combines the flexibility of sieve methods with the robustness of SMM, using a growing set of basis functions to approximate the moment conditions. Sieve-SMM is then used to estimate the parameters and the distribution of the shocks of intractable nonlinear dynamic models. While the sieve-SMM estimator offers significant advantages in handling dynamic models with complex dependence structures, it remains rooted in the SMM framework. Moreover, it considers the characteristic functions, rendering the asymptotic theory less amenable. In contrast, our method directly computes a statistical distance between the probability processes, thereby eliminating potential inefficiencies and identification issues associated with moment-based approaches.

We conceive that our technique delivers several improvements to the literature. First, as already outlined, it is particularly relevant when model-based densities provide richer information than moments alone. Second, by approximating the objective function using a linear combination of K sieve spaces, we can state assumptions on the identifiability and consistency of $\boldsymbol{\theta}$ that allow us: (i) to compute uniform error bounds on the approximation of the objective functions; (ii) to prove consistency of the SMD estimator; (iii) to provide (nonparametric) rates of convergence. Third,

it can be easily adapted to most softwares as the unique additional step is the computation of a polynomial regression. Finally, we do not impose strong distributional restrictions on the model DGP, opening-up the door to a statistical/data-driven approach to simulation-based estimation.

Despite these advances, we acknowledge some drawbacks. Indeed, due to its nonparametric nature, our SNPMD estimator is subject to bias-variance tradeoff and overfitting. However, these problems can be avoided by judiciously choosing the right number of basis functions using specific information criteria such as Aikake Information Criterion (AIC), Bayesian Information Criterion (BIC) or cross-validation. Moreover, SMD remains computationally challenging, particularly in high-dimensional settings. Luckily, recent developments in regularized estimation and machine learning-assisted moment selection have sought to address these issues (Cheng and Liao, 2015; I. Andrews, Gentzkow, and Shapiro, 2017; Chernozhukov et al., 2018). Additionally, improvements in numerical optimization, such as stochastic gradient descent and inexact Newton methods, have enhanced the computational feasibility of minimum-distance estimation in large-scale simulation models (see, e.g., Martinoli, Seri, and Corsi, 2024; Forneron, 2024).

To check the robustness and efficiency of our estimator in finite samples, we perform some Monte Carlo (MC) experiments using a first-order moving-average process and a stochastic volatility model (Harvey, Ruiz, and Shephard, 1994). Moreover, we provide an empirical application exploiting an asset pricing model with heterogeneous beliefs (Brock and Hommes, 1997; Brock and Hommes, 1998), which is a natural candidate to test the failure of the stochastic equicontinuity condition. The outcomes of the MC experiments and the empirical application confirm the theoretical results.

The rest of the paper is structured as follows. In Section 2 we introduce the problem of stochastic equicontinuity violation. In Section 3 we describe the econometric framework. In Section 4 we study the theoretical properties of the estimator. In Section 5 we provide the Monte Carlo experiments useful to study the finite-sample properties of the estimator. The results of the empirical application are discussed in Section 6. Section 7 concludes. Proofs are delegated to Section 8.

2 Violation of the stochastic equicontinuity condition

Many simulation-based estimators rely on the optimization of an objective function depending on both real-world and simulated data. Most proofs of the asymptotic properties of these estimators require an assumption called stochastic equicontinuity (see, e.g., McFadden, 1989; Pakes and Pollard, 1989 and Newey and McFadden, 1994, pp. 2136-2137), a condition of probabilistic continuity of the objective function without which asymptotic properties of simulation-based estimators are not guaranteed.[1]

We try to explain the problem with the following very simple example in which the only parameter is the mean $\mu \in \mathbb{R}$: we have a sample of real data $\{y_1, y_2, \ldots, y_N\}$ and, for a given value of μ, we draw a sample of independent Gaussian random variables $\{z_1(\mu), z_2(\mu), \ldots, z_M(\mu)\}$ with mean μ and variance 1. We compute the estimator $\widehat{\mu}$ minimizing the distance between the sample mean

[1] Stochastic equicontinuity is useful when the proofs of the asymptotic properties use uniform convergence of the objective function. When epigraphical convergence is used instead of the uniform one in the proofs of consistency, this condition can be replaced with a one-sided version (see Hess, 1996; Choirat, Hess, and Seri, 2003; Hess and Seri, 2019), but we do not pursue this topic here.

of the real-world data and the simulated sample mean (see, in the context of a different model, Gouriéroux and Monfort 1996, p. 20):

$$\widehat{f}(\mu) \coloneqq \left(\frac{1}{N} \sum_{n=1}^{N} y_n - \frac{1}{M} \sum_{m=1}^{M} z_m(\mu) \right)^2,$$
$$\widehat{\mu} = \arg\min_{\mu \in \mathbb{R}} \widehat{f}(\mu).$$

We note that $z_m(\mu) = \mu + \varepsilon_m$, where $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M\}$ are independent standard Gaussian random variables.[2] If the variables $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M\}$ are the same for any value of μ or, as often said, they are recycled for different values of μ, $\widehat{f}(\mu)$ is a continuous function of μ and stochastic equicontinuity holds true. If, on the other hand, a new sample $\{\varepsilon_1, \varepsilon_2, \ldots\}$ is drawn for any μ the resulting function $\widehat{f}(\mu)$ will be "rugged" and both the computation and the study of the asymptotic properties of the estimator will be difficult. For this reason, in order to obtain stochastic equicontinuity, the variables $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M\}$ of the model must be recycled for different values of μ, as required in McFadden (1989, p. 999), Gouriéroux and Monfort (1996, p. 16), Kristensen and Shin (2012, p. 78) and Eisenhauer, Heckman, and Mosso (2015, p. 346), among others.

However, several simulation models do not allow for the recycling of errors. This creates two problems. First, as the function to optimize is rugged, optimization routines are often very time-consuming and ad hoc algorithms have to be used. As an example, Gilli and Winker (2003) realize that the objective functions arising in the simulation-based estimation of the parameters of HA models are non-differentiable and devise algorithms combining a simplex search approach with a threshold accepting algorithm in order to identify the optimum. In a different example, Kukacka and Barunik (2017) outline some problems related to the roughness of the objective function (i.e., multiple local minima, identification issues leading to large standard deviations), and perform a graphical inspection of the simulated log-likelihood function (see also Lux, 2024, for a discussion about the lack of parameter identification in behavioral macroeconomic models). Second, the classical asymptotic properties may not hold and the corresponding inferential tools (tests, confidence intervals) are thus not necessarily available. Our method provides solutions to both these problems.

3 Econometric framework

3.1 Distributional distance

We rely on the minimization of a distance between a real-world stationary ergodic time series $y_n = (y_1, \ldots, y_N)$ extracted from a process distribution ρ^y and a stationary ergodic simulated time series $z_m(\boldsymbol{\theta}) = (z_1(\boldsymbol{\theta}), \ldots, z_M(\boldsymbol{\theta}))$ extracted from a process distribution $\rho^z_{\boldsymbol{\theta}}$. Let $D(y_n, z_m(\boldsymbol{\theta}))$ be a distance between the real-world and simulated series, when $N, M \to \infty$, if ergodicity assumptions hold true, we have that $D(y_n, z_m(\boldsymbol{\theta})) \to_{\text{a.s.}} f(\boldsymbol{\theta})$.

Several distances have been proposed to calibrate the marginal distribution of the process, its profile over time or its complete distribution. An attractive choice for our purposes is the distri-

[2] The variables $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M\}$ are often called errors in cross-sectional models and innovations in dynamic models.

butional metric developed by Gray (1988). This metric measures the distance between the process distribution ρ^y and $\rho^z_{\boldsymbol{\theta}}$ and can be defined as follows:

$$D\left(\rho^y, \rho^z_{\boldsymbol{\theta}}\right) \coloneqq \sum_{j=1}^{\infty} w_j \left|\rho^y\left(A_j\right), \rho^z_{\boldsymbol{\theta}}\left(A_j\right)\right|, \tag{3.1}$$

where w_j is a sequence of positive weights, and $\{A_j\}$ is a countable collection of events in $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$, etc. $D\left(\rho^y, \rho^z_{\boldsymbol{\theta}}\right)$ can be estimated by truncating the sum and replacing the probabilities with the empirical frequencies. Therefore, the empirical distributional distance can be defined as follows:

$$D\left(\widehat{\rho}^y, \widehat{\rho}^z_{\boldsymbol{\theta}}\right) \coloneqq \sum_{j=1}^{\infty} w_j \left|\nu\left(y_n, A_j\right), \nu\left(z_m\left(\boldsymbol{\theta}\right), A_j\right)\right|,$$

where $\nu\left(y_n, A_j\right)$ and $\nu\left(z_m\left(\boldsymbol{\theta}\right), A_j\right)$ are the frequencies with which y_n and $z_m\left(\boldsymbol{\theta}\right)$ fall into $\{A_j\}$. It can be shown that, under ergodicity, the empirical distributional distance converges to the distributional distance defined in Equation (3.1). This result is gathered in Lemma (1):

Lemma 1. *Let y_n be a sample generated by the stationary ergodic process ρ^y and let $z_m\left(\boldsymbol{\theta}\right)$ be a sample generated by the stationary ergodic process $\rho^z_{\boldsymbol{\theta}}$, then:*

$$\lim_{N,M\to\infty} D\left(\widehat{\rho}^y, \widehat{\rho}^z_{\boldsymbol{\theta}}\right) \to_{\mathsf{a.s.}} D\left(\rho^y, \rho^z_{\boldsymbol{\theta}}\right) = f\left(\boldsymbol{\theta}\right).$$

Remark 2. Lemma 1 tells us that, for any set A_j belonging to the Borel σ-algebra $\mathcal{A}$, the frequency at which the samples y_n, $z_m\left(\boldsymbol{\theta}\right)$ fall into A_j converges to the probabilities $\rho^y\left(A_j\right)$ and $\rho^z_{\boldsymbol{\theta}}\left(A_j\right)$, respectively. As N, M increase, more and more sets $A_j \in \mathcal{A}$ will exhibit frequencies that have already converged to their corresponding probabilities. Consequently, the combined weight of the sets whose frequencies have not yet converged will converge to zero.

Let us stress that our setting can be applied to several distances. The only requirement is that the latter converge to a function $f\left(\boldsymbol{\theta}\right)$, i.e. the unknown function of the estimation problem. In the next section we will clarify our data generating process.

3.2 Data generating process

Provided that, under suitable assumptions of ergodicity (see Lemma 1), the statistical distance converges almost surely to a function depending on $\boldsymbol{\theta}$, to define our statistical model we select a grid of points $\{\boldsymbol{\theta}_i, i = 1, \ldots, P\} \subset \boldsymbol{\Theta}$ and, for any $\boldsymbol{\theta}_i$ belonging to the grid, we simulate a series $z_m\left(\boldsymbol{\theta}_i\right)$ of length M. Therefore, we can represent our DGP as follows:

$$\widehat{f}\left(\boldsymbol{\theta}_i\right) = f\left(\boldsymbol{\theta}_i\right) + \varepsilon_i, \tag{3.2}$$

where $f\left(\boldsymbol{\theta}_i\right)$ is the unknown average and ε_i is the sampling error of $\widehat{f}\left(\boldsymbol{\theta}_i\right)$ around $f\left(\boldsymbol{\theta}_i\right)$. For each $\boldsymbol{\theta}_i$, one can simulate more than one series $z_{jm}\left(\boldsymbol{\theta}_i\right)$, with $j = 1, \ldots, S$. In that case $\widehat{f}\left(\boldsymbol{\theta}_i\right) = \frac{1}{S}\sum_{j=1}^{S} D\left(y_n, z_{jm}\left(\boldsymbol{\theta}_i\right)\right)$. For our purposes, we will consider $S = 1$. The vector of parameters $\boldsymbol{\theta}_i$

belongs to a parameter space $\boldsymbol{\Theta}$ that is a subset of $\mathbb{R}^d$, where d is the dimension of the parameter space. We state the following regularity conditions on the unknown function and the parameter space.

Assumption 3. *The unknown function f respects the following conditions:*

1. *$\boldsymbol{\Theta}$ is compact;*
2. *f is uniquely minimized at $\boldsymbol{\theta}^\star$;*
3. *f is continuous.*

Remark 4. This assumption provides minimal regularity conditions for consistency. In what follows, we will suppose that the function f belongs to a space $\mathcal{F}$ that will not be specified explicitly. In the discussion of the theoretical results, we will assume that f has s continuous derivatives; in that case, $\mathcal{F}$ will coincide with the Sobolev space $\mathcal{W}_{s,\infty} = \{f : |f|_s < \infty\}$.

We rewrite equation (3.2) using the matrix notation. First of all, we create the $(P \times 1)$-vectors $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_P)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_P)'$. Hence, equation (3.2) becomes:

$$\widehat{f}(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}. \tag{3.3}$$

Then, we make the following assumption concerning the behavior of the errors.

Assumption 5. *The errors respect the following assumptions:*

1. *The sampling error $\boldsymbol{\varepsilon}$ follows a distribution with mean $\mathbb{E}(\boldsymbol{\varepsilon}) = 0$ and variance $\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma}$.*
2. *$\boldsymbol{\Sigma}$ is a $(P \times P)$-positive definite matrix and each element of the covariance matrix is finite.*

Remark 6. The errors $(\varepsilon_1, \ldots, \varepsilon_P)$ are correlated through the series y_n:

$$\mathbf{Cov}(\varepsilon_r, \varepsilon_i) = \mathbf{Cov}\left(\widehat{f}(\theta_r), \widehat{f}(\theta_i)\right) = \mathbf{Cov}\left(D(y_n, z_m(\theta_r)), D(y_n, z_m(\theta_i))\right),$$

with $r, i = 1, \ldots, P$ and $i \neq r$, but this correlation decrease as a function of N (and M). Moreover, we impose standard regularity conditions on the covariance matrix.

To perform estimation, we approximate the function $f(\boldsymbol{\theta})$ through a function $f_K(\boldsymbol{\theta})$ given by an expansion in a series of basis functions:

$$f_K(\boldsymbol{\theta}) = \boldsymbol{\psi}_K(\boldsymbol{\theta})\boldsymbol{\beta},$$

where $\boldsymbol{\psi}_K(\boldsymbol{\theta}) = [\psi_{1K}(\boldsymbol{\theta}), \ldots, \psi_{KK}(\boldsymbol{\theta})]$ is the $(1 \times K)$-vector of known basis functions, while $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]'$ is a $(K \times 1)$-vector of unknown coefficients, with $K \to \infty$, with N and M. Let us discuss the construction of the function f_K using some examples. We first consider some cases when $\boldsymbol{\theta}$ is a scalar, i.e. $\boldsymbol{\theta} = \theta$. Moreover, we suppose that $\theta \in [0, 1]$.

Example 7. [Power series] A solution is to use $\psi_{jK}(\theta) = \theta^{j-1}$, for any j. In this case, $\zeta_s(K) \lesssim K^{1+2s}$.

Example 8. [Orthogonal polynomial series] An alternative is to use orthogonal polynomials as, e.g., Legendre polynomials, that are orthonormal with respect to the Lebesgue measure on $[0,1]$:

$$\boldsymbol{\psi}'_K(\theta) = \left(1, \sqrt{3}\theta, \sqrt{5/4}\left(3\theta^2 - 1\right), \dots\right).$$

The value of $\zeta_s(K)$ does not change.

Example 9. [Spline series] A spline series of order 1 starts from a finite number of equally spaced knots $\ell_1, \dots, \ell_{k-2}$ in $[0,1]$ and defines:

$$\boldsymbol{\psi}_K(\theta) = \left(1, \theta, (\theta - \ell_1)_+, \dots, (\theta - \ell_{k-2})_+\right)'.$$

The cubic splines or spline series of order 3 starts instead from the equally spaced knots $\ell_1, \dots, \ell_{k-4}$ to get:

$$\boldsymbol{\psi}_K(\theta) = \left(1, \theta, \theta^2, \theta^3, (\theta - \ell_1)_+^3, \dots, (\theta - \ell_{k-4})_+^3\right)'.$$

This can be generalized to an arbitrary order s_0. It is often the case that, instead of splines, B-splines are used. In this case, $\zeta_s(K) \lesssim K^{\frac{1}{2}+s}$.

Other examples are in Cox (1988), D. W. K. Andrews (1991), Huang (1998), Chen (2007) and Belloni et al. (2015).

Now we cover the case when $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ is a vector where each component is supposed to belong to $[0,1]$.

Example 10. [Tensor products] In this case, the solution is to take a series $\boldsymbol{\psi}_{K_i}(\theta_i)$ for any $i = 1, \dots, d$. The vector $\boldsymbol{\psi}_K(\boldsymbol{\theta})$ is then built as the tensor product of the previous ones, i.e.:

$$\boldsymbol{\psi}_K(\boldsymbol{\theta}) = \boldsymbol{\psi}_{K_1}(\theta_1) \otimes \cdots \otimes \boldsymbol{\psi}_{K_d}(\theta_d).$$

The number of terms is $K = \prod_{i=1}^d K_i$. The value of $\zeta_s(K)$ is the same of the corresponding method in the scalar case.

Example 11. [Total degree space of monomials] The previous solution contains elements of order higher than each K_i. As an example, if each $\boldsymbol{\psi}_{K_i}(\theta_i)$ is a polynomial series, we will observe a term of order $K_1 + K_2$ like $\theta_1^{K_1}\theta_2^{K_2}$ but we will not observe $\theta_1^{K_1+K_2}$. In some cases, it is possible to build $\boldsymbol{\psi}_K(\boldsymbol{\theta})$ as a union of forms, where a form is a homogeneous polynomial (as in linear or quadratic form). Migliorati (2015) calls total degree space the set of monomials of degree smaller than a certain value. As an example, if $\boldsymbol{\theta} = (\theta_1, \theta_2)$ we have:

$$\boldsymbol{\psi}_K(\boldsymbol{\theta}) = \left(1, \theta_1, \theta_2, \theta_1^2, \theta_1\theta_2, \theta_2^2, \dots\right)'.$$

In the general case, if the degree of each scalar polynomial is k, the number of terms is $K = \frac{(d+k)!}{k!d!} \sim \frac{k^d}{d!}$, where the last relation holds for large k. The corresponding product of tensors is:

$$\boldsymbol{\psi}_K(\boldsymbol{\theta}) = \left(1, \theta_1, \theta_2, \theta_1\theta_2, \theta_1^2, \theta_2^2, \theta_1^2\theta_2^2, \dots\right)'.$$

The value of $\zeta_s(K)$ can be bounded from above by $\zeta_s^d(k)$.

By stacking the vector $\boldsymbol{\psi}_K(\boldsymbol{\theta})$ at every point $\{\boldsymbol{\theta}_i, i = 1, \ldots, P\}$, we can define the $(P \times K)$-matrix

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_{1K}(\theta_1) & \cdots & \psi_{KK}(\theta_1) \\ \vdots & \ddots & \vdots \\ \psi_{1K}(\theta_P) & \cdots & \psi_{KK}(\theta_P) \end{bmatrix},$$

and the DGP can be rewritten as follows:

$$\widehat{f}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\boldsymbol{\beta} + U. \tag{3.4}$$

Rearranging the terms we have $f(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} = \boldsymbol{\Psi}\boldsymbol{\beta} + U$ and $U = f(\boldsymbol{\theta}) - \boldsymbol{\Psi}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

The true value $\boldsymbol{\beta}^\star$ of the unknown coefficients $\boldsymbol{\beta}$ is such that

$$\mathbb{E}\left[\boldsymbol{\Psi}'\left(\widehat{f}(\boldsymbol{\theta}) - \boldsymbol{\Psi}\boldsymbol{\beta}^\star\right)\right] = 0,$$

where the expectation is taken with respect to the distribution of $\widehat{f}(\boldsymbol{\theta})$, that is

$$\boldsymbol{\beta}^\star = \left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\mathbb{E}\left[\widehat{f}(\boldsymbol{\theta})\right].$$

Assuming $\mathbb{E}(U) = 0$ and $\mathbb{V}(U) = \mathbf{I}_P$, where $\mathbf{I}_P$ is the $(P \times P)$ identity matrix, the sieve estimator $\widehat{\boldsymbol{\beta}}$ can be computed using ordinary least squares (OLS):

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\widehat{f}(\boldsymbol{\theta}). \tag{3.5}$$

Finally, we have

$$\widehat{f}_K(\boldsymbol{\theta}) \coloneqq \boldsymbol{\Psi}\widehat{\boldsymbol{\beta}}, \quad f_K^\star(\boldsymbol{\theta}) \coloneqq \boldsymbol{\Psi}\boldsymbol{\beta}^\star,$$

and

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min}\,\widehat{f}_K(\boldsymbol{\theta}).$$

4 Theoretical results

In order to verify the theoretical properties of our estimator, we can build on the results of Seri, Centorrino, and Bernasconi (2019). Let us start by defining the norm:

$$\begin{aligned} |u|_s &\coloneqq \max_{|\lambda|\leqslant s} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\partial^\lambda u(\boldsymbol{\theta})\right|, \\ \|u\|_\infty &\coloneqq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |u(\boldsymbol{\theta})|. \end{aligned}$$

Then, for every $s \geq 0$, we define the following quantity which will be necessary throughout the paper:

$$\zeta_s(K) \coloneqq \max_{|\lambda| \leqslant s} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \partial^{\lambda} \boldsymbol{\psi}_K(\boldsymbol{\theta}) \right\|_F,$$

with $\|\cdot\|_F$ the Frobenius norm. For every integer $s \geq 0$, we define:

$$N_K \coloneqq |f - f_K^{\star}|_s .$$

For $s \geq 0$, we define $\lambda_P = \lambda_{\max}(\boldsymbol{\Sigma})$ as the largest eigenvalue of the covariance matrix, and $\tau_P = tr(\boldsymbol{\Sigma})$ as the trace of the covariance matrix.

The first theoretical result we want to provide concerns an upper bound for the rate of convergence of the sieve estimator $\widehat{f}_K$ to f. To find it and, consequently, to prove consistency, we have to make specific assumptions. In particular, Assumption 12, together with the above-listed definitions, allows to define the bound.

Assumption 12. *For $P \to \infty$, with K fixed, $\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)$ converges in Frobenius norm to a given matrix Q_K, which means that:*

$$\left\| \frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P} - Q_K \right\|_F \to 0,$$

whose smallest eigenvalue $\lambda_{\min}(Q_K) > 0$. For every $s \geq 0$, there exists a finite $\zeta_s(K)$, and $\zeta_s(K) \geq 1$ for K large enough.

Assumption 12 characterizes the asymptotic behavior of the design matrix $\boldsymbol{\Psi}'\boldsymbol{\Psi}$ and, for fixed K, implies that the eigenvalues of $\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}$ converge to the eigenvalues of Q_K. This assumption is needed since we deal with deterministic regressors. If the regressors are supposed to be stochastic as in the case of, e.g., Newey (1997), de Jong (2002) and Belloni et al. (2015), this convergence can be reached by invoking a Law of Large Number.

Remark 13. As in Cox (1988, p. 714), we define the *design measure*, i.e. the discrete uniform distribution supported by the values $\{\boldsymbol{\theta}_i, i = 1, \ldots, P\}$:

$$\mathbb{P}_P(A) := \frac{1}{P} \sum_{i=1}^{P} \mathbb{1}\{\boldsymbol{\theta}_i \in A\}$$

where A is a Borel set in $\mathbb{R}^d$. By choosing the points $\boldsymbol{\theta}_i$ in such a way that their empirical probability converges to the *asymptotic design measure* $\mathbb{P}$, we can obtain an explicit expression for $Q_K = \mathbb{E}\left[\boldsymbol{\psi}_K(\boldsymbol{\theta})' \boldsymbol{\psi}_K(\boldsymbol{\theta})\right]$, where the expectation is taken with respect to $\mathbb{P}$.

The following assumption allows us to obtain consistency.

Assumption 14. *For $P, N, M \to \infty$, we have:*

$$\zeta_s(K) \left(\frac{K\lambda_P \wedge \tau_P}{P} \right)^{\frac{1}{2}} \to 0.$$

For $s = 0$, as $K, P \to \infty$, $N_K \to 0$. For $s > 0$, as $K, P \to \infty$, $\zeta_s(K) N_K \to \infty$.

The quantity $\zeta_s(K)$ is a measure of the complexity of the sieve spaces, while N_K is a measure of how easy it is to approximate f through the sieve spaces. Assumptions 12 and 14 are standard in the sieve literature. Together, these assumptions can be exploited to bound the approximation error and to find a uniform upper bound on the derivative of the vector of basis functions. The following theorem gives an upper bound for the rate of convergence of $\widehat{f}_K$ to f.

Theorem 15. *Under Assumptions 3(i), 5 and 12, we can derive the following uniform bound:*

$$\left|\widehat{f}_K - f\right|_s = O_{\mathbb{P}}\left\{\zeta_s(K)\left[\frac{(K\lambda_P)\wedge\tau_P}{P}\right]^{\frac{1}{2}} + \zeta_s(K)N_K\right\}.$$

If also Assumption 14 holds, the bound is $o(1)$ and the sieve estimator is consistent for f for the norm $|\cdot|_s$.

The following Corollary provides a results on the convergence of $\widehat{\boldsymbol{\theta}}$ to the true value $\boldsymbol{\theta}^\star$.

Corollary 16. *Under Assumptions 3-14:*

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min}\widehat{f}_K(\boldsymbol{\theta}) \xrightarrow{\Pr} \boldsymbol{\theta}^\star = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min} f(\boldsymbol{\theta}).$$

Corollary 16 is useful to obtain the rate of convergence. First of all we state the following assumption:

Assumption 17. *The functions f and $\boldsymbol{\psi}_K$ respect the following differentiability requirements:*

1. *f is twice differentiable on $\operatorname{int}(\boldsymbol{\Theta})$.*
2. *$\boldsymbol{\psi}_K$ is differentiable on $\operatorname{int}(\boldsymbol{\Theta})$ for any K.*
3. *$\lambda_{\min}\{f''(\boldsymbol{\theta})\} \geq \varepsilon > 0$ for $\boldsymbol{\theta} \in N(\boldsymbol{\theta}^\star)$.*

Remark 18. Assumptions 17 (1) and 17 (2) require the function and the vector of basis functions, for any K, to be differentiable on the interior of the parameter space. Assumption 17 (3) states that the smallest eigenvalue of the Hessian must be bounded away from zero in a neighborhood of the true value. This is equivalent to the requirement that the function is strongly convex.

Theorem 19. *Under Assumptions 3-17, the following rate of convergence holds true:*

$$\left\|\boldsymbol{\theta}^\star - \widehat{\boldsymbol{\theta}}\right\| = O\left(\left|\widehat{f}_K - f\right|_1\right).$$

Remark 20. In most cases $\lambda_P \leq \tau_P \sim \frac{P}{N\wedge M}$. Therefore, the bound in Theorem 15 becomes

$$\begin{aligned}\left|\widehat{f}_K - f\right|_s &= O_{\mathbb{P}}\left\{\zeta_1(K)\left[\frac{P}{(N\wedge M)P}\right]^{\frac{1}{2}} + \zeta_1(K)N_K\right\} \\ &= O_{\mathbb{P}}\left\{\zeta_1(K)\left[\frac{1}{\sqrt{N\wedge M}} + N_K\right]\right\}.\end{aligned}$$

Provided that $K/P \to 0$, Assumption 12 ensures consistency, as it implies that $P \to \infty$. For any finite values of K and P, with $K \leq P$, Assumption 12 can be strengthened to ensure that the design matrix is nonsingular. However, since K is finite, the bias component does not vanishes asymptotically, leading to an inconsistent estimator. In our framework, allowing P to approach infinity yields a consistent estimator of the function f. Parametric convergence rates can be achieved for $s = 0$ when both P and K increase sufficiently quickly.

5 Monte Carlo experiments

In this section we report the results of some simulation experiments, of increasing complexity, intended to verify the finite-sample properties of our estimator. We start by providing an example concerning the estimation of the moving-average (MA) parameter θ of a MA(1) process. Although it could appear a simple case, this estimation problem is highly nonlinear and its solution often requires numerical methods. As such, this framework explain well the behavior of our estimator in different situations. Moreover, it allows us to compare the performances of the SNPMD estimator with well-established techniques as II and Maximum Likelihood Estimation through Kalman Filter (KF-MLE).

The second example regards the estimation of the autoregressive parameter ϕ_1 and the variance of the errors σ_ε of a stochastic volatility (SV) model (Harvey, Ruiz, and Shephard, 1994). We decide to adopt this model for two main reasons: (i) the parameters are often estimated using II; (ii) the parameters can be represented as a linear combination of an infinite number of basis functions, which means that the relation between parameters and distribution functions is fully nonparametric.

5.1 Estimation of the MA parameter of a MA(1) model

We consider as benchmark series a first order moving average model of the form

$$y_n = \varepsilon_n + \theta \varepsilon_{n-1},$$

for $n = 1, \ldots, N$, where the innovations $\varepsilon_n \sim \mathcal{N}(0, 1)$ and $\varepsilon_0 = 0$. The true value used to simulate the benchmark series is $\theta^\star = 0.5$.

The estimation algorithm used for the Monte Carlo experiments is summarized in Algorithm 1.

The simulation experiment, described in Algorithm 1, is performed using different configurations. Specifically, to reconcile the theoretical results, we vary the number of polynomials K, the number of points P on the grid and the sample sizes N, M. The values taken by these quantities are the following: $K = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $N, M = \{100, 1000\}$, and the parameter $\theta_i = \{-0.99, \ldots, 0.99\}$, with $i = 1, \ldots, P$, used to simulate the model, is taken on an equispaced grid of cardinality $P = \{201, 401, 801, 1201, 1601\}$.[3] Finally, we have the following DGP for the

[3] We reckon on an equispaced grid for two reasons: (i) it is a "worst-case" scenario since it is well known that other configurations of points of smaller cardinality provide equivalent approximations; (ii) we comply with the design matrix used in so-called computational experiments by the most popular simulation softwares.

Algorithm 1 Computing the SNPMD estimator for the MA(1) model

Set the dimensions for N, M, the MC runs S, and choose the cardinality P
Define the vector $\mathbf{k}$ of polynomials degree
for $k \in \mathbf{k}$ **do**
  for $s = 1, \ldots, S$ **do**
    Simulate the benchmark series $y_n = (y_1, \ldots, y_N)$
    Compute the process distribution ρ^y
    Select a grid of points $\{\boldsymbol{\theta}_i, i = 1, \ldots, P\} \subset \boldsymbol{\Theta}$ and, for any $\boldsymbol{\theta}_i$ belonging to the grid, simulate a series $z_m^s(\boldsymbol{\theta}) = (z_1^s(\boldsymbol{\theta}), \ldots, z_M^s(\boldsymbol{\theta}))$
    For any $\boldsymbol{\theta}_i$, compute the process distribution $\rho_{\boldsymbol{\theta}_i}^z$
    Compute the distance between ρ^y and $\rho_{\boldsymbol{\theta}_i}^z$, i.e. $\widehat{f}(\boldsymbol{\theta})$
    Compute the sieve regression $\widehat{f}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\boldsymbol{\beta} + U$
    Estimate $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\widehat{f}(\boldsymbol{\theta})$ and compute $\widehat{f}_K(\boldsymbol{\theta}) = \boldsymbol{\Psi}\widehat{\boldsymbol{\beta}}$
    Find the parameter values minimizing $\widehat{f}_K(\boldsymbol{\theta})$
  end for
end for

simulation model:

$$z_m^s(\theta_i) = \varepsilon_m^s + \theta_i \varepsilon_{m-1}^s,$$

for $m = 1, \ldots, M$ and $s = 1, \ldots, S$, where the innovations $\varepsilon_m^s \sim \mathcal{N}(0, 1)$ and $\varepsilon_0^s = 0$. We replicate the experiment $S = 10,000$ times, and we compute the bias, standard deviation (SD) and sample root mean squared error (RMSE) of the estimator of θ across Monte Carlo runs. The bias, SD and sample RMSE are compared with the ones of the II estimator that exploits a third order autoregressive process as auxiliary model (Gouriéroux, Monfort, and Renault, 1993), and the maximum likelihood estimator through Kalman Filter, which is known to be efficient.

Figure 5.1 shows the behavior of the bias, SD and RMSE of the SNPMD estimator for different values of K (x-axes), $P = 201$ (solid line), $P = 401$ (dashed line), $P = 801$ (dotted line), $P = 1201$ (dot-dashed line) and $P = 1601$ (long-dashed line). The windows in each panel represent different sample sizes, i.e. $N = M = 100$ (left), $N = 1000$ and $M = 100$ (center), $N = M = 1000$ (right). The horizontal gray dotted lines represent the bias, SD and RMSE of the KF-MLE for different combinations of sample sizes, while the horizontal gray dashed lines refer to the performances of the II estimator.

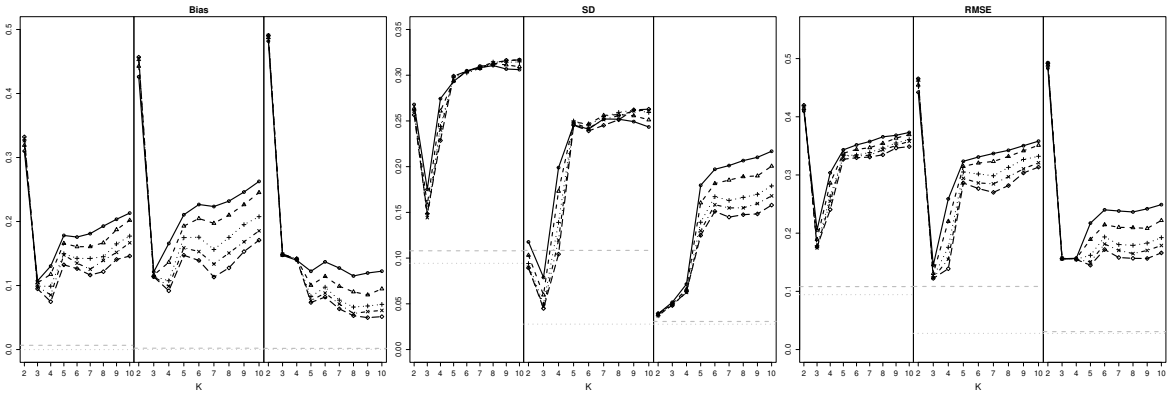

Figure 5.1: Behavior of the Bias, SD and RMSE of the SNPMD estimator for for different combinations of K, P, N and M, with respect to II and KF-MLE.

We can state some considerations about the impacts of K, P, N and M on the performance of the SNPMD estimator. Let us start by studying the role of P. The comparison with KF-MLE and II reveals that the SNPMD estimator exhibits a lower bias for higher values of P. This effect is partially balanced by the behavior of the SD, which is less clear for smaller sample sizes. Overall, the SNPMD RMSE approaches the RMSE of the II estimator and KF-MLE as $P \to \infty$. This is in line with Theorem 19 and Remark 20 as, for finite K and $P \to \infty$, $\widehat{f}_K$ is a consistent estimator of f.

The dependence of the bias, SD and RMSE on K is not monotonic. In fact, the bias seems to decrease as K increases suggesting that the accuracy of the estimator improves with larger K. On the other hand, the SD is lower for smaller values of K and increases for $K > 3$. Finally, the RMSE attains its minimum value when $K = 3$ and increases for $K \geq 4$, reaching a plateau for $K \geq 5$. This suggests that the polynomials degree must be chosen judiciously to avoid overfitting (i.e. relying on some information criteria or cross-validation techniques). Also this situation is compliant with our theoretical results as, for finite K, consistency can be achieved provided that ${}^{K}/_{P} \to 0$.

Finally, for $N, M \to \infty$ the bias, SD and RMSE decrease. In line with the bound in Remark 20, the convergence is driven by $\min\{N, M\}$. Indeed, the case in which $N = 1000$ and $M = 100$ and the case in which $N = M = 100$ provide similar results, meaning that, when $M \neq N$, the shortest sample size influence the convergence of the estimator.

5.2 Estimation of the parameters of a stochastic volatility model

Before discussing the estimation results, we introduce the structure of the SV model. Let us define the series of interest $y_i = \sigma_i \eta_i$, with $i = 1, \ldots, N$, where $\eta_i \sim \mathcal{N}(0, 1)$ and σ_i is the standard deviation. We also define $\log \sigma_i^2 = h_i$, where h_i is modeled according to an autoregressive process of order 1,

$$h_i = -0.5 + \phi_1 h_{i-1} + \varepsilon_i,$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon)$. Since $y_i = \eta_i \exp\left(\frac{1}{2} h_i\right)$, we have

$$\log y_i^2 = h_i + \log \eta_i^2,$$

where the mean and variance of $\log \eta_i^2$ are known to be -1.27 and $\pi^2/2 = 4.93$, respectively (see Harvey, Ruiz, and Shephard, 1994). The model admits the following state-space representation, which is more convenient for estimation:

$$\begin{aligned} \log y_i^2 &= -1.27 + h_i + \xi_i, \\ h_i &= -0.5 + \phi_1 h_{i-1} + \varepsilon_i, \end{aligned}$$

where $\xi_i = \log \eta_i^2 + 1.27$ and $\mathbb{V}(\xi_i) = \pi^2/2$. The series to be matched is $y_n = \log y_i^2$, the parameters to be estimated are the autoregressive parameter ϕ_1 and the standard deviation of the errors σ_ε, and the vector of the true parameters is denoted with $\boldsymbol{\theta}^\star = (\phi_1^\star, \sigma_\varepsilon^\star)$, where $\phi_1^\star = 0.9$ and $\sigma_\varepsilon^\star = 0.35$.

The estimation algorithm used for the Monte Carlo experiments is described in Algorithm 2.

Algorithm 2 Computing the SNPMD estimator for the SV model

1: Set the dimensions for N, M, S and the polynomials degree of K
2: **for** $s = 1, \ldots, S$ **do**
3: Simulate the benchmark series $y_n = (y_1, \ldots, y_N)$
4: Compute the process distribution ρ^y
5: Select a grid of points $\{\boldsymbol{\theta}_i, i = 1, \ldots, P\} \subset \boldsymbol{\Theta}$ and, for any $\boldsymbol{\theta}_i$ belonging to the grid, simulate a series $z_m^s(\boldsymbol{\theta}) = (z_1^s(\boldsymbol{\theta}), \ldots, z_M^s(\boldsymbol{\theta}))$
6: For any $\boldsymbol{\theta}_i$, compute the process distribution $\rho_{\boldsymbol{\theta}_i}^z$
7: Compute the distance between ρ^y and $\rho_{\boldsymbol{\theta}_i}^z$, i.e. $\widehat{f}(\boldsymbol{\theta})$
8: Compute the sieve regression $\widehat{f}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\boldsymbol{\beta} + U$
9: Estimate $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1} \boldsymbol{\Psi}'\widehat{f}(\boldsymbol{\theta})$ and compute $\widehat{f}_K(\boldsymbol{\theta}) = \boldsymbol{\Psi}\widehat{\boldsymbol{\beta}}$
10: Find the parameter values minimizing $\widehat{f}_K(\boldsymbol{\theta})$
11: **end for**

As for the previous exercise, the simulation experiment is performed using different configurations (see Table 1). To simulate the model, we select a grid of P points $\{(\phi_{1i}, \sigma_{\varepsilon i}), i = 1, \ldots, P\} \subset \Theta = (0, +\infty) \times (0, +\infty)$. The values of ϕ_1 are chosen according to an equispaced grid with range $[0, 1]$ and cardinality 21, while the values of σ_ε are chosen according to an equispaced grid with range $[0, 2]$ and cardinality 21. The final values of $\boldsymbol{\theta} = (\phi_1, \sigma_\varepsilon)$ form a two-dimensional grid in $\boldsymbol{\Theta}$, i.e. the parameter values are chosen according to a full factorial design. The experiment is replicated $S = 1000$ times.

In Table 1 we report the bias, SD and RMSE of the estimators of ϕ_1 and σ_ε, together with the coverage probability useful to quantify the small-sample approximation error. The coverage probability is computed by first building the 90% confidence intervals based on Hall (2013) percentiles using the S Monte Carlo runs, and then counting how many times the estimated parameter falls within these confidence intervals. The performances of SNPMD are compared with the ones of the II estimator that exploits a generalized autoregressive conditional heteroskedasticity, GARCH(1,1), model as auxiliary model.

Table 1: Bias, SD and sample RMSE of the estimator of the autoregressive parameter ϕ_1 and the SD of the errors σ_ε in the SV model.

				$\widehat{\phi}_1$				$\widehat{\sigma}_\varepsilon$			
Method	K	N	M	Bias	SD	RMSE	Cov.	Bias	SD	RMSE	Cov.
SNPMD	2										
		250	250	-0.0821	0.1323	0.1556	91.7	0.0876	0.0868	0.1233	87.7
		500	250	-0.0981	0.0630	0.1166	90.9	0.0748	0.0433	0.0864	90.1
		500	500	-0.0573	0.1250	0.1374	89.1	0.1061	0.0965	0.1434	86.9
SNPMD	3										
		500	250	-0.0765	0.1397	0.1592	90.3	0.0891	0.0883	0.1254	88.3
		500	250	-0.0972	0.0649	0.1168	90.7	0.0780	0.0489	0.0921	89.4
		500	500	-0.0567	0.1331	0.1446	89.7	0.1037	0.0894	0.1369	85.1
SNPMD	4										
		250	250	-0.0817	0.1254	0.1496	90.7	0.0896	0.0930	0.1291	87.8
		500	250	-0.1014	0.0628	0.1193	91.0	0.0741	0.0423	0.0853	89.4
		500	500	-0.0560	0.1333	0.1445	89.9	0.1030	0.0914	0.1377	85.9
SNPMD	5										
		250	250	-0.0865	0.1080	0.1383	90.2	0.0827	0.0766	0.1127	88.9
		500	250	-0.0989	0.0573	0.1143	92.0	0.0735	0.0395	0.0834	88.9
		500	500	-0.0571	0.1235	0.1360	88.9	0.1087	0.0956	0.1447	84.8
SNPMD	6										
		250	250	-0.0821	0.1228	0.1476	91.2	0.0880	0.0838	0.1215	88.1
		500	250	-0.0983	0.0595	0.1149	89.5	0.0762	0.0474	0.0898	89.9
		500	500	-0.0543	0.1403	0.1503	89.0	0.1016	0.0885	0.1348	85.6
II											
		250	250	0.0021	0.0628	0.0628	89.0	-0.0508	0.1309	0.1404	88.3
		500	250	0.0076	0.0619	0.0624	87.3	-0.0638	0.1225	0.1380	89.8
		500	500	-0.0035	0.0632	0.0633	89.6	-0.0076	0.1456	0.1458	87.2

Notes: Cov. denotes the coverage probability.

For ϕ_1, the SNPMD method shows a small negative bias across different sample sizes and configurations, indicating a tendency to underestimate the true parameter. However, the bias term seems to decrease as N and M increase. For σ_ε, the SNPMD estimator generally has a small positive bias, which is close to the (absolute) values of the II estimator. The II estimator shows a smaller bias for ϕ_1, but a negative bias for σ_ε.

The SD for the SNPMD estimator of ϕ_1 is lower for $N > M$ (e.g., $N = 500$ and $N = 250$), indicating more precise estimates as $N \to \infty$. The SD of the SNPMD estimator of σ_ε also decreases with larger sample sizes. Furthermore, the latter compares favorably with respect to the II estimator.

The RMSE for the SNPMD estimator of ϕ_1 and σ_ε tends to decrease with larger sample sizes, reflecting improved estimation accuracy. II seems to perform slightly better than our method in the estimation of ϕ_1. On the other hand, the SNPMD technique seems to provide better estimates of σ_ε.

The coverage probability for $\widehat{\phi}_1$ is generally close to the nominal 90% level, indicating that the confidence intervals are accurate. For $\widehat{\sigma}_\varepsilon$, the coverage probability is slightly lower, suggesting some underestimation of the uncertainty in the error standard deviation. Overall, the coverage probability levels of the SNPMD estimator are higher that the ones of the II estimator.

6 Empirical application

In this section we provide an application of our method to the Brock and Hommes (BH) model (Brock and Hommes, 1997; Brock and Hommes, 1998). We decide to adopt this model for three reasons: (i) it is a workhorse in the literature of calibration and estimation of heterogeneous agents models (see, e.g., Boswijk, Hommes, and Manzan, 2007; Recchioni, Tedeschi, and Gallegati, 2015; Lamperti, Roventini, and Sani, 2018); (ii) by construction, it does not allow to recycle innovations leading to the violation of the stochastic equicontinuity hypothesis; (iii) nevertheless its simplicity, the estimation of its parameters is still an open issue.

6.1 The model

The BH model consists in an asset pricing model with heterogeneous beliefs. Agents can choose among a finite number of predictors of future prices of a risky asset and revise their expectation in each period in a bounded rational way, according to a fitness measure based on past realized profits. For our purposes, we refer to the simplified version proposed by Hommes (2006).

Traders have the option to invest in either a risk-free asset, which is supplied perfectly elastically at a gross return $R = (1 + r) > 1$, where r is the risk-free return, or a risky asset that pays an uncertain dividend y and has a price of p. The dynamics of wealth are described by:

$$W_{t+1} = RW_t + (p_{t+1} + y_{t+1} - Rp_t)z_t, \tag{6.1}$$

where W_{t+1}, p_{t+1}, and y_{t+1} are random variables, and z_t represents the number of shares of the risky asset purchased at time t. The market provides a publicly available information set on past prices and dividends, allowing for the definition of the conditional expectation $\mathbb{E}_b$ and variance $\mathbb{V}_b$. Agents have different forecasts (beliefs) about these conditional expectations and variances. Specifically, $\mathbb{E}_{ht}$ and $\mathbb{V}_{ht}$ denote the beliefs of trader h regarding these operators. Each investor is assumed to be a myopic mean-variance maximizer, so her demand z_{ht} for the risky asset is determined by solving:

$$\max_{z_t} \left[\mathbb{E}_{ht}(W_{t+1}) - \frac{\alpha}{2} \mathbb{V}_{ht}(W_{t+1}) \right], \tag{6.2}$$

which leads to:

$$z_{ht} = \frac{\mathbb{E}_{ht}(p_{t+1} + y_{t+1} - Rp_t)}{\alpha \sigma^2}. \tag{6.3}$$

Here, α is the risk aversion parameter, and σ^2 is the conditional variance, assumed to be constant and the same for all traders.

The market equilibrium equation, when there is zero supply of outside shares and different types of traders h, can be expressed as:

$$Rp_t = \sum_{h=1}^{H} n_{ht} \mathbb{E}_{ht}(p_{t+1} + y_{t+1}), \tag{6.4}$$

where n_{ht} represents the fraction of agents of type h at time t. In a scenario where all traders are

identical and homogeneous, the arbitrage market equilibrium equation with rational expectations can be derived from the above equation:

$$Rp_t^\star = \mathbb{E}_t(p_{t+1}^\star + y_{t+1}), \tag{6.5}$$

where $p_t^\star$ denotes the fundamental price, and $\mathbb{E}_t$ is the conditional expectation based on the information set $I_t = [p_{t-1}, p_{t-2}, \ldots; y_{t-1}, y_{t-2}, \ldots]$.

Agents' strategy is driven by an evolutionary dynamics described by a set of equations. The first equation concerns the agents' expectation on the spot price at time t, f_{ht}:

$$f_{ht} = g_h x_{t-1} + b_h, \tag{6.6}$$

where g_h is the trend component and b_h is the bias affecting each trader. In our application we consider only two unbiased agents, "fundamentalists" and "trend followers". Hence, $h = 1, 2$, $b_1 = b_2 = 0$ and, for $h = 1$, $g_h = 0$. The second equation relates to the deviation from the fundamental price, x_t:

$$(1 + r)\, x_t = \sum_{h=1}^{H} n_{ht}\, (f_{ht}) + \varepsilon_t, \tag{6.7}$$

where the random error ε_t due to the uncertainty about the risky asset is i.i.d. uniformly distributed between -0.5 and $+0.5$ (see Brock and Hommes, 1998). The deviation from fundamental price $p_t^\star$ is expressed as $x_t = p_t - p_t^\star$, where, in the special case of i.i.d. dividends with constant mean $\mathbb{E}(d_t) = \overline{d}$, $p_t^\star = \frac{d}{r}$ and p_t is the price of the risky asset at time t. The third equation regards the share of "fundamentalists"/"trend followers" at time t, n_{ht}, modeled according to a multinomial logit:

$$n_{ht} = \frac{\exp\left(\beta U_{h,t-1}\right)}{\sum_{h=1}^{H} \exp\left(\beta U_{h,t-1}\right)}. \tag{6.8}$$

A key role is played by the parameter β, which represents the intensity of choice of the agents. The higher is the value of β the higher is the probability to switch strategy (from "fundamentalists" to "trend followers" and viceversa). Finally, the last equation defines the fitness measure of the strategy, $U_{h,t-1}$:

$$U_{h,t-1} = (x_{t-1} - Rx_{t-2}) \left(\frac{g_h x_{t-3} + b_h - Rx_{t-2}}{\alpha \sigma^2} \right) + wU_{h,t-2}, \tag{6.9}$$

where w is the weight to past profits.

6.2 Estimation results

The model is estimated using the daily closing information of two stock market indices: the USA Standard&Poor's (S&P) 500, which is an index embodying the biggest corporations listed in the New York Stock Exchange, American Stock Exchange and Nasdaq, and the Euro Stoxx 50, which is an index composed by the main corporations in the Euro area. We decide to choose these indices for two reasons: (i) they represent two geographical areas with different characteristics; (ii) they allow for a direct comparison with other calibration exercises (e.g., Recchioni, Tedeschi, and Gallegati,

2015). The sample size of the S&P 500 ranges from January 2, 2012 to December 31, 2012, for a total of $N = 261$ observations. The sample size of the Euro Stoxx 50 ranges from January 3, 2012 to December 30, 2012, for a total of $N = 251$ observations. The real-world series compared with the simulated data is the log-difference of the daily closing prices.

As common in the literature (see, e.g., Boswijk, Hommes, and Manzan, 2007), we choose to estimate the trend component parameter g_h, with $h = 2$, and the intensity of choice β, and we fix the other parameters according to previous studies (Lamperti, 2018; Recchioni, Tedeschi, and Gallegati, 2015). The model parameters and explored ranges are reported in Table 2.

Table 2: Description of parameters and explored ranges.

Parameter	Description	Support	Range S&P 500	Range Euro Stoxx 50
β	Intensity of choice	$[0, +\infty)$	$[1, 30]$	$[1, 30]$
n_1	Initial share of type 1 traders	$[0, 1]$	0.5	0.5
g_2	Trend component of type 2 traders	$(-\infty, +\infty)$	$[-1, 1]$	$[-1, 1]$
w	Weight to past profits	$[0, 1]$	1	1
σ	Asset volatility	$(0, +\infty)$	0.1	0.1
$p_t^\star$	Fundamental price	$[0, +\infty)$	0.8	0.7
α	Risk aversion	$[0, +\infty)$	23	23
r	Daily risk-free return	$(0, +\infty)$	$0.01/250$	$0.01/250$

The (empirical) estimation procedure is described in Algorithm 3.

Algorithm 3 Computing the SNPMD estimator for the BH model

Set the dimensions for N, M and S
Choose the real-world price index p_n and compute $y_n = \log(p_n) - \log(p_{n-1})$
Compute the process distribution ρ^y
for $s = 1, \ldots, S$ **do**
  Select a grid of points $\{\boldsymbol{\theta}_i, i = 1, \ldots, P\} \subset \boldsymbol{\Theta}$ and, for any $\boldsymbol{\theta}_i$ belonging to the grid, obtain the simulated price $p_m^s(\boldsymbol{\theta})$
  Compute $z_m^s(\boldsymbol{\theta}) = \log(p_m^s(\boldsymbol{\theta})) - \log\left(p_{m-1}^s(\boldsymbol{\theta})\right)$
  For any $\boldsymbol{\theta}_i$, compute the process distribution $\rho^z_{\boldsymbol{\theta}_i}$
  Compute the distance between ρ^y and $\rho^z_{\boldsymbol{\theta}_i}$, i.e. $\widehat{f}(\boldsymbol{\theta})$
  for $k \in \mathbf{k}$ **do**
    Compute the sieve regression and its AIC
  end for
  Select the polynomial degree K with the lowest AIC
  Compute the sieve regression $\widehat{f}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\boldsymbol{\beta} + U$
  Estimate $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\widehat{f}(\boldsymbol{\theta})$ and compute $\widehat{f}_K(\boldsymbol{\theta}) = \boldsymbol{\Psi}\widehat{\boldsymbol{\beta}}$
  Find the parameter values minimizing $\widehat{f}_K(\boldsymbol{\theta})$
end for

We evaluate the objective function on a two-dimensional grid $\{(\beta_i, g_i), i = 1, \ldots, P\} \subset \Theta = [0, +\infty) \times (-\infty, +\infty)$, where $\beta_i = \{1, \ldots, 30\}$ and $g_i = \{-1, \ldots, 1\}$ are taken on an equispaced grid of cardinality $P = 101$. As for the previous Monte Carlo experiments, the parameter values are chosen according to a full factorial design. Therefore, the total number of points is $P = 10201$. To better understand the extent of the estimation problem, in Figure 6.1 we report the values of the

objective functions $\widehat{f}(\boldsymbol{\theta})$ (left plot) and $\widehat{f}_K(\boldsymbol{\theta})$ (right plot) for a specific MC run.

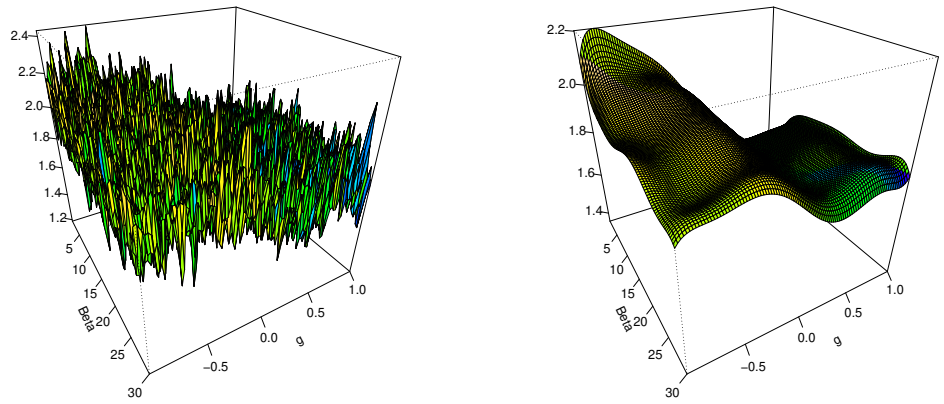

Figure 6.1: Objective function $\widehat{f}(\boldsymbol{\theta})$ (left) and its smoothed approximation $\widehat{f}_K(\boldsymbol{\theta})$ (right).

The mean and standard error, across $S = 10,000$ replications, of the estimated parameters for the S&P 500 and Euro Stoxx 50 indices are reported in Table 3.

Table 3: Mean and standard error, across MC runs, of the estimated parameters.

Parameter	S&P 500	Euro Stoxx 50
$\widehat{\beta}$	4.123	8.292
(S.E.)	(0.1741)	(0.0804)
$\widehat{g}_2$	0.308	0.349
(S.E.)	(0.0097)	(0.0059)

Notes: S.E. denotes the Standard Error.

The estimated $\widehat{\beta}$ values are 4.123 for the S&P 500 and 8.292 for the Euro Stoxx 50. The higher value for the Euro Stoxx 50 suggests a stronger relationship or sensitivity in the model for this index compared to the S&P 500. Specifically, the higher value for Euro Stoxx 50 indicates that agents acting in this market are characterized by a stronger collective behavior. The higher $\widehat{\beta}$ for the Euro Stoxx 50 might also imply a greater susceptibility to speculative behavior or market instability compared to the S&P 500 (Boswijk, Hommes, and Manzan, 2007). In general, the difference in $\widehat{\beta}$ values illustrates that, in contrast with the original BH model, this parameter is not of crucial importance in switching models (Recchioni, Tedeschi, and Gallegati, 2015; Lamperti, Roventini, and Sani, 2018).

The values of $\widehat{g}_2$ are 0.308 for the S&P 500 and 0.349 for the Euro Stoxx 50, suggesting that agents can be identified as pure trend chasers in both markets. Since the values of $\widehat{g}_2$ are relatively close, we can conclude that the model under analysis describes well the empirical data. Furthermore, as also noted by Lamperti, Roventini, and Sani (2018), among others, $\widehat{g}_2$ plays a key role in fitting the empirical distribution of observed returns.

The standard errors for both parameters are relatively low, suggesting precision in the estimation procedure. This precision is crucial for making reliable inferences about market behavior, understanding market dynamics and predicting financial crises.

To sum-up, the results of the empirical estimation shows differences in market behavior between the S&P 500 and Euro Stoxx 50, which could be further explored in the context of market efficiency, volatility, and speculative behavior.

7 Conclusion

In this paper we introduce a nonparametric minimum-distance estimator that combines simulated minimum-distance techniques with sieve estimation to address the challenges of estimating complex simulation models. Our approach avoids the pitfalls of traditional methods, such as the need for auxiliary models or moment selection, and provides a flexible framework for estimating parameters in models where stochastic equicontinuity condition may not hold.

We study the theoretical properties of the SNPMD estimator by stating conditions for consistency of the estimator and providing nonparametric rates of convergence for the estimator and the functional of the parameters. We acknowledge that the asymptotic rate of convergence of $\widehat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}$ is sub-optimal as we consider a nonparametric estimation technique, but the loss of efficiency in terms of convergence speed is compensated by the flexibility deriving from the nonparametric estimation of the objective function.

We perform several Monte Carlo experiments to demonstrate the finite-sample properties of our estimator, showing its ability to accurately estimate parameters in both moving-average and stochastic volatility models. The empirical application to the BH model further highlights the practical relevance of our method in analyzing financial market dynamics. Our results suggest that the proposed estimator is a valuable tool for econometricians and researchers working with complex, heterogeneous models, offering a robust and efficient alternative to existing estimation techniques.

Our framework can be expanded in different directions. First, by considering different statistical metrics to deal with non-ergodic and nonstationary processes. Second, by relying on other smoothing techniques such as splines or Kernel methods. Third, by incorporating machine learning and regularization techniques to mitigate overfitting and curse of dimensionality. We leave these investigations to future research.

8 Proofs

8.1 Proof of distance

Proof of Lemma 1. The proof follows the steps of D. Ryabko and B. Ryabko (2010, pp. 1432-1433). We start noting that for any quantity $\delta > 0$ it is possible to find and index I such that $\sum_{j=I}^{\infty} w_j < \delta/2$. For each ℓ, we also have $\nu\left(\left(y_1, \ldots, y_N\right), A_\ell\right) \to_{\mathsf{a.s.}} \rho^y\left(A_\ell\right)$ so that

$$\left|\nu\left(\left(y_1, \ldots, y_N\right), A_\ell\right) - \rho^y\left(A_\ell\right)\right| < \frac{\delta}{4Iw_\ell} \tag{8.1}$$

for some step N on.

Now, let $K_\ell \coloneqq N$, $K \coloneqq \max_{\ell<I} K_\ell$, which depends on the realization of $(y_1, y_2, \ldots)$, $J_\ell \coloneqq M$, and $J \coloneqq \max_{\ell<I} J_\ell$ that depends on the realization of $\left(z_1\left(\boldsymbol{\theta}\right), z_2\left(\boldsymbol{\theta}\right), \ldots\right)$. Therefore, provided that $N > K$ and $M > J$, we have:

$$\begin{aligned}
&\left|D\left(\widehat{\rho}^y, \widehat{\rho}^z_{\boldsymbol{\theta}}\right) - D\left(\rho^y, \rho^z_{\boldsymbol{\theta}}\right)\right| \\
=&\left|\sum_{j=1}^{\infty} w_j \left(\left|\nu\left(y_n, A_j\right) - \nu\left(z_m\left(\boldsymbol{\theta}\right), A_j\right)\right| - \left|\rho^y\left(A_j\right) - \rho^z_{\boldsymbol{\theta}}\left(A_j\right)\right|\right)\right| \\
\leq&\sum_{j=1}^{\infty} w_j \left(\left|\nu\left(y_n, A_j\right) - \rho^y\left(A_j\right)\right| + \left|\nu\left(z_m\left(\boldsymbol{\theta}\right), A_j\right) - \rho^z_{\boldsymbol{\theta}}\left(A_j\right)\right|\right) \\
\leq&\sum_{j=1}^{I} w_j \left(\left|\nu\left(y_n, A_j\right) - \rho^y\left(A_j\right)\right| + \left|\nu\left(z_m\left(\boldsymbol{\theta}\right), A_j\right) - \rho^z_{\boldsymbol{\theta}}\left(A_j\right)\right|\right) + \frac{\delta}{2} \\
\leq&\sum_{j=1}^{I} w_j \left(\frac{\delta}{(4Iw_\ell)} + \frac{\delta}{(4Iw_\ell)}\right) + \frac{\delta}{2} = \delta,
\end{aligned}$$

where the first inequality comes from the triangle inequality, the second comes from the fact that $\sum_{j=I}^{\infty} w_j < \delta/2$, and the third inequality is obtained by using Equation (8.1). Q.E.D.

8.2 Proof of consistency

We start the proof recalling that, by Assumption 12, we have:

$$\lim_{P\to\infty}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right) = Q_K,$$

where Q_K can be considered an identity matrix $\mathbf{I}_K$ of dimension K. This implies that $\lambda_{\min}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right) \to 1$. We then define the indicator function $\mathbf{1}_P = 1\left\{\lambda_{\min}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right) > c\right\}$, where $0 < c < 1$. Under Assumption 5, we have $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\widehat{f}\left(\boldsymbol{\theta}\right)$, where $\widehat{f}\left(\boldsymbol{\theta}\right) = f\left(\boldsymbol{\theta}\right) + \boldsymbol{\varepsilon}$.

We need the following lemma.

Lemma 21. *Under Assumption 5 and 12:*

$$1_P\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\star\right\| = O_{\mathbb{P}}\left(\left(\frac{(K\lambda_P)\wedge\tau_P}{P}\right)^{\frac{1}{2}}\right)+O\left(N_K\right).$$

Proof. We start demonstrating that:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\star &= \left[\left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\widehat{f}\left(\boldsymbol{\theta}\right)\right]-\left[\left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'f_K^\star\left(\boldsymbol{\theta}\right)\right]\\
&= \left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\left(\widehat{f}\left(\boldsymbol{\theta}\right)-f\left(\boldsymbol{\theta}\right)\right)+\left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\left(f\left(\boldsymbol{\theta}\right)-f_K^\star\left(\boldsymbol{\theta}\right)\right)\\
&= \left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\boldsymbol{\varepsilon}+\left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}'\left(f\left(\boldsymbol{\theta}\right)-f_K^\star\left(\boldsymbol{\theta}\right)\right)\\
&= \frac{1}{P}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\boldsymbol{\Psi}'\boldsymbol{\varepsilon}+\frac{1}{P}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\boldsymbol{\Psi}'\left(f\left(\boldsymbol{\theta}\right)-f_K^\star\left(\boldsymbol{\theta}\right)\right),
\end{aligned}$$

and from the triangle inequality we can write:

$$1_P\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\star\right\| \leq 1_P\left\|\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\varepsilon}}{P}\right)\right\|+1_P\left\|\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\boldsymbol{\Psi}'\left(f\left(\boldsymbol{\theta}\right)-f_K^\star\left(\boldsymbol{\theta}\right)\right)\right\|.$$

Let us consider the first term of the inequality, proceeding in this way:

$$\begin{aligned}
\mathbb{E}\left[1_P\left\|\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\varepsilon}}{P}\right)\right\|^2\right] &= 1_P\mathbb{E}\left[\frac{\boldsymbol{\varepsilon}'\boldsymbol{\Psi}}{P}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\frac{\boldsymbol{\Psi}'\boldsymbol{\varepsilon}}{P}\right]\\
&\leq 1_P\mathbb{E}\mathrm{tr}\left[\frac{\boldsymbol{\varepsilon}'\boldsymbol{\Psi}}{P}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\frac{\boldsymbol{\Psi}'\boldsymbol{\varepsilon}}{P}\right]\lambda_{\max}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\\
&= 1_P\mathbb{E}\mathrm{tr}\left[\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\frac{\boldsymbol{\Psi}'\boldsymbol{\varepsilon}}{P}\frac{\boldsymbol{\varepsilon}'\boldsymbol{\Psi}}{P}\right]\lambda_{\max}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\\
&= 1_P\mathrm{tr}\left[\frac{\boldsymbol{\Psi}}{P}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\frac{\boldsymbol{\Psi}'}{P}\mathbb{E}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)\right]\lambda_{\max}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\\
&= \frac{1_P}{P^2}\mathrm{tr}\left[\boldsymbol{\Psi}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)\boldsymbol{\Psi}'\boldsymbol{\Sigma}\right]\frac{1}{\lambda_{\min}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)}.
\end{aligned}$$

We can derive two different bounds. For the first bound, let us consider the following majorization:

$$\begin{aligned}
\mathbb{E}\left[1_P\left\|\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)^{-1}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\varepsilon}}{P}\right)\right\|^2\right] &\leq \frac{1_P}{P^2}\mathrm{tr}\left[\boldsymbol{\Psi}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)\boldsymbol{\Psi}'\boldsymbol{\Sigma}\right]\frac{1}{\lambda_{\min}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)}\\
&= \frac{1_P}{P^2}\mathrm{tr}\left[\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)\left(\boldsymbol{\Psi}'\boldsymbol{\Sigma}\boldsymbol{\Psi}\right)\right]\frac{1}{\lambda_{\min}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)}\\
&= \frac{1_P}{P^2}\mathrm{tr}\left[\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)\left(\boldsymbol{\Psi}'\boldsymbol{\Psi}\right)\right]\frac{1}{\lambda_{\min}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)}\lambda_{\max}\left(\boldsymbol{\Sigma}\right)\\
&\leq \frac{1_P}{P}\mathrm{tr}\left[\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)\right]\frac{1}{\lambda_{\min}\left(\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{P}\right)}\lambda_{\max}\left(\boldsymbol{\Sigma}\right)
\end{aligned}$$

$$= 1_P \frac{\operatorname{tr}\left(\mathbf{I}_K\right) \lambda_P}{P \lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)} = O\left(\frac{K\lambda_P}{P}\right).$$

The second bound, can be stated as follow:

$$\begin{aligned}
\mathbb{E}\left[1_P \left\| \left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)^{-1} \left(\frac{\mathbf{\Psi}'\boldsymbol{\varepsilon}}{P}\right) \right\|^2\right] &\leq \frac{1_P}{P^2} \operatorname{tr}\left[\mathbf{\Psi}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)\mathbf{\Psi}'\mathbf{\Sigma}\right] \frac{1}{\lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)} \\
&= \frac{1_P}{P^2} \operatorname{tr}\left[\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Psi}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)\mathbf{\Psi}'\mathbf{\Sigma}^{\frac{1}{2}}\right] \frac{1}{\lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)} \\
&= \frac{1_P}{P^2} \operatorname{tr}\left[\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Psi}\left(\mathbf{\Psi}'\mathbf{\Psi}\right)\mathbf{\Psi}'\mathbf{\Sigma}^{\frac{1}{2}}\right] \frac{1}{\lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)} \\
&\leq \frac{1_P}{P^2} \operatorname{tr}\left(\mathbf{\Sigma}\right) \frac{\lambda_{\max}\left(\mathbf{\Psi}\left(\mathbf{\Psi}'\mathbf{\Psi}\right)\mathbf{\Psi}'\right)}{\lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)} \\
&= \frac{1_P \tau_P}{P \lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)} = O\left(\frac{\tau_P}{P}\right),
\end{aligned}$$

where the last inequality is obtained by idempotency of $\mathbf{\Psi}\left(\mathbf{\Psi}'\mathbf{\Psi}\right)\mathbf{\Psi}'$ and the upper bound is a direct result of Markov's inequality.

For the second term of the inequality we have:

$$\begin{aligned}
&\frac{1_P}{P^2}\left\| \left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)^{-1} \mathbf{\Psi}'\left(f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right) \right\| \\
=&\frac{1_P}{P^2}\left[\left(f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right)' \mathbf{\Psi}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)^{-1}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)^{-1} \mathbf{\Psi}'\left(f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right)\right] \\
\leq&\frac{1_P}{P^2}\left[\left(f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right)' \mathbf{\Psi}\left(\mathbf{\Psi}'\mathbf{\Psi}\right)^{-1} \mathbf{\Psi}'\left(f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right)\right] \lambda_{\max}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)^{-1} \\
\leq&\frac{1_P}{P^2}\left[\left(f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right)' \mathbf{\Psi}\left(\mathbf{\Psi}'\mathbf{\Psi}\right)^{-1} \mathbf{\Psi}'\left(f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right)\right] \frac{\lambda_{\max}\left(\mathbf{\Psi}\left(\mathbf{\Psi}'\mathbf{\Psi}\right)\mathbf{\Psi}'\right)}{\lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)} \\
\leq&\frac{1}{\lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right)} \frac{1_P}{P^2}\left[\left(f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right)\left(f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right)\right] \\
=&\frac{1_P}{\lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right) P}\left[f\left(\boldsymbol{\theta}\right) - f_K^\star\left(\boldsymbol{\theta}\right)\right]^2 = \frac{1_P}{\lambda_{\min}\left(\frac{\mathbf{\Psi}'\mathbf{\Psi}}{P}\right) P} N_K^2 = o\left(N_K^2\right).
\end{aligned}$$

Finally, the result of Lemma 21 is the following:

$$\begin{aligned}
1_P\left|\widehat{f}_K - f\right|_s &= 1_P\left|\widehat{f}_K - f_K^\star\right|_s + 1_P\left|f_K^\star - f\right|_s \\
&= 1_P\left|\boldsymbol{\psi}_K\left(\boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\right)\right|_s + 1_P\left|\boldsymbol{\psi}_K\left(\boldsymbol{\theta}\right)\boldsymbol{\beta}^\star - f\right|_s \\
&= 1_P\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\right\| + 1_P N_K \\
&\leq \zeta_s\left(K\right) 1_P\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\right\| + N_K \\
&= O_{\mathbb{P}}\left\{\zeta_s\left(K\right)\left[\left(\frac{\left(K\lambda_P\right) \wedge \tau_P}{P}\right)^{\frac{1}{2}} + N_K\right]\right\} + N_K.
\end{aligned}$$

Q.E.D.

Proof of Theorem 15. The bound in Lemma 21 and Assumption 14 lead to the definition of Theorem 15.

Q.E.D.

Proof of Corollary 16. By Assumption 3 we know that f is continuous, uniquely minimized at $\boldsymbol{\theta}^\star$ and $\boldsymbol{\Theta}$ is compact. By Theorem 15 $\left|\widehat{f}_K - f\right|_s \xrightarrow{\Pr} 0$. These conditions imply that $\widehat{\boldsymbol{\theta}} \xrightarrow{\Pr} \boldsymbol{\theta}^\star$.

Q.E.D.

8.3 Proof of rate of convergence

Proof of Theorem 19. The pseudo-true value $\boldsymbol{\theta}^\star$ is such that $f'(\boldsymbol{\theta}^\star) = 0$ and the estimator $\widehat{\boldsymbol{\theta}}$ is such that $\widehat{f}'\left(\widehat{\boldsymbol{\theta}}\right) = 0$. In order to define the rate of convergence, we expand $f'(\boldsymbol{\theta}^\star)$ around $\widehat{\boldsymbol{\theta}}$:

$$\begin{aligned} f'(\boldsymbol{\theta}^\star) &= f'\left(\widehat{\boldsymbol{\theta}}\right) + f''\left(\widetilde{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}^\star - \widehat{\boldsymbol{\theta}}\right), \\ \boldsymbol{\theta}^\star - \widehat{\boldsymbol{\theta}} &= \left[f''\left(\widetilde{\boldsymbol{\theta}}\right)\right]^{-1}\left[f'(\boldsymbol{\theta}^\star) - f'\left(\widehat{\boldsymbol{\theta}}\right)\right], \end{aligned} \tag{8.2}$$

where $\widetilde{\boldsymbol{\theta}}$ is a point between $\boldsymbol{\theta}^\star$ and $\widehat{\boldsymbol{\theta}}$. As $f'(\boldsymbol{\theta}^\star) = 0 = \widehat{f}'\left(\widehat{\boldsymbol{\theta}}\right)$, we can rewrite equation (8.2) as follows:

$$\begin{aligned} \boldsymbol{\theta}^\star - \widehat{\boldsymbol{\theta}} &= \left[f''\left(\widetilde{\boldsymbol{\theta}}\right)\right]^{-1}\left[\widehat{f}'\left(\widehat{\boldsymbol{\theta}}\right) - f'\left(\widehat{\boldsymbol{\theta}}\right)\right], \\ \left\|\boldsymbol{\theta}^\star - \widehat{\boldsymbol{\theta}}\right\| &\leq \left\|\left[f''\left(\widetilde{\boldsymbol{\theta}}\right)\right]^{-1}\right\| \left\|f'\left(\widehat{\boldsymbol{\theta}}\right) - \widehat{f}'\left(\widehat{\boldsymbol{\theta}}\right)\right\|. \end{aligned} \tag{8.3}$$

By equation (8.3) we can isolate $\left\|f' - \widehat{f}'\right\| \leq \left|\widehat{f}_K - f\right|_1$, which is known by Theorem 15. By Corollary 16 we have that $\widehat{\boldsymbol{\theta}} \xrightarrow{\Pr} \boldsymbol{\theta}^\star$ and $\widetilde{\boldsymbol{\theta}} \xrightarrow{\Pr} \boldsymbol{\theta}^\star$. This implies that, with probability converging to 1, $\widetilde{\boldsymbol{\theta}}$ belongs to the neighborhood $N(\theta^\star)$ of Assumption 17. Therefore:

$$\left\|\left[f''\left(\widetilde{\boldsymbol{\theta}}\right)\right]^{-1}\right\| \leq C\lambda_{\max}\left(\left[f''\left(\widetilde{\boldsymbol{\theta}}\right)\right]^{-1}\right) = C\lambda_{\min}^{-1}\left(f''\left(\widetilde{\boldsymbol{\theta}}\right)\right) \leq C\varepsilon.$$

At last, the rate of convergence is:

$$\left\|\boldsymbol{\theta}^\star - \widehat{\boldsymbol{\theta}}\right\| = O\left(\left|\widehat{f}_K - f\right|_1\right).$$

Q.E.D.

References

Aguirregabiria, Victor and Arvind Magesan (2020). "Identification and estimation of dynamic games when players? beliefs are not in equilibrium". In: *The Review of Economic Studies* 87.2, pp. 582–625.

Altissimo, Filippo and Antonio Mele (2009). "Simulated Non-Parametric Estimation of Dynamic Models". In: *Review of Economic Studies* 76.2, pp. 413–450.

Andrews, Donald W. K. (1991). "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models". In: *Econometrica* 59.2, pp. 307–345.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M Shapiro (2017). "Measuring the sensitivity of parameter estimates to estimation moments". In: *The Quarterly Journal of Economics* 132.4, pp. 1553–1592.

Belloni, Alexandre et al. (2015). "Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results". In: *Journal of Econometrics* 186.2, pp. 345–366.

Bierens, Herman J and Hosin Song (2012). "Semi-nonparametric estimation of independently and identically repeated first-price auctions via an integrated simulated moments method". In: *Journal of Econometrics* 168.1, pp. 108–119.

Boswijk, H. Peter, Cars H. Hommes, and Sebastiano Manzan (2007). "Behavioral Heterogeneity in Stock Prices". In: *Journal of Economic Dynamics and Control* 31.6, pp. 1938–1970.

Brock, William A. and Cars H. Hommes (1997). "A Rational Route to Randomness". In: *Econometrica* 65.5, p. 1059.

— (1998). "Heterogeneous Beliefs and Routes to Chaos in a Simple Asset Pricing Model". In: *Journal of Economic Dynamics and Control* 22.8-9, pp. 1235–1274.

Chen, Xiaohong (2007). "Chapter 76 - Large Sample Sieve Estimation of Semi-Nonparametric Models". In: *Handbook of Econometrics*. Vol. 6. Elsevier, pp. 5549–5632.

Chen, Xiaohong and Timothy M. Christensen (2015). "Optimal Uniform Convergence Rates and Asymptotic Normality for Series Estimators under Weak Dependence and Weak Conditions". In: *Journal of Econometrics* 188.2, pp. 447–465.

Chen, Xiaohong and Demian Pouzo (2012). "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals". In: *Econometrica* 80.1, pp. 277–321.

— (2015). "Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models: Sieve Wald and QLR Inference". In: *Econometrica* 83.3, pp. 1013–1079.

Cheng, Xu and Zhipeng Liao (2015). "Select the valid and relevant moments: An information-based LASSO for GMM with many moments". In: *Journal of Econometrics* 186.2, pp. 443–464.

Chernozhukov, Victor et al. (2018). "Double/debiased machine learning for treatment and structural parameters". In: *Econometrics Journal* 21.1, pp. C1–C68.

Choirat, Christine, Christian Hess, and Raffaello Seri (2003). "A Functional Version of the Birkhoff Ergodic Theorem for a Normal Integrand: A Variational Approach". In: *The Annals of Probability* 31.1, pp. 63–92.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (2005). "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy". In: *Journal of Political Economy* 113.1, pp. 1–45.

Corradi, Valentina and Norman R Swanson (2007). "Evaluation of dynamic stochastic general equilibrium models based on distributional comparison of simulated and historical data". In: *Journal of Econometrics* 136.2, pp. 699–723.

Cox, Dennis D. (1988). "Approximation of Least Squares Regression on Nested Subspaces". In: *The Annals of Statistics* 16.2, pp. 713–732.

de Jong, Robert M. (2002). "A Note on "Convergence Rates and Asymptotic Normality for Series Estimators": Uniform Convergence Rates". In: *Journal of Econometrics* 111.1, pp. 1–9.

Dominicy, Yves and David Veredas (2013). "The Method of Simulated Quantiles". In: *Journal of Econometrics* 172.2, pp. 235–247.

Duffie, Darrell and Kenneth J. Singleton (1993). "Simulated Moments Estimation of Markov Models of Asset Prices". In: *Econometrica* 61.4, pp. 929–952.

Eisenhauer, Philipp, James J. Heckman, and Stefano Mosso (2015). "Estimation of Dynamic Discrete Choice Models by Maximum Likelihood and the Simulated Method of Moments". In: *International Economic Review* 56.2, pp. 331–357.

Forneron, Jean-Jacques (2023). "A Sieve-SMM Estimator for Dynamic Models". In: *Econometrica* 91.3, pp. 943–977.

— (2024). "Estimation and inference by stochastic optimization". In: *Journal of Econometrics* 238.2, p. 105638.

Gallant, A. Ronald and George Tauchen (1996). "Which Moments to Match?" In: *Econometric Theory* 12.04, p. 657.

Gilli, Manfred and Peter Winker (2003). "A Global Optimization Heuristic for Estimating Agent Based Models". In: *Computational Statistics & Data Analysis* 42.3, pp. 299–312.

Goettler, Ronald L and Brett R Gordon (2011). "Does AMD spur Intel to innovate more?" In: *Journal of Political Economy* 119.6, pp. 1141–1200.

Gospodinov, Nikolay, Ivana Komunjer, and Serena Ng (2017). "Simulated Minimum Distance Estimation of Dynamic Models with Errors-in-Variables". In: *Journal of Econometrics* 200.2, pp. 181–193.

Gouriéroux, Christian and Alain Monfort (1996). *Simulation-Based Econometric Methods*. CORE Lectures. New York, N.Y.: Oxford University Press.

Gouriéroux, Christian, Alain Monfort, and Eric Renault (1993). "Indirect Inference". In: *Journal of Applied Econometrics* 8.S1, S85–S118.

Gray, Robert M (1988). *Probability, Random Processes, and Ergodic Properties*. New York, N.Y.: Springer.

Grenander, Ulf (1981). *Abstract Inference*. Wiley Series in Probability and Mathematical Statistics. New York, N.Y.: Wiley.

Hall, Peter (2013). *The bootstrap and Edgeworth expansion*. New York, N.Y.: Springer Science & Business Media.

Harvey, A., E. Ruiz, and N. Shephard (1994). "Multivariate Stochastic Variance Models". In: *The Review of Economic Studies* 61.2, pp. 247–264.

Hess, Christian (1996). "Epi-Convergence of Sequences of Normal Integrands and Strong Consistency of the Maximum Likelihood Estimator". In: *The Annals of Statistics* 24.3, pp. 1298–1315.

Hess, Christian and Raffaello Seri (2019). "Generic Consistency for Approximate Stochastic Programming and Statistical Problems". In: *SIAM Journal on Optimization* 29.1, pp. 290–317.

Hnatkovska, Viktoria, Vadim Marmer, and Yao Tang (2012). "Comparison of Misspecified Calibrated Models: The Minimum Distance Approach". In: *Journal of Econometrics* 169.1, pp. 131–138.

Hommes, Cars H. (2006). "Chapter 23 - Heterogeneous Agent Models in Economics and Finance". In: *Handbook of Computational Economics*. Vol. 2. Elsevier, pp. 1109–1186.

Huang, Jianhua Z. (1998). "Projection Estimation in Multiple Regression with Application to Functional ANOVA Models". In: *The Annals of Statistics* 26.1, pp. 242–272.

Kristensen, Dennis and Yongseok Shin (2012). "Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood". In: *Journal of Econometrics* 167.1, pp. 76–94.

Kukacka, Jiri and Jozef Barunik (2017). "Estimation of Financial Agent-Based Models with Simulated Maximum Likelihood". In: *Journal of Economic Dynamics and Control* 85, pp. 21–45.

Lamperti, Francesco (2018). "Empirical Validation of Simulated Models through the GSL-div: An Illustrative Application". In: *Journal of Economic Interaction and Coordination* 13.1, pp. 143–171.

Lamperti, Francesco, Andrea Roventini, and Amir Sani (2018). "Agent-Based Model Calibration Using Machine Learning Surrogates". In: *Journal of Economic Dynamics and Control* 90, pp. 366–389.

Lee, Lung-Fei (1992). "On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models". In: *Econometric Theory* 8.4, pp. 518–552.

Lux, Thomas (2024). "Lack of identification of parameters in a simple behavioral macroeconomic model". In: *Journal of Economic Dynamics and Control* 168, p. 104972.

Martinoli, Mario (2021). "Essays on Estimation, Calibration and Inference for Simulation Models". PhD thesis. Varese: Università degli Studi dell'Insubria.

Martinoli, Mario, Raffaello Seri, and Fulvio Corsi (2024). *Generalized Optimization Algorithms for Complex Objective Functions*. LEM Papers Series 2024/18. Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy.

McFadden, Daniel (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration". In: *Econometrica* 57.5, pp. 995–1026. JSTOR: 1913621.

Migliorati, Giovanni (2015). "Multivariate Markov-type and Nikolskii-type Inequalities for Polynomials Associated with Downward Closed Multi-Index Sets". In: *Journal of Approximation Theory* 189, pp. 137–159.

Newey, Whitney K. (1997). "Convergence Rates and Asymptotic Normality for Series Estimators". In: *Journal of Econometrics* 79.1, pp. 147–168.

— (2001). "Flexible simulated moment estimation of nonlinear errors-in-variables models". In: *Review of Economics and statistics* 83.4, pp. 616–627.

Newey, Whitney K. and Daniel McFadden (1994). "Chapter 36 - Large Sample Estimation and Hypothesis Testing". In: *Handbook of Econometrics*. Vol. 4. Elsevier, pp. 2111–2245.

Nikolov, Boris and Toni M Whited (2014). "Agency conflicts and cash: Estimates from a dynamic model". In: *The Journal of Finance* 69.5, pp. 1883–1921.

Pakes, Ariel and David Pollard (1989). "Simulation and the Asymptotics of Optimization Estimators". In: *Econometrica* 57.5, pp. 1027–1057.

Recchioni, Maria Cristina, Gabriele Tedeschi, and Mauro Gallegati (2015). "A Calibration Procedure for Analyzing Stock Price Dynamics in an Agent-Based Framework". In: *Journal of Economic Dynamics and Control* 60, pp. 1–25.

Ryabko, Daniil and Boris Ryabko (2010). "Nonparametric Statistical Inference for Ergodic Processes". In: *IEEE Transactions on Information Theory* 56.3, pp. 1430–1435.

Seri, Raffaello, Samuele Centorrino, and Michele Bernasconi (2019). "Nonparametric Estimation and Inference in Economic and Psychological Experiments". In: *arXiv preprint arXiv:1904.11156*.

Seri, Raffaello, Mario Martinoli, et al. (2021). "Model Calibration and Validation via Confidence Sets". In: *Econometrics and Statistics* 20, pp. 62–86.

Smith Jr., Anthony A. (1993). "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions". In: *Journal of Applied Econometrics* 8.S1, S63–S84.

Voena, Alessandra (2015). "Yours, mine, and ours: Do divorce laws affect the intertemporal behavior of married couples?" In: *American Economic Review* 105.8, pp. 2295–2332.